

# Beschluss

## des Gemeinsamen Bundesausschusses über die Freigabe des Berichts des IQTIG zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern zur Veröffentlichung

Vom 16. Juni 2022

Der Gemeinsame Bundesausschuss (G-BA) hat in seiner Sitzung am 16. Juni 2022 beschlossen, den Bericht des Instituts für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) *Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern. Abschlussbericht zu Stufe 1 und Stufe 2* zur Beauftragung vom 18. Januar 2018 gemäß **Anlage 1** gemeinsam mit einer Kommentierung gemäß **Anlage 2** für die Veröffentlichung auf den Internetseiten des IQTIG ([www.iqtig.org](http://www.iqtig.org)) freizugeben.

Berlin, den 16. Juni 2022

Gemeinsamer Bundesausschuss  
gemäß § 91 SGB V  
Der Vorsitzende

Prof. Hecken



Institut für Qualitätssicherung und  
Transparenz im Gesundheitswesen

# Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern

Abschlussbericht zu Stufe 1 und Stufe 2

Erstellt im Auftrag des  
Gemeinsamen Bundesausschusses

Stand: 11. Februar 2020

---

# Impressum

**Thema:**

Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern. Abschlussbericht zu Stufe 1 und Stufe 2

**Ansprechpartnerin:**

Andrea Wolf

**Auftraggeber:**

Gemeinsamer Bundesausschuss

**Datum des Auftrags:**

18. Januar 2018

**Datum der Abgabe:**

31. Januar 2020; redaktionell geändert am 11. Februar 2020

**Herausgeber:**

IQTIG – Institut für Qualitätssicherung  
und Transparenz im Gesundheitswesen

Katharina-Heinroth-Ufer 1  
10787 Berlin

Telefon: (030) 58 58 26-0  
Telefax: (030) 58 58 26-999

[info@iqtig.org](mailto:info@iqtig.org)

<https://www.iqtig.org>

# Inhaltsverzeichnis

Tabellenverzeichnis.....	8
Abbildungsverzeichnis.....	9
Abkürzungsverzeichnis.....	12
1 Einleitung.....	14
1.1 Hintergrund.....	14
1.2 Schnittstellen zu anderen Beauftragungen und Richtlinien .....	15
1.3 Auftragsverständnis .....	16
1.4 Einbeziehung der auf Landesebene zuständigen LAG- und LQS-Vertreterinnen und -Vertreter .....	19
1.5 Einbeziehung weiterer externer Expertise.....	19
1.6 Stellungnahmeverfahren zum Vorbericht.....	20
2 Analyse der Ausgangssituation .....	21
2.1 Funktionen des Strukturierten Dialogs .....	21
2.2 Anforderungen an die Qualitätsmessung und -bewertung .....	23
2.3 Strukturelle Voraussetzungen des Strukturierten Dialogs.....	26
2.4 Durchführung des bisherigen Strukturierten Dialogs .....	26
2.4.1 Einstieg in den Strukturierten Dialog .....	28
2.4.2 Bewertung der Versorgungsqualität .....	28
2.4.3 Bewertung der Dokumentationsqualität von Qualitätsindikatoren .....	31
2.4.4 Qualitätsförderung.....	32
2.5 Biometrische Analyse der Heterogenität in der Vorgehensweise und Bewertung im Strukturierten Dialog.....	33
2.6 Biometrische Analyse zu Aufwand und Effizienz des Strukturierten Dialogs .....	38
2.6.1 Aufwandsüberblick QS-Verfahren <i>Hüftendoprothesenversorgung</i> .....	38
2.6.2 Rechnerische Auffälligkeiten je Qualitätsindikator.....	38
2.6.3 Qualitative Auffälligkeiten je Qualitätsindikator.....	42
2.6.4 Zusammenhang zwischen Fallzahl und Auffälligkeitseinstufung .....	45
2.6.5 Zwischenfazit.....	53
3 Methodischer Hintergrund.....	55
3.1 Qualitätsmessung mittels Qualitätsindikatoren .....	55
3.2 Die methodische Funktion von Stellungnahmen .....	56

3.2.1	Einflussfaktoren auf das Indikatorergebnis.....	56
3.2.2	Prüfung der Validität je Messung.....	57
3.3	Reichweite der Qualitätsaussage von Stellungnahmen.....	58
3.4	Gütekriterien für Bewertungsprozesse.....	60
3.5	Ableitung von Anforderungen an das Stellungsnahmeverfahren .....	61
3.6	Vor- und Nachteile explizierter Beurteilungsregeln im Stellungsnahmeverfahren ...	62
3.7	Methodische Einordnung der Qualitätsförderung.....	64
3.8	Zusammenfassung .....	67
4	Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und -förderung .....	69
4.1	Modulare Betrachtungsweise .....	69
4.2	Qualitätsindikatoren als kleinste Bewertungseinheit .....	70
4.3	Grundlegender Ablauf des Moduls „Qualitätsbewertung“ .....	71
4.4	Beschränkung der Informationsgrundlage auf Stellungnahmen .....	73
5	Statistische Auswertungsmethodik.....	74
5.1	Rahmenkonzept der Auswertungsmethodik .....	74
5.1.1	Herangehensweise .....	76
5.1.2	Stichprobenart für die vorgegebene Auswertungsebene.....	80
5.1.3	Berechnungsart .....	80
5.1.4	Bewertungsart.....	81
5.1.5	Ziele der Auswertung für einen Leistungserbringer.....	81
5.2	Entscheidungstheoretische Modellierung des Bewertungsprozesses.....	83
5.2.1	Bayesianische Netzwerke und Einflussdiagramme .....	84
5.2.2	Einflussdiagramm für den Bewertungsprozess.....	85
5.2.3	Diskussion des entscheidungstheoretischen Ansatzes .....	89
5.3	Methoden für die quantitative 1-Jahres-Auffälligkeitseinstufung.....	89
5.3.1	Lösung des Entscheidungsdiagramms.....	90
5.3.2	Bayesianische Modellierung .....	99
5.3.3	Sensitivität und Spezifität der Einstufungsmethoden.....	111
5.3.4	Umgang mit verteilungsabhängigen Referenzwerten .....	113
5.3.5	Vor- und Nachteile der Klassifikationsmethoden .....	115
5.4	Berücksichtigung der Daten mehrerer Erfassungsjahre .....	116
5.4.1	Zusammenhang mit Strukturbruchproblemen, statistischer Prozesskontrolle und sequentiellen Entscheidungsproblemen.....	118

5.4.2	Untersuchte Methoden.....	122
5.4.3	Vergleich anhand von Sensitivität und Spezifität.....	130
5.4.4	Vergleich durch Entscheidungsdiagramm mit zwei Zeitpunkten.....	136
5.4.5	Weitere Aspekte.....	139
5.4.6	Fazit .....	142
5.5	Illustration der Auswertungsmethodik .....	142
5.6	Zusammenfassung und Empfehlungen.....	145
6	Fachliche Bewertung .....	151
6.1	Bewertungsschema zur Einstufung der Indikatorergebnisse .....	151
6.1.1	Kategorien des Bewertungsschemas .....	152
6.1.2	Differenzierung der Kategorie „kein Hinweis auf Qualitätsdefizit“ .....	153
6.1.3	Differenzierung der Kategorie „Qualitätsdefizit“ .....	155
6.1.4	Bewertungsschema im Überblick.....	156
6.1.5	Empfehlung für die zukünftige Einstufung der Indikatorergebnisse .....	157
6.2	Methodik der fachlichen Bewertung .....	158
6.3	Ablauf der fachlichen Bewertung.....	161
6.4	Bewertungsalgorithmus .....	162
6.4.1	Analyse der in den Stellungnahmen genannten Gründe .....	165
6.4.2	Festlegung von Referenzbereichen für die partielle Nachberechnung .....	168
6.4.3	Methodische Limitationen verteilungsbezogener Referenzbereiche.....	170
6.4.4	Vergabe der Unterkategorien für die Einstufung „Qualitätsdefizit“ .....	171
6.4.5	Vergabe der Unterkategorien bei Entkräftung des Hinweises auf ein Qualitätsdefizit .....	172
6.5	Mindestanforderungen an Stellungnahmen.....	175
6.5.1	Formale Kriterien .....	175
6.5.2	Inhaltlicher Fokus .....	176
6.6	Umsetzung und Aufwand.....	179
6.7	Zusammensetzung von Fachkommissionen .....	182
6.7.1	Kriterien für die Auswahl der Expertinnen und Experten .....	183
6.7.2	Kriterien für die Zusammensetzung der Fachkommission.....	185
6.8	Fazit .....	188
7	Modul Qualitätsförderung .....	189
7.1	Empfehlungen zur Einleitung qualitätsverbessernder Maßnahmen .....	190

7.2	Zielvereinbarungen .....	193
7.2.1	Formulierung von Zielvereinbarungen.....	193
7.2.2	Übersicht über den Stand der Zielvereinbarungen.....	193
8	Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens .....	195
8.1	Das Verfahren gemäß DeQS-RL.....	195
8.2	Empfehlungen zur Verkürzung des Stellungnahmeverfahrens .....	196
8.2.1	Aufwandsreduktion durch Volumenreduktion .....	197
8.2.2	Effizienzsteigerung durch Standardisierung im Vorgehen.....	197
8.2.3	Auswirkungen der Empfehlungen auf den Beginn der Bewertung von Stellungnahmen .....	198
8.2.4	Auswirkungen der Empfehlungen auf den Abschluss des Stellungnahmeverfahrens.....	198
8.3	Vorschläge der Vertreterinnen und Vertreter der LAG und der LQS.....	200
8.4	Verkürzung des Gesamtverfahrens durch früheren Start des Stellungnahmeverfahrens.....	202
9	Berichterstattung .....	204
9.1	Berichterstattung zur Versorgungsqualität.....	206
9.2	Berichterstattung über die Maßnahmen der Qualitätsförderung.....	206
10	Empfehlungen zur Evaluation des neuen Konzepts.....	209
10.1	Evaluationsdesign.....	209
10.2	Datenquellen und Zeitschiene .....	211
10.3	Zusammenfassung der Empfehlungen zur Evaluation des vorliegenden Konzepts	211
10.4	Evaluation der quantitativen Qualitätsbewertung .....	212
10.4.1	Limitationen der Evaluation .....	213
10.4.2	1-Jahres-Einstufung.....	214
10.4.3	Fazit .....	220
10.4.4	Mehrjahreseinstufung.....	221
10.4.5	Fazit .....	226
11	Zusammenfassungen der Empfehlungen und Konsequenzen .....	228
11.1	Anforderungen an und Weiterentwicklung von Qualitätsindikatoren .....	228
11.2	Gleichsetzung quantitativer Auffälligkeit mit qualitativer Auffälligkeit .....	229
11.3	Indizes als Aufgreifkriterien .....	231
11.4	Leistungsbereichübergreifende Qualitätsindikatorensets.....	232

11.5 Zusammenfassung der zentralen Empfehlungen .....	233
11.5.1 Überblicksliste der Empfehlungen .....	237
11.5.2 Fazit .....	240
Literatur.....	242

## Tabellenverzeichnis

Tabelle 1: Sensitivität und Spezifität des Gesamtverfahrens zur Qualitätsmessung bezogen auf einen Qualitätsindikator .....	22
Tabelle 2: Indikatorenauswahl für exemplarische Analyse der Heterogenität des Strukturierten Dialogs. Alle Angaben beziehen sich auf den Strukturierten Dialog zum Erfassungsjahr 2016. ....	35
Tabelle 3: Rechnerisch auffällige Standorte im HEP-Verfahren für alle 14 Qualitätsindikatoren. Daten aus dem Erfassungsjahr 2017 .....	39
Tabelle 4: Vier-Felder-Schema zur Darstellung der Aufwände jeder Entscheidung bei der Klassifikation eines Leistungserbringers anhand dessen beobachteter Ergebnisse des Leistungserbringers, d. h. $(o, J)$ . Exemplarisch dargestellt ist die Situation, wenn nur Fehlklassifikationen zu Aufwänden führen.....	88
Tabelle 5: Vier-Felder-Tafel der Aufwandsannahmen des zweischrittigen Klassifikationsproblems .....	91
Tabelle 6: Aufwandsannahmen für die statistisch relevante Einstufungsmethode .....	96
Tabelle 7: Übersichtstabelle über die Annahmen und Entscheidungsregeln der drei vorgestellten quantitativen Auffälligkeitseinstufungsmethoden.....	97
Tabelle 8: Kritische Zahl interessierender Ereignisse, ab der Leistungserbringer in den drei Auffälligkeitseinstufungsmethoden als quantitativ auffällig bewertet werden.....	104
Tabelle 9: Kritische Anzahl an interessierenden Ereignissen, die zur quantitativen Auffälligkeit führt, je nach erwarteter Anzahl an interessierenden Ereignissen und Einstufungsstrategie .....	109
Tabelle 10: Wahl der Übergangswahrscheinlichkeiten für den Kompetenzparameter .....	137
Tabelle 11: Wahl der mit verschiedenen Fehlklassifikationen assoziierten Aufwände.....	137
Tabelle 12: Erwarteter Verlust der verschiedenen 2-Jahres-Methoden für das konkrete Einflussdiagramm mit zwei Zeitpunkten.....	138
Tabelle 13: Rechenregeln eines Beispielindikators, angelehnt an den Qualitätsindikator 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ .....	142
Tabelle 14: Bewertungsschema .....	157
Tabelle 15: Vorlage für den Bericht an das Lenkungsgremium zum aktuellen Stand aktiver Zielvereinbarungen .....	194
Tabelle 16: Stellungnahmeverfahren (STNV), quantitative Auffälligkeiten, entdeckte qualitative Auffälligkeiten und PPV für ausgewählte Werte von $\alpha$ .....	220
Tabelle 17: Quantitative Auffälligkeiten, entdeckte qualitative Auffälligkeiten und PPV für ausgewählte Werte von $\alpha$ .....	226

## Abbildungsverzeichnis

Abbildung 1: Schematische Darstellung von Maßnahmen im Rahmen des Strukturierten Dialogs.....	27
Abbildung 2: Beispielhafte Darstellung der Analyseergebnisse zum Indikator „Präoperative Verweildauer bei endoprothetischer Versorgung einer Hüftgelenknahen Femurfraktur“ (QI-ID 54003).....	36
Abbildung 3: Standortergebnisse im Qualitätsindikator QI-ID 54017 „Allgemeine Komplikationen bei Hüftendoprothesen-Wechsel bzw. -Komponentenwechsel“ differenziert nach Fallzahl.....	41
Abbildung 4: Qualitative Auffälligkeiten in Relation zu rechnerischen Auffälligkeiten (oben) und angeforderten Stellungnahmen (unten). Daten des Erfassungsjahres 2017. ....	43
Abbildung 5: Rechnerische Auffälligkeitseinstufung und Stellungnahmen für Nicht-Sentinel-Event-Indikatoren im HEP-Verfahren differenziert in Dezilen der Fallzahlverteilung.....	47
Abbildung 6: Rechnerische und qualitative Auffälligkeitseinstufung zum Sentinel-Event-Indikator „Todesfälle während des akut-stationären Aufenthaltes bei geringer Sterbewahrscheinlichkeit“ (QI-ID 54013) differenziert in Dezilen der Fallzahlverteilung.....	48
Abbildung 7: Anteil qualitativ auffälliger Standortergebnisse unter allen Ergebnissen, für die eine Stellungnahme angefordert wurde, differenziert nach Fallzahldezilen .....	49
Abbildung 8: Fallzahlabhängigkeit der Behandlungsqualität auf Bundesebene (oben) und der rechnerischen Auffälligkeitseinstufung (unten) zum Qualitätsindikator „Allgemeine Komplikationen bei endoprothetischer Versorgung einer Hüftgelenknahen Femurfraktur“ .....	52
Abbildung 9: Grundlegender Ablauf der Qualitätsbewertung und -förderung .....	71
Abbildung 10: Bayesianisches Netzwerk zur Illustration der Zusammenhänge zwischen Fallzahl $J$ , Kompetenzparameter $\theta$ und Anzahl unerwünschter Qualitätsergebnisse $O$ für einen Ratenindikator. ....	84
Abbildung 11: Schematisches Einflussdiagramm für das zweistufige Entscheidungsverfahren des Bewertungsprozesses für einen Ratenindikator. ....	85
Abbildung 12: Vereinfachtes Entscheidungsdiagramm für die quantitative Auffälligkeitseinstufungsmethode bei einem Ratenindikator.....	88
Abbildung 13: links: Beta-Verteilung mit Parametern $a = b = \frac{1}{2}$ ; Mitte: resultierende A-posteriori-Verteilung bei 7 von 10 Ereignissen; rechts: resultierende A-posteriori-Verteilung bei 70 von 100 Ereignissen.....	101
Abbildung 14: Funnelpplot für die quantitative Auffälligkeitseinstufung der drei Klassifikationsmethoden für einen Indikator mit Referenzbereich [0–10 %].....	103
Abbildung 15: Wahrscheinlichkeitsdichte einer Gamma-Verteilung mit $\theta = 1$ bei verschiedenen A-priori-Werten von $\alpha$ .....	107
Abbildung 16: Funnelpplot-Darstellung verschiedener Auffälligkeitseinstufungsmethoden für risikoadjustierte Qualitätsindikatoren.....	108

Abbildung 17: Vergleich von Sensitivität und 1-Spezifität für die drei Methoden der quantitativen Auffälligkeitseinstufung. Die Fallzahl auf der x-Achse ist dabei auf einer logarithmischen Skala dargestellt. ....	112
Abbildung 18: Beispielhafter Verlauf des QI-Ergebnis pro Erfassungsjahr eines hypothetischen Leistungserbringers für einen Ratenindikator. Der Referenzbereich $\leq 10\%$ ist grau schattiert. Um die Ergebnisse sind bayesianische 90%-Unsicherheitsintervalle eingezeichnet. ....	117
Abbildung 19: Entscheidungsdiagramm für das sequentielle Entscheidungsproblem der quantitativen Auffälligkeitseinstufung. ....	119
Abbildung 20: Verhalten der Laufregel bei verschiedenen Ergebnissen aus zwei Jahren für einen Leistungserbringer mit 25 Fällen in jedem Jahr. ....	124
Abbildung 21: Verhalten der statistischen Auffälligkeit auf bis zu zwei Jahren bei verschiedenen Ergebnissen für einen Leistungserbringer mit 25 Fällen in jedem Jahr. ....	126
Abbildung 22: Verhalten des GLR-Binom-Verfahrens bei verschiedenen Ergebnissen aus zwei Jahren für einen Leistungserbringer mit 25 Fällen in jedem Jahr. ....	128
Abbildung 23: Verhalten der statistisch relevanten Auffälligkeitseinstufung auf den Daten zweier Jahre bei verschiedenen Ergebnissen für einen Leistungserbringer mit 25 Fällen in jedem Jahr. ....	130
Abbildung 24: Die Auslösewahrscheinlichkeit verschiedener Methoden bei $\theta=R$ als Funktion von $J$ . ....	132
Abbildung 25: Die Auslösewahrscheinlichkeit verschiedener Methoden bei $\theta=2R$ als Funktion von $J$ . ....	133
Abbildung 26: Die Auslösewahrscheinlichkeit verschiedener Methoden bei $\theta=R/2$ als Funktion von $J$ . ....	134
Abbildung 27: Die Auslösewahrscheinlichkeit verschiedener Methoden als Funktion von $\theta$ gemittelt über $J=3, \dots, 7$ . Die gestrichelten Linien markieren den Referenzwert $R$ sowie den Schwellenwert $\alpha=0,05$ . ....	135
Abbildung 28: Die Auslösewahrscheinlichkeit verschiedener Methoden als Funktion von $\theta$ gemittelt über $J=11, \dots, 20$ . Die gestrichelten Linien markieren den Referenzwert $R$ sowie den Schwellenwert $\alpha=0,05$ . ....	135
Abbildung 29: Optimale Strategie im gewählten Entscheidungsdiagramm in Abhängigkeit von $\sigma$ , $\sigma - 1$ und $S_0$ . ....	138
Abbildung 30: Dichte der A-posteriori-Wahrscheinlichkeit für die zugrunde liegende Rate $\theta$ bei einem Leistungserbringer mit $J = 20$ Fällen im Nenner und $\sigma = 5$ Fällen im Zähler. Anhand des Referenzwerts von 15 % wird diese Dichte in Dichte für Werte innerhalb und außerhalb des Referenzbereiches unterteilt. ....	144
Abbildung 31: Mögliche Kombinationen von quantitativem Indikatorergebnis und Ergebnis des Stellungnahmeverfahrens. ....	153
Abbildung 32: Ablauf der fachlichen Bewertung von Stellungnahmen im Überblick. ....	161
Abbildung 33: Grundlegender Algorithmus der fachlichen Bewertung. ....	163
Abbildung 34: Beispiel für die Nachberechnung von Indikatorergebnis und Referenzbereich zur fachlichen Bewertung. ....	169

Abbildung 35: Algorithmus zur Differenzierung der Gründe für die Einstufung als Qualitätsdefizit.....	171
Abbildung 36: Algorithmus für die quantitative Differenzierung der Gründe für einen entkräfteten Hinweis auf ein Qualitätsdefizit.....	173
Abbildung 37: Heuristische Differenzierung der Gründe für einen entkräfteten Hinweis auf ein Qualitätsdefizits .....	174
Abbildung 38: Empfehlungen zur Einleitung qualitätsverbessernder Maßnahmen (Ausschnitt aus Abbildung 9 in Kapitel 4) .....	191
Abbildung 39: Phasen und Fristen des Verfahrens gemäß DeQS-RL .....	195
Abbildung 40: Schematischer Ablauf der Bewertung von Auffälligkeiten gemäß DeQS-RL.....	200
Abbildung 41: Schematischer Ablauf der Bewertung von Auffälligkeiten gemäß Empfehlungen.....	201
Abbildung 42: Mögliche Phasen und Änderungen des Verfahrens nach Erfahrung mit der Umsetzung der Empfehlungen. ....	203
Abbildung 43: Über die vom Referenzwert hinaus tolerierte Anzahl an Fällen mit unerwünschtem Ereignis in Abhängigkeit des Tuning-Parameters ( $M(\alpha)$ ) .....	215
Abbildung 44: Anzahl an quantitativen Auffälligkeiten in Abhängigkeit des Tuning-Parameters.....	217
Abbildung 45: Entdeckte qualitative Auffälligkeiten in Abhängigkeit des Tuning-Parameters	218
Abbildung 46: PPV in Abhängigkeit des Tuning-Parameters .....	219
Abbildung 47: Über die vom Referenzwert hinaus tolerierte Anzahl an Fällen mit unerwünschtem Ereignis in Abhängigkeit vom Tuning-Parameter .....	222
Abbildung 48: Anzahl an quantitativen Auffälligkeiten in Abhängigkeit vom Tuning-Parameter .....	223
Abbildung 49: Entdeckte qualitative Auffälligkeiten in Abhängigkeit vom Tuning-Parameter	224
Abbildung 50: PPV in Abhängigkeit vom Tuning-Parameter .....	225

## Abkürzungsverzeichnis

Abkürzung	Bedeutung
aQua-Institut	Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH
BQB	Bundesqualitätsbericht
DeQS-RL	Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung
EJ	Erfassungsjahr
EJ + 1	Jahr, das ein Jahr nach dem Erfassungsjahr endet (usw. EJ + 2, EJ + 3)
esQS	externe stationäre Qualitätssicherung
G-BA	Gemeinsamer Bundesausschuss
GKV-FQWG	GKV-Finanzstruktur- und Qualitätsweiterentwicklungsgesetz (Gesetz zur Weiterentwicklung der Finanzstruktur und der Qualität in der gesetzlichen Krankenversicherung)
ID	Identifikationsnummer
IQTIG	Institut für Qualitätssicherung und Transparenz im Gesundheitswesen
KHSG	Krankenhausstrukturgesetz (Gesetz zur Reform der Strukturen der Krankenhausversorgung)
LAG	Landesarbeitsgemeinschaft(en) für sektorenübergreifende Qualitätssicherung
LE	Leistungserbringer
LQS	Landesgeschäftsstelle(n) für Qualitätssicherung
O/E	Verhältnis aus der beobachteten und der erwarteten Rate ( <i>observed to expected ratio</i> )
plan. QI-RL	Richtlinie zu planungsrelevanten Qualitätsindikatoren
PPV	Positive-Predictive-Value
QB	Qualitätsbewertung
Qesü-RL	Richtlinie zur einrichtungs- und sektorenübergreifenden Qualitätssicherung
QFD-RL	Qualitätsförderungs- und Durchsetzungs-Richtlinie
QI	Qualitätsindikator
QIDB	Qualitätsindikatorendatenbank
QS	Qualitätssicherung

<b>Abkürzung</b>	<b>Bedeutung</b>
QS-Verfahren	Qualitätssicherungsverfahren
<i>QS CHE</i>	<i>QS-Verfahren Cholezystektomie</i>
<i>QS PCI</i>	<i>QS-Verfahren Perkutane Koronarintervention (PCI) und Koronarangiographie</i>
<i>QS WI</i>	<i>QS-Verfahren Vermeidung nosokomialer Infektionen – postoperative Wundinfektionen</i>
QSEB	Qualitätssicherungsergebnisberichte
QSKH-RL	Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern
RL	Richtlinie
SD	Strukturierter Dialog
SGB V	Fünftes Buch Sozialgesetzbuch
SIR	standardisiertes Inzidenzverhältnis
SMR	standardisiertes Morbitätsverhältnis
SOP	Standardvorgehensweise ( <i>Standard Operating Procedure</i> )

# 1 Einleitung

Mit Beschluss vom 18. Januar 2018 hat der Gemeinsame Bundesausschuss (G-BA) das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) mit der Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern beauftragt. Dieser Weiterentwicklungsauftrag ist in zwei Stufen mit unterschiedlichen Inhalten und Abgabefristen gegliedert. Der Abschlussbericht zu Stufe 1 wurde dem G-BA am 30. November 2018 übermittelt. Der vorliegende Bericht beinhaltet sowohl die weiterentwickelten Empfehlungen des Berichts zu Stufe 1 als auch die für Stufe 2 beauftragten Inhalte. Aus Gründen der Lesbarkeit wurde davon abgesehen, Veränderungen im Vergleich zu dem Bericht zu Stufe 1 hervorzuheben.

Im Folgenden werden zunächst der Hintergrund der Beauftragung, Schnittstellen zu anderen Beauftragungen sowie das Auftragsverständnis des IQTIG für beide Stufen der Beauftragung dargestellt. Anschließend wird auf die Einbeziehung der auf Landesebene zuständigen LAG- und LQS-Vertreterinnen und -Vertreter und weiterer externer Expertise sowie den Ablauf des Stellungnahmeverfahrens zum Abschlussbericht eingegangen.

## 1.1 Hintergrund

Als Strukturierter Dialog werden diejenigen Prozesse bezeichnet, die gemäß §§ 11 bis 14 QSKH-RL eingeleitet werden sollen, wenn ein Leistungserbringer in einem Qualitätsindikator oder einem Auffälligkeitskriterium der QSKH-RL vom Referenzwert abweicht. Laut § 12 QSKH-RL soll dadurch geprüft werden, ob die dokumentierten Leistungen in der fachlich gebotenen Qualität erbracht und korrekt dokumentiert wurden.

Die Verantwortung für die Durchführung des Strukturierten Dialogs liegt bei der jeweils mit der Durchführung des Verfahrens beauftragten Stelle. Dies ist für einige Qualitätssicherungsverfahren (QS-Verfahren) der Unterausschuss Qualitätssicherung des G-BA (sog. direkte bzw. bundesbezogene Verfahren) und für den Großteil der QS-Verfahren (sog. indirekte bzw. landesbezogene Verfahren) das verantwortliche Gremium der Landesgeschäftsstellen für Qualitätssicherung (LQS) bzw. der Landesarbeitsgemeinschaften (LAG) in den Bundesländern.

Analysen der Ergebnisse des Strukturierten Dialogs, die das IQTIG dem G-BA jährlich im Mai vorlegt, deuten auf eine deutliche Heterogenität im Vorgehen und in den Ergebnissen zwischen den beauftragten Stellen hin (IQTIG 2018a; siehe auch Abschnitt 2.5). Faire Leistungserbringervergleiche und aussagekräftige Ergebnisse erfordern jedoch ein einheitliches, standardisiertes Vorgehen aller beauftragten Stellen. Gleichzeitig wächst der Aufwand der beauftragten Stellen mit steigender Anzahl von QS-Verfahren und Indikatoren. Um die Prozesse des Strukturierten Dialogs stärker zu standardisieren und den Aufwand zu reduzieren wurden 2013 in einem gemeinsamen Workshop des G-BA mit den LQS Schwachstellen und Verbesserungspotenziale identifiziert. Dieser Weiterentwicklungsbedarf wurde im Auftrag des G-BA im Rahmen von drei Projektgruppen weiter konkretisiert. Die Ergebnisse dieser Projektgruppen mündeten 2014 in

Vorschläge für Aufträge für die damalige Institution nach § 137a SGB V<sup>1</sup>. Eine Beauftragung erfolgte jedoch aufgrund der Beendigung des Vertrags mit dem AQUA-Institut damals nicht.

## 1.2 Schnittstellen zu anderen Beauftragungen und Richtlinien

Am 19. Juli 2018 hat der G-BA die Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung (DeQS-RL)<sup>2</sup> beschlossen (G-BA 2018b), in die die indikatorbasierten QS-Verfahren sowohl der bestehenden Richtlinie zur sektorenübergreifenden Qualitätssicherung (Qesü-RL)<sup>3</sup> als auch der sektorenspezifischen Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern (QSKH-RL) überführt werden sollen.

Seit dem 1. Januar 2019 ist die DeQS-RL in Kraft und hat die Qesü-RL abgelöst. Damit sind jetzt die bisher in der Qesü-RL verorteten Indikatorensets den Regelungen der DeQS-RL unterworfen. Gleiches wird zukünftig sukzessive für die QS-Verfahren im Rahmen der QSKH-RL erfolgen. Das bedeutet, dass in der Übergangszeit einige QS-Verfahren noch nach den Vorgaben der QSKH-RL und einige QS-Verfahren bereits nach den Vorgaben der DeQS-RL durchgeführt werden. Darüber hinaus werden seit 2017 einige Qualitätsindikatoren der QSKH-RL auch nach den Regelungen der Richtlinie für planungsrelevante Qualitätsindikatoren (plan. QI-RL) verwendet, um den für die Krankenhausplanung zuständigen Landesbehörden qualitätsorientierte Entscheidungen zu ermöglichen. Gemeinsam ist allen vier Richtlinien, dass Qualitätsindikatoren verwendet werden, um leistungserbringerbezogene Aussagen zur Versorgungsqualität in bestimmten Themenbereichen zu machen.

Mit Beschluss vom 17. Mai 2018 wurde das IQTIG durch den G-BA auch mit der „Entwicklung von Kriterien für die Datenbewertung und die Einleitung und Durchführung von Qualitätssicherungsmaßnahmen“ im Rahmen der Qesü-RL beauftragt. Dieser Auftrag beinhaltete Empfehlungen zu entwickeln für bundesweit einheitliche Bewertungskriterien für die Ergebnisse der Qualitätsindikatoren aus den QS-Verfahren *Perkutane Koronarintervention (PCI) und Koronarangiographie (QS PCI)* und *Vermeidung nosokomialer Infektionen – postoperative Wundinfektionen (QS WI)*. In § 17 Qesü-RL ist geregelt, dass bei Auffälligkeiten in den Indikatorergebnissen durch die mit der Durchführung des Verfahrens zuständige Stelle ein Stellungnahmeverfahren durchzuführen ist. Dieses Stellungnahmeverfahren wird im vorliegenden Bericht als Pendant zum Strukturierten Dialog der QSKH-RL verstanden. In der Qesü-RL ist jedoch weder definiert, was eine Auffälligkeit in einem Indikatorergebnis ist, noch ist der Ablauf des

---

<sup>1</sup> AQUA – Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH; bis 31. Dezember 2015 die zuständige Institution nach § 137a SGB V.

<sup>2</sup> Richtlinie des Gemeinsamen Bundesausschusses gemäß § 136 Abs. 1 SGB V i. V. m. § 135a SGB V über Maßnahmen der Qualitätssicherung für nach § 108 SGB V zugelassene Krankenhäuser (Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern / QSKH-RL). In der Fassung vom 15. August 2006, zuletzt geändert am 16. März 2018, in Kraft getreten am 27. April 2018. URL: <https://www.g-ba.de/informationen/richtlinien/38/> (abgerufen am: 05.07.2018).

<sup>3</sup> Richtlinie des Gemeinsamen Bundesausschusses nach § 92 Abs. 1 Satz 2 Nr. 13 i. V. m. § 136 Abs. 1 Nr. 1 SGB V über die einrichtungs- und sektorenübergreifenden Maßnahmen der Qualitätssicherung (Richtlinie zur einrichtungs- und sektorenübergreifenden Qualitätssicherung – Qesü-RL). In der Fassung vom 19. April 2010, zuletzt geändert am 19. Oktober 2017, in Kraft getreten am: 8. Februar 2018. URL: <https://www.g-ba.de/informationen/richtlinien/72/> (abgerufen am: 28.11.2018).

Stellungnahmeverfahrens näher beschrieben. Die Beauftragung vom 17. Mai 2018 beinhaltet also eine Ausgestaltung des Stellungnahmeverfahrens nach § 17 Qesü-RL.

Um ein einheitliches Vorgehen bei der Auswertung und Bewertung von Indikatorergebnissen über verschiedene Richtlinien hinweg sicherzustellen, wurden die Empfehlungen für das Stellungnahmeverfahren nach § 17 Qesü-RL aus dem Bericht zu Stufe 1 der Beauftragung zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern abgeleitet. Zu dieser Beauftragung hat das IQTIG am 15. März 2019 den Abschlussbericht an den G-BA übermittelt. Die Qesü-RL trat jedoch schon mit Wirkung zum 1. Januar 2019 außer Kraft.

Die Beauftragung des IQTIG zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern beinhaltet die Entwicklung eines Rahmenkonzepts, das mit anderen relevanten Richtlinien abzustimmen ist; dies beinhaltet aus Sicht des IQTIG die DeQS-RL sowie die plan. QI-RL.<sup>4</sup> Da die Qesü-RL zum Zeitpunkt der Berichtlegung nicht mehr in Kraft ist und die QSKH-RL wie oben beschrieben auch mittelfristig außer Kraft gesetzt werden soll, erscheint es aus Sicht des IQTIG sinnvoll, ein einheitliches Rahmenkonzept für ein Stellungnahmeverfahren zu den quantitativen Ergebnissen der datengestützten Qualitätssicherung zu entwickeln, das prinzipiell in allen indikatorbasierten QS-Verfahren zur Anwendung kommen kann. Dies ist auch im Einklang mit der Beauftragung, die vorsieht, dass „Elemente des Strukturierten Dialogs [...] zukünftig auch in der Rahmenrichtlinie zur datengestützten Qualitätssicherung [DeQS-RL] und in der Richtlinie zur qualitätsabhängigen Vergütung eingesetzt werden können [sollen].“ Mit diesem Ziel legt das IQTIG daher in diesem Abschlussbericht seine Empfehlungen für die Gestaltung eines Stellungnahmeverfahrens für die DeQS-RL<sup>5</sup> vor. Er beinhaltet konkrete Empfehlungen für die Auswertung und Bewertung von leistungserbringerbezogenen Qualitätsindikatorergebnissen auf Basis von Dokumentationsdaten der Leistungserbringer, auf Basis von Sozialdaten sowie auf Basis durch Patientenbefragungen<sup>6</sup> erhobener Daten.

### 1.3 Auftragsverständnis

Das IQTIG wurde beauftragt, Empfehlungen für die Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern zu entwickeln. Laut Auftrag verfolgt der G-BA damit die folgenden drei Ziele: Reduktion der Heterogenität in der Vorgehensweise zwischen den beauftragten Stellen, Erhöhung der Transparenz und Nachvollziehbarkeit der Entscheidungen sowie Steigerung der Effizienz des Strukturierten Dialogs.

Unter „1. Optimierung der Einheitlichkeit der Vorgehensweise“ ist das IQTIG beauftragt, einheitliche Kriterien für die Einholung von Stellungnahmen sowie Empfehlungen zu Form und Inhalt der Stellungnahmen zu entwickeln. Darüber hinaus sollen Kriterien erarbeitet werden, die die Bewertung von Stellungnahmen vereinheitlichen und erleichtern sollen. In diesem Kontext soll

---

<sup>4</sup> Eine Abstimmung mit einer „Richtlinie zur qualitätsabhängigen Vergütung“ ist auch gefordert. Diese lag zum Zeitpunkt der Berichtserstellung allerdings noch nicht vor.

<sup>5</sup> Aus diesem Grund werden ab Kapitel 4 „Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und -förderung“ nicht mehr die Begriffe der QSKH-RL sondern die der DeQS-RL verwendet und der Begriff Strukturierter Dialog durch Stellungnahmeverfahren ersetzt.

<sup>6</sup> Die Auswertungsmethodik für Qualitätsindikatoren auf Basis von Patientenbefragungsdaten ist in den jeweiligen Abschlussberichten beschrieben.

auch das bestehende „Schema zur Einstufung und Bewertung von Indikatorergebnissen“ überarbeitet werden.

In diesem Zusammenhang sollen auch Empfehlungen für die Anforderungen an die Fachexpertinnen und Fachexperten sowie die Zusammensetzung der Expertengremien als auch Möglichkeiten, um die Auswirkung von Interessenkonflikten der Fachexpertinnen und Fachexperten zu minimieren, entwickelt werden.

Die Empfehlungen für „1. Optimierung der Einheitlichkeit der Vorgehensweise“ finden sich in Abschnitt 5.6 und Kapitel 6.

Das IQTIG ist weiterhin beauftragt, Empfehlungen zur Evaluierung des erarbeiteten Konzepts zu geben. Obwohl unter 1. aufgeführt, wird dies als Auftrag verstanden, Evaluationsempfehlungen für das gesamte Konzept zu erarbeiten. Die Empfehlungen zur Evaluation finden sich in Kapitel 10.

Unter „2. Optimierung der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung“ soll die Berichterstattung zu den Ergebnissen des Strukturierten Dialogs weiterentwickelt werden. Dies beinhaltet gemäß der Beauftragung Vorschläge, wie die gleiche Datengrundlage für den Qualitätsreport und für den Bericht zum Strukturierten Dialog, den das IQTIG regelmäßig im Mai auf Basis der Berichte der Stellen auf Landesebene erstellt, gewährleistet werden kann. Zusätzlich sollen Mindestanforderungen für Berichte der Landesstellen formuliert werden. Während im Bericht zu Stufe 1 der Beauftragung noch Empfehlungen für die Berichte der QSKH-RL gemacht wurden, wird in dem vorliegenden Abschlussbericht davon abgesehen, diese Empfehlungen weiterzuentwickeln. Vor dem Hintergrund des absehbaren Außerkrafttretens der QSKH-RL erscheint der Aufwand, der eine Umgestaltung der Berichterstattung zu den Qualitätsergebnissen im Rahmen dieser Richtlinie bedeutet, nicht mehr gerechtfertigt. Daher werden auch an dieser Stelle die Empfehlungen mit Blick auf die DeQS-RL gegeben. Die Empfehlungen zu „2. Optimierung der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung“ finden sich in Kapitel 9.

Unter 3. „Optimierung der Effizienz des Verfahrens“ wird die Optimierung der „Diskriminationsfähigkeit“ der Indikatoren unter Prüfung verschiedener statistischer Verfahren beauftragt. Unter Diskriminationsfähigkeit versteht das IQTIG in diesem Kontext die Güte, mit der Leistungserbringer auf Basis eines Indikatorergebnisses in solche mit Qualitätsdefizit im Sinne des Qualitätsziels des jeweiligen Indikators und solche Leistungserbringer ohne Qualitätsdefizit unterteilt werden können. Aus Sicht des IQTIG spielen allerdings neben dem statistischen Verfahren auch andere Faktoren eine Rolle für die Effizienz des Verfahrens: Da mit zunehmender Standardisierung von Prozessen typischerweise der für die Durchführung benötigte Aufwand sinkt, versteht das IQTIG die zur Vereinheitlichung der Vorgehensweise zu entwickelnden Empfehlungen als weitere Faktoren zur Steigerung der Effizienz. Die Empfehlungen zur Optimierung der Effizienz finden sich dementsprechend im Abschnitt 5.6 und Kapitel 6.

Außerdem sollen Möglichkeiten zur Verkürzung des Verfahrens geprüft werden. Hierunter versteht das IQTIG Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens als Teil des Gesamtverfahrens. Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens finden sich in Kapitel 8.

Unter „3. Optimierung der Effizienz des Verfahrens“ sind außerdem die Prüfung der Verwendbarkeit von Indizes als Aufgreifkriterium für die Qualitätsbewertung von Leistungserbringern innerhalb eines Leistungsbereichs sowie Optionen für leistungsbereichübergreifende Indikatorensets beauftragt. Unter Indizes versteht das IQTIG das Ergebnis einer gewichteten Aggregation von Variablen, in diesem Fall Qualitätsindikatoren. Dazu erfolgt in Abschnitt 2.2 eine Diskussion des Verständnisses von Qualitätsindikatoren als Aufgreifkriterien gegenüber Qualitätsindikatoren als statistische Kenngrößen für Qualität. Die Empfehlungen zu diesen Punkten der Beauftragung finden sich in den Abschnitten 11.3 und 11.4.

Darüber hinaus wurde die methodische Weiterentwicklung aller Indikatoren mit beauftragt. Erste Empfehlungen, wie eine solche Weiterentwicklung durchgeführt werden sollte, finden sich in Abschnitt 11.1.

Unter Punkt 4. des Auftrags wird die Entwicklung eines Rahmenkonzepts beauftragt, das die Elemente des Strukturierten Dialogs weiterentwickelt und mit anderen relevanten Richtlinien abstimmt. Das IQTIG versteht diesen Punkt der Beauftragung als den konzeptuellen Rahmen, in dem auf Basis des Grundverständnisses von Qualitätsmessungen und deren Gütekriterien die zentralen Festlegungen zusammengefasst werden, aus denen sich die konkreten Empfehlungen ableiten lassen. Wie in Abschnitt 1.2 dargestellt, wurden aufgrund des Auslaufens der QSKH-RL und der außer Kraft gesetzten Qesü-RL ein richtlinienunabhängiges Rahmenkonzept und Empfehlungen entwickelt, die sich an den Begrifflichkeiten der DeQS-RL orientieren und damit auch die vertragsärztliche Versorgung betreffen. Die Eckpunkte dieses Rahmenkonzepts finden sich in Kapitel 4.

Die sich an die Qualitätsbewertung anschließenden Maßnahmen zur Qualitätsverbesserung sind nach Verständnis des IQTIG nicht Teil der Beauftragung. Weder in den Zielen der Beauftragung noch in den Beauftragungspunkten 1 bis 4 wird eine Weiterentwicklung der Qualitätsförderungskomponente des Strukturierten Dialogs angesprochen. Da die Steigerung der Versorgungsqualität allerdings ein zentrales Ziel der gesetzlichen Qualitätssicherung darstellt, legt das IQTIG auch erste Empfehlungen für das Einleiten von qualitätsfördernden Maßnahmen vor. Dies beinhaltet Empfehlungen für Kriterien zum Abschluss und zum Inhalt von Zielvereinbarungen sowie Instrumente zur mittel- bis längerfristigen Überprüfung der Zielerreichung. Da qualitätsfördernde Maßnahmen an die identifizierten Verbesserungspotenzialen der betroffenen Leistungserbringer individuell angepasst werden sollten, sind diese jedoch schwieriger zu standardisieren als Qualitätsmessungen. Da individuell gestaltete qualitätsverbessernde Maßnahmen erfahrungsgemäß zeit- und ressourcenintensiv sind, versteht das IQTIG den Auftrag auch dahingehend, dass durch eine Verschlankung des Vorgehens bei der Qualitätsbewertung mehr Ressourcen für anschließende Maßnahmen der Qualitätsverbesserung zur Verfügung stehen können. Diese Empfehlungen für qualitätsverbessernde Maßnahmen finden sich in Kapitel 7.

Obwohl nach § 9 Abs. 7 QSKH-RL auch bei Auffälligkeiten in der Datenvalidierung der Strukturierte Dialog eingeleitet werden soll, sieht die Beauftragung nicht explizit das Erarbeiten eines Konzepts zur Bewertung der Daten- und Dokumentationsqualität vor. Nach Verständnis des IQTIG ist dies damit nicht Teil des Auftrags. Die Weiterentwicklung des Vorgehens bei Auffälligkeiten in der Datenvalidierung im Rahmen der DeQS-RL wird jedoch empfohlen.

#### **1.4 Einbeziehung der auf Landesebene zuständigen LAG- und LQS-Vertreterinnen und -Vertreter**

Die Beauftragung verlangt auch, die auf Landesebene mit der Durchführung des Strukturierten Dialogs nach QSKH-RL beauftragten Stellen (LQS) und soweit wie möglich die Landesarbeitsgemeinschaften (LAG) in die Weiterentwicklung einzubeziehen. Dies wurde durch einen Workshop mit den Vertreterinnen und Vertretern der oben genannten Stellen umgesetzt<sup>7</sup>.

Der Workshop fand am 27. Mai 2019 in den Räumlichkeiten des IQTIG statt. Ziel des Workshops war es, die Eckpunkte des Konzepts vorzustellen und die Expertise der zuständigen Teilnehmerinnen und Teilnehmer in der Durchführung des Strukturierten Dialogs sowie deren Rückmeldungen zum Konzept einzuholen, um diese in die weitere Bearbeitung von Stufe 2 der Beauftragung einfließen zu lassen. Der Workshop wurde sieben Wochen vor Beginn allen LQS- und LAG-Vertreterinnen und -Vertretern angekündigt. Insgesamt nahmen 19 LAG- und LQS-Vertreterinnen und -vertreter aus 14 Bundesländern sowie die beim IQTIG für die direkten Verfahren zuständigen Personen am Workshop teil. Zur Vorbereitung auf den Workshop wurden allen Teilnehmerinnen und Teilnehmer zwei Wochen vor dem Workshop eine schriftliche Zusammenfassung des Rahmenkonzepts und die Tagesordnung zugesandt.

Im Workshop präsentierte das IQTIG das Rahmenkonzept und diskutierte mit den Teilnehmerinnen und Teilnehmern mögliche Herausforderungen sowie Optionen zur Verkürzung des Verfahrens. Der Verlauf und die Ergebnisse des Workshops wurden in einem Protokoll festgehalten und die Ergebnisse für die Weiterentwicklung berücksichtigt (siehe Anhang, Kapitel 1).

Das IQTIG präsentierte das Rahmenkonzept und die Ergebnisse des Workshops unter Beteiligung der LAG-Vertreterinnen und -Vertreter auch im darauffolgenden LQS-IQTIG-Treffen am 18. Juni 2019 (siehe Anhang, Kapitel 2).

#### **1.5 Einbeziehung weiterer externer Expertise**

Gemäß der in den Methodischen Grundlagen V1.1 festgehaltenen Vorgehensweise zur „Einbeziehung externen Wissens und externer Erfahrung“ wurde zusätzliche externe Expertise für die Entwicklung der statistischen Methoden (Kapitel 5) herangezogen. Als Fachexperte für statistische Prozesskontrolle wurde Prof. Dr. Sven Knoth vom Lehrstuhl für Rechnergestützte Statistik der Helmut-Schmidt-Universität, Hamburg, konsultiert. Im Rahmen eines eintägigen Workshops

---

<sup>7</sup> In den unterschiedlichen Kapiteln wird auf die seitens der Ländervertreterinnen und -vertreter genannten Herausforderungen (Ergebnisse des Workshops) Bezug genommen.

wurden zusammen mit Prof. Dr. Sven Knoth die unter Punkt 3.a des Auftrags genannten Fragestellungen diskutiert und statistische Lösungsansätze evaluiert. Als Ergebnis dieser Beratung stehen die Empfehlungen in Abschnitt 5.6.

## 1.6 Stellungnahmeverfahren zum Vorbericht

Gemäß § 137a Abs. 7 SGB V sind bei Entwicklungsarbeiten des IQTIG verschiedene Organisationen, Verbände und Interessenvertretungen des Gesundheitswesens zu beteiligen. Um dieser gesetzlichen Pflicht nachzukommen, führte das IQTIG ein Stellungnahmeverfahren für den gemeinsamen Abschlussbericht zu Stufe 1 und Stufe 2 der Beauftragung durch. Die folgenden Organisationen, Verbände und Interessenvertretungen erhielten vom 10. Oktober 2019 bis zum 21. November 2019 die Gelegenheit, schriftlich Stellung zu den Inhalten dieses Berichts zu nehmen:

- die Kassenärztliche Bundesvereinigung
- die Deutsche Krankenhausgesellschaft
- der Spitzenverband Bund der Krankenkassen
- der Verband der Privaten Krankenversicherung
- die Bundesärztekammer
- die Bundeszahnärztekammer
- die Bundespsychotherapeutenkammer
- die Berufsorganisationen der Krankenpflegeberufe
- die wissenschaftlichen medizinischen Fachgesellschaften
- das Deutsche Netzwerk Versorgungsforschung
- die für die Wahrnehmung der Interessen der Patientinnen und Patienten und der Selbsthilfe chronisch kranker und behinderter Menschen maßgeblichen Organisationen auf Bundesebene
- der oder die Beauftragte der Bundesregierung für die Belange der Patientinnen und Patienten
- zwei von der Gesundheitsministerkonferenz der Länder zu bestimmende Vertreter
- die Bundesoberbehörden im Geschäftsbereich des Bundesministeriums für Gesundheit, soweit ihre Aufgabenbereiche berührt sind
- die mit der Durchführung der QS-Verfahren gemäß QSKH-RL sowie DeQS-RL beauftragten Stellen auf Landesebene

Die eingegangenen Stellungnahmen werden vom IQTIG schriftlich gewürdigt und der Bericht wurde auf Basis der Stellungnahmen auf Anpassungsbedarf hin geprüft. Der vorliegende Abschlussbericht beinhaltet alle Anpassungen, die sich aus den Stellungnahmen zu dem Vorbericht ergeben haben.

## 2 Analyse der Ausgangssituation

Um die Weiterentwicklung des Strukturierten Dialogs an den Zielen der Qualitätssicherung auszurichten, wird im vorliegenden Kapitel eine Ist-Analyse vorgenommen, in der die aktuelle Vorgehensweise gemäß QSKH-RL beschrieben und methodisch eingeordnet wird. Obwohl sich die Empfehlungen in den folgenden Kapiteln auf ein richtlinienunabhängiges Stellungnahmeverfahren beziehen (vgl. Abschnitt 1.3) und diese aufgrund des Auslaufens der QSKH-RL vor allem mit Blick auf das in § 17 DeQS-RL beschriebene Stellungnahmeverfahren entwickelt wurden, wird eine Analyse des Strukturierten Dialogs gemäß QSKH-RL als sinnvoll erachtet. Einerseits finden sich die in Abschnitt 2.1 identifizierte Funktionen des Strukturierten Dialogs auch in ähnlicher Form in dem Stellungnahmeverfahren der DeQS-RL. Andererseits lässt § 17 DeQS-RL ähnlich wie die QSKH-RL Spielräume in der Ausgestaltung dieses Stellungnahmeverfahrens, welche potenzielle Quellen für Heterogenität in der Umsetzung der Richtlinie darstellen.

### 2.1 Funktionen des Strukturierten Dialogs

In den §§ 11 bis 14 QSKH-RL sind die Regelungen zum Strukturierten Dialog mit Krankenhäusern festgehalten. Allerdings beinhaltet die QSKH-RL keine Definition des Begriffs „Strukturierter Dialog“. Im Beschluss zur Beauftragung des IQTIG wird der Strukturierte Dialog beschrieben als Verfahren, „in dem unter Einbeziehung von Experten auffällige Ergebnisse von Einrichtungen bewertet und erforderlichenfalls geeignete Maßnahmen zur Qualitätsverbesserung und -förderung eingeleitet werden“ (G-BA 2018a). Zudem sieht die QSKH-RL vor, dass im Rahmen des Strukturierten Dialogs auch eine Überprüfung der Validität der Dokumentation erfolgt (§ 12 Abs. 1 QSKH-RL). Aus den §§ 11 bis 14 sowie § 9 Abs. 7 QSKH-RL lassen sich für den Begriff „Strukturierter Dialog“ mehrere Prozesse mit unterschiedlichen Funktionen und Zielen ableiten und zusammenfassen:

- **Bewertung der Versorgungsqualität**

Der Strukturierte Dialog dient bisher dazu, anhand der Ergebnisse von Qualitätsindikatoren und schriftlichen Stellungnahmen der Leistungserbringer sowie aufgrund von Begehungen und kollegialen Gesprächen die Versorgungsqualität der Leistungserbringer zu bewerten.

- **Bewertung der Dokumentationsqualität**

Außerdem erfolgt eine Beurteilung der Dokumentationsqualität im Rahmen des Datenvalidierungsverfahrens und der Prüfung der schriftlichen Stellungnahmen hinsichtlich angegebener Dokumentationsfehler. Der Strukturierte Dialog erfüllt demnach auch die Funktion der Bewertung der Dokumentationsqualität der einzelnen Leistungserbringer.

- **Qualitätsförderung**

Die Steigerung der Qualität von Krankenhausleistungen ist darüber hinaus ebenso ein Ziel des Strukturierten Dialogs. Zu diesem Zweck kennt die QSKH-RL Maßnahmen, die zur Steigerung der Versorgungsqualität eingeleitet werden können, wie etwa Zielvereinbarungen.

## Sensitivität und Spezifität

Neben diesen explizit in der QSKH-RL festgehaltenen Funktionen hat der Strukturierte Dialog auch die implizite Funktion, die Spezifität der Qualitätsmessung zu steigern. Zur Illustration wird diese Funktion anhand eines Qualitätsindikators und unter folgenden Annahmen erläutert:

- Jeder Leistungserbringer weist bezogen auf das Qualitätsziel eines Indikators eine zugrundeliegende „tatsächliche“ Qualität auf.
- Das Vorgehen besteht aus einem ersten indikatorbasierten Prüfschritt und einem nachgelagerten zweiten Prüfschritt (Strukturierter Dialog).
- Das Gesamtverfahren bestehend aus Prüfschritt 1 und 2 attestiert am Ende entweder unauffällige oder auffällige Qualität bezogen auf das Qualitätsziel des betreffenden Indikators.

Daraus ergeben sich am Ende des Gesamtverfahrens die in Tabelle 1 dargestellten Möglichkeiten von richtigen und falschen Klassifikationen. Als Sensitivität des Vorgehens wird dabei der Anteil der mit „auffällige Qualität“ bewerteten Leistungserbringerergebnisse an allen Leistungserbringerergebnissen mit „tatsächlich“ unzureichender Qualität bezeichnet. Als Spezifität des Vorgehens wird der Anteil der mit „unauffällige Qualität“ bewerteten Leistungserbringerergebnissen an allen Leistungserbringerergebnissen mit „tatsächlich“ zureichender Qualität bezeichnet. Diese Betrachtung ist allerdings hypothetisch, da einerseits kein Goldstandard für die Detektion „tatsächlicher“ Qualitätsdefizite vorliegt und andererseits eine Definition „tatsächlicher“ Qualität nur schwer möglich ist.<sup>8</sup>

Tabelle 1: Sensitivität und Spezifität des Gesamtverfahrens zur Qualitätsmessung bezogen auf einen Qualitätsindikator

		„tatsächliche“ Qualität	
		unzureichend	zureichend
Qualitätsbewertung	auffällig	richtig positiv	falsch positiv
	unauffällig	falsch negativ	richtig negativ

Ein auffälliges Indikatorergebnis (Ergebnis erster Prüfschritt) kann im Rahmen des Strukturierten Dialogs (zweiter Prüfschritt) zwar widerlegt werden, bei einem unauffälligen Indikatorergebnis kann dieses jedoch nicht widerlegt werden, da nur auffällige Ergebnisse einem zweiten Prüfschritt unterzogen werden. In Tabelle 1 entspricht dies einer Verschiebung einiger Ergebnisse des ersten Prüfschritts von den „falsch positiven“ hin zu den „richtig negativen“

<sup>8</sup> Dahinter steht die Frage, ob es eine von der konkreten Messung unabhängige, „tatsächliche“ Qualität der Versorgung theoretisch geben kann, oder ob Qualität nur mit Bezug zu der jeweiligen Messung definiert werden kann.

durch den zweiten Prüfschritt. Daher wird durch den Strukturierten Dialog die Spezifität des Vorgehens erhöht, im Vergleich zu einem Vorgehen ohne Strukturierten Dialog. Die Sensitivität des Gesamtverfahrens wird dagegen maßgeblich durch den ersten Prüfschritt bestimmt.

Eine Steigerung der Spezifität durch einen zweiten Prüfschritt ohne Steigerung der Sensitivität ist dann sinnvoll, wenn der erste Prüfschritt eine hohe Sensitivität aber eine vergleichsweise niedrigere Spezifität aufweist. Bei der Konstruktion von Qualitätsindikatoren kann die Vielfältigkeit der Gesundheitsversorgung zu unterschiedlichem Ausmaß berücksichtigt werden. Je stärker diese Vielfältigkeit schon bei der Definition von Zähler, Nenner sowie Ein- und Ausschlusskriterien des Indikators berücksichtigt wird, desto höher ist die Spezifität des Qualitätsindikators im obigen Sinn.

Damit einher geht jedoch auch eine Steigerung in der Komplexität des Indikators, was häufig auch größeren Erhebungsaufwand bedeutet. Dies wird am Beispiel eines Qualitätsindikators „Thromboseprophylaxe“ illustriert, der den Anteil von Patientinnen und Patienten mit Thromboseprophylaxe an allen operierten Patientinnen und Patienten abbildet. Als Ausschlusskriterium könnte die Einnahme von Medikamenten zur Blutverdünnung entweder direkt miterfasst werden, oder erst im zweiten Prüfschritt für die dokumentierten Fälle überprüft werden. Wird dieses Ausschlusskriterium erst im zweiten Prüfschritt berücksichtigt, ist die Klassifikation von Leistungserbringern in diejenigen mit unauffälliger und diejenigen mit auffälliger Qualität im ersten indikatorbasierten Schritt weniger spezifisch im obigen Sinn als wenn dies bereits durch den Indikator berücksichtigt würde. Eine Steigerung der Spezifität des Vorgehens kann also aufgrund einer als zu niedrig empfundenen Spezifität der indikatorbasierten Klassifikation als erforderlich erachtet werden. Anhand dieses Beispiels wird auch deutlich, dass eine Aufwands- und Spezifitätsverschiebung vom ersten Prüfschritt hin zu dem zweiten erfolgt, sofern das Ausschlusskriterium erst im zweiten Prüfschritt berücksichtigt wird. Indikatoren, die wie in diesem Beispiel weniger „spezifisch“ sind, eignen sich damit eher als „Aufgreifkriterien der Versorgungsqualität“ für einen darauffolgenden Prüfschritt und weniger als abschließende Qualitätsmessungen. Denn wie im Folgenden erläutert wird, unterscheiden sich die methodischen Anforderungen an Qualitätsmessungen je nach Verwendungszweck (IQTIG 2019a, S. 38).

## **2.2 Anforderungen an die Qualitätsmessung und -bewertung**

Die Anforderungen, denen die Prozesse und Ergebnisse der gesetzlich verpflichtenden Qualitätssicherung genügen müssen, haben sich von ihren Anfängen bis heute deutlich gesteigert. Während in den Anfängen der Qualitätssicherungsbemühungen im Gesundheitswesen das Lernen von den Besseren in einem geschützten Raum im Vordergrund stand, wurden durch verschiedene Gesetzgebungen mehr und mehr die Transparenz, öffentliche Berichterstattung und Steuerungsfunktion von Versorgungsqualität in den Vordergrund gerückt (siehe auch I. Hintergrund zur Beauftragung im Beschluss zur Beauftragung).

So wurde im Jahr 2014 mit dem GKV-Finanzstruktur- und Qualitäts-Weiterentwicklungsgesetz (GKV-FQWG) beschlossen, dass eine leistungserbringervergleichende Übersicht über die Qualität der stationären Versorgung im Internet veröffentlicht werden soll. Mit dem im darauffolgenden Jahr beschlossenen Krankenhausstrukturgesetz (KHSG) wurde geregelt, dass Ergebnisse

zur Struktur-, Prozess- und Ergebnisqualität der stationären Gesundheitsversorgung von den Landesplanungsbehörden für Zwecke der qualitätsorientierten Krankenhausplanung Verwendung finden sollen. Es findet demnach eine Erweiterung der Verwendung der gesetzlich verpflichtenden Qualitätsmessungen statt, die neben *improvement* (z. B. Peer-Reviews, eigenständige qualitätsorientierte Umgestaltung der Versorgung durch die Leistungserbringer) nun auch im Rahmen von *accountability* (z. B. öffentliche Berichterstattung, Planungsentscheidungen) eingesetzt werden (Solberg et al. 1997, Berwick et al. 2003). Sowohl der Einsatz für *improvement* als auch für *accountability* sind Strategien, die das gleiche Ziel einer besseren Gesundheitsversorgung verfolgen, jedoch über andere Wirkmechanismen (Berwick et al. 2003). So gehen etwa Berwick et al. davon aus, dass durch die öffentliche Verfügbarkeit aussagekräftiger, leistungserbringerbezogener Qualitätsinformationen sowohl Patientinnen und Patienten als auch andere Entscheidungsträger (z. B. Landesplanungsbehörden) eine Selektion von Leistungserbringern mit höherer Qualität vornehmen. Dies wiederum sollte Leistungserbringern mit niedrigerer Qualität einen Anreiz geben, durch eigene Anstrengungen die Qualität ihrer Versorgung zu steigern. Auf der anderen Seite gehen Berwick et al. davon aus, dass durch die Rückmeldung von Qualitätsergebnissen an die Leistungserbringer selbst, diese weitergehende Einblicke in ihre Versorgungsqualität erhalten und damit in die Position versetzt werden, die Versorgung so umzugestalten, dass sie eine höhere Qualität erreicht (*improvement*). Das bisherige Vorgehen des Strukturierten Dialog kann damit eher letzterer Strategie zugeordnet werden, während die neueren Verwendungszwecke der Qualitätsmessungen der gesetzlich verpflichtenden Qualitätssicherung eher die erstgenannte Strategie verfolgen.

Berücksichtigt man diese Erweiterung der Verwendung von Qualitätsindikatoren für öffentliche Berichterstattung oder um Planungsentscheidungen (*accountability*), steigen auch die methodischen Anforderungen an Qualitätsindikatoren (Solberg et al. 1997, Chassin et al. 2010, Berwick et al. 2003, Gardner et al. 2018, Freeman 2002). Um im geschützten Raum in einen Dialog über Verbesserungsmöglichkeiten der Versorgungsqualität zu führen, genügen „Aufgreifkriterien“. Um jedoch die Ergebnisse von indikatorbasierten Qualitätsmessungen für beispielsweise Planungs- und Vergütungsentscheidungen oder Patienteninformationen zu verwenden, müssen die Indikatorergebnisse aussagekräftiger sein. Sollen nämlich beispielsweise quantitative Indikatorergebnisse miteinander verglichen werden, müssen numerische Unterschiede zwischen Leistungserbringerergebnissen auch mindestens tendenziell Unterschiede<sup>9</sup> in der Qualität der Versorgung widerspiegeln.<sup>10</sup> Es genügt damit nicht mehr, diese als bloße Aufgreifkriterien der Versorgungsqualität zu verstehen. Stattdessen müssen sie auf Basis hoher Datenqualität zuverlässige Rückschlüsse auf die Versorgungsqualität zulassen. Entsprechend dieser gestiegenen Anforderungen an die datengestützte Qualitätssicherung ist ein Wandel im Verständnis von Qualitätsindikatoren in der gesetzlich verpflichtenden Qualitätssicherung erforderlich, weg von Aufgreifkriterien hin zu quantitativen Größen für Qualität (vgl. Solberg et al. 1997, IQTIG 2019a,

---

<sup>9</sup> Hier muss auch die Frage nach der Rolle statistischer Unsicherheit beantwortet werden.

<sup>10</sup> Dies kann auch als Frage des Skalenniveaus der Messung verstanden werden: Während Aufgreifkriterien eher Nominalskalenniveau (auffällig vs. unauffällig) zugeschrieben wird, setzt der Vergleich zweier Messwerte auf einem Indikator metrisches Skalenniveau voraus.

Carter 1989). Dies bedeutet nicht, dass Qualitätsmessungen grundsätzlich nicht als Aufgreifkriterien verwendet werden sollten, sondern dass deren Verwendung im Rahmen von *accountability* höhere methodische Anforderungen an die Messung stellt als im Rahmen von *improvement*.

Der bisherige Strukturierte Dialog mit Krankenhäusern im Rahmen der QSKH-RL weist auch Aspekte eines Peer-Review Verfahrens auf, in dem „auffällige Ergebnisse“ zunächst in einem zweiten Prüfschritt im Dialog mit dem Leistungserbringer „geklärt werden“. Diese sprachliche Trennung von quantitativem Messergebnis („rechnerisches Ergebnis“) und Qualitätsbewertung im Rahmen des dialogischen Klärungsprozesses in der QSKH-RL könnte als implizites Verständnis von Qualitätsindikatoren als Aufgreifkriterien verstanden werden, deren numerischer Wert eher geringere Aussagekraft hat und eher die Funktion erfüllt, einen darauffolgenden Prüfschritt auszulösen.

Andererseits werden schon heute die quantitativen Indikatorergebnisse in den Strukturierten Qualitätsberichten der Krankenhäuser veröffentlicht. Dies bedeutet, dass den Indikatorergebnissen selbst auch eine Funktion (und Information) über die als Aufgreifkriterien für den Strukturierten Dialog hinaus zugeschrieben wird. Im Rahmen der Beauftragung zur Veröffentlichung von leistungserbringervergleichenden Übersichten über die Qualität in der stationären Versorgung (sog. G-BA-Qualitätsportal), soll das IQTIG unter anderem auch die Frage beantworten, ob metrische Ergebnisse mehrerer Indikatoren verdichtet werden sollen. Eine solche Verdichtung oder Aggregation von Indikatorergebnissen setzt voraus, dass die Indikatorergebnisse mehr als bloße Aufgreifkriterien darstellen und ihre numerischen Ergebnisse selbst eine Aussagekraft haben.

Des Weiteren ergibt sich ein Zielkonflikt zwischen dem bisherigen, ressourcenintensiven dialogischen Vorgehen im zweiten Schritt und dem Wunsch nach einer „Verschlankung des Verfahrens“ (vgl. Beauftragung). Zukünftig wird die Anzahl an Qualitätsindikatoren eher steigen, da weitere Leistungsbereiche einschließlich Patientenbefragungen für die QS erschlossen werden. Wie in Abschnitt 2.1 erläutert, geht mit der Verlagerung von Spezifität von der indikatorbasierten Klassifikation hin zum nachgelagerten Prüfschritt auch eine Aufwandsverlagerung einher. So bedeutet eine größere Anzahl an Indikatoren, die nach dem hier beschriebenen Verständnis eher als Aufgreifkriterien mit eher niedriger Spezifität konstruiert sind, auch einen erhöhten Aufwand im nachgelagerten Prüfschritt.

In der Zusammenschau der gestiegenen (methodischen) Anforderungen an die gesetzliche Qualitätssicherung und dem Wunsch nach einem ressourcenschonenderen Vorgehen ergibt sich die Notwendigkeit, den Prozess der Qualitätsmessung stärker, auch mittels automatisierbarer statistischer Methoden, zu standardisieren und die aufwendigen dialogischen Anteile des Verfahrens weg von der Qualitätsbewertung hin zur Qualitätsförderung zu verlagern. Dies bedeutet jedoch nicht, dass Maßnahmen zur Steigerung der Versorgungsqualität (*improvement*) weniger wichtig werden, nur weil die Ergebnisse der Qualitätsmessungen zusätzlich im Kontext von *accountability* Verwendung finden. Wie später (siehe Kapitel 1) erläutert wird, sollten diese Maßnahmen jedoch von den Prozessen der Qualitätsmessung und -bewertung aus mehreren Gründen getrennt werden und ihren sinnvollen Einsatz bei der Qualitätsförderung (*improvement*) finden.

### 2.3 Strukturelle Voraussetzungen des Strukturierten Dialogs

Im Folgenden wird zunächst der Ist-Zustand der strukturellen Voraussetzungen für die Durchführung des Strukturierten Dialogs beschrieben. Der bisherige Strukturierte Dialog findet zwischen dem jeweiligen Leistungserbringer (Fachabteilung eines Krankenhausstandorts und dem abteilungsübergreifenden internen Qualitätsmanagement der Einrichtung) und der jeweiligen beauftragten Stelle (verantwortliche Einrichtungen der Landesebene bzw. dem IQTIG auf Bundesebene) statt.

Die beauftragten Stellen auf Bundes- und Landesebene richten zusätzlich Fachgruppen<sup>11</sup> mit fachkundigen ehrenamtlichen Expertinnen und Experten ein, um die Bewertung der Qualitätsindikatorergebnisse vorzunehmen und um Qualitätssicherungsmaßnahmen anzustoßen. Laut QSKH-RL werden die fachkundigen Mitglieder unter anderem von den Trägerorganisationen des G-BA entsendet. In Bezug auf die Qualifikation der Mitglieder werden ausschließlich Kenntnisse aus dem Bereich Qualitätsmanagement und der Qualitätssicherung gefordert. Auf welchem Gebiet die Mitglieder darüber hinaus fachkundig sein sollten, ist nicht näher beschrieben. Folglich ist unklar, über welche genaue Qualifikation die einzelnen Mitglieder verfügen sollen. Bundesweit existieren derzeit insgesamt 114 auf Landesebene und durchschnittlich 7 Fachgruppen pro beauftragte Stelle, die für die Umsetzung der Maßnahmen der externen stationären Qualitätssicherung zuständig sind (siehe Anhang, Kapitel 3, Tabelle 1). Die Anzahl der Mitglieder pro Fachgruppe variiert stark sowohl zwischen den Bundesländern als auch zwischen den Fachgruppen. Laut einer Befragung des G-BA aus dem Jahr 2012, an der alle Landesgeschäftsstellen für Qualitätssicherung sowie die Institution nach § 137a SGB V teilnahmen, schwankt die Anzahl der Fachgruppenmitglieder je nach Bundesland zwischen 3 und 18 Mitgliedern. Innerhalb einer Fachgruppe sind durchschnittlich 7 Mitglieder vertreten [(siehe Anhang, Kapitel 3, Tabelle 1; Winkler-Komp et al. (2014)].

Bei allen indirekten Verfahren existieren zusätzlich zu den Fachgruppen auf Landesebene auch Fachgruppen auf Bundesebene, die jedoch andere Aufgaben wahrnehmen und daher hier nicht thematisiert werden.

Auch das IQTIG richtet verschiedene Fachgruppen für die Umsetzung des Strukturierten Dialogs für die direkten Verfahren ein.

### 2.4 Durchführung des bisherigen Strukturierten Dialogs

Die Vorgaben in der QSKH-RL legen lediglich Eckpunkte des Ablaufs des Strukturierten Dialogs fest, sodass die konkrete Ausgestaltung der oben genannten drei Funktionen und hinsichtlich des Zeitraums, in welchem die Leistungserbringer eine Stellungnahme abzugeben haben, ein großer Gestaltungsspielraum existiert.

---

<sup>11</sup> Im Folgenden wird die Bezeichnung Fachgruppe aus der QSKH-RL verwendet, da sich die Analyse der Ausgangssituation auf den Strukturierten Dialog im Rahmen der QSKH-RL bezieht. Neben dem Begriff Fachgruppe existieren z. B. auf Ebene der Länder oder in anderen Richtlinien wie der DeQS-RL auch andere Bezeichnungen wie z. B. Arbeitsgruppe, Fachbereich, Fachkommission, Fachausschüsse usw. Ab Kapitel 4 wird für den Begriff Fachgruppe der Begriff Fachkommission gemäß DeQS-RL verwendet.

Abbildung 1 veranschaulicht schematisch den Ablauf des Strukturierten Dialogs. Weicht das auf Grundlage der übermittelten Daten für einen Krankenhausstandort errechnete Ergebnis vom bundesweit einheitlich festgelegten Referenzbereich ab (*rechnerische Auffälligkeit*), ist der Leistungserbringer darauf hinzuweisen oder es ist eine Stellungnahme vom Leistungserbringer einzuholen (§ 11 QSKH-RL). Auf einen Hinweis an den Leistungserbringer oder die Einholung einer Stellungnahme kann laut § 10 QSKH-RL verzichtet werden, wenn für den Standort nur ein Fall im Qualitätsindikator auftrat. Werden bei einem rechnerisch auffälligen Ergebnis keine Maßnahmen eingeleitet, ist der Grund dafür darzustellen und anzugeben (z. B. bei der Verwendung von Qualitätsindikatorensets oder von landesindividuellen Referenz- und Vertrauensbereichen, § 10 QSKH-RL).

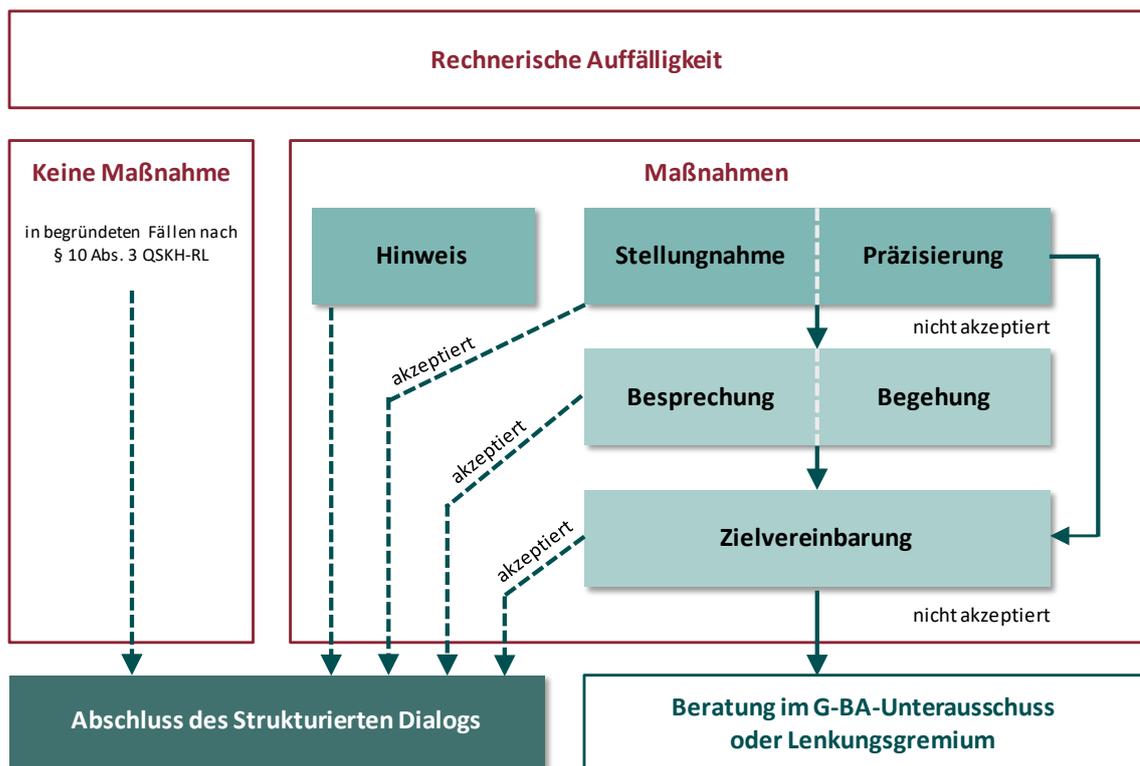


Abbildung 1: Schematische Darstellung von Maßnahmen im Rahmen des Strukturierten Dialogs

Welche Form oder welchen Inhalt eine Stellungnahme haben soll, ist nicht definiert. In § 12 QSKH-RL ist weiterhin geregelt, dass die Stellungnahme dahingehend geprüft werden soll, ob die dokumentierten Leistungen in der fachlich gebotenen Qualität erbracht wurden. Wie genau und nach welchen Kriterien diese Prüfung stattfinden soll, bleibt offen. In der Praxis werden mit unbekannter Häufigkeit Präzisierungsanfragen zu der ursprünglichen Stellungnahme des Leistungserbringers gestellt.

Am Ende des Strukturierten Dialogs werden alle auffälligen Ergebnisse in das bundeseinheitliche Einstufungsschema (siehe Anhang, Kapitel 3, Tabelle 2) eingeordnet. Es beinhaltet 7 Kategorien (AQUA 2013: 55). Eine Ausdifferenzierung mit diversen Ziffern führt in der Summe zu insgesamt

21 unterschiedlichen Kategorien.<sup>12</sup> Die Kategorien „N“ und „R“ des Einstufungsschemas beziehen sich auf rechnerisch unauffällige Ergebnisse, bei welchen kein Strukturiertes Dialog stattfindet. Daher wird auf diese Kategorien im Folgenden nicht weiter eingegangen.

#### 2.4.1 Einstieg in den Strukturierten Dialog

Der Strukturierte Dialog kann durch zwei Arten von rechnerisch auffälligen Ergebnissen ausgelöst werden.<sup>13</sup> Zum einen kann ein Leistungserbringer ein rechnerisch auffälliges Ergebnis in einem *Qualitätsindikator* aufweisen. Zum anderen führt auch ein rechnerisch auffälliges Ergebnis in einem *Auffälligkeitskriterium* zum Einstieg in den Strukturierten Dialog. Auffälligkeitskriterien sind Kennzahlen zur Messung der Dokumentationsqualität. Sie prüfen die Plausibilität, Vollständigkeit und Vollzähligkeit der Dokumentation zum Zwecke der Qualitätssicherung. Zur Bewertung dieser Auffälligkeitskriterien existieren im Kategorienschema (vgl. Anhang, Kapitel 3, Tabelle 2) zwei Ziffern (*U30* und *A40*, grün hinterlegt), die exklusiv im Kontext der Datenvalidierung zu verwenden sind.

In Abhängigkeit davon, ob sich das rechnerisch auffällige Ergebnis auf einen Qualitätsindikator oder auf ein Auffälligkeitskriterium bezieht, hat auch die Bewertung des Ergebnisses einen anderen Fokus. Handelt es sich um ein rechnerisch auffälliges Ergebnis in einem Qualitätsindikator, liegt der Fokus auf der Bewertung der Versorgungsqualität. Handelt es sich um ein rechnerisch auffälliges Ergebnis in einem Auffälligkeitskriterium, liegt der Fokus auf der Bewertung der Dokumentationsqualität. Wie bereits in Abschnitt 1.3 beschrieben liegt der Fokus dieses Berichts auf der Bewertung der Versorgungsqualität, daher wird im Folgenden auf die Bewertung der Dokumentationsqualität nicht weiter eingegangen.

Wird ein rechnerisch auffälliges Ergebnis in einen Qualitätsindikator festgestellt, hat die mit der Durchführung des Strukturierten Dialogs beauftragte Stelle (LQS oder IQTIG) gemäß QSKH-RL zwei Handlungsoptionen: Das Krankenhaus ist „unter Beschreibung des Sachverhalts und Bezeichnung des betroffenen Standorts auf die Auffälligkeit hinzuweisen oder [es ist] eine Stellungnahme innerhalb einer festzusetzenden Frist anzufordern“ (§ 11 QSKH-RL). Entscheidet sich die mit der Durchführung des Strukturierten Dialogs beauftragte Stelle (LQS oder IQTIG), das Krankenhaus zu einer schriftlichen Stellungnahme aufzufordern, hat der Leistungserbringer die Gelegenheit, in der schriftlichen Stellungnahme darzulegen, warum die Leistungen trotz rechnerisch auffälligem Ergebnis in der fachlich gebotenen Qualität erbracht worden sind (§ 12 Abs. 1 QSKH-RL).

#### 2.4.2 Bewertung der Versorgungsqualität

Ob bei einem rechnerisch auffälligen Qualitätsindikatorergebnis zunächst nur ein Hinweis versendet oder eine schriftliche Stellungnahme angefordert wird, ist entscheidend für die Quali-

---

<sup>12</sup> Die resultierende Bezifferung (z. B. *A41: Hinweise auf Struktur- und Prozessmängel*) impliziert auch die verpflichtende Darstellung für die Veröffentlichung der Ergebnisse in den jährlichen Qualitätsberichten der Krankenhäuser.

<sup>13</sup> Unabhängig von einem rechnerisch auffälligen Ergebnis kann der Strukturierte Dialog im Rahmen der Zweiterfassung der Datenvalidierung eingeleitet werden.

tätsbeurteilung. Eine Qualitätsbeurteilung der Leistungserbringung wird nur nach Einholung einer schriftlichen Stellungnahme vorgenommen. Im Fall der Versendung eines Hinweises erfolgt keine Beurteilung des rechnerisch auffälligen Indikatorergebnisses. Bisher existieren keine Regularien dafür, anhand welcher Kriterien diese erste wichtige Entscheidung zu treffen ist.

Eine im Jahr 2012 stattgefundene Befragung des G-BA der Landesgeschäftsstellen für Qualitätssicherung gibt Auskunft darüber, auf welcher Grundlage die Landesebene entscheiden. Die Befragungsergebnisse zeigen, dass die Entscheidung der Fachgruppen, welche Maßnahme ergriffen wird, oft anhand sehr unterschiedlicher Kriterien und häufig anhand mehrerer Kriterien getroffen wird. In einigen Ländern erfolgt die Entscheidung anhand:

- der Relevanz des Indikators
- der statistischen Signifikanz von Abweichungen
- des Vorliegens einer Risikoadjustierung

Aus den Ergebnissen der Befragung des G-BA geht weder hervor, wie die Relevanz des Indikators bestimmt oder welches Signifikanzniveau gewählt wird noch inwieweit dies einheitlich zwischen den beauftragten Stellen erfolgt (Winkler-Komp et al. 2014).

Das IQTIG holt in den direkten QS-Verfahren für jedes rechnerisch auffällige Ergebnis (Qualitätsindikator und Auffälligkeitskriterium) eine schriftliche Stellungnahme ein.

### **Hinweise**

Einen Überblick über den Anteil der eingesetzten Hinweise in den einzelnen Bundesländern gibt die Tabelle 3 im Anhang, Kapitel 3. Es ist ersichtlich, dass der Anteil an versandten Hinweisen an allen rechnerisch auffälligen Ergebnissen stark zwischen den Bundesländern variiert. In Bayern und Baden-Württemberg liegt der Anteil bei weit über 50 %, während in Bremen und im Saarland kaum Hinweise versandt werden. Bundesweit wird im Durchschnitt für mehr als ein Drittel aller rechnerisch auffälligen Indikatorergebnisse ein Hinweis (Kategorie „H“) an den Leistungserbringer versendet. Durch die fehlende Qualitätsbeurteilung wird das rechnerisch auffällige Ergebnis zwar nicht als unauffällig deklariert, dennoch kann die fehlende Qualitätsbeurteilung den Eindruck erwecken, dass das rechnerisch auffällige Ergebnis unauffällig sei. Tatsächlich ist es jedoch so, dass zu rechnerisch auffälligen Ergebnissen, bei welchen ein Hinweis versandt wurde, keinerlei Aussagen zu möglichen qualitativen Versorgungsproblemen gemacht werden können. Das bedeutet, dass in vielen Bundesländern für eine große Anzahl an rechnerisch auffälligen Ergebnissen keine Qualitätsbewertung vorgenommen wird. In den direkten Verfahren wird aufgrund der kleineren Anzahl an teilnehmenden Standorten und Abteilungen jedes rechnerisch auffällige Ergebnis mindestens durch die Anforderung einer Stellungnahme überprüft.

### **Schriftliche Stellungnahmen**

Fordert die beauftragte Stelle eine schriftliche Stellungnahme vom Leistungserbringer an, wird diese dahingehend geprüft, ob die medizinischen Leistungen mit der fachlich gebotenen Qualität erbracht und valide dokumentiert wurden (§ 11 Abs. 1 QSKH-RL). Das bedeutet, dass anhand der schriftlichen Stellungnahmen auch eine Einstufung der Dokumentationsqualität stattfindet (vgl. Kategorie „D“ in Tabelle 2 im Anhang, Kapitel 3), die unabhängig von der bereits erwähnten

Bewertung der Auffälligkeitskriterien durchgeführt wird. Für die Bewertung der Versorgungsqualität ist die Einstufung der Dokumentationsqualität insofern relevant, da eine *echte* Qualitätsbewertung des Indikatorergebnisses im Hinblick auf die Qualität medizinischer Leistungen (Kategorie „A“ oder „U“) nur stattfinden kann, wenn die Angaben der Leistungserbringer korrekt sind. Wurden in der schriftlichen Stellungnahme des Leistungserbringers keine Angaben zu Dokumentationsfehlern gemacht, kann geprüft werden, ob die medizinische Leistung in der fachlich gebotenen Qualität erbracht wurde. Kann der Leistungserbringer in seiner Stellungnahme darlegen, dass sich das rechnerisch auffällige Ergebnis nicht durch ein Qualitätsdefizit begründet, kann das rechnerisch auffällige Ergebnis von der beauftragten Stelle als qualitativ unauffällig eingestuft werden (Kategorie „U“). Verbindliche Kriterien für diesen Entscheidungsprozess existieren bisher nicht.

Aus dem Bericht zum Strukturierten Dialog zum Erfassungsjahr 2015 geht hervor, dass der Anteil an schriftlichen Stellungnahmen bezogen auf alle rechnerisch auffälligen Ergebnisse bei 61,8 % lag. Das entspricht 9.797 schriftlichen Stellungnahmen deutschlandweit, die von den LQS und vom IQTIG angefordert, von den Leistungserbringern erstellt und anschließend von den beauftragten Stellen und ihren Fachgruppen bewertet wurden (IQTIG 2017a: 54). Obwohl die schriftliche Stellungnahme die Grundlage für die Bewertung der Versorgungsqualität bildet, existieren hier ebenso wenig bundeseinheitliche Vorgaben, die die Leistungserbringer bei der Erstellung der schriftlichen Stellungnahmen einzuhalten haben. Wie detailliert die schriftliche Stellungnahme ausfällt, liegt in der Verantwortung des jeweiligen Leistungserbringers und beeinflusst den Aufwand und die Bewertung der Indikatorergebnisse durch die Fachgruppe. Gehen beispielsweise Stellungnahmen mit ungenauen, unzureichenden oder gar fehlenden Angaben ein, kann die beauftragte Stelle eine Präzisierung der Stellungnahme einholen, um weitergehende Informationen zu erhalten, die dann ggf. eine abschließende Bewertung erst ermöglichen. Allerdings ist der Umgang mit Präzisierungen in den Rahmenvorgaben der Richtlinie nicht verankert. Die Einholung einer Präzisierung sowie die damit zusammenhängende erneute Prüfung und Bewertung des gesamten Sachverhalts beinhaltet einen hohen Arbeitsaufwand für die beauftragte Stelle. Wie häufig Präzisierungen eingeholt werden ist außerdem nicht bekannt und wird bisher nicht berichtet. Folglich existieren auch für Präzisierungen weder auf Bundes- noch auf Landesebene inhaltliche oder formale einheitliche Kriterien.

### **Besprechungen<sup>14</sup> und Begehungen im Rahmen der Qualitätsbewertung**

Laut QSKH-RL können auch weiterführende Maßnahmen veranlasst werden, wenn trotz Stellungnahme noch Zweifel bestehen. Bleiben also anhand der schriftlichen Stellungnahme und einer möglichen Präzisierung noch inhaltliche Fragen offen, sodass eine abschließende Bewertung des rechnerisch auffälligen Ergebnisses nicht vorgenommen werden kann, können weiterführende Maßnahmen eingeleitet werden. Weiterführende Maßnahmen, die als Grundlage für die Qualitätsbeurteilung dienen können, sind z. B. Besprechungen mit dem Leistungserbringer oder auch Vor-Ort-Begehungen. Letztere sind allerdings lediglich mit *Einverständnis des Krankenhauses* möglich (§ 12 Abs. 3 QSKH-RL). Ein bundesweit einheitliches Vorgehen zum Einsatz

---

<sup>14</sup> Die Begriffe Besprechungen und kollegiales Gespräch werden synonym verwendet.

und zur Durchführung dieser weiterführenden Maßnahmen mit anschließender Bewertung existiert nicht. Deutschlandweit fanden in allen QS-Verfahren zum Erfassungsjahr 2015 insgesamt 282 Besprechungen und 19 Begehungen statt (IQTIG 2017a: 54).

Aus den Erfahrungen des IQTIG ist zu berichten, dass diese Maßnahmen für die beauftragte Stelle, die entsprechende Fachgruppe und auch für den Leistungserbringer sehr ressourcenintensiv sind. In der Praxis wird nicht nur bei verbleibenden Zweifeln nach Sichtung der Stellungnahmen, sondern auch bei wiederholt rechnerisch auffälligen Ergebnissen eine Begehung durchgeführt. Der direkte Kontakt zu den Krankenhäusern in einem kollegialen Gespräch oder während einer Begehung ist eine effektive, allerdings auch eine ressourcenintensive Maßnahme im Strukturierten Dialog.

### **Prozess der Entscheidungsfindung**

Die Grundlage für eine abschließende Bewertung des Indikatorergebnisses bildet bei der Beurteilung von rechnerischen Auffälligkeiten in der Regel die schriftliche Stellungnahme des Leistungserbringers. Für diese finale Entscheidung, ob ein rechnerisch auffälliges Indikatorergebnis als qualitativ auffällig oder unauffällig eingestuft wird, gibt es keine Vorgaben oder Kriterien in der QSKH-RL.

In der Befragung der Landesgeschäftsstellen für Qualitätssicherung zeigte sich, dass die Landesgeschäftsstellen ihre Fachgruppen in unterschiedlichem Umfang in die Bewertung des Indikatorergebnisses anhand der schriftlichen Stellungnahmen einbinden (Winkler-Komp et al. 2014). Einerseits gibt es Landesgeschäftsstellen, die ihren Fachgruppen nicht immer alle schriftlichen Stellungnahmen der Leistungserbringer zur Bewertung vorlegen. Beispielsweise bindet eine Landesgeschäftsstelle die Fachgruppen beim erstmaligen Vorliegen von Dokumentationsfehlern nicht in die Bewertung ein. Andererseits geht aus der Befragung hervor, dass die meisten Landesgeschäftsstellen ihren Fachgruppen keine Bewertungsvorschläge für das Indikatorergebnis zusammen mit den schriftlichen Stellungnahmen vorlegen, sodass die Bewertung direkt in der Sitzung vorgenommen wird. Es wird zudem beschrieben, wie der Prozess der Entscheidungsfindung im Rahmen der anschließenden Bewertung von Stellungnahmen (qualitativ auffälliges oder unauffälliges Ergebnis) auf der Landesebene getroffen wird. Am häufigsten wird darüber von den beauftragten Stellen im Konsens entschieden (siehe Tabelle 4 im Anhang, Kapitel 3). Ob die Landesgeschäftsstellen Informationen, die nicht Bestandteil der schriftlichen Stellungnahmen sind, für die Bewertung der Indikatorergebnisse heranziehen, ist nicht bekannt.

#### **2.4.3 Bewertung der Dokumentationsqualität von Qualitätsindikatoren**

Wird aufgrund eines rechnerisch auffälligen Ergebnisses eine schriftliche Stellungnahme angefordert, wird diese neben der inhaltlichen Überprüfung der Angaben zur medizinischen Versorgung auch dahingehend geprüft, ob der Leistungserbringer Angaben gemacht hat, die er aufgrund einer fehlerhaft vorgenommenen Dokumentation nicht zu verantworten hat. Hierbei ist zwischen zwei Arten der „Entlastung“ des Leistungserbringers aufgrund der Angabe von Dokumentationsfehlern zu unterscheiden. Leistungserbringer können im Rahmen der schriftlichen Stellungnahmen einerseits das Vorhandensein von Dokumentationsfehlern anführen, die im

Hinblick auf eine mögliche qualitative Auffälligkeit zu einer indirekten Entlastung führen, da in der Folge keine Qualitätsbeurteilung vorgenommen werden kann. Das Bewertungsschema sieht für diese Konstellationen vor, das rechnerisch auffällige und potenziell qualitativ auffällige Ergebnis in die Kategorie „D“ („Bewertung nicht möglich wegen fehlerhafter Dokumentation“) einzustufen. Bundesweit wurden für das Erfassungsjahr 2015 durchschnittlich 10,4 % der rechnerisch auffälligen Indikatorergebnisse aufgrund nachträglicher Angaben von Dokumentationsfehlern nicht bewertet (IQTIG 2017a: 64 ff.).

Andererseits können derzeit rechnerisch auffällige Ergebnisse als qualitativ unauffällig bewertet werden und damit direkt zu einer Entlastung des Leistungserbringers führen, wenn die beauftragte Stelle nach der Prüfung der Stellungnahme resümiert, dass vereinzelt Dokumentationsprobleme vorlagen. Bundesweit wurden für das Erfassungsjahr 2015 durchschnittlich 7,4 % der Ergebnisse aufgrund vereinzelter Dokumentationsprobleme als unauffällig eingestuft (Kategorie „U33“). Die Spannweite der Häufigkeit von Kategorie „U33“ reicht in Abhängigkeit vom Bundesland von 0 % bis 24,1 % (IQTIG 2017b: 52 ff.).

Fasst man die Kategorien „D“ und „U33“ zusammen, zeigt sich, dass in Deutschland im Jahr 2015 durchschnittlich 17,9 % aller rechnerisch auffälligen Ergebnisse aufgrund von nachträglich angegebenen Dokumentationsproblemen entweder indirekt oder direkt zu einer Entlastung des Leistungserbringers führten (IQTIG 2017b: 52 ff.).

#### **2.4.4 Qualitätsförderung**

In der Beauftragung zur Weiterentwicklung des Strukturierten Dialogs wird die Qualitätsförderung von Krankenhausleistungen als eindeutiges Ziel des Strukturierten Dialogs verstanden. In der QSKH-RL werden jedoch keine konkreten Maßnahmen zur Qualitätsförderung unter den *Maßnahmen der Externen stationären Qualitätssicherung* (QSKH-RL, Abschnitt B) genannt. Aufgrund der notwendigen und engen Abstimmung zwischen beauftragter Stelle und dem Leistungserbringer können vor allem Besprechungen, Begehungen und Zielvereinbarungen als qualitätsfördernde Maßnahmen und weniger als Instrumente der Qualitätsbewertung verstanden werden.

#### **Besprechungen und Begehungen im Rahmen der Qualitätsförderung**

Im Hinblick auf *Besprechungen* legt die QSKH-RL u. a. Folgendes fest: „Eine Besprechung dient der Aufklärung von Zweifeln und der erforderlichen, ggf. vom Krankenhaus erbetenen, Beratung“ (§ 12 Abs. 2 QSKH-RL). Demnach können gemeinsam qualitätsverbessernde Maßnahmen erarbeitet werden, insbesondere, wenn vonseiten des Leistungserbringers eine Beratung erwünscht und eine Zusammenarbeit mit der beauftragten Stelle gewollt ist. Nach einer Begehung, deren Fokus auf der Prüfung von Qualitätsmängeln liegt, soll eine Besprechung angeschlossen werden. Es kann davon ausgegangen werden, dass Besprechungen und Begehungen im Vergleich zur Einholung einer schriftlichen Stellungnahme beim Leistungserbringer durch den direkten Austausch eine stärkere qualitätsfördernde Funktion erfüllen. Da eine abschließende Bewertung des Indikatorergebnisses zum Zeitpunkt der Begehung jedoch ggf. noch nicht stattgefunden haben muss, ist zu vermuten, dass die noch ausstehende Bewertung möglicherweise

einen ungünstigen Einfluss auf die qualitätsfördernden Maßnahmen hat. Wird in diesem Zusammenhang das Ergebnis eines Leistungserbringers in einem Qualitätsindikator als qualitativ auffällig eingestuft, kann der in der Besprechung oder Begehung gemeinsam erkannte Verbesserungsbedarf in einer Zielvereinbarung festgehalten werden (§ 12 Abs. 2 QSKH-RL).

Besprechungen und Begehungen stellen für alle Beteiligten einen hohen Aufwand dar. Es ist anzunehmen, dass sie aus diesem Grund im Vergleich zu schriftlichen Stellungnahmen relativ selten angewandt werden. Bundesweit wurden im Jahr 2016 zum Erfassungsjahr 2015 282 Besprechungen und 19 Begehungen durchgeführt (IQTIG 2017a: 54).

### **Zielvereinbarungen**

In einer Zielvereinbarung sollen qualitätsfördernde Maßnahmen vereinbart werden, die dem Leistungserbringer helfen sollen, sein Qualitätsproblem zu beheben. Die gemeinsame Erarbeitung von Zielen zur Qualitätsverbesserung und das schriftliche Festhalten einer diesbezüglichen Vereinbarung (§ 12 Abs. 2 QSKH-RL) ist ein wesentliches Element der Qualitätsförderung im Strukturierten Dialog. Zum Erfassungsjahr 2015 wurden 1.121 Ziele zwischen den beauftragten Stellen und Leistungserbringern vereinbart. Das entspricht einem Anteil an 7,1 % an allen rechnerisch auffälligen Ergebnissen. Da bundesweit deutlich mehr Zielvereinbarungen als Besprechungen und Begehungen (insgesamt 301) stattgefunden haben, kann geschlussfolgert werden, dass Zielvereinbarungen teilweise ausschließlich auf Basis schriftlicher Stellungnahmen abgeschlossen werden (IQTIG 2017a: 54).

## **2.5 Biometrische Analyse der Heterogenität in der Vorgehensweise und Bewertung im Strukturierten Dialog**

Im Folgenden soll die Heterogenität in der Einstufung von Leistungserbringern durch die für die Durchführung des Strukturierten Dialogs beauftragten Stellen quantitativ analysiert werden. Speziell sollen Unterschiede zwischen den Bewertungsstellen bei der Einleitung eines Stellungnahmeverfahrens und bei der Qualitätsbewertung anhand der Daten des Strukturierten Dialogs zum Erfassungsjahr 2016 quantifiziert werden. Auch wenn die jährlichen Berichte zum Strukturierten Dialog hinreichend belegen, dass es etwa im Anteil an versendeten Hinweisen Unterschiede zwischen den Bewertungsstellen gibt (vgl. IQTIG (2017a), IQTIG (2018a)), bleibt zum einen die Frage, wie sich diese Unterschiede zusammenfassend quantifizieren lassen und zum anderen, wie stark sich etwaige Unterschiede zwischen den Bewertungsstellen auch durch Unterschiede in der Zusammensetzung der Leistungserbringer in den jeweiligen Bundesländern erklären lassen. Da in den Bewertungsstellen für jedes QS-Verfahren unterschiedliche Fachgruppen existieren, welche die Bewertung von Stellungnahmen vornehmen, ist auch innerhalb einer beauftragten Stelle Heterogenität bei der Bewertung von Stellungnahmen zwischen verschiedenen QS-Verfahren zu erwarten. Im Folgenden soll zunächst jedoch nur die Heterogenität zwischen den Bewertungsstellen exemplarisch anhand ausgewählter Qualitätsindikatoren analysiert werden.

Analog zur Risikoadjustierung bei Qualitätsindikatoren muss auch bei der Analyse der Unterschiede im Strukturierten Dialog der beauftragten Stellen gefragt werden: Inwiefern lassen sich

die beobachteten Unterschiede beispielsweise im Anteil qualitativ auffälliger Ergebnisse zwischen den Bewertungsstellen durch tatsächliche Qualitätsunterschiede der zugeordneten Leistungserbringer erklären und wie viel davon geht zurück auf eine heterogene Bewertungsweise der beauftragten Stellen?

Für die Analyse der Unterschiede zwischen den Bewertungsweisen der beauftragten Stellen sollen folgende Fragestellungen beantwortet werden:

- **Fragestellung 1:** Wie groß ist die bewertungsstellenabhängige Heterogenität bezüglich der Aufnahme eines Stellungnahmeverfahrens bei rechnerischer Auffälligkeit eines Leistungserbringers?
- **Fragestellung 2:** Wie groß ist die bewertungsstellenabhängige Heterogenität bezüglich der Bewertung als qualitativ auffällig nach Aufnahme eines Stellungnahmeverfahrens?

Diese Fragestellungen sind jeweils unter Berücksichtigung der rechnerischen Ergebnisse der Leistungserbringer in ihrem Bundesland zu beantworten, um auszuschließen, dass sich die beobachteten Unterschiede durch Unterschiede zwischen den Leistungserbringern erklären. Es geht also darum, zu charakterisieren, wie die jeweiligen Bewertungsstellen *vergleichbare rechnerische Ergebnisse* bewerten.<sup>15</sup>

Zur Beantwortung dieser Fragen wurde eine Methodik entwickelt, die mithilfe von sog. gemischten logistischen Regressionsmodellen (vgl. z. B. Agresti (2013), Stroup (2013)) in der Lage ist, den Einfluss von Bewertungsstelle und Leistungserbringerergebnis modellbasiert zu trennen. Im Anhang wird diese Methodik in Kapitel 4 ausführlich beschrieben. Dabei werden die Bewertungsergebnisse des Strukturierten Dialogs mit den rechnerischen Ergebnissen der Leistungserbringer verknüpft und ausgewertet. Die Methodik eignet sich allgemein zur Messung von Heterogenität in den Bewertungen des Strukturierten Dialogs und kann künftig auch im Rahmen weiterer Evaluationen verwendet und verallgemeinert werden.

Im konkreten Fall wurde die Methodik auf exemplarisch ausgewählte Qualitätsindikatoren und QS-Verfahren angewendet. Um ein aussagekräftiges Analyseergebnis zu erhalten, wurden für diese Analyse drei Qualitätsindikatoren ausgewählt, für welche zum Erfassungsjahr 2016 eine besonders hohe absolute Zahl an Stellungnahmen zu verzeichnen war. Dabei wurden jeweils ein Prozess-, Indikations- und Ergebnisindikator ausgewählt (vgl. Tabelle 2), die aus den QS-Verfahren bzw. Auswertungsmodulen *Herzschríttmacher-Implantation*, *Hüftendoprothesenversorgung* und *Pflege: Dekubitusprophylaxe* stammen.

---

<sup>15</sup> Dabei dienen die rechnerischen Ergebnisse der Leistungserbringer für die Analyse der qualitativ auffälligen Bewertungen (Fragestellung 2) lediglich als Approximation, um tatsächliche Qualitätsunterschiede zwischen den Leistungserbringern zu modellieren.

Tabelle 2: Indikatorenauswahl für exemplarische Analyse der Heterogenität des Strukturierten Dialogs. Alle Angaben beziehen sich auf den Strukturierten Dialog zum Erfassungsjahr 2016.

QI-ID	Modul-Kürzel	QI-Bezeichnung	Bewertungsart	Anz. Standorte	Anz. rechn. auff. Standorte	Anz. Stellungn.	Anz. qual. auff. Standorte
52010	DEK	Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4	Sentinel Event	1867	544	477	98
54003	HEP	Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur	Ratenbasiert	1231	625	384	134
54139	09/1	Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen	Ratenbasiert	1106	498	279	43

Eine detaillierte Darstellung der Ergebnisse für jeden der drei ausgewählten Qualitätsindikatoren ist im Anhang unter Abschnitt 4.2 zu finden. Beispielhaft werden die Ergebnisse der Analyse zum Indikator „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ anhand von Abbildung 2 dargestellt. Dargestellt wird für Fragestellung 1 (oben) die modellbasierte Wahrscheinlichkeit  $P_{lb}$ , dass von einem rechnerisch auffälligen Leistungserbringer  $l$  durch Bewertungsstelle  $b$  eine Stellungnahme angefordert wird. Jedem Punkt in der Grafik entspricht eine Bewertungsstelle. Dargestellt wird für Fragestellung 2 (unten) die modellbasierte Wahrscheinlichkeit, dass ein Leistungserbringer  $l$  nach Anfordern einer Stellungnahme durch Bewertungsstelle  $b$  als qualitativ auffällig eingestuft wird. Dabei wird in den vier Spalten der Grafik dargestellt, wie sich die Wahrscheinlichkeiten für Fragestellung 1 und 2 in Abhängigkeit des rechnerischen Ergebnisses des Leistungserbringers  $l$  verhalten. Beispielsweise zeigen die Bewertungswahrscheinlichkeiten zu Fragestellung 1 von Leistungserbringern im 1. Quartil der Ergebnisse (links) eine größere bewertungsstellenabhängige Streuung als bei Leistungserbringern im 4. Quartil der Ergebnisse (rechts).

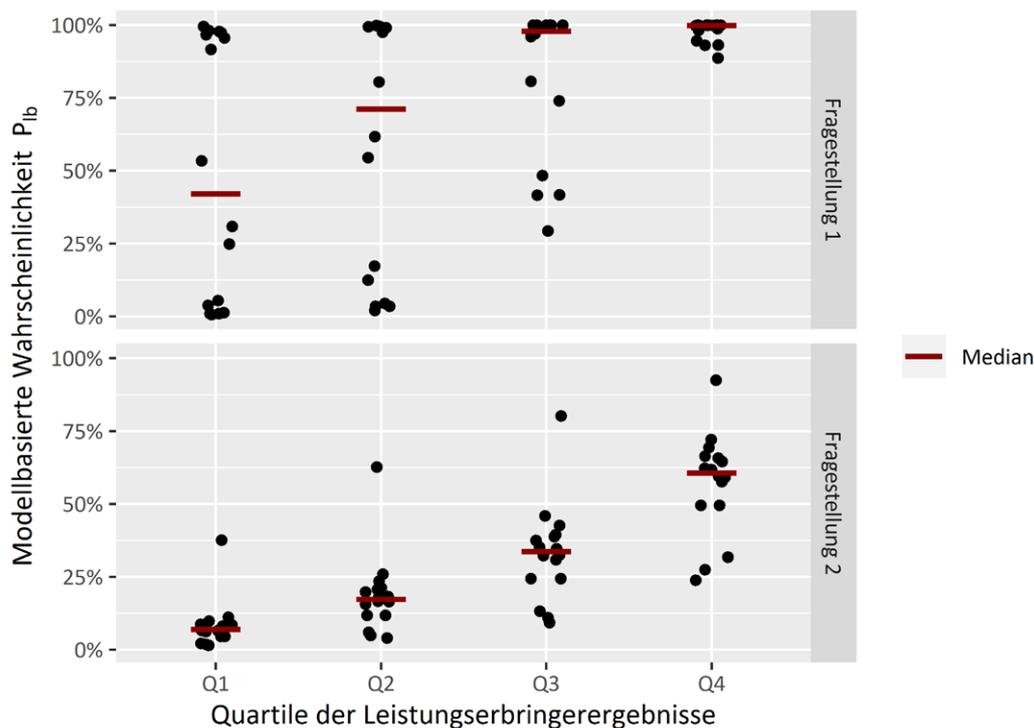


Abbildung 2: Beispielhafte Darstellung der Analyseergebnisse zum Indikator „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ (QI-ID 54003)

Die wichtigsten Erkenntnisse, die sich anhand der Analyseergebnisse belegen lassen, sind dabei folgende:

Zunächst lässt sich zeigen, dass die Wahrscheinlichkeit für das Anfordern einer Stellungnahme abhängig sowohl von der Bewertungsstelle als auch vom Ausmaß der Überschreitung des indicatorspezifischen Referenzbereichs durch den Leistungserbringer ist. Extreme Überschreitungen des Referenzbereichs führen häufiger zum Anfordern einer Stellungnahme als knappe Überschreitungen des Referenzbereichs. Analoge Beobachtungen lassen sich nach dem Beginn des Stellungnahmeprozesses auch für die Wahrscheinlichkeit einer Bewertung eines Leistungserbringers als qualitativ auffällig machen. Auch hier unterscheiden sich die Bewertungswahrscheinlichkeiten je nach Bewertungsstelle und Ausmaß der Überschreitung des Referenzbereichs. Leistungserbringer, die den Referenzbereich deutlich überschreiten, werden häufiger als qualitativ auffällig bewertet als Leistungserbringer, die den Referenzbereich nur knapp verfehlen.

Zu den Eingangs formulierten Fragestellungen 1 und 2 nach der Heterogenität der Bewertungsergebnisse lässt sich zeigen, dass eine stark ausgeprägte bewertungsstellenabhängige Heterogenität unabhängig von den numerischen Indikatorergebnissen der Leistungserbringer in den jeweiligen Bundesländern besteht. Im Detail ist dabei anzumerken, dass diese Effekte je nach Qualitätsindikator und nach Fragestellung (Anfordern von Stellungnahmen (1) oder Bewertung als qualitativ auffällig (2)) unterschiedlich ausgeprägt sind, vgl. Anhang, Abschnitt 4.2. So ist bei den beiden ratenbasierten Indikatoren „Präoperative Verweildauer bei endoprothetischer Ver-

sorgung einer hüftgelenknahen Femurfraktur“ und „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“ die Heterogenität im Versenden von Hinweisen/Anfordern von Stellungnahmen (Fragestellung 1) deutlich ausgeprägter als beim Sentinel-Event-Indikator zu einem stationär erworbenen Dekubitalulcus Grad 4, welcher wiederum eine höhere Heterogenität bei der Bewertung als qualitativ auffällig (Fragestellung 2) zeigt als die beiden anderen Indikatoren.

Eine weitere Erkenntnis bezüglich der Heterogenität der Bewertungen ist, dass die größten Unterschiede zwischen den beauftragten Stellen bei der Bewertung von Leistungserbringern auftreten, deren numerische Indikatorergebnisse den Referenzbereich nur knapp überschreiten. Dies lässt sich anhand der Abbildung 2 (oben) bzw. anhand der Abbildungen 2 bis 4 im Anhang unter Abschnitt 4.2 nachvollziehen.

Für das Ziel der Vereinheitlichung des Strukturierten Dialogs lässt sich aus der stark ausgeprägten Heterogenität bezüglich Fragestellung 1 ableiten, dass eine einheitliche Handhabung beim Versenden von Hinweisen bzw. Anfordern von Stellungnahmen durch die beauftragten Stellen als ausschlaggebender Schritt anzusehen ist – im Hinblick auf die daraus resultierenden Unterschiede in den Ergebnissen. Insbesondere bei Leistungserbringern mit Indikatorergebnissen, die den Referenzbereich nur knapp verletzen, sind einheitliche Regelungen erforderlich, wann eine Stellungnahme verpflichtend anzufordern ist. So könnte beispielsweise durch die Berücksichtigung statistischer Unsicherheit bei der Auffälligkeitseinstufung eine Regelung getroffen werden, die diesen besonders heterogen gehandhabten Wertebereich numerischer Indikatorergebnisse einheitlich regelt. Es ist wichtig darauf hinzuweisen, dass an diesem Punkt die beiden Auftragsaspekte „Vereinheitlichung“ und „Effizienzsteigerung“ des Strukturierten Dialogs zusammen zu denken sind. Eine große Anzahl an rechnerisch auffälligen Leistungserbringern führt bei vielen Bewertungsstellen zu einer Priorisierung: Von Leistungserbringern, die den Referenzbereich stark überschreiten, wird häufiger eine Stellungnahme angefordert als von Leistungserbringern, die den Referenzbereich nur schwach überschreiten. Gelingt es, durch eine veränderte quantitative Auffälligkeitseinstufung eine sinnvolle Reduktion der anzufordernden Stellungnahmen zu erreichen, können mehr Ressourcen zur Bearbeitung der einzelnen Stellungnahmen zur Verfügung stehen.

Eine weitere Konsequenz der vorliegenden Analyseergebnisse ist, dass die Vereinheitlichung des Strukturierten Dialogs nicht auf die Frage der Handhabung von Hinweisen zu reduzieren ist: In den Ergebnissen zu Fragestellung 2 wurde ebenfalls eine deutliche Heterogenität bei der qualitativen Bewertung von Stellungnahmen festgestellt, die belegt, dass auch hier Vereinheitlichungsbedarf besteht.

## 2.6 Biometrische Analyse zu Aufwand und Effizienz des Strukturierten Dialogs

Auftragsgegenstand der Weiterentwicklung des Strukturierten Dialoges ist eine „Optimierung der Effizienz des Verfahrens“ (G-BA 2018a). Ausgangspunkt für eine Effizienz-Steigerung muss dabei eine Feststellung des Ist-Zustandes der Effizienz sein. Im folgenden Abschnitt wird exemplarisch anhand des QS-Verfahrens *Hüftendoprothesenversorgung* (HEP) dargestellt, welcher Ressourcenaufwand – gemessen anhand der Anzahl zu führender Strukturierter Dialoge – in der aktuellen Form der rechnerischen Auffälligkeitseinstufung für den Strukturierten Dialog entsteht. Diesem Aufwand werden die durch den Strukturierten Dialog identifizierten Qualitätsdefizite als „Nutzen“ des Strukturierten Dialoges gegenübergestellt. Dies entspricht einem eindimensionalen Verständnis von Effizienz als Verhältnis von Aufwand und Nutzen.

Durch diese deskriptive Analyse können Probleme und Einsparungspotenzial in der gegenwärtigen Methodik zur rechnerischen Auffälligkeitseinstufung gezielt beschrieben und adressiert werden. Die hier vorgestellten Zahlen sind in weiten Teilen bekannt aus dem jährlichen Bericht zum Strukturierten Dialog (IQTIG 2018a) – exemplarisch werden die Analysen für das Erfassungsjahr 2017 durchgeführt, welche das aktuellste Erfassungsjahr ist, für welches die Ergebnisse des Strukturierten Dialogs vorliegen. Ein besonderer Fokus der folgenden Aufwandsanalyse ist die Darstellung von rechnerischen und qualitativen Auffälligkeiten in Abhängigkeit von der Art des Referenzbereiches und der dem Indikatorergebnis zugrunde liegenden Fallzahl. Die Fallzahl ist dabei die Kardinalität der Grundgesamtheit des Indikators (n), welche zur Berechnung eines Standortergebnisses verwendet wird. Je nach Qualitätsindikator ist dies die Anzahl an Patientinnen und Patienten, die Anzahl durchgeführter Prozeduren oder die Anzahl stationärer Behandlungsfälle. Als Art des Referenzbereiches wird zwischen Qualitätsindikatoren mit festem Referenzbereich, Qualitätsindikatoren mit perzentilbasiertem Referenzbereich und Sentinel-Event-Indikatoren unterschieden.

### 2.6.1 Aufwandsüberblick QS-Verfahren *Hüftendoprothesenversorgung*

Im QS-Verfahren *Hüftendoprothesenversorgung* (HEP) gab es im Erfassungsjahr 2017 ca. 1280 Standorte, die Daten zu hüftendoprothetischen Eingriffen übermittelt haben. In 14 Qualitätsindikatoren gab es auf Standortebene ca. 17.000 zu bewertende Indikatorergebnisse. Mehr als 2.000 dieser Ergebnisse wurden als rechnerisch auffällig eingestuft. Das HEP-Verfahren verursacht, gezählt über alle QSKH-Verfahren, ca. 19 % aller rechnerischen Auffälligkeiten. Von diesen rechnerisch auffälligen Ergebnissen wurden 223 Ergebnisse im Strukturierten Dialog als qualitativ auffällig bewertet.

### 2.6.2 Rechnerische Auffälligkeiten je Qualitätsindikator

Tabelle 3 zeigt, wie sich die rechnerisch auffälligen Ergebnisse des HEP-Verfahrens auf die 14 Qualitätsindikatoren verteilen.

Tabelle 3: Rechnerisch auffällige Standorte im HEP-Verfahren für alle 14 Qualitätsindikatoren. Daten aus dem Erfassungsjahr 2017

QI-ID	QI-Titel	Art des Referenzbereiches	Anzahl Standorte	rechn. auff. Standorte	Anteil
54003	Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur	Fester Referenzwert	1226	527	43 %
54013	Todesfälle während des akut-stationären Aufenthaltes bei geringer Sterbewahrscheinlichkeit	Sentinel-Event	1251	320	25,6 %
54002	Indikation zum Hüftendoprothesen-Wechsel bzw. -Komponentenwechsel	Fester Referenzwert	1149	193	16,8 %
54017	Allgemeine Komplikationen bei Hüftendoprothesen-Wechsel bzw. -Komponentenwechsel	95. Perzentil	1149	146	12,7 %
54010	Beweglichkeit bei Entlassung	Fester Referenzwert	1193	140	11,7 %
54120	Spezifische Komplikationen bei Hüftendoprothesen-Wechsel bzw. -Komponentenwechsel	95. Perzentil	1149	133	11,6 %
54001	Indikation zur elektiven Hüftendoprothesen-Erstimplantation	Fester Referenzwert	1195	104	8,7 %
54016	Allgemeine Komplikationen bei elektiver Hüftendoprothesen-Erstimplantation	95. Perzentil	1195	102	8,5 %
54019	Spezifische Komplikationen bei elektiver Hüftendoprothesen-Erstimplantation	95. Perzentil	1195	92	7,7 %
54015	Allgemeine Komplikationen bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur	95. Perzentil	1228	87	7,1 %
54012	Verhältnis der beobachteten zur erwarteten Rate (O/E) an Patienten mit Gehunfähigkeit bei Entlassung	95. Perzentil	1267	82	6,5 %
54018	Spezifische Komplikationen bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur	95. Perzentil	1228	76	6,2 %
10271	Verhältnis der beobachteten zur erwarteten Rate (O/E) an Hüftendoprothesen-Wechsel bzw. -Komponentenwechsel im Verlauf	95. Perzentil	1279	69	5,4 %
54004	Sturzprophylaxe	Fester Referenzwert	1272	67	5,3 %
<b>Alle</b>	-		<b>16976</b>	<b>2138</b>	<b>12,6 %</b>

Unter den 5 Indikatoren mit den meisten rechnerischen Auffälligkeiten befinden sich drei Indikatoren mit festem Referenzbereich, darunter der Qualitätsindikator „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“, der im Erfassungsjahr 2017 als Qualitätsindikator mit besonderem Handlungsbedarf eingestuft wurde. Etwa 25 % aller rechnerischen Auffälligkeiten dieses QS-Verfahrens fallen allein auf diesen Qualitätsindikator. Auch der Sentinel-Event-QI „Todesfälle während des akut-stationären Aufenthaltes bei geringer Sterbewahrscheinlichkeit“ löst mehr als 300 Strukturierte Dialoge aus. Alle anderen Qualitätsindikatoren bewegen sich im Bereich von 67 bis 193 rechnerisch auffälligen Standortergebnissen.

Der Anteil der rechnerisch auffälligen Standorte bei Indikatoren mit perzentilbasierten Referenzwerten (in allen Fällen gilt in HEP das 95. Perzentil der Standortergebnisse als Referenzwert), liegt immer deutlich über dem nominell angestrebten Wert von 5 % auffälligen Standorten. Dies ist auf die besondere Art der Perzentilberechnung für den Referenzwert zurückzuführen, die den Perzentil-Referenzbereich nur auf Standorten mit wenigstens 20 Nenner-Fällen berechnet.<sup>16</sup> So ist beispielsweise die Anzahl der rechnerischen Auffälligkeiten bei den Komplikationsindikatoren zu Hüftendoprothesen-Wechseln (QI 54120 und QI 54017) mehr als doppelt so hoch wie das nominell suggerierte Niveau von 5 %. Diese Indikatoren haben einen sehr hohen Anteil an Standorten mit weniger als 20 Fällen, die nicht in die Berechnung des Perzentilwertes einfließen. Sind in dieser Gruppe mehr als 5 % der Standorte oberhalb des berechneten Perzentilwertes, so können auch insgesamt mehr als 5 % der Standorte rechnerisch auffällig werden. In den vorliegenden Extremfällen (QI 54120 und QI 54017) werden sogar mehr als doppelt so viele Standorte auffällig, wie nominell durch den Perzentilreferenzwert angestrebt. Abbildung 3 illustriert diesen Sachverhalt für den Qualitätsindikator mit QI-ID 54017.

---

<sup>16</sup> vgl. RAW-10- bzw. RAW-20-Methode in (Paddock 2014)

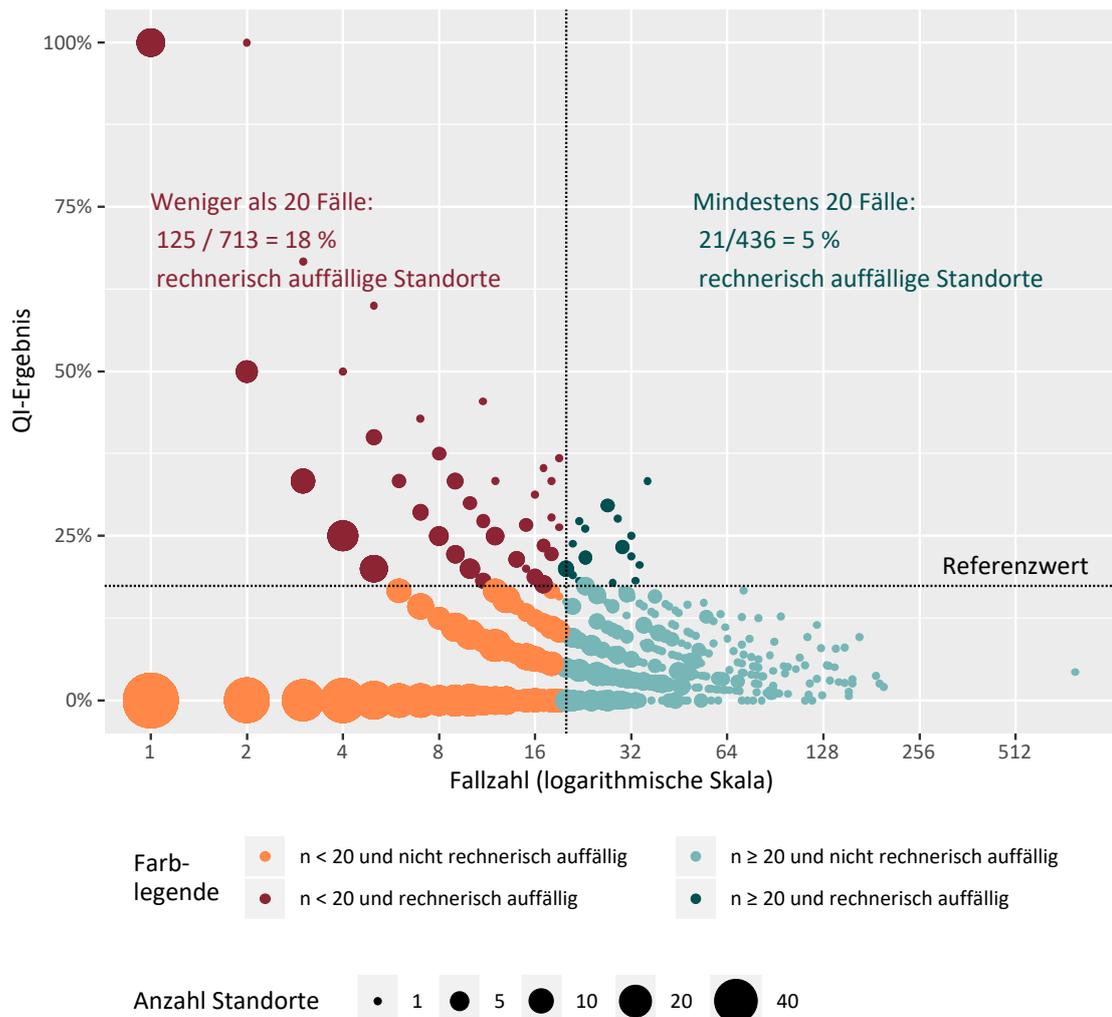


Abbildung 3: Standortergebnisse im Qualitätsindikator QI-ID 54017 „Allgemeine Komplikationen bei Hüftendoprothesen-Wechsel bzw. -Komponentenwechsel“ differenziert nach Fallzahl.

Gezeigt werden Standortergebnisse differenziert nach Fallzahl. Zur besseren Darstellung der Fallzahl wurde eine logarithmische Achsenskalierung zur Basis 2 gewählt, d. h. der Abstand zwischen zwei vertikalen Linien auf der x-Achse entspricht immer einer Verdopplung der Fallzahl. Die Größe der Punkte ist proportional zur Anzahl an Standorten mit gleicher Fallzahl und Ergebnis. Horizontal eingezeichnet ist der Referenzwert, der als 95. Perzentil aller Standorte mit mindestens 20 Fällen in der Grundgesamtheit des Indikators berechnet wird. 146 von 1.150 Standorten (12,7 %) wurden in diesem Indikator als rechnerisch auffällig eingestuft. Typisch für diese Funnelplot-Darstellung sind die erkennbaren Hyperbeläste  $1/n$ ,  $2/n$ ,  $3/n$  etc. entlang derer die Standortergebnisse mit steigender Fallzahl  $n$  verlaufen. Die Grafik ist in vier Farbsegmente nach rechnerischer Auffälligkeit (ja/nein) und Fallzahlgruppe ( $< 20$ ,  $\geq 20$ ) unterteilt. Während der Referenzbereich so bestimmt wird, dass genau 5 % der Standorte mit mindestens 20 Fällen auffällig werden, führt die Anwendung des gleichen Referenzbereiches auf die Standorte mit weniger als 20 Fällen dazu, dass in dieser Fallzahlgruppe ein deutlich über 5 % liegender Teil von 18 % der Standorte als rechnerisch auffällig eingestuft wird.

### 2.6.3 Qualitative Auffälligkeiten je Qualitätsindikator

Charakterisiert man den Aufwand des Strukturierten Dialoges anhand der absoluten Zahl rechnerischer Auffälligkeiten, so stellt sich die Frage wodurch bzw. ob ein geeignetes Maß für den „Nutzen“ des Strukturierten Dialoges im Sinne einer Aufwand-Nutzen-Gegenüberstellung definiert werden kann. In Abschnitt 5.2 werden Ansätze einer strukturierten Aufwand-Nutzen-Betrachtung bzw. Verlust-Betrachtung für den Strukturierten Dialog vorgestellt. Im Folgenden wird der Anteil an durch den Strukturierten Dialog identifizierten Qualitätsdefiziten, bezogen auf alle rechnerischen Auffälligkeiten bzw. alle Stellungnahmen, als eindimensionales Maß für den „Nutzen“ des Strukturierten Dialoges dargestellt. Dies ermöglicht eine Betrachtung von Effizienz im Sinne einer Aufwand-Nutzen-Abwägung, klammert allerdings aus, dass nicht identifizierte Qualitätsdefizite „immaterielle Kosten“ des Gesundheitssystems sind, die im Strukturierten Dialog nicht messbar sind. Eine wichtige Einschränkung der folgenden Analyse ist, dass die qualitativen Bewertungen des Strukturierten Dialogs bezogen auf die betrachteten rechnerischen Auffälligkeiten i. d. R. keine Sensitivität und Spezifität von 100 % aufweisen, d. h. es ist nicht sicher, dass Qualitätsdefizite nach einer rechnerischen Auffälligkeit immer entdeckt werden und nicht jede Bewertung als qualitativ auffällig identifiziert sicher ein Qualitätsdefizit.<sup>17</sup> Die qualitativen Bewertungen des Strukturierten Dialogs sind somit kein Goldstandard, der sich zum Benchmarking verschiedener statistischer Methoden zur quantitativen Auffälligkeitseinstufung eignet.

Abbildung 4 (oben) zeigt für jeden Qualitätsindikator die Anzahl an rechnerischen Auffälligkeiten (x-Achse) und den Anteil an qualitativen Auffälligkeiten unter allen rechnerisch auffälligen Standorten. Da nicht jeder rechnerischen Auffälligkeit im Strukturierten Dialog nachgegangen wird, wird in Abbildung 4 (unten) der Anteil an qualitativen Auffälligkeiten auf die Zahl an angeforderten Stellungnahmen bezogen.

---

<sup>17</sup> Andernfalls, bei perfekter Sensitivität und Spezifität, gäbe es auch keine Heterogenität der Bewertungen, wie sie in 2.5 deutlich nachgewiesen ist.

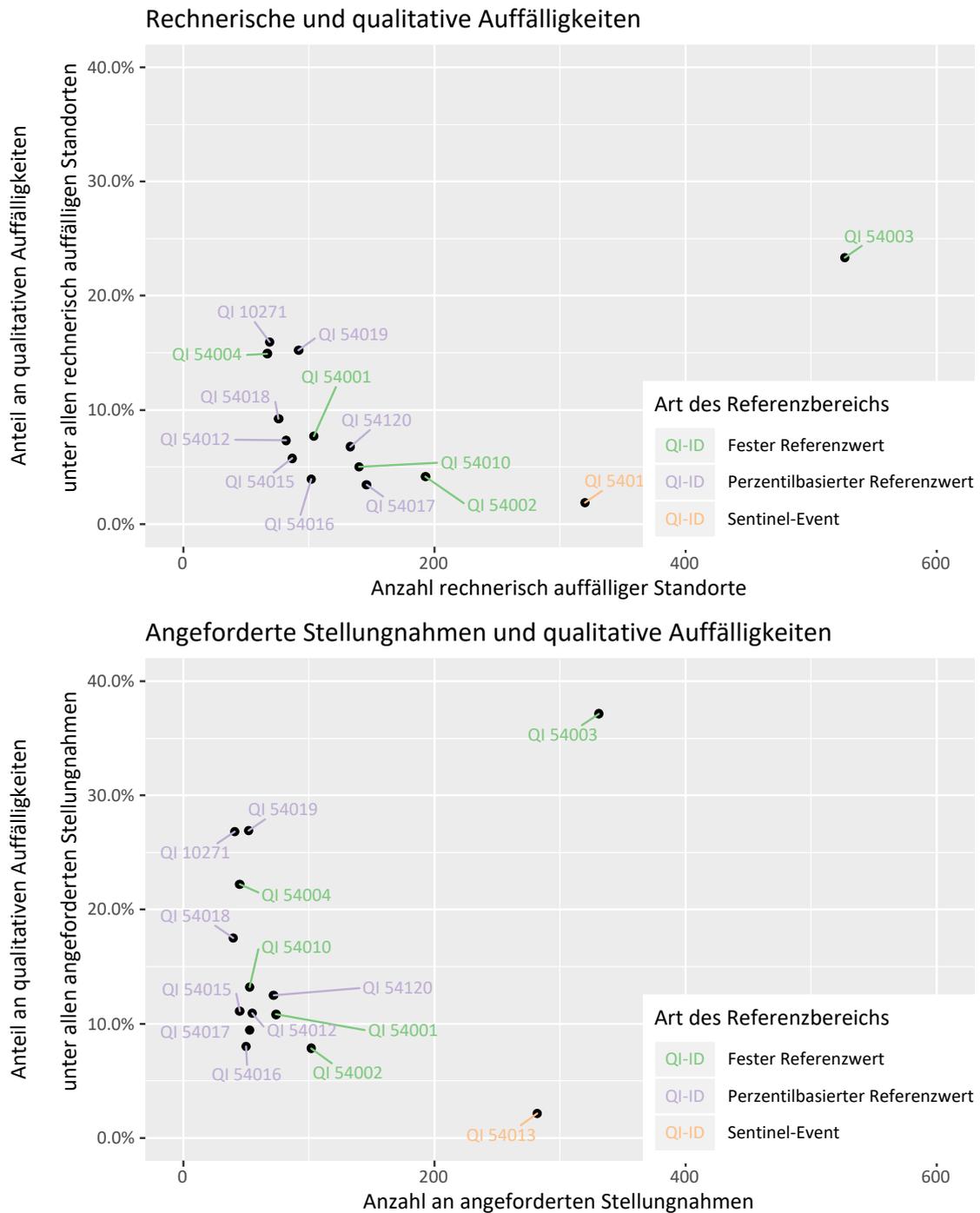


Abbildung 4: Qualitative Auffälligkeiten in Relation zu rechnerischen Auffälligkeiten (oben) und angeforderten Stellungnahmen (unten). Daten des Erfassungsjahres 2017.

Es fällt auf, dass der Anteil der qualitativen Auffälligkeiten an den rechnerischen Auffälligkeiten je nach Qualitätsindikator sehr unterschiedlich ist. Während eine rechnerische Auffälligkeit im Sentinel-Event-Indikator zur Sterblichkeit (QI-ID 54013) nur sehr selten (< 2 %) qualitative Auffälligkeiten nach sich zieht, wird im Indikator zur präoperativen Verweildauer (QI-ID 54003) mehr als jede vierte rechnerische Auffälligkeit auch als qualitativ auffällig bewertet. Die Mehrheit aller Indikatoren des HEP-Verfahrens erreicht weniger als 10 % qualitative Auffälligkeiten

unter allen rechnerischen Auffälligkeiten. Bezieht man die Zahl der qualitativen Auffälligkeiten auf die Anzahl an Stellungnahmen, zeigt sich für die meisten Qualitätsindikatoren eine höhere „Trefferquote“ im Sinne der obigen Auswertung. Insgesamt wurden bei ca. 61 % aller rechnerisch auffälligen Ergebnisse Stellungnahmen vom Leistungserbringer angefordert. Betrachtete man die qualitative Einstufung des Strukturierten Dialoges im Sinne eines diagnostischen Tests als „Goldstandard“<sup>18</sup>, an der die Güte der rechnerischen Auffälligkeitseinstufung zu messen wäre, dann gäbe es im HEP-Verfahren über alle 14 Indikatoren berechnet ca. 90 % falsch positive rechnerische Auffälligkeiten. Ein Grund für diese hohe Zahl falsch positiver Ergebnisse ist möglicherweise darin zu sehen, dass das HEP Verfahren aus sehr vielen Ergebnisindikatoren besteht. Angesichts zum Teil nicht-risikoadjustierter Ergebnisindikatoren bzw. Risikoadjustierungen bei denen wichtige aus der Literatur bekannte Risikofaktoren nicht eingehen können, weil sie nicht erhoben werden, ist die Bewertung der Ergebnisqualität z. B. bei Todesfällen und Komplikationen eine besondere Herausforderung für den Strukturierten Dialog. Ein Grund dafür ist, dass strukturelle Probleme beim Leistungserbringer in diesem Fall deutlich schwieriger zu erkennen sind als beispielsweise bei Prozessindikatoren – besonders bei kleinen Fallzahlen.

Ein weiterer Grund für falsch positive Bewertungen ist auch in mangelnder Dokumentations- und Datenfeldqualität zu sehen. Beispielsweise wurde unter allen in HEP bewerteten Stellungnahmen ca. 13 % Dokumentationsprobleme im Sinne der Bewertungskategorien „D50“, „D51“ und „D99“ festgestellt. Angesichts des absoluten Aufwands, den das HEP-Verfahren durch die hohe Anzahl rechnerisch auffälliger Ergebnisse für den Strukturierten Dialog verursacht, ist ein Anteil von 90 % falsch positiver Auffälligkeiten als sehr hoch anzusehen. Es stellt sich die Frage, ob jeder auf Landesebene beauftragten Stelle für die gründliche Bearbeitung und Bewertung dieser Vielzahl von Stellungnahmen ausreichende Ressourcen zur Verfügung stehen. Gegebenenfalls könnte ein Mangel an für den Strukturierten Dialog zur Verfügung stehenden Ressourcen auch eine Erklärung für hohe falsch-positive Bewertungen sein. Weiterhin ist anzuführen, dass in der aktuellen Form des Strukturierten Dialogs die Beweislast bei den Bewertungsstellen liegt, ein bestehendes Qualitätsdefizit anhand der Stellungnahme des Leistungserbringers genau benennen zu können, was wiederum eine weitere Erklärung für den hohen Anteil rechnerischen Auffälligkeitseinstufungen sein kann, die dann im Strukturierten Dialog als qualitativ unauffällig bewertet werden. Die Ursachen für hohe falsch-positive Zahlen sind letztlich nicht klar und eindeutig: Sensitivität und Spezifität des Strukturierte Dialogs, Sensitivität und Spezifität des Einstufungsverfahrens zur rechnerischen Auffälligkeit, Datenqualität, aber auch Inhaltsvalidität und Operationalisierung jedes einzelnen Qualitätsindikators haben Einfluss auf die Effizienz des QS-Verfahrens insgesamt.

---

<sup>18</sup> Dabei sind die in Abschnitt 2.5 benannten Einschränkungen hinsichtlich der bewertungsstellenabhängigen Heterogenität der qualitativen Bewertungen zu berücksichtigen.

#### 2.6.4 Zusammenhang zwischen Fallzahl und Auffälligkeitseinstufung

Um das Zustandekommen der sehr hohen Anzahl rechnerischer Auffälligkeiten, die als qualitativ unauffällig bewertet werden, zu untersuchen, wird im Folgenden eine nach Fallzahl differenzierte Analyse durchgeführt. Damit soll gezeigt werden, dass es im gegenwärtigen Einstufungsalgorithmus fallzahlabhängige Verzerrungen gibt, die zu überdurchschnittlich hohen Raten rechnerisch auffälliger Leistungserbringer in bestimmten Fallzahlgruppen führen.

Eine Sonderrolle spielt dabei der Sentinel-Event-Indikator mit der QI-ID 54013 (Todesfälle während des akut-stationären Aufenthaltes bei geringer Sterbewahrscheinlichkeit). Da die Fallzahl in die rechnerische Auffälligkeitseinstufung nicht mit einfließt, wie es z. B. bei der Berechnung einer Rate der Fall ist, sind auch die Zusammenhänge zwischen Fallzahl und Auffälligkeitseinstufung andere, als z. B. bei Raten- oder *O/E*-Indikatoren. Wegen dieser Sonderkonstruktion wird der Qualitätsindikator im Folgenden einzeln analysiert, während alle anderen Qualitätsindikatoren nur zusammengefasst dargestellt werden. Der Indikator ist nicht als Beispiel für den gesamten Leistungsbereich zu verstehen.

Die folgenden Abbildungen zeigen jeweils Anzahl und Anteile an Indikatorergebnissen in den folgenden vier sich ausschließenden Kategorien:

- „nicht rechnerisch auffällig“: Indikatorergebnisse, die als nicht rechnerisch auffällig bewertet wurden.
- „rechnerisch auffällig und keine Stellungnahme angefordert“: Indikatorergebnisse, die als rechnerisch auffällig bewertet wurden, zu welchen allerdings keine Stellungnahme angefordert wurde und daher auch keine Qualitätsbewertung erfolgte
- „rechnerisch auffällig und Stellungnahme angefordert“: Indikatorergebnisse, die als rechnerisch auffällig bewertet wurden, zu welchen eine Stellungnahme angefordert wurde, die aber nicht als qualitativ auffällig im Sinne der Bewertungen „A40“, „A41“, „A42“, „A99“ eingestuft wurden.
- „qualitativ auffällig“: Indikatorergebnisse, die als rechnerisch auffällig bewertet wurden, zu welchen eine Stellungnahme angefordert wurde, und die als qualitativ auffällig im Sinne der Bewertungen „A40“, „A41“, „A42“, „A99“ eingestuft wurden.

Diese vier Kategorien werden farblich kodiert und in Dezilen der Standortfallzahlen getrennt dargestellt. Durch die Kategorisierung nach Dezilen ist gesichert, dass in jeder der unten abgebildeten Kategorie gleich viele Leistungserbringer enthalten sind, und somit die berechneten Anteile vergleichbarer statistischer Unsicherheit unterliegen.

Zunächst werden der Zusammenhang zwischen Fallzahl und rechnerischer Auffälligkeitseinstufung und der Zusammenhang zwischen Fallzahl und qualitativer Bewertung deskriptiv beschrieben. Anschließend werden in Abschnitt 2.6.4.3 mögliche Ursachen diskutiert. Bei der Interpretation der dargestellten Zusammenhänge zwischen Fallzahl und rechnerischer Auffälligkeit, zwischen Fallzahl und Einholung einer Stellungnahme, sowie zwischen Fallzahl und qualitativer Bewertung ist zu beachten, dass sich systematische Eigenschaften des Einstufungsverfahrens und reale Zusammenhänge zwischen Fallzahl und Ergebnisqualität anhand dieser Zahlen nicht trennen lassen. Auf diesen Aspekt wird ebenfalls im Abschnitt 2.6.4.3 eingegangen.

### 2.6.4.1 Zusammenhang zwischen Fallzahl und rechnerischer Auffälligkeit

#### Nicht Sentinel-Event-QI

Im Folgenden stellen wir nach Fallzahl differenzierte Auswertungen für die 13 nicht-Sentinel-Event Indikatoren des HEP-Verfahrens dar. Dabei wird auf eine Einzeldarstellung jedes Indikators verzichtet und eine aggregierte Darstellung der Ergebnisse vorgenommen. Abbildung 5 zeigt die Häufigkeit rechnerischer Auffälligkeiten und in Folge dessen angeforderte Stellungnahmen aufgeteilt in die zehn Dezile der Fallzahlverteilung. Die obere Grafik zeigt die absoluten Zahlen an Indikatorergebnissen, die als rechnerisch auffällig eingestuft wurden und in Dunkelgrün die Anzahl solcher Auffälligkeiten, für die auch eine Stellungnahme angefordert wurde. Die rote Horizontale markiert dabei die durchschnittliche Anzahl an rechnerischen Auffälligkeiten pro Dezil. Man erkennt, dass überdurchschnittlich viele Leistungserbringer aus den ersten Dezilen als rechnerisch auffällig eingestuft wurden, während die Anzahl der rechnerischen Auffälligkeiten mit steigender Fallzahl sinkt. Die untere Grafik stellt den Anteil der rechnerischen Auffälligkeiten dar, für die auch eine Stellungnahme angefordert wurde. Der Anteil an Stellungnahmen an den rechnerischen Auffälligkeiten liegt im untersten Dezil bei 31 % und steigt bis zum 4. Dezil auf 71 % an. Dies kann in Teilen durch die in der QSKH-Richtlinie vorgesehene Anwendung der sog. 1-Fall-Regel erklärt werden, wonach bei rechnerisch auffälligen Ergebnissen mit nur einem auffälligen Zähler-Fall ein Hinweis versandt werden darf. Vom 5. bis 10. Dezil ist kein monotoner Zusammenhang im Anteil angeforderter Stellungnahmen zu beobachten. Im 8. Dezil liegt der Anteil bei 53 %, im 10. Dezil bei 70 %.

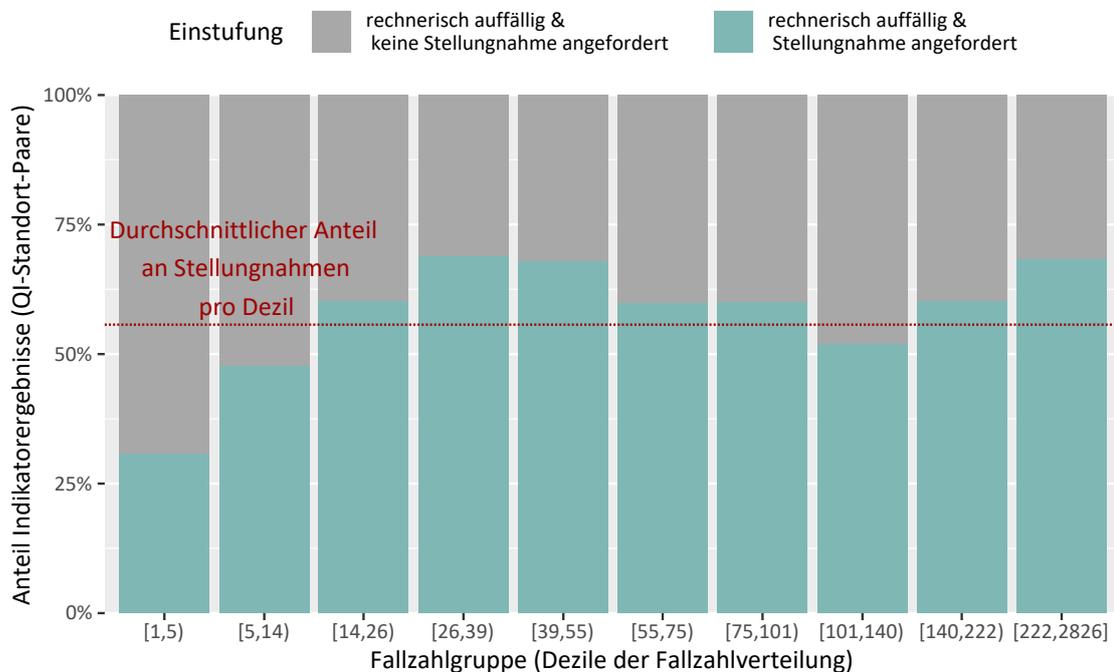
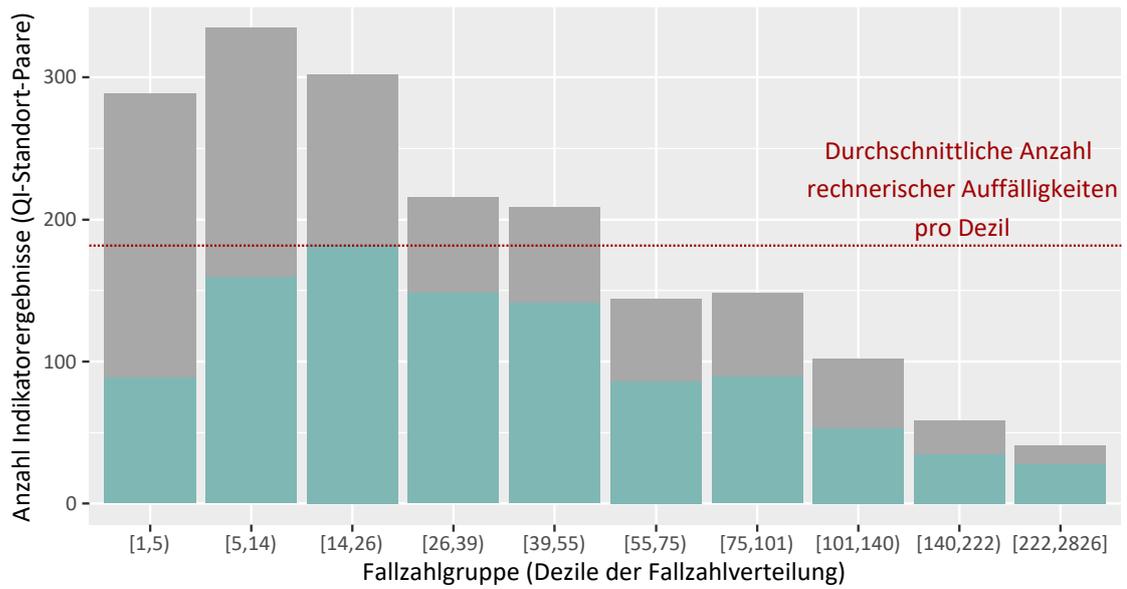


Abbildung 5: Rechnerische Auffälligkeitseinstufung und Stellungnahmen für Nicht-Sentinel-Event-Indikatoren im HEP-Verfahren differenziert in Dezilen der Fallzahlverteilung.

**Sentinel-Event-QI**

Abbildung 6 zeigt den Zusammenhang zwischen Fallzahl und rechnerischer Auffälligkeit beim Sentinel-Event-Indikator mit der QI-ID 54013 „Todesfälle während des akut-stationären Aufenthaltes bei geringer Sterbewahrscheinlichkeit“.

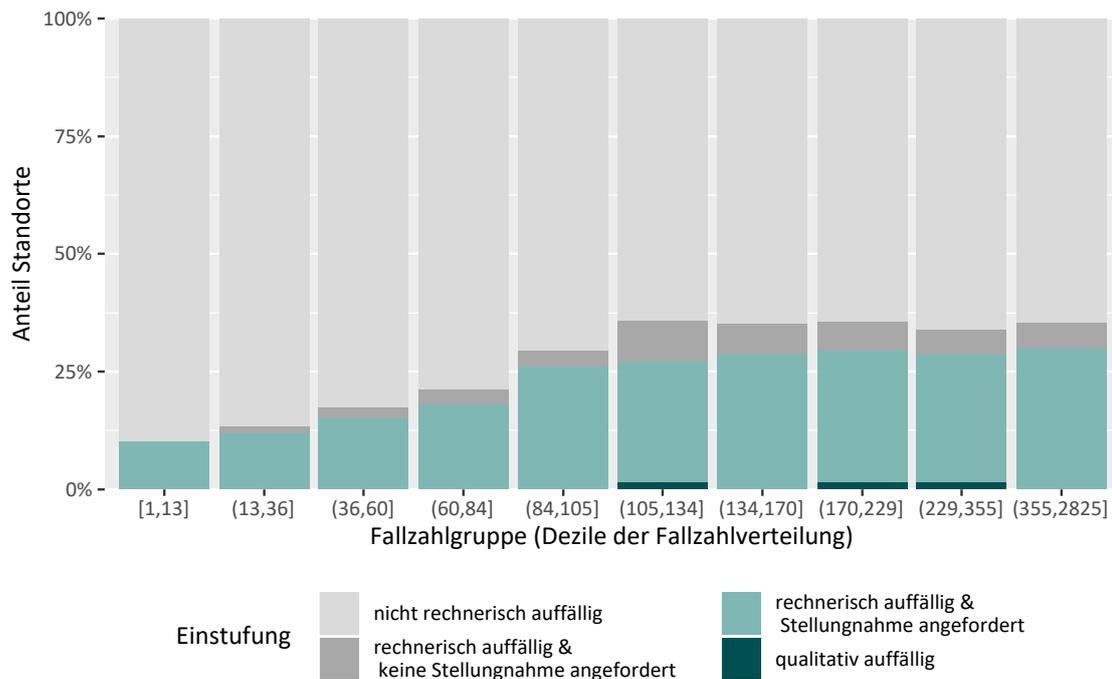


Abbildung 6: Rechnerische und qualitative Auffälligkeitseinstufung zum Sentinel-Event-Indikator „Todesfälle während des akut-stationären Aufenthaltes bei geringer Sterbewahrscheinlichkeit“ (QI-ID 54013) differenziert in Dezilen der Fallzahlverteilung.

Die Darstellung zeigt, dass der Anteil rechnerisch auffälliger Standorte vom 1. bis zum 6. Dezil der Fallzahlverteilung von 10 % auf 36 % steigt. Eine mögliche Erklärung für diesen Zusammenhang von Fallzahl und rechnerischer Auffälligkeit liegt in der speziellen Konstruktion des Qualitätsindikators. Als geringe Sterbewahrscheinlichkeit wird in der entsprechenden Datenbank der Qualitätsindikatoren ein Wert von höchstens 1,9 % definiert (IQTIG 2018e). Nur Fälle, deren Sterbewahrscheinlichkeit durch die Risikoadjustierung als gering eingestuft wird, gehen in den Qualitätsindikator ein. Mit steigender Fallzahl steigt allerdings die Wahrscheinlichkeit, dass sich mindestens ein Todesfall pro Standort zu ereignet, auch wenn das Risiko für den einzelnen Fall als gering eingestuft wird.<sup>19</sup>

#### 2.6.4.2 Zusammenhang zwischen Fallzahl und qualitativer Auffälligkeit

##### Nicht-Sentinel-Event-QI

Neben der Analyse des Zusammenhangs zwischen Fallzahlen und rechnerischen Auffälligkeiten ist auch eine nach Fallzahl differenzierte Auswertung der qualitativen Auffälligkeiten interessant. Abbildung 7 zeigt für die 13 Nicht-Sentinel-Event-Indikatoren den Anteil qualitativ auffälliger Standorte unter allen Standorten, von denen eine Stellungnahme angefordert wurde, und zwar aufgeteilt in den Dezilen der Fallzahlverteilung der QI-Standort-Paare. Zu erkennen ist, dass

<sup>19</sup> Nimmt man vereinfachend an, dass alle Fälle unabhängig voneinander mit einer „geringen“ Sterbewahrscheinlichkeit von  $\pi$  auftreten, so ist die Wahrscheinlichkeit eines Leistungserbringers mit  $n$  Fällen „zufällig auffällig“ zu werden durch  $1 - (1 - \pi)^n$  gegeben, also eine mit  $n$  monoton wachsende Funktion. Mit  $\pi = 0,5\%$  und  $n = 100$  ergäbe sich schon eine Wahrscheinlichkeit von ca. 39 %, ohne dass es Anhaltspunkte für ein besonderes Qualitätsdefizit gäbe.

der Anteil qualitativ auffälliger Ergebnisse vom 1. bis zum 6. Dezil monoton von 4 % auf 37 % steigt. Im 7. und 8. Dezil bleibt der Anteil bei über 30 % und fällt im 9. und 10. Dezil auf unter 20 % ab. Während also rechnerische Auffälligkeitseinstufungen in den kleinen Fallzahlbereichen überproportional häufig sind, sind die qualitativen Auffälligkeiten im mittleren Fallzahlbereich besonders häufig.

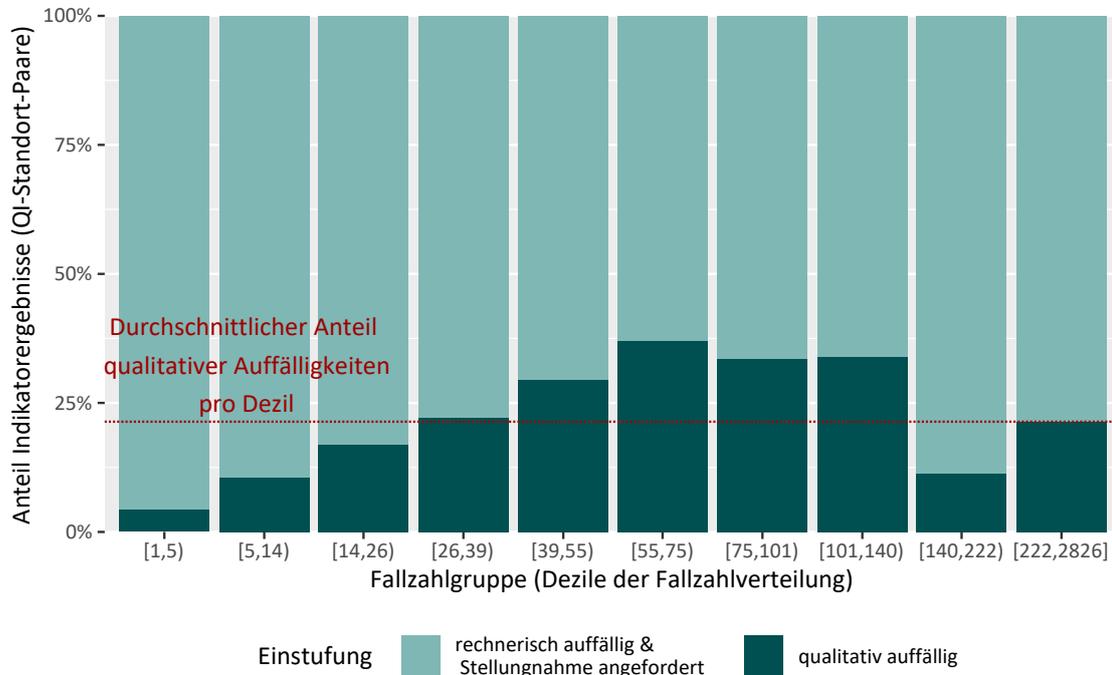


Abbildung 7: Anteil qualitativ auffälliger Standortergebnisse unter allen Ergebnissen, für die eine Stellungnahme angefordert wurde, differenziert nach Fallzahldezilen

### Sentinel-Event-QI

Da es sich um einen Sentinel-Event-Indikator handelt, wird bei einem großen Teil (zwischen 84 % und 100 %) der rechnerischen Auffälligkeiten auch eine Stellungnahme angefordert. Qualitative Auffälligkeiten gibt es nur vereinzelt (vgl. Abbildung 6). Ein potentieller Zusammenhang zwischen qualitativer Auffälligkeit und Fallzahl ist wegen der Seltenheit der qualitativen Auffälligkeiten in diesem Indikator nicht zu bewerten.

#### 2.6.4.3 Ursachen für den Zusammenhang zwischen Fallzahl und Auffälligkeitseinstufung

Die markanteste Beobachtung in Abbildung 5 zum Zusammenhang von Fallzahl und rechnerischer Auffälligkeitseinstufung bei Nicht-Sentinel-Event-Indikatoren ist die überproportionale Häufung von rechnerischen Auffälligkeiten bei kleiner Fallzahl. Wählt man wie in Abbildung 3 die Darstellung eines sog. Funnelplots für die nach Fallzahl differenzierte Darstellung der Standortergebnisse eines einzelnen Qualitätsindikators, so ist typischerweise eine breitere Streuung der Ergebnisse bei kleiner Fallzahl zu beobachten. Beispielsweise kann ein Leistungserbringer mit einer Fallzahl von fünf in einem ratenbasierten Indikator nur die Ergebniswerte 0 %, 20 %, 40 %, 60 %, 80 % und 100 % erreichen. Ein zusätzlicher Fall mit unerwünschtem Behandlungsergebnis (d. h. für den die Zählerbedingung des Indikators erfüllt ist) macht bei dieser Fallzahl somit immer einen Ergebnisunterschied von 20 % aus, während bei steigender Fallzahl der Beitrag

jedes Einzelfalls sinkt. Allerdings haben neben der Qualität der Versorgung durch einen Leistungserbringer auch andere, wie in Abschnitt 12.3 und Kapitel 15 von (IQTIG 2019a) beschriebene, Faktoren einen Einfluss auf das Indikatorergebnis eines Leistungserbringers. Diese werden im Folgenden unter „Stochastizität“ bzw. „statistische Unsicherheit“ subsumiert. Die Konsequenzen des Einflusses dieser stochastischen Komponente auf das beobachtete Indikatorergebnis ist für Leistungserbringer mit kleiner Fallzahl größer als für Leistungserbringer mit großer Fallzahl.

Prinzipiell wird die Möglichkeit weiterer Einflussfaktoren auf das Indikatorergebnis bereits durch die Wahl des Referenzbereiches berücksichtigt. Ein fester Referenzwert von 15 % im Indikator zur präoperativen Verweildauer sieht beispielsweise bereits vor, dass ein Leistungserbringer die Vorgaben zur präoperativen Verweildauer bei 15 % der Patientinnen und Patienten überschreiten darf. Es wird also angenommen, dass für bis zu 15 % der Patientinnen und Patienten Einflussfaktoren vorliegen, die die Überschreitung der präoperativen Verweildauer notwendig machen. Für einen Leistungserbringer mit weniger als sieben Fällen bedeutet der Referenzbereich dennoch, dass ein einziger Fall zur rechnerischen Auffälligkeit führen kann. Die gegenwärtige rechnerische Auffälligkeitseinstufung der QSKH-Richtlinie berücksichtigt nicht, dass die statistische Unsicherheit wesentlich von der Fallzahl abhängt. Es findet somit keine adäquate Berücksichtigung statistischer Unsicherheit statt,<sup>20</sup> was in Teilen die Häufung rechnerischer Auffälligkeiten bei kleinen Fallzahlen erklären kann. Tatsächlich kann man unter einfachen Modellannahmen zeigen, dass die Wahrscheinlichkeit trotz ausreichender Qualität rechnerisch auffällig zu werden (Fehler 1. Art) mit der Fallzahl des Standortes sinkt.<sup>21</sup>

Geht man davon aus, dass ein Grund für die Häufung rechnerischer Auffälligkeiten bei kleinen Fallzahlen im nicht Berücksichtigen statistischer Unsicherheit bei der rechnerischen Auffälligkeitseinstufung liegt, so wäre die Konsequenz eine höhere Zahl falsch positiver Ergebnisse bei kleinen Fallzahlen, was wiederum den beobachteten Zusammenhang zwischen Fallzahl und qualitativer Auffälligkeitseinstufung in Abbildung 7 teilweise erklären könnte.

Die bisherigen Darstellungen zur Fallzahlabhängigkeit der rechnerischen Auffälligkeitseinstufung können nicht unterscheiden, ob die beobachteten Tendenzen reale Zusammenhänge zwischen Fallzahl und Versorgungsqualität darstellen, oder ob es sich um Artefakte des Einstufungsalgorithmus bzw. des Einstufungsprozesses handelt. Die folgende Analyse zeigt, dass die beschriebenen Phänomene maßgeblich mit dem Einstufungsverfahren an sich zusammenhängen. Dafür wird exemplarisch der Qualitätsindikator „Allgemeine Komplikationen bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ (QI-ID 54015) betrachtet. Ziel der Analyse ist es, Differenzen zwischen dem Fallzahl-Ergebnis-Zusammenhang des Einstufungsverfahrens und dem Fallzahl-Ergebnis-Zusammenhang tatsächlicher Versorgungsqualität aufzudecken. Um dies zu erreichen, wird der Fallzahl-Ergebnis-Zusammenhang über alle Einzelfälle im Verfahren mit dem Fallzahl-Ergebnis-Zusammenhang der rechnerischen Auffälligkeitseinstufung

---

<sup>20</sup> Durch die in der §10(2) Satz 4 der QSKH-Richtlinie festgehaltene 1-Fall-Ausnahmeregel wird versucht dies auszugleichen. Die Richtlinie erlaubt in Fällen für die sich die rechnerische Auffälligkeit durch einen einzigen auffälligen Vorgang begründet, auf das Anfordern einer Stellungnahme zu verzichten.

<sup>21</sup> Vgl. Abschnitt 5.3.3

der Standorte gegenübergestellt. Dazu werden die gleichen Daten<sup>22</sup> auf unterschiedlichen Ebenen aggregiert. Durch eine Aggregation aller Einzelfälle, die in Standorten gleicher Fallzahl behandelt wurden, kann eine durchschnittliche Tendenz des tatsächlichen Zusammenhangs zwischen Fallzahl und Versorgungsqualität ermittelt werden. Aggregiert man die rechnerische Auffälligkeitseinstufung aller Standorte gleicher Fallzahl, so lässt sich der Zusammenhang zwischen Fallzahl und rechnerischer Auffälligkeitseinstufung ermitteln. Folgen die so ermittelten Fallzahl-Ergebnis-Zusammenhänge der beiden Aggregationsebenen nicht dem gleichen qualitativen Verlauf, so ist davon auszugehen, dass der Einstufungsalgorithmus systematische Verzerrungen des realen Zusammenhangs von Fallzahl und Versorgungsqualität bewirkt, welcher durch die Aggregation auf Bundesebene approximiert wird.

In Abbildung 8 werden oben die Ergebnisse des Qualitätsindikators auf Fallebene differenziert nach Fallzahl des behandelnden Standortes dargestellt. Grüne Punkte am oberen Rand der Grafik symbolisieren Behandlungsfälle, bei denen allgemeine Komplikationen während des stationären Aufenthaltes aufgetreten sind („Fälle mit interessierendem Ereignis“). Blaue Punkte am unteren Rand der Grafik stellen Vorgänge ohne Komplikationen dar. Die Größe der Punkte ist dabei proportional zur Anzahl an Behandlungsfälle je Fallzahl und Behandlungsergebnis. Durch die gelbe Linie wird dargestellt, wie hoch der modellbasiert geschätzte Anteil an Patientinnen und Patienten mit allgemeinen Komplikation ist, die in Standorten einer gegebenen Fallzahl behandelt wurden. Dieser Grafik liegt dabei ein additives logistisches Regressionsmodell zu Grunde, welches den Zusammenhang zwischen Fallzahl und dem Behandlungsergebnis durch einen Spline modellbasiert interpoliert und glättet (vgl. Wood (2006)). Man erkennt, dass der Anteil an Patientinnen und Patienten mit allgemeinen Komplikationen unabhängig von der Fallzahl bei einem gleichmäßigen Niveau von ca. 12 % liegt.

---

<sup>22</sup> QI-Ergebnisse aus dem Erfassungsjahr 2017 zum Indikator mit der QI-ID 54015

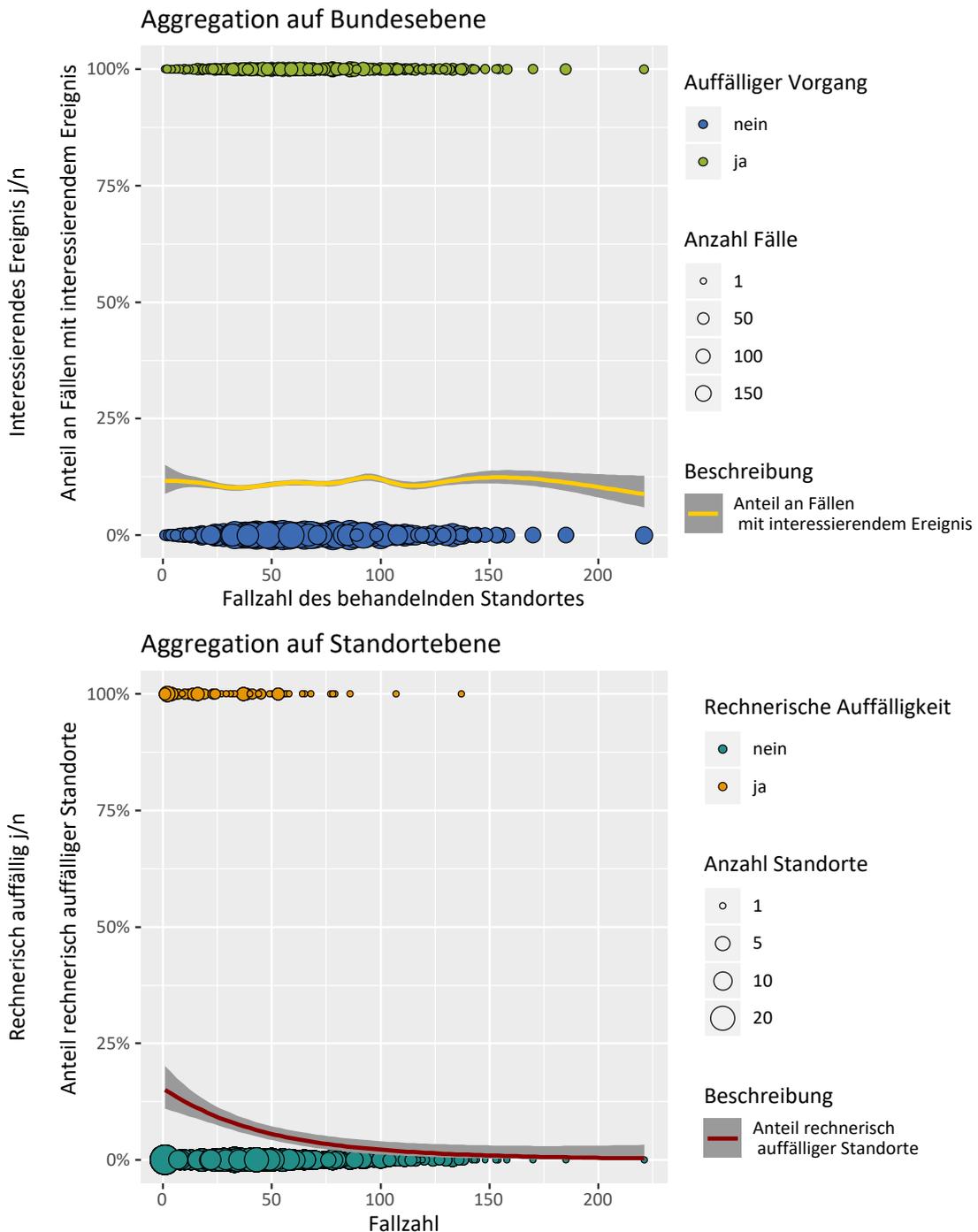


Abbildung 8: Fallzahlabhängigkeit der Behandlungsqualität auf Bundesebene (oben) und der rechnerischen Auffälligkeitseinstufung (unten) zum Qualitätsindikator „Allgemeine Komplikationen bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“

Demgegenüber steht die rechnerische Auffälligkeitseinstufung in Abbildung 8 unten. In dieser Grafik symbolisieren orange Punkte am oberen Rand Standorte, die als rechnerisch auffällig eingestuft wurden. Türkise Punkte am unteren Rand stellen Standorte dar, die nicht rechnerisch auffällig wurden. Die Größe der Punkte ist dabei proportional zur Anzahl an Standorten mit der

gleichen Fallzahl und rechnerischen Auffälligkeitseinstufung. Die rote Linie beschreibt den modellbasiert geschätzten Anteil rechnerisch auffälliger Standorte pro Fallzahl. Ähnlich wie in der oberen Grafik wird diese Linie durch ein additives logistisches Regressionsmodell geschätzt, welches den Zusammenhang zwischen Fallzahl und rechnerischer Auffälligkeitseinstufung durch einen Spline interpoliert und glättet. Zu erkennen ist, dass der Anteil rechnerisch auffälliger Standorte monoton mit steigender Fallzahl fällt. Dies steht im Widerspruch zum Fallzahl-Ergebnis-Zusammenhang auf Bundesebene, der in der oberen Grafik dargestellt ist. Dieses Phänomen wird von George et al. (2017) als „Mismatch of general and individual advice“ bezeichnet. Die obere Grafik legt nahe, dass die Fallzahl bei der Wahl eines Krankenhauses irrelevant ist (*general advice*). Jedoch sind Standorte mit kleinerer Fallzahl häufiger rechnerisch auffällig als Standorte mit größerer Fallzahl, sodass die quantitative Bewertung des einzelnen Standortes bei kleinerer Fallzahl tendenziell schlechter ausfällt als für größere Standorte (*individual advice*). Theoretisch ließe sich dieser Widerspruch durch stark ausgeprägte Heteroskedastizität erklären, d. h. wenn ein starker Zusammenhang zwischen Fallzahl und *Heterogenität* der Versorgungsqualität vorläge, ohne dass die durchschnittliche Versorgungsqualität selbst fallzahlabhängig wäre. Beispielsweise könnte es hypothetische Konstellationen geben, in denen bei Standorten kleiner Fallzahl besonders viele Standorte mit überdurchschnittlicher und unterdurchschnittlicher Versorgungsqualität existieren, während Versorgungsunterschiede bei großer Fallzahl weniger ausgeprägt sind. In diesem Szenario wäre die mittlere Behandlungsqualität für alle Fallzahlen ähnlich, ein erhöhter Anteil auffälliger Standorte bei kleiner Fallzahl aber gerechtfertigt. Dies müsste sich aber auch in einer erhöhten Anzahl qualitativer Auffälligkeiten bei kleiner Fallzahl widerspiegeln, was sich aber nicht in den empirischen Ergebnissen des Strukturierten Dialogs zeigt (vgl. Abbildung 7).

Des Weiteren kann einfach gezeigt werden (vgl. Abschnitt 5.3.3), dass auch ohne Zusammenhang zwischen Fallzahl und Versorgungsqualität die Wahrscheinlichkeit rechnerisch auffällig zu werden mit der Fallzahl sinkt. Daher wird im Folgenden davon ausgegangen, dass der Zusammenhang zwischen Fallzahl und Auffälligkeitseinstufung auf ein Artefakt des Einstufungsalgorithmus zurück zu führen ist, der statistische Unsicherheit nicht berücksichtigt. Bei der Evaluation anderer möglicher Einstufungsverfahren ist zu fragen, ob die Eigenschaft, dass der Einstufungsalgorithmus qualitativ den gleichen Fallzahl-Ergebnis-Zusammenhang wiedergibt, der sich auf Fallebene nachweisen lässt („Fallzahlgerechtigkeit“), als Gütekriterium einzubeziehen ist.

#### 2.6.5 Zwischenfazit

Der Aufwandsüberblick über die 14 Qualitätsindikatoren des HEP-Verfahrens macht deutlich, dass die Effizienzsteigerung im Strukturierten Dialog aus zweierlei Richtungen angegangen werden muss. Einerseits ist zu prüfen, bei welchen Qualitätsindikatoren es indikatorspezifischen Weiterentwicklungsbedarf gibt. Beispielsweise wird im Rahmen eines durch den G-BA erteilten Auftrags die Evidenzgrundlage des Qualitätsindikators zur präoperativen Verweildauer bereits geprüft (G-BA 2019). Auch hat sich im Strukturierten Dialog gezeigt, dass für den Qualitätsaspekt „Sterblichkeit in der Hüftendoprothesenversorgung“ die angewandte Sentinel-Event-Methodik nicht zielführend ist, sodass die Bundesfachgruppe *Orthopädie und Unfallchirurgie* bereits eine

Änderung des Qualitätsindikators vorgeschlagen hat. Andererseits kann eine Effizienzsteigerung durch grundsätzliche methodische Weiterentwicklungen in der rechnerischen Auffälligkeitseinstufung erreicht werden. Hierzu haben die vorangegangenen Analysen gezeigt, dass die derzeitige verwendete Methode der rechnerischen Auffälligkeitseinstufung tendenziell zu viele Standorte mit kleiner Fallzahl als auffällig einstuft, weil statistische Unsicherheit in der Bestimmung der Auffälligkeitseinstufung nicht berücksichtigt wird. Dies führt zu einer hohen Zahl an rechnerischen Auffälligkeiten, die sich dann im Strukturierten Dialog nicht bestätigen – insbesondere bei kleiner Fallzahl.

Problematisch ist auch die Berechnung von perzentilbasierten Referenzwerten und die damit verbundene Auffälligkeitseinstufung, da die gegenwärtig verwendete Methodik dazu führt, dass insbesondere bei Qualitätsindikatoren mit einem hohen Anteil von Standorten mit einer Fallzahl kleiner 20 weit mehr Standorte auffällig werden, als durch das nominell angestrebte Niveau des Perzentilreferenzbereiches.

Grundsätzlich ist die absolute Zahl der im Strukturierten Dialog zu prüfenden Indikatorergebnisse und Stellungnahmen im Vergleich zur Anzahl der Krankenhausstandorte sehr hoch. In Abschnitt 2.5 dieses Berichtes wird thematisiert, dass die Landesstellen für Qualitätssicherung mit dieser Herausforderung sehr unterschiedlich umgehen, was sich beispielsweise darin ausdrückt, dass in einigen Bewertungsstellen häufiger auf das Anfordern einer Stellungnahme verzichtet wird und statt dessen nur ein Hinweis an rechnerisch auffällige Leistungserbringer versendet wird. Die Reduktion des Aufwands im Strukturierten Dialog kann unter diesem Gesichtspunkt auch einen wesentlichen Beitrag zur Reduktion der Heterogenität in den Bewertungen des Strukturierten Dialoges leisten. Gleichzeitig können mit der Reduktion des Aufwands mehr Ressourcen in die Durchführung jedes einzelnen Strukturierten Dialoges fließen.

## 3 Methodischer Hintergrund

Die Qualitätsmessung und -bewertung stellt ein zentrales Element der indikatorgestützten, gesetzlich verpflichtenden Qualitätssicherung dar. Nur auf Basis aussagekräftiger Qualitätsbewertungen sind die wichtigen weiterführenden Elemente, wie etwa qualitätsfördernde Maßnahmen oder die Veröffentlichung von Qualitätsergebnissen, sinnvoll. Ist die Aussagekraft der Qualitätsbewertung reduziert, werden darauf aufbauende Maßnahmen fehlgeleitet. Daher werden in diesem Kapitel die methodischen Kriterien für die Bewertung der Versorgungsqualität mittels Qualitätsindikatoren beschrieben sowie Kriterien für erfolgreiche Qualitätsförderung hergeleitet. Dazu werden zunächst einige Grundannahmen dargestellt.

### 3.1 Qualitätsmessung mittels Qualitätsindikatoren

Gemäß den Methodischen Grundlagen (IQTIG 2019a, Kap. 5) unterscheidet das IQTIG grundsätzlich zwischen Methoden der indikatorbasierten Messung der Versorgungsqualität und qualitativen Verfahren der Qualitätsmessung. Qualitätsindikatoren werden dabei als quantitative Größen verstanden, mittels derer Merkmale der Versorgung gemessen und dargestellt werden (IQTIG 2019a: 42). Sie erlauben eine statistische Aussage über die Erfüllung der Qualitätsanforderungen auf Aggregatebene in Bezug auf diese Merkmale, indem die Messergebnisse des Indikators mit einem Referenzbereich, der die Qualitätsanforderungen repräsentiert, verglichen werden (Soll-Ist-Vergleich<sup>23</sup>). Im Rahmen der Verfahren der datengestützten Qualitätssicherung und unabhängig von ihrer Richtliniengrundlage (QSKH-RL, Qesü-RL und DeQS-RL) repräsentiert die Referenzbereichsgrenze eines Indikators die von jedem Leistungserbringer erwartbare Versorgungsqualität – sie stellt Anforderungen dar, deren Erfüllung von jedem Leistungserbringer erwartet wird. Sofern ein Qualitätsindikator das Qualitätsmerkmal korrekt abbildet, bedeutet die Feststellung einer Abweichung vom Soll daher, dass das geforderte Qualitätsniveau der Versorgung nicht erreicht wurde.

Die indikatorbasierte Messung der Versorgungsqualität hat den Vorteil hoher Objektivität und Effizienz, da auch große Datenmengen anhand einheitlicher Regeln automatisiert aggregiert und ausgewertet werden können. Dementsprechend eignet sich diese Methode der Qualitätsmessung insbesondere für Situationen, in denen die Qualität der Versorgung bei einer Vielzahl von Leistungserbringern nach einheitlichen Bewertungskriterien beurteilt werden soll. Alternativ können auch qualitative Verfahren wie die des Peer-Review, Visitationen oder Audits für die Messung der Versorgungsqualität herangezogen werden. Diese Methoden basieren häufig auf Detailanalysen einzelner Behandlungsfälle (kasuistische Analyse), beispielsweise mittels Dokumentenanalysen (Begutachtung von Patientenakten), Gruppeninterviews (Dialogverfahren) oder Kombinationen davon (z. B. bei Peer-Review-Verfahren). Ein Vorteil solcher Verfahren ist, dass die besonderen Umstände des Einzelfalls berücksichtigt werden können, da eine potenziell

---

<sup>23</sup> Hier ist nicht der Soll-Ist-Abgleich im Rahmen der Soll-Ist-Statistik gemeint. Soll-Ist-Vergleich meint hier ganz allgemein den Vergleich zwischen dem geforderten und dem tatsächlichen Ergebnis des Qualitätsindikators.

größere Informationsgrundlage für die Beurteilung zur Verfügung steht als im Rahmen der indikatorbasierten Messung. Durch den größeren Aufwand, der bei diesen Verfahren aufgrund der weniger spezifischen Auswahl der zu betrachtenden Behandlungsfälle durch die geringeren Automatisierungsmöglichkeiten sowie die größere zu verarbeitende Informationsmenge resultiert, eignen sie sich eher in Fällen, in denen die Qualität weniger Leistungserbringer gemessen werden soll. Auch wenn die Ursachen für ein bestimmtes Indikatorergebnis oder wenn mögliche Maßnahmen zur Qualitätsverbesserung ermittelt werden sollen, eignen sich qualitative Bewertungsmethoden. Durch die geringeren Standardisierungsmöglichkeiten dieser Verfahren spielen allerdings implizite, subjektive Beurteilungskriterien der Durchführenden eine größere Rolle als bei der indikatorbasierten Qualitätsmessung (Donabedian 2003: 64 ff, Berwick 1991: 1218 f.), was zu einer niedrigeren Objektivität der Verfahren und damit eine zu einer Einschränkung der Vergleichbarkeit der Qualitätsmessungen unterschiedlicher Leistungserbringer z. B. für Auswahlentscheidungen.

Sollen die Ergebnisse von Qualitätsmessung auch im Kontext von *accountability* verwendet werden (vgl. Abschnitt 2.2) und Auswahlentscheidungen zwischen Leistungserbringer möglich sein, sind standardisierte Messmethoden wie die indikatorbasierte Qualitätsmessung besser geeignet. Steht ein *improvement*-Ansatz im Vordergrund der Verwendung der Qualitätsmessungen und stehen die Ressourcen für qualitative Verfahren zur Verfügung, beispielsweise, weil die Qualität nur weniger Leistungserbringer zu messen ist und sind Kriterien-basierte Auswahlentscheidungen von geringerer Bedeutung (bspw. in Notfallsituationen), sind qualitative Verfahren, wie Audits oder Peer-Reviews geeignete Methoden. Gemäß Beauftragung adressiert die in den folgenden Kapiteln beschriebenen Methodik die Qualitätsmessung mittels Qualitätsindikatoren.

## **3.2 Die methodische Funktion von Stellungnahmen**

Die Ergebnisse von Messungen mit Qualitätsindikatoren werden nicht nur von der Versorgungsqualität der Leistungserbringer bestimmt, sondern auch von weiteren Einflussfaktoren, die nicht von den Leistungserbringern beeinflussbar sind, wie etwa Patientenrisikoprofile, stochastische Einflüsse oder im Indikator unberücksichtigte Versorgungssituationen (IQTIG 2019a: 181 f.). Um diese Einflussfaktoren auf das Indikatorergebnis bei der Bewertung der tatsächlichen Versorgungsqualität zu berücksichtigen, stehen unterschiedliche Maßnahmen zur Verfügung, die bei der Konstruktion eines Indikators, bei der Festlegung des quantitativen Bewertungskonzepts (Referenzbereich und statistisches Verfahren für Vergleiche mit diesem) oder bei der Anwendung der indikatorgestützten Bewertung auf den einzelnen Leistungserbringer angewandt werden können.

### **3.2.1 Einflussfaktoren auf das Indikatorergebnis**

Bei der Konstruktion von Qualitätsindikatoren wird durch eine angemessene Festlegung der Berechnungsvorschrift und durch Maßnahmen der Risikoadjustierung der Einfluss bekannter, nicht qualitätsbezogener Faktoren auf das Indikatorergebnis so weit wie möglich reduziert. Für die Berücksichtigung von stochastischen Einflüssen auf die Indikatorergebnisse stehen beispielsweise verschiedene statistische Methoden zur Verfügung (siehe Abschnitt 5.1).

In der Praxis kann es aber selbst bei sorgfältiger Entwicklung eines Indikators entlang der Eignungskriterien (IQTIG 2019a: 135 ff. (Kapitel 10) ) dazu kommen, dass das Indikatorergebnis bei manchen Leistungserbringern die Versorgungsqualität nicht angemessen abbildet. Dies kann durch besondere Konstellationen in der Versorgung geschehen, die in der Berechnungsvorschrift des Qualitätsindikators nicht berücksichtigt sind. Solche Konstellationen können zum einen dadurch zustandekommen, dass nicht alle Einflussfaktoren, die zu systematischen Abweichungen vom Qualitätsziel führen können, schon bei der Entwicklung des Indikators bekannt sind. Beispielsweise könnte erst im Routinebetrieb auffallen, dass bestimmte Begleiterkrankungen der Patientinnen und Patienten systematisch zu schlechteren Ergebnissen führen. Zum anderen können auch bekannte Einflussfaktoren im Qualitätsindikator nicht berücksichtigt sein, da ihre Ausprägung nicht für den Indikator erhoben wird. Ein Grund dafür, diese Informationen nicht zu erheben, kann z. B. der hohe Aufwand sein, der bei Erhebung sämtlicher bekannter, auch seltener, Risikofaktoren entstünde, und der dem Gebot der Praktikabilität und Datensparsamkeit entgegensteht (vgl. Abschnitt 2.1). Auch eine unzureichende Objektivität und Reliabilität bestimmter Daten kann dazu führen, dass auf die Erhebung dieser Daten verzichtet wird.

Bei der Festlegung des Referenzbereichs eines Qualitätsindikators werden diese unbekanntes oder nicht gemessenen Einflussfaktoren, die nicht vom Leistungserbringer zu verantworten sind, berücksichtigt. Dies geschieht dadurch, dass die Grenze eines Referenzbereichs nicht auf den Wert auf der Indikatorkala festlegt wird, der idealerweise erreicht würde, d. h. nicht auf 0 % oder 100 % (bei anteilsbasierten Indikatoren). Vielmehr sollte im Referenzbereich eine Abweichung von diesem Idealwert toleriert werden, die mindestens so groß ist wie das geschätzte Ausmaß, das die unbekanntes und nicht gemessenen Einflussfaktoren im Mittel auf die Indikatorergebnisse haben.<sup>24</sup>

### 3.2.2 Prüfung der Validität je Messung

Treten unbekanntes oder nicht gemessene Einflussfaktoren, die nicht in der Verantwortung der Leistungserbringer liegen, bei einem Leistungserbringer häufiger auf, als dies im Referenzbereich berücksichtigt ist, kann der Indikatorwert dieses Leistungserbringers den Referenzbereich verfehlen, obwohl die Versorgungsqualität zureichend ist. Es liegt dann eine Konstellation vor, in der der Abgleich zwischen Indikatorwert und Referenzbereich die Erfüllung der Qualitätsanforderungen für diesen Leistungserbringer nicht angemessen abbildet – die Validität des Indikators ist damit in diesem Fall in Zweifel gezogen. Um solche besonderen Konstellationen zu berücksichtigen, die nicht schon durch die Konstruktion des Indikators und sein Bewertungskonzept berücksichtigt sind, kann ein Stellungnahmeverfahren durchgeführt werden. Aus dieser Perspektive besteht die Funktion eines solchen Stellungnahmeverfahrens darin zu klären, ob das Verfehlen des Referenzbereichs durch einen Leistungserbringer nicht von diesem zu verantworten war. Diese Funktion eines Stellungnahmeverfahrens entspricht damit der Prüfung der Validität des Indikatorergebnisses im Einzelfall. Leistungserbringerübergreifend

---

<sup>24</sup> Zusätzlich wird meist auch toleriert, dass nicht immer optimale Qualität von den Leistungserbringern verlangt wird, sondern es werden nur von allen erwartbare Standards gefordert. Dies ist für die fachliche Bewertung von Bedeutung, siehe Abschnitt 6.2 (wird dort näher erläutert).

wird die Validität der Messung mittels Qualitätsindikatoren dadurch sichergestellt, dass ihre Entwicklung entsprechend den Eignungskriterien für Qualitätsmessung erfolgt (IQTIG 2019a: 135 ff. [Kapitel 10]). Das Stellungnahmeverfahren erfüllt damit die in Abschnitt 2.1 beschriebene Funktion, die Spezifität des Gesamtverfahrens (indikatorbasierte Messung einschließlich Prüfschritt 2) zur Feststellung von Qualitätsdefiziten durch Erhöhung der Validität zu steigern.

Aus dieser Perspektive ist es nicht Funktion des Stellungnahmeverfahrens, die Indikatorwerte für Leistungserbringervergleiche, zum Beispiel in Form von Rangreihenfolgen, zu korrigieren. Eine solche Korrektur der Indikatorwerte ist nicht sinnvoll, da dies nur für diejenigen Leistungserbringer möglich wäre, die eine Stellungnahme abgegeben haben. Damit ein fairer Vergleich zwischen Leistungserbringern anhand ihrer Indikatorwerte vorgenommen werden kann, müssten sowohl die Indikatorwerte der Leistungserbringer, die eine Stellungnahme abgegeben haben, als auch diejenigen, die keine Stellungnahme abgegeben haben, korrigiert werden. Zudem stehen in den QS-Verfahren nach QSKH-RL, DeQS-RL und plan. QI-RL Aussagen über die Erfüllung von Standards durch einen Leistungserbringer im Vordergrund.

Da es bei der Anwendung der Qualitätsindikatoren auf einzelne Leistungserbringer besondere Konstellationen geben kann, die in der standardisierten Qualitätsdarstellung anhand des Indikators nicht berücksichtigt sind, besteht die Funktion des Stellungnahmeverfahrens für die indikatorgestützte Qualitätsmessung darin, die Angemessenheit der Indikatoraussage für den jeweiligen Leistungserbringer zu prüfen. Die Prüfung der Validität der Messung für einen Leistungserbringer im Rahmen des Stellungnahmeverfahrens muss also die Frage beantworten, ob die Schlussfolgerung vom Indikatorergebnis auf die Erfüllung des Qualitätsmerkmals auch für die konkrete Konstellation dieses Leistungserbringers angemessen ist und sie alle wichtigen Einflussfaktoren, die nicht vom Leistungserbringer beeinflussbar sind, berücksichtigt. Ausgehend von dieser Funktion werden im Folgenden Kriterien für die Durchführung eines solchen Stellungnahmeverfahrens hergeleitet.

Stellungnahmen zu Indikatorergebnissen können prinzipiell und wurden vermutlich im bisherigen Strukturierten Dialog auch als Instrument der Qualitätsförderung verstanden. Durch die Aufforderung zur Stellungnahme zu einem „auffälligen“ Indikatorergebnis werden die Leistungserbringer aufgefordert, sich mit der Qualität ihrer Versorgung selbstkritisch auseinanderzusetzen. Im Sinne der *improvement* Strategie (Berwick et al. 2003) kann dies zur selbstständigen Identifikation von Qualitätsdefiziten führen und eine Umgestaltung der Versorgung durch die Leistungserbringer anregen, um Versorgungsqualität zu verbessern. Da jedoch die Indikatorergebnisse nicht nur für diesen Zweck, sondern auch im Rahmen von *accountability* verwendet werden, ist es wichtig, dass diese fair und aussagekräftig sind. Daher wird empfohlen, das Stellungnahmeverfahren als Schritt zur Prüfung der Validität der Messung des jeweiligen Indikatorergebnisses zu verstehen. Aus diesem Verständnis leiten sich die in Abschnitt 3.4 dargelegten Anforderungen an das Stellungnahmeverfahren ab.

### 3.3 Reichweite der Qualitätsaussage von Stellungnahmen

Gemäß der oben hergeleiteten Funktion von Stellungnahmen bezieht sich die fachliche Bewertung im Stellungnahmeverfahren auf das vom Indikator adressierte Qualitätsmerkmal. Es wird

nicht als Aufgabe des Stellungnahmeverfahrens verstanden, eine umfassende Qualitätsbewertung des gesamten Behandlungsprozesses vorzunehmen, zu dem das dem jeweiligen Indikator zugrundeliegende Qualitätsmerkmal gehört. So ist es nicht Ziel der fachlichen Bewertung, in den Behandlungsfällen, die zu einem auffälligen Indikatorergebnis geführt haben, grobe Behandlungsfehler zu finden, wenn diese keinen Bezug zum dargestellten Qualitätsmerkmal haben. Beispielsweise ist es im Rahmen der fachlichen Bewertung von Stellungnahmen zu einem Indikator zur Komplikationshäufigkeit nicht Ziel, unterlassene Aufklärungen in die Bewertung miteinzubeziehen, selbst wenn sie anhand einer Stellungnahme zu Tage treten. Dieses Vorgehen entspricht eher dem Verständnis von Qualitätsindikatoren als Aufgreifkriterien für einen nachfolgenden, gegebenenfalls umfassenden Qualitätsbewertungsprozess.

Für diese Beschränkung auf die jeweiligen Qualitätsmerkmale gibt es mehrere Gründe. Zum einen können Qualitätsmessungen Qualität nie vollumfänglich abbilden, sie können sich immer nur auf bestimmte Aspekte fokussieren. In der externen Qualitätssicherung sind dies die Qualitätsaspekte des Qualitätsmodells (Kap. 5.2 IQTIG 2019a: 44 ff.). Nach dem oben geschilderten Verständnis von Qualitätsindikatoren soll ein Qualitätsindikator zur Sterblichkeit bei akutem Herzinfarkt nicht die Qualität der Infarktbehandlung insgesamt beschreiben, sondern nur darstellen, ob die Sterblichkeit von Patientinnen und Patienten mit diesem Krankheitsbild bei einem Leistungserbringer zu hoch ist. Im Gegensatz zu Qualitätsindikatoren, für die das geforderte Qualitätsniveau über den Referenzbereich eindeutig definiert ist und durch Beschluss des G-BA normativen Charakter hat, liegen für die Gesamtqualität einer Behandlung keine standardisierten Beurteilungskriterien vor. Die Gesamtqualität der Behandlung korreliert auch nicht zwingend mit der Qualität in dem von einem einzelnen Indikator abgebildeten Merkmal. Vielmehr können verschiedene Aspekte des Behandlungsprozesses von einem Leistungserbringer mit unterschiedlicher Qualität durchgeführt werden (Shwartz et al. 2011). Die quantitative Auffälligkeit eines Indikatorergebnisses würde zudem bei einer über das Qualitätsmerkmal hinausgehenden Bewertung durch die Expertinnen und Experten möglicherweise als erhöhte A-priori-Wahrscheinlichkeit für eine unzureichende Gesamtqualität der Behandlung wahrgenommen und könnte daher zu einer Verzerrung bei der Beurteilung der Gesamtqualität führen.

Würden mit dem Stellungnahmeverfahren eine eigenständige Bewertung der Gesamtqualität angestrebt und würden Indikatoren lediglich als Aufgreifkriterium genutzt (siehe Abschnitt 2.2), wäre ein separates, von der indikatorbasierten Qualitätsmessung weitgehend unabhängiges Verfahren nötig (z. B. ein Auditverfahren). Die Bewertungskriterien, die für ein solches Verfahren benötigt würden, hätten nicht unbedingt einen direkten Bezug zu den konsentierten Anforderungen und Qualitätsmodellen der bisherigen indikatorbasierten QS-Verfahren. Die Qualitätsmessung mit Indikatoren ist zudem als ein quantitatives Verfahren bevorzugt für Themen mit nennenswerten Fallzahlen und Ereigniszahlen geeignet (IQTIG 2019a, S. 42 f.). Daraus folgt, dass die im Stellungnahmeverfahren von den Einrichtungen zur Verfügung gestellten Informationen über die Behandlung das Behandlungsgeschehen nicht vollumfänglich beschreiben können, sondern nur eine Auswahl an Informationen je geschilderten Behandlungsfall darstellen können. Diese Einschränkung der Informationsgrundlage steht einer angemessenen Gesamtbeurteilung eines Behandlungsfalls entgegen. Darüber hinaus müsste eine zuverlässige Gesamtaussage über die Behandlungsqualität bei einem Leistungserbringer alle Fälle des Leistungserbringers beim

interessierenden Behandlungsprozess oder wenigstens eine repräsentative Zufallsstichprobe dieser Fälle betrachten. Dieses Vorgehen würde im Vergleich zum bisherigen Vorgehen, bei dem nur die Fälle mit unerwünschtem Ereignis betrachtet werden (siehe Abschnitt 2.4), den Aufwand vervielfachen.

### 3.4 Gütekriterien für Bewertungsprozesse

Der Prozess der Bewertung in einem Stellungnahmeverfahren kann analog zum diagnostischen Prozess der Urteilsbildung verstanden werden, bei dem eine Vielzahl von Informationen von einem Experten, einer Expertin oder einem Expertengremium<sup>25</sup> berücksichtigt und zu einem Urteil verarbeitet werden (Schmidt-Atzert und Amelang 2012, Krohne und Hock 2015). Beispiele außerhalb des hier betrachteten Bewertungsprozesses des Stellungnahmeverfahrens sind die klinische Diagnosestellung oder die Entscheidung, ob ein Bewerber für eine Stelle geeignet ist. In diesen beiden Situationen fällen Expertinnen und Experten jeweils ein Urteil auf Basis einer gegebenen Informationsgrundlage und nach bestimmten, expliziten oder impliziten Entscheidungsregeln. Da die wissenschaftliche Literatur zur Urteilsbildung unter dem Begriff „klinische Urteilsbildung“ Fragestellungen untersucht hat, die mit dem Vorgehen der Qualitätsbewertung im Stellungnahmeverfahren verwandt sind, soll die Ähnlichkeit beider Sachverhalte genutzt werden, um auf Basis der Erkenntnisse aus der Urteilsforschung eine methodisch begründete Empfehlung für das Stellungnahmeverfahren zu entwickeln.

In der wissenschaftlichen Literatur zur Urteilsbildung wird zwischen sogenannter klinischer oder intuitiver Urteilsbildung und sogenannter statistischer oder „mechanischer“ Urteilsbildung unterschieden (Krohne und Hock 2015, Westen und Weinberger 2004). Im Gegensatz zur mechanischen Urteilsbildung, bei der Informationen nach formalen Regeln zu einem Urteil verarbeitet werden, integriert die Expertin oder der Experte bei der klinischen/intuitiven Urteilsbildung die Informationen in ihren/seinen Denkprozess. Hier zeigte die Urteilsforschung, dass die Güte der intuitiven Urteilsbildung, die auf impliziten Regeln beruht, der Güte der mechanischen Urteilsbildung auf Basis expliziter und konsistent angewandter Regeln unterlegen ist (Westen und Weinberger 2004, Krohne und Hock 2015, Dawes et al. 1989). Als Gründe werden einerseits die begrenzte und fehleranfällige menschliche Informationsverarbeitung und andererseits die flexible Anpassung von impliziten Regeln auf den Einzelfall angeführt. Es wird vermutet, dass sich bei der intuitiven Urteilsbildung individuelle Besonderheiten aufdrängen, die zu einem Abweichen von der (impliziten) Regel führen bzw. die dazu führen, dass die Auswahl der Regel von der Entscheidungssituation und von Eigenschaften des beurteilenden Individuums abhängt (Kruglanski und Gigerenzer 2011). Im Rahmen von beispielsweise Auswahl- und Zulassungsentscheidungen für Studienplatz- und Stellenbewerberinnen und -bewerber scheint diese intuitive Auswahl der Entscheidungsregel über viele Fälle hinweg zu ungünstigeren Ergebnissen zu führen als die mechanische Urteilsbildung (Kuncel et al. 2013).

---

<sup>25</sup> In der DeQS-RL wird der Begriff Fachkommission verwendet. Hier wird jedoch der allgemeinere Begriff „Expertengremium“ verwendet.

Bei der wissenschaftlichen Analyse des Prozesses der Urteilsbildung stellte sich die Unterscheidung zwischen der Art der zugrunde gelegten Informationen und der Art der Informationsverknüpfung als hilfreich heraus (Krohne und Hock 2015, Westen und Weinberger 2004). Durch explizite Vorgabe und konsistente Anwendung einer einheitlichen Entscheidungsregel auf der Basis einer standardisierten Informationsgrundlage kann die Güte von Entscheidungen erhöht werden (Westen und Weinberger 2004). Übertragen auf das Stellungnahmeverfahren bezieht sich die Art der zugrunde gelegten Informationen auf die Inhalte, die von den Expertinnen und Experten bei ihrer Qualitätsbewertung berücksichtigt werden dürfen. Die Art der Informationsverknüpfung bezieht sich auf die Regeln und Kriterien, nach denen die vorliegenden Informationen von dem Expertengremium zu einem Urteil (hinreichender vs. kein hinreichender Hinweis für unzureichende Qualität) verknüpft werden.

### 3.5 Ableitung von Anforderungen an das Stellungnahmeverfahren

Bislang existieren weder in der QSKH-RL noch in der DeQS-RL Regelungen, auf welche Art das Stellungnahmeverfahren (schriftliche Stellungnahme, Gespräch oder Begehung) durchzuführen ist und nach welchen Regeln die Bewertungsentscheidungen getroffen werden sollen. Abgeleitet aus der Forschung zur Urteilsbildung ist davon auszugehen, dass sowohl die Art des Stellungnahmeverfahrens (durch die unterschiedliche Informationsgrundlage) als auch die Abwesenheit expliziter Bewertungsregeln die Bewertung der Indikatorergebnisse beeinflusst. Um ein einheitliches Vorgehen bei der Qualitätsbewertung für alle Leistungserbringer auf der Grundlage gleicher Informationen zu gewährleisten, sollte daher eine Standardisierung des Stellungnahmeverfahrens erfolgen. Dies bedeutet, eine einheitliche Informationsgrundlage für die Bewertung der Stellungnahmen sicherzustellen. Des Weiteren sollten die Entscheidungsregeln für die Bewertung möglichst explizit sein, um zu verhindern, dass diese innerhalb und zwischen den urteilenden Personen bzw. Gruppen variieren.

Diese Anforderungen eines möglichst standardisierten, expliziten Vorgehens bei der Einholung, des Inhaltes und der Bewertungsregeln von und für Stellungnahmen lassen sich auch aus dem Gütekriterium der Objektivität ableiten. „Objektivität bedeutet, dass die Ergebnisse eines diagnostischen Verfahrens unabhängig davon zustande kommen, wer die Untersuchung, die Auswertung und die Interpretation durchführt“ (Schmidt-Atzert und Amelang 2012, S. 133). Sind Messergebnisse nach dieser Definition nicht objektiv, sind sie wenigstens zum Teil davon abhängig, wer die Messung, die Auswertung oder die Interpretation der Ergebnisse vornimmt. Objektivität ist darüber hinaus eine notwendige, aber nicht hinreichende Bedingung für die Reliabilität von Messergebnissen. Wird also (ausreichende) Reliabilität von dem Vorgehen im Strukturierten Dialog gefordert, muss zunächst eine ausreichende Objektivität sichergestellt werden.

Wird nun an der bisherigen Umsetzung des Strukturierten Dialogs eine starke Heterogenität bemängelt (siehe Beschluss zur Beauftragung sowie Abschnitt 2.5), bedeutet dies im Sinne dieser Definition, dass die Ergebnisse des Strukturierten Dialogs nicht objektiv genug sind. Dies bedeutet, dass neben der tatsächlichen Versorgungsqualität eines Leistungserbringers auch eine Rolle spielt, welche Stelle die Stellungnahme einholt, wie diese formuliert ist und welche Fachkommission diese bewertet. Diese Interpretation wird gestützt durch die Ergebnisse der empirischen

Analysen in Abschnitt 2.5, die zeigen, dass bewertungsstellenabhängige Heterogenität im Vorgehen vorliegt. Um die in der Beauftragung genannten Ziele einheitlicher, transparenter und nachvollziehbarer Qualitätsergebnisse zu erreichen, sollten demzufolge für die Prozessschritte der Qualitätsbewertung explizite Regeln vorgegeben werden (Standardisierung).

Es wird deutlich, dass ein explizites Vorgehen bei der Urteilsbildung wichtige Vorteile gegenüber einem impliziten Vorgehen hat und dass das explizite Vorgehen eine Voraussetzung für das in der Beauftragung angestrebte Ziel einer einheitlichen und nachvollziehbaren Bewertung ist.

### **3.6 Vor- und Nachteile explizierter Beurteilungsregeln im Stellungnahmeverfahren**

Für das Stellungnahmeverfahren ist vor allem die Bewertung der zu einem Indikatorergebnis vorgebrachten Argumente von zentraler Bedeutung. Ein implizites Vorgehen würde bedeuten, dass die Argumente qualitativ durch die Expertinnen und Experten abgewogen würden. Für ein explizites Vorgehen würden vorab eine Rechenvorschrift und Gewichte für die Einflussfaktoren festgelegt werden, anhand derer eine Nachberechnung des Indikatorwerts für diese Einrichtung und ein erneuter Vergleich mit einem ggf. angepassten Referenzbereich durchgeführt würde. Explizites und implizites Vorgehen gehen mit unterschiedlichen Konsequenzen für die Umsetzung einher.

#### **Vorteile expliziter Beurteilungsregeln bei der fachlichen Bewertung**

Ein explizites Vorgehen bei der Bewertung der Einflussfaktoren in einem Stellungnahmeverfahren hat folgende Vorteile:

- Ein explizites Vorgehen einschließlich Nachberechnung des Indikatorergebnisses würde die Beurteilung als „Qualitätsdefizit“ bzw. „kein Hinweis auf ein Qualitätsdefizit“ stark standardisieren und damit die Objektivität des Bewertungsverfahrens steigern.
- Durch ein möglichst explizites Vorgehen kann ggf. die implizite Tendenz minimiert werden, in der fachlichen Beurteilung über das eigentlich zu beurteilende Qualitätsmerkmal hinauszugehen (siehe Abschnitt 3.3).
- Durch den Vergleich eines bereinigten Indikatorwerts mit einem (angepassten) Referenzbereich würde berücksichtigt, dass der Referenzbereich eines Indikators in der externen Qualitätssicherung durch Beschluss des G-BA normative Wirkung hat.
- Beim expliziten Vorgehen sind die Entscheidungsregeln transparent und ihr Einhalten prinzipiell überprüfbar (Dawes et al. 1989).
- Die Berechnung eines bereinigten Indikatorwerts ist auch mit statistischer Unsicherheit behaftet, die durch den Ausschluss von Behandlungsfällen aus der Grundgesamtheit im Vergleich zum ursprünglichen Indikatorwert noch erhöht sein kann. Diese Unsicherheit liegt impliziten Beurteilungen gleichermaßen zugrunde, würde aber durch ein explizites Vorgehen transparent gemacht.

### Limitationen expliziter Beurteilungsregeln bei der fachlichen Bewertung

Gegen ein explizites Vorgehen in der Beurteilung im Rahmen eines Stellungnahmeverfahrens sprechen folgende Überlegungen:

- Durch explizite Regeln in der Beurteilung wird das Vorgehen deutlich komplexer: Wo vorher ein summarisch-implizites Urteil gefällt wurde, müssten dann mehrere Prozesse durchgeführt werden, wie eine explizite Bewertung jedes genannten Einflussfaktors, eine Nachberechnung des Indikatorwerts, ein Vergleich mit dem Referenzwert etc. Viele dieser Schritte können nicht ohne weitere technische Unterstützung durchgeführt werden.
- Für ein explizites Vorgehen müssten bereits A-priori-Bewertungsregeln für alle möglichen Einflussfaktoren vorliegen. Es ist aber charakteristisch für die zu bewertenden besonderen Konstellationen, dass eben nicht alle möglichen Einflussfaktoren vor Beginn des Stellungnahmeverfahrens bekannt sind.
- Es gibt Konstellationen, in denen die verfügbaren Informationen nicht ausreichen, um Entscheidungsregeln und Berechnungsvorschriften festzulegen oder diese sinnvoll anzuwenden.
- Eine Umstellung von einem eher impliziten auf ein expliziteres Vorgehen könnte für klinische Fachexpertinnen und Fachexperten, die berufsbedingt eher nach einem implizit-erfahrungsbasierten Kognitionsmuster vorgehen (Djulfbegovic et al. 2014), ungewohnt sein.

Mit einer Steigerung der Objektivität des Stellungnahmeverfahrens geht auch eine gewisse Steigerung der Komplexität des Verfahrens einher. Je mehr Regeln explizit formuliert werden, desto mehr Regeln müssen auch umgesetzt werden. Dies weist auf einen Konflikt hin zwischen den in der Beauftragung genannten Zielen einer einheitlichen und transparenten (d. h. objektiven) Vorgehensweise einerseits und einem schlanken (d. h. aufwandsarmen) Verfahren andererseits hin. Eine aufwandsarme Qualitätsmessung kann im Allgemeinen vor allem mittels (automatisierbarer) quantitativer Methoden, wie der Indikator-basierten Qualitätsmessung im Vergleich zu qualitativen Verfahren, wie etwa Peer-Reviews oder Begehungen, erreicht werden (vgl. IQTIG 2019a, Kapitel 5.1)<sup>26</sup>. Sollen die Ergebnisse quantitativer Verfahren in einem zweiten qualitativen Prüfschritt analysiert werden, wird damit der Effizienzgewinn des quantitativen Verfahrens wenigstens teilweise aufgegeben. Werden an den zweiten qualitativen Prüfschritt auch noch ähnlich hohe Anforderungen an Objektivität angelegt wie an das quantitative Verfahren, erhöht sich der Aufwand zusätzlich. Außerdem setzt die qualitative Informationsgrundlage von Stellungnahmen Grenzen für die Standardisierung des Vorgehens. Aus dieser Gegenüberstellung wird deutlich, dass die methodischen Anforderungen, die sich aus dem beauftragten Ziel der Standardisierung ergeben, sowie das Ziel einer „Verschlankung des Verfahrens“ unter Beibehaltung eines qualitativen Prüfschritts gegeneinander abgewogen werden müssen. Dies erfolgt in Form von konkreten Empfehlungen in Kapitel 6.

---

<sup>26</sup> Neben dem hier gemeinten Aufwand für die Datenaus- und -bewertung fällt auch Aufwand für die Datenerhebung an. Für alle Daten, die nicht in Form von Sozialdaten, Patientenbefragungsdaten oder Daten, die automatisiert aus dem KIS oder der PVS auslesbar sind, fällt ein bedeutsamer Aufwand für die Erhebung bei den Leistungserbringern an.

### 3.7 Methodische Einordnung der Qualitätsförderung

Neben der Qualitätsbewertung und auf ihr aufbauend ist die Qualitätsförderung ein weiteres wichtiges Element des Strukturierten Dialogs. Wird ein Leistungserbringer nach der abschließenden Qualitätsbewertung als qualitativ auffällig eingestuft, ist das vorrangige Ziel der Qualitätsförderung, das einrichtungsinterne Qualitätsmanagement darin zu unterstützen, den Qualitätsmangel zu beheben.

Zu den Instrumenten externer Qualitätsförderung zählen neben dem Strukturierten Dialog klinische Audits, Visitationen, Audit-Feedback-Verfahren und Qualitätszirkel. Alle Instrumente ähneln dem Instrument des Peer-Review-Verfahrens. Auch der Strukturierte Dialog kommt dem Peer-Review am nächsten (Hensen 2016, Griem et al. 2013). Eine Evaluation des Strukturierten Dialogs wurde bislang nicht durchgeführt, sodass derzeit keine wissenschaftlichen Daten zu sogenannten „Erfolgsfaktoren“<sup>27</sup> für den Strukturierten Dialog vorliegen. Um dennoch eine methodische Einordnung der qualitätsfördernden Maßnahmen des Strukturierten Dialogs vornehmen zu können, wird auf die Literatur zu Erfolgsfaktoren für die Gruppe der Peer-Review-Verfahren zurückgegriffen (BÄK 2014, Chop und Eberlein-Gonska 2012, Gerlach 2001, Hudson et al. 2012, Rink 2013). Diese sollen hier herangezogen werden, um eine Einordnung des Strukturierten Dialogs im Hinblick auf den Erfolg der qualitätsfördernden Maßnahmen vorzunehmen und daraus Empfehlungen für die Weiterentwicklung abzuleiten. Bei dieser Einordnung und der Ableitung von Empfehlungen ist zu berücksichtigen, dass der Strukturierte Dialog durch die Richtlinienvorgaben einigen grundsätzlichen Regelungen unterlegen ist. Daher können manche Erfolgsfaktoren nur bedingt auf den Strukturierten Dialog übertragen werden. Die Erfolgsfaktoren beziehen sich nicht nur auf die Umsetzung der jeweils gezielten Maßnahme zur Behebung eines Qualitätsdefizits (kollegiales Gespräch oder Begehung), sondern auf alle Phasen des Peer-Review und betreffen damit alle Elemente des Strukturierten Dialogs. Im Folgenden werden Erfolgsfaktoren beschrieben, die durch Optimierung der Qualitätsbewertung eine Qualitätsförderung positiv beeinflussen.

In der Literatur zu Erfolgsfaktoren für Peer-Review-Verfahren werden die fachliche Kompetenz der Peers sowie die Zusammensetzung unabhängiger, interdisziplinärer Peer-Review-Teams als wichtige Erfolgsfaktoren für qualitätsfördernde Maßnahmen genannt (Eberlein-Gonska et al. 2017, Chop und Eberlein-Gonska 2012, Gerlach 2001, BÄK 2014, Rink 2013, Hudson et al. 2012). Um für die Qualitätssicherung ein Höchstmaß an Glaubwürdigkeit und Akzeptanz zu erreichen, ist die Glaubwürdigkeit und Kompetenz der Peers von Bedeutung (DeGEval 2008). Zudem ist davon auszugehen, dass durch eine entsprechende Expertise leichter eine Verknüpfung zwischen dem Einsatz der Qualitätsindikatoren und den Prozessen und Strukturen der Leistungserbringer hergestellt werden kann. Übertragen auf den Strukturierten Dialog und in Bezug auf seine Akzeptanz bei allen Beteiligten bedeutet das zum einen, dass die medizinische bzw. fach-

---

<sup>27</sup> Neben dem Begriff Erfolgsfaktor kommen in der Literatur auch die Begriffe Erfolgs- oder Qualitätskriterien vor. Diese Begriffe werden in der entsprechenden Literatur nicht trennscharf verwendet, ihnen allen ist jedoch gemein, dass es sich um Faktoren handelt, die einen positiven Einfluss auf das Gelingen qualitätsfördernder Maßnahmen haben.

liche Qualifikation und die praktische Erfahrung im jeweiligen Versorgungsbereich zentrale Auswahlkriterien für die Mitglieder der Fachkommission sein sollten. Zum anderen sollte die Zusammensetzung der Fachkommission alle wesentlichen, an der Versorgungsqualität beteiligten Berufsgruppen im zu bewertenden QS-Verfahren (Fachärztinnen und Fachärzte, Gesundheitsfachberufe und ggf. Hygieniker) involvieren, um eine interdisziplinäre Zusammensetzung zu gewährleisten. Da Interessenkonflikte das Urteilsvermögen verzerren können, kann es zu Fehleinschätzungen kommen (Lieb et al. 2011). Daher sollten für einen möglichst unabhängigen Bewertungsprozess bei der Wahl der Expertinnen und Experten deren Interessenkonflikte bekannt und geprüft worden sein.

Aus der Analyse der Ausgangssituation geht hervor, dass aktuell nicht einheitlich oder gar nicht geregelt ist, welche spezifischen Qualifikationen und wieviel Erfahrung notwendig sind, um als Expertin oder Experte ausgewählt zu werden. In Bezug auf die Zusammensetzung ist in der QSKH-RL festgelegt, dass die Fachkommissionsmitglieder von den jeweiligen Interessenvertretungen zu berufen sind. Durch diese Regelung alleine kann jedoch nicht von einer unabhängigen und interdisziplinären Zusammensetzung ausgegangen werden. Um den Erfolg und die Akzeptanz des Verfahrens durch eine möglichst objektive Bewertung der Versorgungsqualität zu steigern, sollten die genannten Aspekte bei dem Auswahlprozess der Mitglieder berücksichtigt werden. Zudem sollte für eine größere Akzeptanz des Verfahrens die interdisziplinäre Perspektive durch die Perspektive der Patientinnen und Patienten, die die eigentlichen Adressaten der externen Qualitätssicherung sind, ergänzt werden. Das bedeutet, dass auch Vertreterinnen und Vertreter der Patientinnen und Patienten als Mitglieder in das Expertengremium aufgenommen werden sollten.

Eine vorhandene *Bewertungssystematik* mit klaren Festlegungen z. B. zu den Bewertungsinhalten und zum Bewertungsvorgehen ist ein weiterer wesentlicher Erfolgsfaktor für qualitätsfördernde Maßnahmen (BÄK 2014, Chop und Eberlein-Gonska 2012, Rink 2013, Hudson et al. 2012). Welche methodischen Anforderungen an beide Faktoren zu stellen sind, wurde bereits in den vorherigen Abschnitten beschrieben. Zu ergänzen ist, dass die Bewertungssystematik einen großen Einfluss auf die Qualitätsförderung besitzt. Die Analyse der Ausgangssituation zeigt, dass bei der Umsetzung qualitätsfördernder Maßnahmen (wie z. B. dem kollegialen Gespräch) die Bewertung der Versorgungsqualität oft noch nicht abgeschlossen ist. Wurde noch keine Bewertung vorgenommen ist davon auszugehen, dass die offene Selbstreflexion des Leistungserbringers bezüglich bestehender Qualitätsdefizite unter dem Druck der noch ausstehenden Bewertung eingeschränkt sein könnte. Somit können qualitätsfördernde Maßnahmen nicht an den relevanten Problemen ansetzen und würden folglich fehlgeleitet. Für eine effiziente Qualitätsförderung ist demnach der Bewertungsvorgang abzuschließen, bevor qualitätsfördernde Maßnahmen eingeleitet werden.

Ein weiterer Erfolgsfaktor, der als Voraussetzung für den Erfolg von Peer-Reviews hervorgehoben wird, ist die *Sanktionsfreiheit* innerhalb des Verfahrens. In der Literatur wird beschrieben, dass durch ein *bedrohungsfreies Klima* die Bereitschaft der Betroffenen erhöht wird, die im Praxisalltag auftretenden Probleme offen anzusprechen und konstruktiv damit umzugehen. Zudem wird die Motivation, Veränderungen einzuleiten, erhöht (BÄK 2014, Hudson et al. 2012). Diese

Voraussetzung ist für den Strukturierten Dialog nicht gegeben, da die Bewertung eines Leistungserbringers bei Versorgungsmängeln eine Art *Sanktion* darstellt, insbesondere da qualitativ auffällige Ergebnisse für die Öffentlichkeit zugänglich und verpflichtend zu berichten sind. Zudem gibt die Richtlinie vor, dass bei einem qualitativ auffälligen Ergebnis weiterführende Maßnahmen einzuleiten sind. Dementsprechend ist der Strukturierte Dialog kein sanktionsfreies Verfahren. Einheitliche Regelungen, die die Bewertung der Versorgungsqualität strikt von der Qualitätsförderung trennen, würden jedoch eine freiere Qualitätsförderung im Anschluss an die Bewertung ermöglichen und könnten damit den Erfolg der Maßnahmen steigern.

Des Weiteren ist eine Regelung zur *Vertraulichkeit und zum Datenschutz* ein relevanter Erfolgsfaktor (BÄK 2014). Im Hinblick z. B. auf kollegiale Gespräche und Begehungen ist davon auszugehen, dass nur in einem vertraulichen Rahmen die Leistungserbringer eine selbstkritische Analyse vornehmen werden. Zudem wird auch durch die Einhaltung klarer Regelungen zum Datenschutz die Akzeptanz des Gesamtverfahrens gesteigert. Das führt dazu, dass qualitätsfördernde Maßnahmen gewinnbringender umgesetzt werden können. Klare und einheitliche Regelungen zum Datenschutz sind im Rahmen des Strukturierten Dialogs bereits für die Qualitätsbewertung anhand von Stellungnahmen von Bedeutung. In der Analyse der Ausgangssituation wurde beschrieben, dass im Hinblick auf die Qualitätsbewertung bei den direkten Verfahren der Datenschutz durch das Festlegen von formalen Kriterien (z. B. Wahrung der Anonymität der Patientinnen und Patienten und der Leistungserbringer) für die Stellungnahmen teilweise abgedeckt ist. Regeln im Umgang mit sensiblen Daten sollten in Abhängigkeit von den jeweiligen Prozessen des Strukturierten Dialogs festgelegt werden. Hier sollte zwischen der Bewertung der Versorgungsqualität anhand von Stellungnahmen einerseits und qualitätsfördernden Maßnahmen andererseits, die einen persönlichen Kontakt mit dem Leistungserbringer erfordern, unterschieden werden. Für beide Vorgänge ist die Verarbeitung unterschiedlicher Informationen (z. B. Stellungnahmen vs. Patientenakten, Operationsberichte, SOP des Leistungserbringers) notwendig. Des Weiteren ist das Setting, in welchem die Qualitätsbewertung und qualitätsfördernde Maßnahmen stattfinden, ein anderes. Aufgrund dieser Unterschiede ergeben sich für die Vertraulichkeit und den Datenschutz in Abhängigkeit davon, ob es sich um den Prozess der Qualitätsbewertung oder um qualitätsfördernde Maßnahmen handelt, unterschiedliche Anforderungen. Vor dem Hintergrund unterschiedlicher Anforderungen sollten folglich Vertraulichkeitserklärungen und Datenschutzregeln zwischen der Qualitätsbewertung und der Qualitätsförderung differenzieren.

Die *Freiwilligkeit der Teilnahme am Verfahren* wird in der Literatur als eine weitere Voraussetzung für den Erfolg qualitätsfördernder Maßnahmen beschrieben (BÄK 2014, Chop und Eberlein-Gonska 2012, Gerlach 2001, Nimptsch et al. 2016). Freiwilligkeit setzt die Überzeugung der Teilnehmer vom Nutzen des Verfahrens voraus, wodurch die Akzeptanz als auch die Veränderungsbereitschaft erhöht wird. Die Ziele des Verfahrens werden idealerweise die eigenen Ziele der teilnehmenden Krankenhäuser bzw. Leistungserbringer, sodass das Verfahren als selbstbestimmte Qualitätsentwicklung wahrgenommen wird (BÄK 2014). Die Freiwilligkeit ist im Kontext der externen stationären Qualitätssicherung nicht gegeben, dennoch kann die Überzeugung der Teilnehmer von Sinn und Nutzen des Verfahrens günstig beeinflusst werden.

Die *Überzeugung der Teilnehmer vom Nutzen des Verfahrens* gilt in der Literatur selbst als Erfolgsfaktor (BÄK 2014). Dieser steht in engem Zusammenhang mit den Erfolgsfaktoren *Kenntnis und Einhaltung der Verfahrensregeln durch alle Teilnehmer* (BÄK 2014). Sind die Teilnehmerinnen und Teilnehmer eines Verfahrens von diesem überzeugt, werden sie die Verfahrensregeln kennen und einhalten. Teilnehmerinnen und Teilnehmer des Verfahrens sind die Leistungserbringer, die Mitarbeiterinnen und Mitarbeitern der beauftragten Stelle sowie das Expertengremium. Existieren für die entsprechenden Akteure klare Verfahrensregeln und Voraussetzungen für die Qualitätsbewertung, führt das zu einer transparenteren und fairen Beurteilung der Versorgungsqualität und dadurch zu einer größeren Akzeptanz des Verfahrens. Zudem kann eine von der Qualitätsbewertung abgekoppelte Qualitätsförderung den Erfolg und die Akzeptanz der externen qualitätsfördernden Maßnahmen steigern. Insgesamt können im Rahmen der externen Qualitätssicherung dadurch möglichst optimale Bedingungen geschaffen werden, um das einrichtungsinterne Qualitätsmanagement beim Abbau von Qualitätsmängeln zu unterstützen.

### **3.8 Zusammenfassung**

In den vorherigen Abschnitten wurden aus der wissenschaftlichen Literatur und den Methodischen Grundlagen V1.1 Kriterien für die Güte von Bewertungs- bzw. Beurteilungsprozessen und für den Erfolg von Qualitätsfördermaßnahmen hergeleitet. Ziel dieser Herleitung ist es, methodische Begründungen für die in den folgenden Kapiteln beschriebenen Empfehlungen für die Weiterentwicklung des Strukturierten Dialogs zu geben.

Aus dieser Herleitung ergibt sich, dass die quantitative Beurteilung der Versorgungsqualität mittels Indikatoren und die Beurteilung im Stellungnahmeverfahren unterschiedliche Rollen bei der Qualitätsbewertung haben, die aufeinander abgestimmt sein müssen. Quantitative Untersuchungsmethoden sind zu bevorzugen, wenn standardisierte Bewertungen eines Sachverhalts vorgenommen werden sollen (IQTIG 2019a, S. 42 f.), da sie eine hohe Objektivität ermöglichen und zumindest bei der Datenaus- und –bewertung ressourcenschonender sind als qualitative Untersuchungsmethoden. Dies ist der Fall in der externen Qualitätssicherung nach QSKH-RL, DeQS-RL und plan. QI-RL, wo einerseits Aussagen über die Erfüllung von Standards durch die Leistungserbringer im Mittelpunkt stehen, sowie die Ergebnisse für Auswahlentscheidungen von Patientinnen und Patienten Verwendung finden sollen.

Die qualitative Beurteilung im Stellungnahmeverfahren ist gemäß den hier hergeleiteten Kriterien nicht als ein eigenständiges Bewertungsinstrument zu verstehen, sondern sie erfüllt eine klar umschriebene Funktion im Gesamtverfahren (vgl. auch Abschnitt 2.1). Diese besteht in der Prüfung der Validität der Qualitätsaussage eines Qualitätsindikators in dem speziellen Fall eines Leistungserbringers, der auffällig wurde – in Abgrenzung zur Validität des Qualitätsindikators im Allgemeinen, die durch die Entwicklung entlang der Eignungskriterien sichergestellt wird. Dabei soll sich die Prüfung auf das Qualitätsmerkmal des Indikators und auf das durch den Referenzbereich operationalisierte Qualitätsniveau beziehen. Die Vorteile einer standardisierten Erhebung und Bewertung durch Qualitätsindikatoren gingen verloren bzw.

wären nicht erforderlich, wenn die Qualitätsindikatoren lediglich als Aufgreifkriterien zur Ermittlung auffälliger Leistungserbringer genutzt würden, das Stellungnahmeverfahren aber eine eigene, vom Qualitätsziel des Indikators unabhängige Bewertung vornähme.

Die aus dem Ziel der Beauftragung, ein möglichst einheitliches (objektives) Vorgehen im Stellungnahmeverfahren zu gewährleisten, abgeleiteten methodischen Anforderungen werden in den folgenden Kapiteln in Form von Empfehlungen für ein möglichst explizites Vorgehen ausgearbeitet. Wie in den vorangehenden Kapiteln deutlich wird, ergibt sich jedoch ein Zielkonflikt, wenn an ein qualitatives Verfahren, wie ein Stellungnahmeverfahren ähnlich hohe Anforderungen an die Objektivität angelegt werden wie an ein quantitatives Verfahren, wie die Indikatorbasierte Qualitätsmessung. Denn die zur Standardisierung des qualitativen Verfahrens notwendigen Maßnahmen erhöhen einerseits wenigstens teilweise den Aufwand für die Durchführung des Verfahrens und schränken andererseits notwendigerweise die Freiheit ein, unterschiedliche Informationen für das qualitative Verfahren heranzuziehen.

## 4 Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und -förderung

Ein zentrales Motiv für die Beauftragung zur Weiterentwicklung des Strukturierten Dialogs ist eine Steigerung der Objektivität und eine Reduktion des Aufwands (siehe Beauftragung S. 1). Wie in Abschnitt 3.3 dargelegt entspricht dieses Ziel im Verständnis des IQTIG einer stärkeren Standardisierung von Prozessen, die teilweise auch mit einer Aufwandsreduktion einhergeht. In der Analyse des bisher geführten Strukturierten Dialogs nach QSKH-RL wurden die drei Funktionen Qualitätsbewertung, Qualitätsförderung und Bewertung der Dokumentationsqualität identifiziert (siehe auch Abschnitt 2.1). Diese drei Funktionen sind auch in den Regelungen der DeQS-RL zu finden. Allerdings sind diese Bestandteile weder in der QSKH-RL noch in der DeQS-RL eindeutig voneinander abgegrenzt. Dialogische Elemente wie Begehungen und kollegiale Gespräche können beispielsweise gleichzeitig für die Bewertung der Versorgungsqualität und für die Vereinbarung von qualitätsverbessernde Maßnahmen eingesetzt werden. Durch die Vermischung beider Funktionen leiden jedoch die Objektivität und Effizienz von Qualitätsbewertung sowie die Effektivität der Qualitätsförderung. Beispielsweise ist es wahrscheinlich, dass Empfehlungen im Rahmen eines Peer Reviews weniger gut angenommen werden können, wenn gleichzeitig die Qualitätsbewertung noch aussteht. Demnach sollten Verbesserungspotenziale im kollegialen Gespräch erst identifiziert werden, wenn die Bewertung abgeschlossen ist (vgl. BÄK 2014: S. 35). Zudem könnte die Qualitätsbewertung dadurch beeinflusst werden, wie offen der Leistungserbringer für Vorschläge zur Qualitätsverbesserung ist. So könnte die Qualitätsbewertung beispielsweise strenger ausfallen, wenn sich ein Leistungserbringer Vorschlägen zur Qualitätsverbesserung verschließt, und weniger streng, wenn er sich offen für solche Vorschläge zeigt. All dies sind Beispiele für Einflüsse, die die Objektivität des Verfahrens reduzieren können (vgl. Abschnitt 3.3).

### 4.1 Modulare Betrachtungsweise

Aus diesen Gründen ist eine zentrale Eigenschaft im Rahmenkonzept für die beauftragte Weiterentwicklung die Trennung von Qualitätsbewertung und Qualitätsförderung: Es wird empfohlen, zukünftig den Prozess der Qualitätsbewertung sowohl konzeptuell als auch zeitlich vom Prozess der Qualitätsförderung abzugrenzen. Dies bedeutet, dass unter dem Begriff „Stellungnahmeverfahren“ ausschließlich die Prozesse von der Einholung bis zur abschließenden Bewertung der Stellungnahmen und des Indikatorergebnisses verstanden werden sollen. Auf die Qualitätsbewertung aufbauende Maßnahmen sollen nicht als Teil des Stellungnahmeverfahrens angesehen werden. Im Anschluss an die Bewertung bestehen neben der Qualitätsförderung prinzipiell verschiedene Handlungsoptionen (IQTIG 2019a: S. 38), z. B. im Rahmen von planungsrelevanten Qualitätsindikatoren oder Qualitätszu- und -abschlägen. Eine Trennung findet sich

ähnlich auch in die Regelungen zum Stellungnahmeverfahren nach § 11 der Richtlinie zu planungsrelevanten Qualitätsindikatoren (plan. QI-RL),<sup>28</sup> in der keine Fördermaßnahmen im Rahmen des Stellungnahmeverfahrens vorgesehen sind. Daher wird im vorliegenden Bericht die Unterscheidung der drei Funktionen des bisherigen Strukturierten Dialogs in Form von drei unterschiedlichen Modulen vorgeschlagen:

- Das **Modul „Qualitätsbewertung“** subsumiert alle Schritte, die von der Berechnung eines Indikatorergebnisses bis hin zu der abschließenden Qualitätsbewertung durchgeführt werden
- Das **Modul „Qualitätsförderung“** beinhaltet diejenigen Maßnahmen, die zur Steigerung der Versorgungsqualität bei einem Leistungserbringer nach abgeschlossener Qualitätsbewertung eingesetzt werden.
- Das **Modul „Daten- und Dokumentationsqualität“** beinhaltet das Vorgehen im Zusammenhang mit den sogenannten Auffälligkeitskriterien. Wie in Abschnitt 1.3 erläutert, ist dieses Modul nicht Teil der Beauftragung und daher auch nicht Teil des vorliegenden Konzepts.

Die folgenden Empfehlungen zielen darauf ab, dass im ersten Schritt aussagekräftige (valide) und nachvollziehbare (objektive) Qualitätsbewertungen vorgenommen werden, auf die weiterführende Maßnahmen, wie etwa qualitätssteigernde Maßnahmen bei den Leistungserbringern oder die Information von Patientinnen und Patienten über die Versorgungsqualität (Public Reporting), sinnvoll aufbauen können. Gemäß Beauftragung wird damit eine Standardisierung der Prozesse des Moduls „Qualitätsbewertung“ angestrebt. Die Empfehlungen orientieren sich an den methodischen Erfordernissen für aussagekräftige Qualitätsbewertungen (siehe Abschnitt 3.1).

## 4.2 Qualitätsindikatoren als kleinste Bewertungseinheit

Da ein Qualitätsindikator die kleinste Bewertungseinheit der gesetzlich verpflichtenden Qualitätssicherung ist, bezieht sich das Rahmenkonzept immer auf das Ergebnis eines einzigen Indikators. Dies bedeutet auch, dass sich die Qualitätsbewertung immer ausschließlich auf das Qualitätsmerkmal bezieht, das durch den betreffenden Indikator abgebildet werden soll. Eine Aussage über die „Gesamtqualität“ der Leistungserbringung einer Fachabteilung oder eines Standorts kann grundsätzlich nur in der Zusammenschau der Ergebnisse eines für diesen Zweck geeigneten QI-Sets getroffen werden (siehe auch Abschnitt 3.3).

Das grundsätzliche Vorgehen bei der Qualitätsbewertung gemäß des vorliegenden Konzepts ist identisch für alle Indikator Typen (Struktur, Prozess, Ergebnis) jeder Datenquelle (Dokumentation der Leistungserbringer, Sozialdaten, Daten aus Patientenbefragungen). Damit wird ein möglichst einheitliches Vorgehen für alle Indikatoren gewährleistet. Auch für so genannte Follow-up-Indikatoren, deren Auswertung sich auf Ereignisse (z. B. Revisionseingriffe oder der Tod einer Patientin oder eines Patienten) innerhalb eines bestimmten Beobachtungszeitraumes im Anschluss an ein interessierendes Ereignis (z. B. eine Organtransplantation) bezieht, kann und soll nach

<sup>28</sup> Richtlinie zu planungsrelevanten Qualitätsindikatoren gemäß § 136 Absatz 1 SGB V i. V. m. § 136c Absatz 1 und Absatz 2 SGB V. In der Fassung vom 15. Dezember 2016, zuletzt geändert am 18. Januar 2018, in Kraft getreten am 12. Mai 2018. URL: <https://www.g-ba.de/informationen/richtlinien/91/> (abgerufen am 14.03.2019).

diesem Konzept verfahren werden. Denn Follow-up-Indikatoren unterscheiden sich nicht prinzipiell von anderen Indikatortypen. Fragen der Zuschreibbarkeit der Verantwortung und der Datenqualität, die sich bei Follow-up-Indikatoren besonders stellen, müssen im Rahmen der Eignungsbeurteilung der Indikatoren beantwortet werden (siehe auch Eignungskriterien, IQTIG 2019a, S. 150).

### 4.3 Grundlegender Ablauf des Moduls „Qualitätsbewertung“

In Abbildung 9 werden die Empfehlungen für die grundlegenden Prozesse der Module „Qualitätsbewertung“ und „Qualitätsförderung“ dargestellt und für das Modul „Qualitätsbewertung“ im Folgenden erläutert.

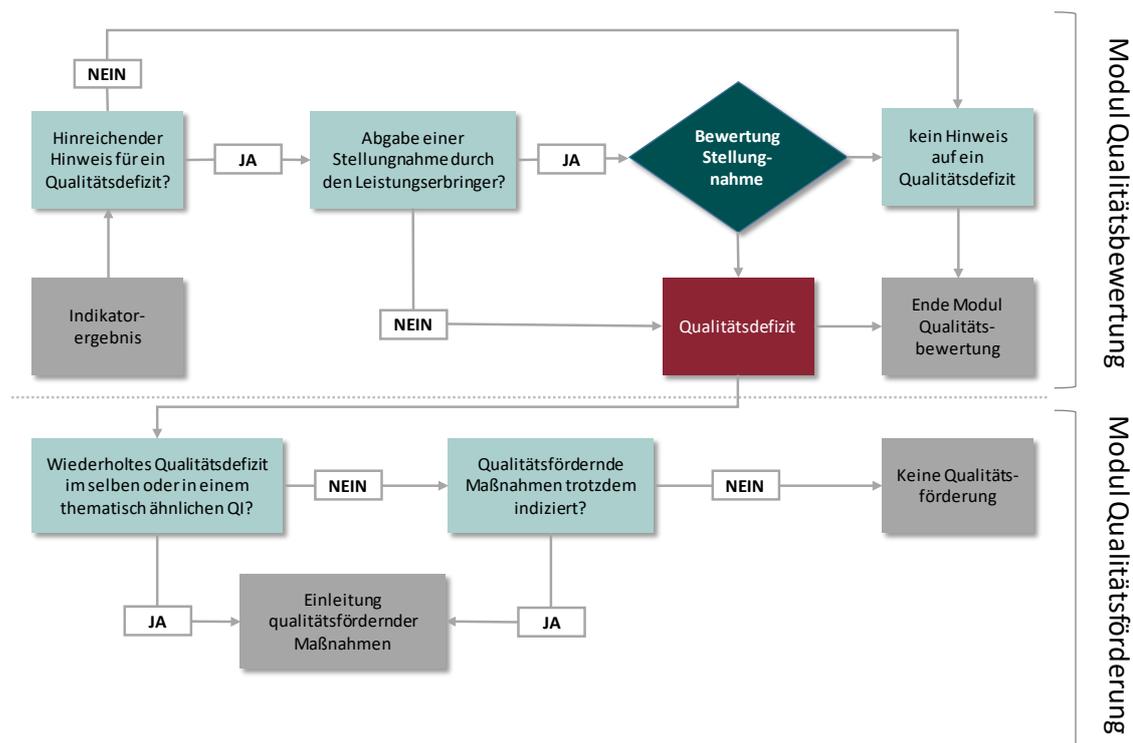


Abbildung 9: Grundlegender Ablauf der Qualitätsbewertung und -förderung

Legt ein Indikatorergebnis einen hinreichenden Hinweis für ein Qualitätsdefizit nahe, so ist dem Leistungserbringer von der LAG die befristete Möglichkeit zur schriftlichen Stellungnahme einzuräumen. Das zunächst abstrakte Kriterium des hinreichenden Hinweises für ein Qualitätsdefizit kann auf unterschiedliche Weise operationalisiert werden. Im Rahmen der QSKH-RL wurde bisher ein numerischer Vergleich zwischen Indikatorwert und Referenzbereichsgrenze vorgenommen („rechnerische Auffälligkeit“). Dieses Vorgehen wird jedoch für einige Prozess- und viele Ergebnisindikatoren als ungeeignet angesehen, da es mögliche stochastische, also fallzahlabhängige, Einflüsse auf das Indikatorergebnis unberücksichtigt lässt (vgl. Abschnitt 15.1, IQTIG 2019a). In Kapitel 2 wird daher das abstrakte Kriterium des hinreichenden Hinweises für ein Qualitätsdefizit unter Berücksichtigung stochastischer Einflüsse für die verschiedenen Indikatortypen (ratenbasierte Indikatoren, O/E-Indikatoren etc.) statistisch operationalisiert.

Mit der Aufforderung zur Stellungnahme wird dem Leistungserbringer die Möglichkeit eingeräumt, Einflussfaktoren auf das Indikatorergebnis, die nicht von ihm zu verantworten sind und nicht bereits im Qualitätsindikator oder dessen Risikoadjustierung oder mittels einer anderen statistischen Methode berücksichtigt sind, bei der Qualitätsbewertung geltend zu machen. Damit hat die Einholung und Bewertung einer Stellungnahme eine genau umschriebene Funktion, auf die hin die Kriterien für eine Einholung sowie für die fachliche Bewertung ausgerichtet sind. Im Rahmen der fachlichen Bewertung der Stellungnahme prüft eine Fachkommission die in der Stellungnahme angeführten Gründe für das Verfehlen des Referenzbereichs und beurteilt, ob diese tatsächlich nicht vom Leistungserbringer zu verantworten sind und damit der Hinweis aus dem Indikatorergebnis entkräftet ist. In Kapitel 6 werden der fachliche Bewertungsprozess sowie die Kriterien dafür genauer erläutert.

Die Abgabe einer Stellungnahme durch den Leistungserbringer soll freiwillig sein. Wird seitens des Leistungserbringers auf eine Stellungnahme verzichtet, gilt gemäß des Verständnisses von Qualitätsindikatoren (siehe Abschnitt 3.1), dass ein hinreichender Hinweis auf ein Qualitätsdefizit vorliegt, der Indikator in seiner Validität nicht angezweifelt wird und somit ein Qualitätsdefizit vorliegt (siehe Abbildung 9). Daher soll in diesem Fall die Bewertung „Qualitätsdefizit“ erfolgen.

Gemäß § 17 DeQS-RL kann die LAG auch bei „besonders guten Ergebnissen“ eine Stellungnahme anfordern. Hier wird jedoch empfohlen, die Einholung von Stellungnahmen auf Situationen zu beschränken, in denen ein hinreichender Hinweis auf ein Qualitätsdefizit gemäß den statistischen Kriterien besteht. Aus der Funktion von Stellungnahmen im Rahmen von *accountability* (vgl. Abschnitt 3.1) ergibt sich, dass Stellungnahmen für die Qualitätsbewertung nur dann indiziert sind, wenn die Validität des Indikatorergebnisses einer weiteren Prüfung unterzogen werden soll. Vor dem Hintergrund begrenzter Ressourcen bei den LAG und den Leistungserbringern wird daher empfohlen, diesen zweiten Prüfschritt nur im Sinne einer Steigerung der Spezifität des Vorgehens einzusetzen (s. u.).

Legt das Indikatorergebnis gemäß der empfohlenen biometrischen Auswertungsmethodik keinen hinreichenden Hinweis auf ein Qualitätsdefizit nahe, soll das Indikatorergebnis automatisch mit „kein Hinweis auf ein Qualitätsdefizit“ bewertet werden. Dies stellt eine begründete Asymmetrie im Vorgehen zugunsten der Leistungserbringer dar: Bei einem hinreichenden Hinweis auf ein Qualitätsdefizit auf Basis des quantitativen Indikatorergebnisses wird im Rahmen des Stellungnahmeverfahrens eine weitere Prüfung des Ergebnisses vorgenommen, während bei Abwesenheit eines solchen Hinweises kein weiterer Prüfschritt vorgesehen ist. Durch das Stellungnahmeverfahren wird somit die Spezifität des Vorgehens für „tatsächliche Qualitätsdefizite“ erhöht, im Vergleich zu einem Vorgehen ohne Stellungnahmeverfahren, da bei einem hinreichenden Hinweis für ein Qualitätsdefizit dieser Hinweis widerlegt werden kann, aber das Fehlen eines hinreichenden Hinweises für ein Qualitätsdefizit nicht widerlegt werden kann. Dies bedeutet, dass Einflüsse wie Fehler in der Datengrundlage und nicht vom Leistungserbringer zu vertretende Faktoren, die sich günstig für diesen auswirken (z. B. eine in diesem Jahr gesündere Patientenpopulation, die in der Risikoadjustierung nicht berücksichtigt ist), zu einem positiveren Indikatorergebnis führen können als die tatsächliche Versorgungsqualität hergibt.

Die Methodik der einzelnen Elemente und Prozessschritte wird in den folgenden Kapiteln ausführlich beschrieben.

#### **4.4 Beschränkung der Informationsgrundlage auf Stellungnahmen**

Die in der Beauftragung genannten Ziele einer möglichst einheitlichen Vorgehensweise – also einer hohen Objektivität – sowie einer Verschlankung des Verfahrens sind nur durch Standardisierung des Verfahrens erreichbar. Demgegenüber steht möglicherweise der Wunsch nach Berücksichtigung einer möglichst flexiblen Informationsgrundlage im zweiten, qualitativen Schritt. Durch (telefonische) Nachfragen, Begehungen, Präzisierungen der Stellungnahmen und die Möglichkeit korrigierte Stellungnahmen nachzureichen, kann die Informationsgrundlage nach Bedarf erweitert werden und damit möglicherweise die Wahrscheinlichkeit erhöht werden, bei inhaltlich zunächst unzureichenden Stellungnahmen doch noch entlastende Gründe zu finden (= Steigerung der Spezifität des Vorgehens, vgl. Abschnitt 2.1). Eine möglichst flexible Informationsgrundlage ist damit nur eingeschränkt vereinbar mit den in der Beauftragung genannten Zielen eines aufwandsarmen und objektiven Verfahrens. Erfolgt die Hinzuziehung und Bewertung zusätzlicher Informationen jedoch nicht nach einheitlichen Regeln, mindert dies die Objektivität des Verfahrens. Aus methodischer Sicht ist eine möglichst flexible Informationsgrundlage für das Stellungnahmeverfahren jedoch gar nicht notwendig, da wie in Abschnitt 3.1 hergeleitet, die methodische Funktion des Stellungnahmeverfahrens darin gesehen wird, Einflussfaktoren zu identifizieren, die die Aussagekraft eines Indikatorergebnisses in Zweifel ziehen. Für diese Funktion erscheinen schriftliche Stellungnahmen, in denen Leistungserbringer aus ihrer Sicht vorliegende Einflussfaktoren benennen können, zweckmäßig und praktikabel. Des Weiteren erhöht sich durch eine flexiblere Informationsgrundlage auch der Aufwand und die Dauer des Verfahrens. Auch deshalb wird empfohlen, keine zusätzlichen Informationen über das Indikatorergebnis und die (kriterienkonforme) Stellungnahme hinaus für das Modul „Qualitätsbewertung“ hinzuzuziehen.

Im Gegensatz dazu sollte für das Modul „Qualitätsförderung“ eine möglichst breite und flexible Informationsgrundlage geschaffen werden können. Eine Notwendigkeit zur Standardisierung ergibt sich hier nicht, weder aus dem Auftrag noch aus methodischen Gründen, denn anders als die Ergebnisse von Qualitätsmessungen sind die Fördermaßnahmen sinnvollerweise nicht für Leistungserbringervergleiche vorgesehen. Im Gegenteil – wie in Kapitel 7 dargelegt – sollten Fördermaßnahmen auf die individuelle Situation des jeweiligen Leistungserbringers angepasst sein, um eine optimale Wirkung zu entfalten.

## 5 Statistische Auswertungsmethodik

Ziel dieses Kapitels ist die Darstellung eines einheitlichen Rahmenkonzepts zur statistischen Berechnung von Indikatorergebnissen anhand von Daten. Ein solches, übergreifendes Konzept für die externe Qualitätssicherung existiert bislang nicht, ist jedoch notwendig, um später fachlich zu verschiedenen Methoden der rechnerischen Auffälligkeitseinstufung diskutieren zu können. Diese umfassende Darstellung der Auswertungsmethodik bietet den notwendigen Rahmen, um fundierte Vorschläge für die Operationalisierung des Begriffs „hinreichender Hinweis auf ein Qualitätsdefizit“ aus Kapitel 4 zu generieren.

Betrachtet wird in diesem Kapitel ein bestehender Qualitätsindikator, welcher entsprechend den Schritten in den „Methodischen Grundlagen“ des IQTIG (IQTIG 2019a) entwickelt worden ist und der nun im Rahmen der DeQS-RL berechnet werden soll. Die statistische Auswertungsmethodik für die Berechnung von QI-Ergebnissen wird im Folgenden grundlegend dargestellt: dies enthält den prinzipiellen Aspekt der Berechnung von QI-Ergebnissen anhand der beobachteten Daten und den spezielleren Aspekt der Auslösung des Stellungnahmeverfahrens für Leistungserbringer anhand deren QI-Ergebnisse im Rahmen einer statistischen Klassifikation. Fokus des vorliegenden Kapitels ist dieser Klassifikationsaspekt.

Der Abschnitt 5.1 liefert ein Rahmenkonzept für die statistische Auswertungsmethodik für Qualitätsindikatoren. Danach richtet sich der Fokus in Abschnitt 5.2 auf den Klassifikationsaspekt der Auswertungsmethodik, welcher formal im Rahmen eines Entscheidungsproblems dargestellt wird. Der Abschnitt 5.3 beschreibt dann verschiedene Formen der rechnerischen Auffälligkeitseinstufung für die Daten eines Erfassungsjahres. Dabei werden die Methoden statistisch formal hergeleitet und jeweils das Entscheidungssetting diskutiert, in dem die Einstufungsmethode die optimale Klassifikationsgüte hat. Abschnitt 5.4 erweitert diese Methoden dann, damit Daten aus mehreren Erfassungsjahren bei der Auffälligkeitseinstufung berücksichtigt werden können. Insgesamt sind die Abschnitte 5.2 bis 5.4 anspruchsvoller statistischer Natur und setzen fortgeschrittene Kenntnisse inferenzstatistischer Methoden voraus. Eine Darstellung auf diesem Niveau ist jedoch notwendig, um später fachlich fundiert über die Optimierung der Effizienz des SD-Prozesses diskutieren zu können. Eine Illustration der Einstufungsmethodik findet sich im Abschnitt 5.5 und eine etwas allgemeinverständlichere Zusammenfassung der Inhalte der Abschnitte 5.2 bis 5.4 im Abschnitt 5.6.

### 5.1 Rahmenkonzept der Auswertungsmethodik

Jede neu entwickelte statistische Auswertungsmethodik im Bereich der externen Qualitätssicherung sollte eine belastbare quantitative Aussage über die Qualität der Leistungserbringer liefern. Die Ergebnisse der statistischen Auswertung sollen hilfreiche Informationen für Patientinnen und Patienten und für die Leistungserbringer zur Verfügung stellen und damit eine erhöhte Transparenz über die Qualität der Versorgung schaffen. Damit die Methodik sinnvoll zum Einsatz kommen kann, muss sichergestellt werden, dass sie für die vorliegenden bzw. erwarteten Fallzahlen, in Kombination mit den interessierenden Fragestellungen, geeignet ist.

Grundlage für die Auswertung sind die Daten, die gemäß der Spezifikation für das zugehörige QS-Verfahren im Rahmen der DeQS-RL erhoben werden. In der einfachsten Form wird der Indikator durch Mengendefinitionen der zu betrachtenden Population und des interessierenden Ereignisses in diesen Daten operationalisiert. Ein wichtiger Bestandteil dieser Mengendefinitionen ist die Festlegung, aus welchen Entitäten<sup>29</sup> die Population besteht, dies könnten z. B. Patientinnen und Patienten, aber auch Prozeduren sein. Prinzipiell könnte die Entität auch ein Fragebogen aus einer Einrichtungsbefragung sein. Im Folgenden wird jedoch der Begriff „Fall“ als generischer Begriff für die betrachtete Entität verwendet.

Grundsätzlich sind folgende weitere Festlegungen für eine konkrete Auswertung eines Indikators zu treffen: Die betrachtete Gruppierungseinheit für die Auswertung sowie der zu bewertende Zeitraum. Prinzipiell wird die Auswertung für mehrere Gruppierungseinheiten vorgenommen, z. B. auf Ebene der Leistungserbringer (entlassender oder leistungserbringender Standort), der Landesebene und der Bundesebene. Für die Auslösung des Stellungnahmeverfahrens ist jedoch nur die Einheit Leistungserbringer relevant, weshalb andere Auswertungseinheiten im Folgenden nicht weiter diskutiert werden. Der zu bewertende Zeitraum ist der temporale Aspekt der Datengrundlage. Eine klare Definition der Datengrundlage für die Auswertung wird dabei vorausgesetzt, es geht hier nur um den zeitlichen Aspekt der Datengrundlage. In den einfachsten Fällen ist dies das Erfassungsjahr, d. h. z. B. alle Entlassungen in der betrachteten Gruppierungseinheit im Kalenderjahr *X*. Im Rahmen der gleichen Auswertung kann es durchaus zu mehreren unterschiedlichen Berechnungen des gleichen Indikators bzgl. unterschiedlicher Gruppierungseinheiten und Zeiträume kommen, z. B. Leistungserbringerergebnisse im Vergleich zum Bundesergebnis, Ergebnis mit den aktuellen Rechenregeln für das aktuelle Erfassungsjahr vs. Ergebnis mit aktuellen Rechenregeln für das vorherige Erfassungsjahr.

Die quantitative Auffälligkeitseinstufung<sup>30</sup> der Leistungserbringer, für die ein Stellungnahmeverfahren eingeleitet werden soll, sollte nachvollziehbar sein und erlauben, dass daraus Schlüsse gezogen und Konsequenzen abgeleitet werden können. Grundlegende Voraussetzung dafür und für die Akzeptanz der Auswertungsergebnisse ist, dass die Methodik fachlich fundiert ist und zugrunde liegende Annahmen explizit gemacht werden, damit die Adäquatheit der Annahmen und deren Konsequenzen diskutiert werden können. Des Weiteren ist die Methodik so konkret darzustellen, dass die Leistungserbringer die entsprechenden Ergebnisse nachvollziehen können. Die quantitative Einstufung soll zudem eine hohe Treffsicherheit<sup>31</sup> gewährleisten, um einen

---

<sup>29</sup> Entität wird hier als Sammelbegriff für die Eigenschaften der Fälle verwendet, die Gegenstand der Qualitätssicherung sind. Die Festlegung der Entität ist vor allem für QS-Verfahren mit Unterbogenstruktur relevant, da hier die Festlegung der Entität die Zählweise bei der Bestimmung der Populationen aus den zusammengeführten Bögen festlegt. Als Beispiel könnte Entität in der Geburtshilfe entweder die Mütter oder die Kinder bezeichnen.

<sup>30</sup> In diesem Kapitel wird der Begriff „quantitative Auffälligkeit“ als Überbegriff für verschiedene statistische Verfahren zur Klassifikation des Leistungserbringerergebnisses verwendet. Es wird bewusst ein neuer Begriff hierfür verwendet, um sich vom Begriff der QSKH-RL, der „rechnerischen Auffälligkeit“, zu unterscheiden, welches den direkten Vergleich zwischen Indikatorergebnis und Referenzbereich vorschreibt, d. h. bei QIs mit unerwünschten interessierenden Ereignissen, ob Indikatorergebnis > Referenzwert.

<sup>31</sup> Mit „Treffsicherheit“ ist die Genauigkeit der Klassifikation gemeint, d. h. die Eigenschaft des Klassifikationsverfahrens, tatsächliche Qualitätsdefizite anhand der QI-Ergebnisse zu identifizieren und umgekehrt auch das Fehlen tatsächlicher Qualitätsdefizite als solche zu identifizieren.

effizienten Verfahrensablauf zu sichern, bei dem jedoch keine Qualitätsdefizite übersehen werden. Die oben erwähnten Kriterien machen schnell klar, dass die gewählte Methodik sich im Spannungsfeld zwischen anspruchsvollen Berechnungen und Kommunizierbarkeit befinden, zwischen denen behutsam abgewogen werden muss.

Für die Wahl einer geeigneten statistischen Auswertungsmethodik spielen sowohl die indikatorübergreifende Herangehensweise und Stichprobenart sowie die Berechnungsart des betrachteten Indikators eine wichtige Rolle. Zusammen mit der Bewertungsart stellen die Festlegungen eine Taxonomie für die Auswertungsmethodik eines Qualitätsindikators dar und sind vorab für jede Auswertung des Qualitätsindikators festzulegen. Im Folgenden werden Herangehensweise, Stichprobenart für die vorgegebene Auswertungsebene und Berechnungsart kurz vorgestellt, da sie die Grundlage für eine fundierte Diskussion der Bewertungsart als Klassifikation der Leistungserbringerergebnisse sind.

### 5.1.1 Herangehensweise

Hinsichtlich des Zwecks einer statistischen Datenauswertung lassen sich zwei Herangehensweisen – die analytische und die enumerative – differenzieren (vgl. z. B. Deming (1953)). Die folgende Erläuterung dieser beiden Herangehensweisen wurde (in modifizierter Form) bereits im Rahmen der Abschlussberichte zur Entwicklung von Patientenbefragungen der Qualitätssicherungsverfahren *Perkutane Koronarintervention (PCI) und Koronarangiographie* sowie *Schizophrenie* angeführt (IQTIG 2018c, IQTIG 2018d). Die analytische Herangehensweise wurde auch implizit bei der Entwicklung der statistischen Auswertungsmethodik für die plan. QI-RL vorausgesetzt (IQTIG 2016).

#### Analytische Herangehensweise

Bei der analytischen Herangehensweise liegt das Interesse am zugrunde liegenden Prozess. Im Rahmen der externen Qualitätssicherung bedeutet dies, dass im Gegensatz zum reinen Auszählen der dokumentierten interessierenden Ereignisse in einem Erfassungsjahr anhand eines statistischen Modells Schlüsse über die Systematik bezüglich der betrachteten Qualitätsmerkmale beim Leistungserbringer gezogen werden sollen. Das Interesse liegt somit in der *zugrunde liegenden Kompetenz* des Leistungserbringers bezüglich des im Indikator abgebildeten Qualitätsmerkmals, welches durch die Rechenvorschrift des Indikators statistisch operationalisiert wird. Dabei muss es sich nicht um eine spezifische, benennbare Kompetenz handeln. Vielmehr abstrahiert der im statistischen Modell angenommene zugrunde liegende Kompetenzparameter in diesem Zusammenhang das komplexe Zusammenspiel vieler Eigenschaften des Leistungserbringers, die sich zusammen mit weiteren Faktoren, wie beispielsweise patientenseitigen Faktoren, auf die in dem Indikator betrachtete Behandlung der Patientinnen und Patienten auswirken. Bei einem Ratenindikator stellt der Kompetenzparameter die zugrunde liegende Wahrscheinlichkeit, mit der das interessierende Ereignis bei den Fällen der betrachteten Grundgesamtheit eintritt, dar. Patientenseitige Faktoren bleiben bei Ratenindikatoren unberücksichtigt. Da die Operationalisierung immer ein Kompromiss zwischen Präzision und Aufwand der Datenerhebung darstellt und darüber hinaus i. d. R. nicht alle für den Indikator relevanten Informationen erhoben werden können, ist der Kompetenzparameter immer im Kontext der Operationalisierung,

d. h. der formulierten Rechenregeln des Indikators zu interpretieren. Zum Beispiel können besondere Versorgungssituationen oft nicht in den Rechenregeln des Indikators abgebildet werden. Somit kann der Kompetenzparameter des verwendeten statistischen Modells diesen Aspekt der „Kompetenz“ auch nicht berücksichtigen. Diese Berücksichtigung ist stattdessen Aufgabe im Zuge der fachlichen Bewertung. Für die Berechnung und Bewertung der Ergebnisse wird anhand von statistischer Inferenz basierend auf den zu diesem Zweck erhobenen Daten eines Erfassungsjahres mittels der Rechenvorschriften der Indikatoren jeweils ein Schätzwert für den jeweiligen latenten Kompetenzparameter des Leistungserbringers ermittelt. Diese Schätzwerte, d. h. die tatsächlich gemessenen Ergebnisse der Leistungserbringer in den Qualitätsindikatoren können als komplexe Funktion verschiedener Einflussgrößen aufgefasst werden (IQTIG 2019a: 181). Es wird angenommen, dass es unbekannte oder nicht erfasste Einflüsse auf die tatsächlich gemessenen Ergebnisse der Leistungserbringer gibt, die im Folgenden unter dem Konzept der Stochastizität subsumiert werden (vgl. auch IQTIG (2019a)). Da zur Bewertung der Leistungserbringer nur durch ihn beeinflussbare Einflüsse eingehen sollen (IQTIG 2017d, Kapitel 10), wird die statistische Unsicherheit resultierend aus der Stochastizität bei der Bewertung berücksichtigt. Bei der analytischen Herangehensweise wird, im Gegensatz zur enumerativen Herangehensweise, somit selbst dann statistische Unsicherheit angegeben und bei der Bewertung von Leistungserbringerergebnissen berücksichtigt, wenn für alle Fälle eines Leistungserbringers im betrachteten Erfassungsjahr Dokumentationen vorliegen. Da bei der analytischen Herangehensweise keine Schlüsse über eine endliche Population gezogen werden, sondern eine Aussage über den zugrunde liegenden Prozess getroffen wird, ist weniger entscheidend, ob es sich um eine Stichprobe oder eine Vollerhebung im klassischen Sinn handelt. Vielmehr ist wichtig, dass eine ausreichende Anzahl an Fällen vorliegt um eine präzise Schätzung des zugrunde liegenden Kompetenzparameters – also der Systematik – zu ermöglichen. Die zugrunde liegenden Kompetenzparameter der Indikatoren werden jeweils anhand der für ein Erfassungsjahr vorliegenden Daten und basierend auf mathematisch-statistischen Annahmen an den zugrunde liegenden stochastischen Prozess geschätzt, wirken sich letztlich jedoch auch auf die Fälle über das Ende des Erfassungsjahres hinaus aus, sofern angenommen werden kann, dass sich die zugrunde liegende Kompetenz für den jeweiligen Indikator nicht oder nur langsam verändert. Für die analytische Herangehensweise gibt die Einteilung in Erfassungsjahre lediglich vor, welche Daten zur Schätzung der Kompetenzparameter verwendet werden. Es ist jedoch nicht davon auszugehen, dass sich die zugrunde liegende Kompetenz abrupt nach Abschluss des Erfassungsjahres verändert, sodass sich annehmen lässt, dass sie sich auch auf zukünftig behandelte Patientinnen und Patienten auswirkt.

Die analytische Herangehensweise ist insbesondere dann angebracht, wenn Verallgemeinerungen über das Geschehene hinaus von Interesse sind (Deming 1953). Diese Herangehensweise findet in einigen internationalen Arbeiten im Rahmen der externer Qualitätssicherung Anwendung (siehe z. B. Spiegelhalter et al. 2012, Ash et al. 2012)<sup>32</sup> und bildet die Basis für die Auswertungen von Qualitätsindikatoren im Rahmen der QSKH-RL, der DeQS-RL und der plan. QI-RL, da

---

<sup>32</sup> Insbesondere kann immer von einer analytische Herangehensweise ausgegangen werden, wenn trotz Vollerhebung im klassischen Sinn statistische Unsicherheit berücksichtigt wird.

es für die meisten Prozessindikatoren und alle Outcome-Indikatoren nicht um das reine Zahlenwesen im Sinne eines statistischen Jahrbuchs geht, sondern darum, Rückschlüsse auf die Qualitätsprozesse des Leistungserbringers zu ziehen, um die Ergebnisse als Teil der Qualitätssicherung und zur Patienteninformation prospektiv einzusetzen. Die Berücksichtigung stochastischer Einflüsse bei analytischer Herangehensweise lässt sich umsetzen, indem die daraus resultierende statistische Unsicherheit nicht ignoriert, sondern in der Bewertung der Kompetenz explizit berücksichtigt wird. Eine Möglichkeit, statistische Unsicherheit zu berücksichtigen, ist, neben dem Schätzwert für den zugrunde liegenden Kompetenzparameter auch ein zugehöriges Unsicherheitsintervall anzugeben und dieses bzw. einen dazugehörigen statistischen Test auch in der Einstufung der Leistungserbringer zu verwenden (IQTIG 2019a). Dies führt zu einer Unterteilung der statistischen Auswertungsmethodik in drei Bestandteile:

1. Indikatorwert
2. zugehöriges Unsicherheitsintervall
3. quantitative Auffälligkeitseinstufung

Im Rahmen der analytischen Herangehensweise stellt der Indikatorwert im Allgemeinen einen Punktschätzer eines im Rahmen eines statistischen Modells definierten, zugrunde liegenden Parameters des Leistungserbringers dar (IQTIG 2019b, Kapitel 12). Dieser Parameter ist zunächst unbekannt und wird aus den (für diesen Indikator relevanten) vorliegenden Daten für den Leistungserbringer geschätzt. Vereinfachend wird dieser latente Parameter im Folgenden auch der QI-spezifische Kompetenzparameter des Leistungserbringers genannt. Da verschiedene Indikatoren ggf. eine Aussage über die Kompetenz der Leistungserbringer bezüglich unterschiedlicher Patientenpopulationen bzw. Behandlungsarten ermöglichen sollen, werden für jeden Indikator Bedingungen formuliert, unter denen die Patientinnen und Patienten in die Berechnung eingeschlossen werden. Patientinnen und Patienten, die diese Bedingungen erfüllen, werden im Folgenden auch als Teil der Grundgesamtheit, d. h. der Zielpopulation des Indikators, bezeichnet, wobei diese klar von dem Konzept der endlichen Grundgesamtheit in der enumerativen Herangehensweise zu unterscheiden ist. Diese Tatsache zeigt erneut, dass der Kompetenzparameter immer im Kontext der formulierten Rechenregeln des Indikators zu interpretieren ist.

Eine Berücksichtigung der oben beschriebenen stochastischen Einflüsse, die im Rahmen der analytischen Herangehensweise angenommen werden, findet im Rahmen von parametrischen statistischen Modellen über Verteilungsannahmen für die beobachtete Anzahl an interessierenden Ereignissen statt. Konkret bedeutet dies, dass die Anzahl an interessierenden Ereignissen als Zufallsvariable modelliert wird, dessen Verteilung von einem zugrunde liegenden Parameter abhängt, der als zugrunde liegende Kompetenz des Leistungserbringers im betrachteten QI interpretiert werden kann. Die Berechnung des Indikatorwerts für einen Leistungserbringer kann somit als eine Schätzung – im inferenzstatistischen Sinne – für die zugrunde liegende Kompetenz des Leistungserbringers im betrachteten QI gesehen werden. Da dieser statistische Schätzwert für die Kompetenz nicht die Unsicherheit bei der Bestimmung enthält, sollte zu jedem Indikatorwert auch ein zugehöriges Unsicherheitsintervall bestimmt werden, dessen Breite angibt, wie belastbar der berechnete Indikatorwert im Rahmen der Annahmen ist.

Allgemein dienen Unsicherheitsintervalle der Quantifizierung von statistischer Unsicherheit bezüglich der aus den Daten gewonnenen Informationen über den unbekannt Parameter von Interesse. Unter den getroffenen Annahmen über den datengenerierenden Prozess gibt die Breite eines Unsicherheitsintervalls an, wie stark die stochastischen Einflüsse bei der Bestimmung des Schätzwertes ins Gewicht fallen. Je schmaler das Intervall, desto belastbarer sind die Informationen, die aus den erhobenen Daten über den zugrunde liegenden Parameter gewonnen werden können. Die Breite von Unsicherheitsintervallen hängt u. a. von der Fallzahl ab, die dem Schätzwert zugrunde liegt. Die Verwendung von Unsicherheitsintervallen beschränkt sich nicht nur auf die reine Information über statistische Unsicherheit: die dahinterliegende Unsicherheit wird in der quantitativen Auffälligkeitseinstufung explizit berücksichtigt – teilweise direkt über die Unsicherheitsintervalle.

Bei der quantitativen Auffälligkeitseinstufung der Leistungserbringer handelt es sich um eine statistische Klassifikation. Ziel ist es, die Qualität der Leistungserbringer anhand der vorliegenden Daten auf Auffälligkeiten zu prüfen und im Falle von Auffälligkeiten Stellungnahmeverfahren einzuleiten, die dann ggf. Maßnahmen zur Qualitätsverbesserung initiieren. Da sich die der quantitativen Auffälligkeit anschließenden Prozesse sowohl in deren Konsequenzen als auch in Art und Umfang in einem sensiblen Kontext befinden, ist es notwendig, dass die quantitative Klassifikation treffsicher ist. Zur Definition und Bestimmung der Treffsicherheit, d. h. der Sensitivität und Spezifität einer statistischen Klassifikation, wird jedoch ein Außenkriterium, ein sogenannter Goldstandard, benötigt, den es in diesem Kontext nicht wirklich gibt (vgl. Seite 49 in IQTIG (2019b)).

Basiert die Einstufung ausschließlich auf dem berechneten Indikatorwert eines Leistungserbringers, so besteht die Gefahr, dass insbesondere wenn wenig Information über die interessierende Größe vorhanden ist, zufällige Einflüsse und Konstellationen dazu führen, dass eine Fehlentscheidung getroffen wird. Eine Fehleinschätzung kann dabei in beide Richtungen, d. h. falsch positiv bzw. falsch negativ, erfolgen. Um dies zu vermeiden und zu einer belastbaren Aussage zu gelangen, sollte die statistische Unsicherheit in der Bestimmung des Indikatorwerts bei der quantitativen Einstufung berücksichtigt werden.

### **Enumerative Herangehensweise**

Im Gegensatz dazu liegt das Interesse bei der *enumerativen* Herangehensweise in einer interessierenden Größe in einer endlichen Grundgesamtheit, beispielsweise gegeben durch die vom Leistungserbringer im betrachteten Erfassungsjahr behandelten Patientinnen und Patienten, die die Grundgesamtheitsbedingungen erfüllen. Es wird angenommen, dass die interessierende Größe in der Grundgesamtheit deterministisch ist, d. h. es keine unbekannt oder nicht erfassten Prozesse gibt, die die Messung beeinflussen. Im Kontext der enumerativen Herangehensweise ist es daher im Gegensatz zur analytischen Herangehensweise sinnvoll zu unterscheiden, ob es sich bei der Stichprobenart (siehe nächster Abschnitt) um eine Vollerhebung handelt oder nicht. Handelt es sich um eine Vollerhebung, so wird keine statistische Unsicherheit berücksichtigt und das beobachtete Indikatorergebnis des Leistungserbringers wird direkt mit dem Referenzwert verglichen. Liegt das Leistungserbringerergebnis über dem Referenzwert, dann ist der Leistungserbringer quantitativ auffällig. Liegt hingegen keine Vollerhebung vor, so ist das Ziel,

Rückschlüsse von der Stichprobe auf die endliche Grundgesamtheit zu ziehen, und es muss die statistische Unsicherheit, die durch das Betrachten der Stichprobe entsteht, berücksichtigt werden, um korrekte Inferenz zu betreiben (Kauermann und Küchenhoff 2011). Die enumerative Herangehensweise ist dann angebracht, wenn das Ziel der Auswertung ist, eine Bestandsaufnahme zu machen und die interessierende, deterministische Größe für eine vorgegebene, endliche Population zu beschreiben. Diese Herangehensweise liegt beispielsweise den Auswertungen zu besonders häufigen Dokumentationsfehlern im Rahmen der Datenvalidierung nach QSKH-RL zugrunde (vgl. Kapitel 2 im Abschlussbericht zu „Kriterien für den gezielten Datenabgleich in der Datenvalidierung nach QSKH-RL“ (IQTIG 2018f)).

Zusammenfassend liegen die statistisch-methodischen Herausforderungen in der Auswertungsmethodik hauptsächlich in der analytischen Herangehensweise bzw. in der enumerativen Herangehensweise, wenn (z. B. aus Effizienzgründen) die vorliegenden Daten nur Stichproben der eigentlich interessierenden endlichen Grundgesamtheit ausmachen.

### **5.1.2 Stichprobenart für die vorgegebene Auswertungsebene**

Im Rahmen der externen Qualitätssicherung finden bereits unterschiedliche Stichprobenarten Anwendung. Für die meisten Verfahren der stationären Qualitätssicherung liegen Vollerhebungen vor, weil alle Leistungserbringer, die im betrachteten Erfassungszeitraum Patientinnen und Patienten behandeln, die durch den QS-Filter ausgelöst werden, QS-Dokumentationsdaten zu all diesen Fällen liefern müssen. Hingegen werden im Rahmen der Datenvalidierung und zukünftig auch im Rahmen der Patientenbefragungen Zufallsstichproben von Leistungserbringern bzw. Patientinnen und Patienten beim Leistungserbringer gezogen.

Auch wenn die Unterscheidung, ob eine Vollerhebung vorliegt oder nicht, bei der analytischen Herangehensweise keine Rolle spielt, ist die Art der Stichprobenziehung dennoch für die statistische Inferenz relevant. Wichtig bei der Betrachtung der Stichprobenart ist, dass sie von der Auswertungsebene abhängt. Im Fall der Patientenbefragungen beispielsweise handelt es sich pro Leistungserbringer um eine einfache Zufallsstichprobe. Ist jedoch eine Auswertung aller Fragebögen auf Bundesebene von Interesse, so muss beachtet werden, dass die Gesamtheit der Fragebögen nicht als zufällige Stichprobe aus allen behandelten Patientinnen oder Patienten interpretiert werden kann, weil unterschiedlich viele Fragebögen pro Leistungserbringer eingehen und diese ggf. auch nicht den gleichen Anteil an jeweils bei den Leistungserbringern behandelten Patientinnen und Patienten darstellen.

### **5.1.3 Berechnungsart**

Die Berechnungsart eines Indikators gibt an, wie das Indikatorergebnis (Punktschätzer) konkret basierend auf den Mengendefinitionen der Grundgesamtheit und der interessierenden Ereignisse sowie ggf. zusätzlicher Informationen (wie Patienteneigenschaften bei risikoadjustierten Indikatoren) bestimmt wird. Beispielsweise gibt die Berechnungsart „Rate“ an, dass der zugrunde liegende Kompetenzparameter als Anteil an Fällen mit interessierendem Ereignis an Fällen in der Grundgesamtheit berechnet wird. Neben Indikatoren mit Bewertungsart „Rate“, die in den bisherigen Auswertungen im Rahmen der externen Qualitätssicherung den größten Anteil

ausmachen, finden auch weitere Berechnungsarten Verwendung, wie beispielsweise das standardisierte Mortalitätsverhältnis (SMR, auch oft einfach nur O/E genannt), Kaplan-Meier-adjustierte Rate, standardisiertes Inzidenzverhältnis (SIR) und verschiedene Index-Konstruktionen.

#### 5.1.4 Bewertungsart

Für die Bewertung von Indikatorergebnissen eines Leistungserbringers ist zusätzlich die Bewertungsart entscheidend, welche die statistische Klassifikationsmethode für den Vergleich mit dem Referenzbereich unter Berücksichtigung von Herangehensweise, Stichprobenart und Berechnungsart festlegt und als Operationalisierung des Begriffs „hinreichender Hinweis auf ein Qualitätsdefizit“ verstanden werden soll. Um hervorzuheben, dass es sich bei dem betrachteten Hinweis um statistische Evidenz aus den Ergebnissen des Qualitätsindikators anhand der vorliegenden Daten handelt, wird für die Klassifikation in diesem Bericht, wie oben erläutert, der Überbegriff „quantitative Auffälligkeitseinstufung“ verwendet.

#### 5.1.5 Ziele der Auswertung für einen Leistungserbringer

Zusammenfassend stellen die vier vorgestellten Komponenten Herangehensweise, Stichprobenart für die vorgegebene Auswertungsebene, Berechnungsart und Bewertungsart eine Taxonomie für die Auswertungsmethodik eines Qualitätsindikators dar, und sind vorab für jede Auswertung des Qualitätsindikators festzulegen. Im Folgenden wird nun im Rahmen dieser Taxonomie auf die relevanten Ziele für die Berechnung und Bewertung für Leistungserbringer bei den Auswertungen fokussiert. Folgende Ziele sollen mit der Auswertung für einen Leistungserbringer mit den vorliegenden Daten erreicht werden:

- Zahlenergebnisse, die im Rahmen eines Public Reporting Transparenz über das Versorgungsgeschehen des Leistungserbringers herstellen können (vgl. Kapitel 11 in IQTIG (2019a)) oder für andere Verwendungszwecke weiterverwendet werden können. Dies wird im Folgenden das Leistungserbringerergebnis des QI genannt. Für einen Ratenindikator besteht das Leistungserbringerergebnis z. B. aus: Anzahl an Fällen in der Grundgesamtheit des Leistungserbringers im QI für das relevante Erfassungsjahr, Anzahl an interessierenden Fällen, QI-Wert und das zugehörige Unsicherheitsintervall.
- Eine zusätzliche quantitative Bewertung des beobachteten QI-Ergebnisses des Leistungserbringers für den zu bewertenden Zeitraum als Klassifikation in einer der beiden Kategorien „quantitativ unauffällig“ bzw. „quantitativ auffällig“. Dabei bedeutet „quantitativ auffällig“, dass das QI-Ergebnis hinreichende statistische Hinweise auf ein Qualitätsdefizit enthält, so dass diese im Sinne des Moduls „Qualitätsbewertung“ in Kapitel 4 im Rahmen eines Stellungnahmeverfahrens zu prüfen ist. Analog bedeutet „quantitativ unauffällig“, dass nicht genug bzw. keine Hinweise für ein solches Qualitätsdefizit vorliegen. Die quantitative Einstufung ist somit als erster Schritt eines Prozesses zu sehen, welcher im zweiten Schritt quantitativ auffälligen Leistungserbringern die Möglichkeit zur Reflexion ihrer Kompetenz und ggf. zur Erklärung dieser quantitativen Auffälligkeit im Rahmen einer Stellungnahme bietet, die dann fachlich bewertet wird.

Obwohl die Zahlen des Public Reportings nicht zwingend die gleiche Datengrundlage verwenden müssen<sup>33</sup> wie die quantitative Auffälligkeitseinstufung, ist es wichtig, dass die beiden Ziele mit einem einheitlichen statistisch-methodischen Ansatz verfolgt werden. Dies ist notwendig, um spätere Inkonsistenzen soweit möglich zu vermeiden, z. B. wenn im Rahmen eines Stellungnahmeverfahrens über die Ergebnisse der einzelnen Jahre diskutiert wird. In Abschnitt 5.1 wird daher ein einheitlicher Rahmen dargestellt, welcher die Komponenten der statistischen Auswertung eines Qualitätsindikators benennt und kategorisiert. Da für die vorliegende Beauftragung aus statistischer Sicht vor allem die Methodik zur quantitativen Auffälligkeitseinstufung von Interesse ist, wird diese in Abschnitt 5.2 in den Gesamtrahmen des Bewertungsprozesses eingebettet. Es werden verschiedene Vorschläge für eine Methodik für die quantitative Auffälligkeitseinstufung gemacht.

Das Ziel der nächsten Abschnitte ist die Entwicklung einer mathematisch-statistischen Methodik, welche fair und transparent und ggf. unter Berücksichtigung stochastischer Unsicherheit die Ergebnisse eines Leistungserbringers mit dem Referenzwert vergleicht. Dies soll die Basis darstellen für die Entscheidung, ob anhand der beobachteten Daten hinreichende statistische Hinweise auf ein Qualitätsdefizit vorliegen.

Bei der Operationalisierung einer solchen Methodik ist zu bedenken, dass jede quantitative Auffälligkeit prinzipiell einen qualitativen Prozess nach sich ziehen wird. Der Vorschlag aus Kapitel 4 ist es, diesen qualitativen Prozess im Rahmen eines Stellungnahmeverfahrens durchzuführen, vgl. Abbildung 9. Zusammenfassend kann der Bewertungsprozess daher als zweistufiges Verfahren mit den folgenden beiden Stufen angesehen werden:

- Schritt 1: Es werden Leistungserbringer identifiziert, bei denen anhand der vorliegenden Daten genügend statistische Hinweise auf ein Qualitätsdefizit bestehen (quantitative Auffälligkeitseinstufung).
- Schritt 2: Nur wenn die beobachteten Ergebnisse zu einer solchen statistisch belastbaren Abweichung vom Referenzwert des betrachteten Qualitätsindikators führen, kommt es zur Initiierung eines Stellungnahmeverfahrens wie es in Kapitel 4 beschrieben wird, welches die statistischen Hinweise überprüft und entsprechend fachlich bewertet (qualitativer Schritt).

Damit diese qualitative Überprüfung sorgfältig, einheitlich, transparent und fair durchgeführt werden kann, sind entsprechende Ressourcen für Schritt 2 notwendig. Stehen diese Ressourcen nicht im Einklang mit dem Aufwand, kann dies schnell die Belastbarkeit der entsprechend getroffenen Aussagen und Bewertungen infrage stellen. Ein wichtiger praktischer Aspekt für die Effizienz des gesamten Bewertungsprozesses ist dabei, wie oft es zu Abweichungen zwischen der Klassifikation des Leistungserbringerergebnisses in Schritt 1 und der anschließenden fach-

---

<sup>33</sup> Beispielsweise könnten die QI-Ergebnisse sich auf das Erfassungsjahr 2019 beziehen, die quantitative Auffälligkeitseinstufung jedoch sowohl auf die Daten des Erfassungsjahres 2018 als auch 2019. Der Qualitätsbericht der Krankenhäuser würde dann im Rahmen des Public Reportings nur die Ergebnisse für das Erfassungsjahr 2019 enthalten und das Ergebnis des Bewertungsprozesses. Aus Gründen der Transparenz müssten im Rückmeldebericht an den Leistungserbringer jedoch alle verwendeten Leistungserbringerergebnisse mitgeteilt werden.

lichen Bewertung in Schritt 2 kommt. Wird ein sehr starker statistischer Hinweis für die Auslösung des Stellungnahmeverfahrens gefordert, kann dies dazu führen, dass wichtige Qualitätsdefizite unentdeckt bleiben.

Die zentrale statistische Frage ist daher, wie genau und anhand welcher Informationen der in Kapitel 4 definierte Begriff „hinreichender Hinweis auf ein Qualitätsdefizit“ im Rahmen der Bewertungsart der statistischen Auswertungsmethodik operationalisiert werden soll, damit die verfügbaren Ressourcen zur fachlichen Bewertung effizient eingesetzt werden können.

Aus biometrischer Perspektive umfasst die Beantwortung dieser Frage gleichzeitig auch eine Beantwortung von Punkt 3.a des Auftrags, d. h. der „Optimierung der Diskriminationsfähigkeit der Indikatoren unter Prüfung verschiedener (auch derzeit noch nicht genutzter) statistischer Verfahren, Abwägung von Vor- und Nachteilen, z. B. in Anlehnung an das Verfahren „Planungsrelevante Qualitätsindikatoren“. Unter dem Begriff „Diskriminationsfähigkeit“ wird dabei die Klassifikationsgüte des Verfahrens zur quantitativen Auffälligkeitseinstufung verstanden. Um diese Frage im Rahmen der „Optimierung der Effizienz des Verfahrens“ diskutieren zu können, ist es notwendig, die quantitative Auffälligkeitseinstufung in einen formalen Rahmen einzubetten. Innerhalb eines solchen entscheidungstheoretischen Rahmens könnten verschiedene Methoden für die Operationalisierung des Begriffs „hinreichender Hinweis“ statistisch formuliert werden und deren „Effizienz“ mittels Verlustfunktionen verglichen werden. Die Formulierung der Methoden wird dabei in zwei Teile aufgeteilt:

- Teil 1: Optimierung der quantitativen Auffälligkeitseinstufung anhand der existierenden Datengrundlage (d. h. jährliche Erfassung) – siehe Abschnitt 5.3.
- Teil 2: Erweiterung von Teil 1, indem die Daten- bzw. Entscheidungsgrundlage auf einen 2-Jahres-Zeitraum erweitert wird – siehe Abschnitt 5.4. Der Grund für die Erweiterung ist, dass viele der prinzipiellen Festlegungen generisch für den 1-Jahres-Zeitraum diskutiert werden können. Die Erweiterung auf einen 2-Jahres-Zeitraum führt zu einer Reihe an zusätzlichen Festlegungen, die dann gesondert diskutiert werden.

## 5.2 Entscheidungstheoretische Modellierung des Bewertungsprozesses

Betrachtet wird die Situation für einen Qualitätsindikator (QI), der als Rate ausgedrückt werden kann und für den eine analytische Herangehensweise festgelegt wird. Es wird angenommen, dass der Referenzbereich für den QI  $[0, R]$  ist, d. h. die im Indikator interessierenden Ereignisse sind unerwünschte Ereignisse, z. B. Todesfälle. Der Entscheidungsprozess der Bewertung eines Leistungserbringers für einen solchen Ratenindikator kann im Rahmen einer Spezifikationsprache für Entscheidungsprobleme als sogenanntes *Entscheidungsgraph* bzw. *Einflussdiagramm* (Englisch: *Influence Diagram*) grafisch dargestellt werden, vgl. z. B. Kapitel 9 in Jensen und Nielsen (2007). Entscheidungsgraphen erweitern die sogenannten *Bayesianischen Netzwerke* um Entscheidungen und illustrieren, welche Informationen für konkrete Entscheidungen vorliegen und welche Kriterien bzgl. der Entscheidungsfindung optimiert werden sollen.

### 5.2.1 Bayesianische Netzwerke und Einflussdiagramme

Bayesianische Netzwerke<sup>34</sup> sind statistische Modelle, welche probabilistische Abhängigkeitsstrukturen zwischen (Zufalls-)Variablen mittels gerichteter azyklischer Graphen illustrieren. Wie z. B. in Jensen und Nielsen (2007) oder Pearl (2009) erläutert, werden hierbei Variablen als runden Knoten im Netzwerk dargestellt, Pfeile symbolisieren bedingte Abhängigkeiten in der gemeinsamen Verteilung der Zufallsvariablen und das Fehlen von Pfeilen zwischen Variablen illustriert bedingte Unabhängigkeiten. Im Rahmen der analytischen Auswertung eines Ratenindikators können beispielsweise die Zusammenhänge zwischen der zugrunde liegenden Rate  $\theta$  des Leistungserbringers (später auch Kompetenzparameter genannt), Fallzahl  $J$  und der Anzahl an unerwünschten Qualitätsereignissen  $O$  eines Leistungserbringers unter Berücksichtigung der innewohnenden Stochastizität wie in Abbildung 10 über Zufallsvariablen dargestellt werden:

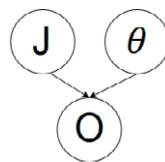


Abbildung 10: Bayesianisches Netzwerk zur Illustration der Zusammenhänge zwischen Fallzahl  $J$ , Kompetenzparameter  $\theta$  und Anzahl unerwünschter Qualitätsereignisse  $O$  für einen Ratenindikator.

Zum Beispiel illustriert das Netzwerk in Abbildung 10, dass die Kenntnis von  $J$  und  $\theta$  die Variable  $O$  probabilistisch festlegen und dass  $J$  und  $\theta$  voneinander stochastisch unabhängig sind, d. h. die gemeinsame Wahrscheinlichkeitsverteilung  $f(J, \theta, O) = f(O|J, \theta)f(J)f(\theta)$  ist. Die Darstellung alternativer Zusammenhänge, z. B. dass die Fallzahl auch die Kompetenz beeinflusst, ist im Diagramm über einen Pfeil von  $J$  zu  $\theta$  möglich. Für alle Knoten ohne „Eltern“ im Graphen (d. h. Knoten, bei denen keine Pfeile in Richtung des Knotens laufen) muss im bayesianischen Netzwerk eine A-priori-Verteilung gewählt werden, welche das Vorwissen über die Zufallsgröße über eine Verteilung parametrisiert. Da die Wahl einer geeigneten A-priori-Verteilung ein wichtiger Aspekt der bayesianischen Inferenz darstellt, wird in Abschnitt 5.3.2.3 auf verschiedene Möglichkeiten der A-priori-Wahl eingegangen. Da die Fallzahl  $J$  in der Auswertungsmethodik, wie in Abschnitt 5.1.1 erläutert, als gegeben angesehen wird, und somit die A-priori-Verteilung für  $J$  entsprechend eine Punktmasse hat, muss nur für den Kompetenzparameter  $\theta$  eine A-priori-Verteilung angenommen werden.

Wichtig ist an dieser Stelle, dass sich grafische Modelle dieser Art aus Sicht des IQTIG hervorragend dafür eignen, die komplexe Zusammenhangsstruktur zwischen den relevanten Variablen für die Auswertungsmethodik vereinfachend darzustellen. Grafische Modelle sind für diesen Zweck bereits bei der Beschreibung der statistischen Auswertungsmethodik bei Patientenbefragungen zum Einsatz gekommen (IQTIG 2018c, IQTIG 2018d). Das Einflussdiagramm stellt die entscheidungstheoretische Erweiterung des bayesianischen Netzwerkes dar. Es erweitert die Notation der bayesianischen Netzwerke mit zusätzlichen Symbolen für Nutzen- bzw. Verlustfunk-

<sup>34</sup> In der Literatur auch „Kausaldiagramm“ oder „Kausalgraph“ genannt (Pearl 2009).



und die Annahme, dass die Behandlungsergebnisse – gegeben dem Kompetenzparameter – unabhängig voneinander sind, führt dazu, dass  $O \sim \text{Bin}(J, \theta)$ . Das heißt, dass die Anzahl an interessierenden Ereignissen binomialverteilt ist mit Anzahlparameter  $J$  und zugrunde liegender Wahrscheinlichkeit<sup>35</sup>  $\theta$ . Zur besseren Veranschaulichung der Rolle des Referenzwertes  $R$  wird dieser auch als Knoten im Diagramm repräsentiert, obwohl er keine Zufallsvariable darstellt – rein technisch entspricht die Verteilung einer Ein-Punkt-Verteilung mit der kompletten Wahrscheinlichkeitsmasse im Wert.

Die erste Entscheidung, die im Bewertungsprozess eines Leistungserbringers getroffen wird, ist die Beantwortung der Frage, ob anhand der beobachteten Daten für den Leistungserbringer ein Signal ( $S$ ) in Form einer quantitativen Auffälligkeit für den betrachteten Indikator generiert werden soll. Anhand der vorliegenden Informationen soll also entschieden werden, ob man sich für  $S = \text{ja}$  oder  $S = \text{nein}$  entscheidet. Dies wird in Form einer sogenannten Strategie für die Entscheidung dargestellt, d. h. einer Funktion, die für jede vorhandene Information – hier ein Tripel aus beobachteter Anzahl an unerwünschten Qualitätsergebnissen, der Fallzahl und dem Referenzwert – eine entsprechende Entscheidungsalternative auswählt:

$$s: (o, J, R) \rightarrow \{\text{ja, nein}\}.$$

Das heißt,  $s$  stellt die Strategie der Auffälligkeitseinstufungsmethode da, wohingegen die Variable  $S$  das Ergebnis der Auffälligkeitseinstufung darstellt. Für die Entscheidung spielt der Referenzbereich  $[0, R]$  des QI eine wichtige Rolle: gilt für den Leistungserbringer  $\theta \leq R$ , dann erfüllt der Leistungserbringer die QI-spezifischen Qualitätsanforderungen. Ist dagegen  $\theta > R$ , dann ist dies ein Hinweis auf ein mögliches Qualitätsdefizit. Die Schwierigkeit im analytischen Kontext ist jedoch, dass  $\theta$  für den Leistungserbringer nicht bekannt ist, sondern nur anhand der beobachteten Daten  $J$  und  $o$  im inferenzstatistischen Sinne geschätzt werden kann, vgl. z. B. IQTIG (2016: Abschnitt 3.4) bzw. IQTIG (2019a: Abschnitt 13.1). Die Art und Weise, wie die beobachteten Daten unter Berücksichtigung von Unsicherheit Rückschlüsse auf  $\theta$  erlauben, ist eine der Herausforderungen bei der Entwicklung der statistischen Auswertungsmethodik und spiegelt sich in der genauen Form der Funktion  $s$  wieder. Damit die Entscheidung, ob das Signal gegeben wird oder nicht, in einem Rahmen betrachtet wird, welcher es erlaubt, Optimalitätseigenschaften herzuleiten, wird die geeignete Wahl der Strategie  $s$  als Minimierung einer vordefinierten Verlustfunktion gesehen. Der formale Rahmen dafür wird im folgenden Abschnitt vorgestellt; in Abschnitt 5.3.1 wird das Konzept dann auf einen Ratenindikator für zwei Verlustfunktionen angewandt, um unterschiedliche Methoden für die Auffälligkeitseinstufung formal herzuleiten. Die Einbettung in diesen Rahmen erlaubt eine strukturierte Abwägung der Vor- und Nachteile beider Konzepte, welche auch qualitative Dimensionen und Aspekte wie Kommunizierbarkeit beinhalten.

Anhand des in Abbildung 11 dargestellten Entscheidungsdiagramms sind Strategien für die Entscheidungen bzgl. der quantitativen Auffälligkeit (Signal = ja oder nein) und der anschließenden fachlichen Prüfung zu entwickeln. Falls  $s(o, J, R) = \text{nein}$ , wird dies nicht weiter fachlich geprüft

<sup>35</sup> Im Rahmen der externen Qualitätssicherung werden solche Anteils- bzw. Quoten-Indikatoren auch gerne Ratenindikatoren genannt. Somit wird teilweise auch die Bezeichnung „zugrunde liegende Rate“ für den Parameter  $\theta$  verwendet.

und die Qualitätsbewertung im zweiten Schritt entspricht automatisch dem Ergebnis der quantitativen Klassifikation. Falls  $s(o, J, R) = \text{ja}$ , kommt es in einem zweiten Schritt zur fachlich-inhaltlichen Überprüfung des statistischen Hinweises. Dieser zweite Prozessschritt wird im Weiteren unter der Überschrift „Stellungnahmeverfahren“ zusammengefasst. Die konkreten Schritte der Informationssuche für die Überprüfung sind stark indikatorabhängig, da sie die Prüfung von Einzelfällen, die z. B. nicht in den Rechenregeln abgebildete Ausnahmen darstellen (z. B. medizinische Ausnahmetatbestände) sowie eine Prüfung der Datenqualität beinhalten (vgl. Kapitel 4). Die genaue Abhängigkeits- und Informationsstruktur der Entscheidungsfindung im zweiten Schritt lässt sich daher nicht wirklich formalistisch im Rahmen eines Einflussdiagramms abbilden. Um dies zu verdeutlichen, werden die Knoten und Pfeile im Diagramm für diesen Teil bewusst vage als Wolken und geschwungene Pfeile dargestellt. Dies soll verdeutlichen, dass es sich um schwer formalisierbare Größen bzw. schwierig zu definierende Abhängigkeiten handelt. Wichtig im Diagramm ist aber die Feststellung, dass das Stellungnahmeverfahren sich auf zusätzliche Informationen stützt, die über die für den ersten Schritt der quantitativen Einstufung erfassten Daten hinausgehen. Diese zusätzliche Information wird im Diagramm über zusätzliche Pfeile in den Knoten, der die Entscheidung der Qualitätsbewertung (QB) repräsentiert, dargestellt.

Somit wird deutlich, dass die optimale Vorgehensweise sowohl vom Vorgehen der quantitativen Auffälligkeitseinstufung in Schritt 1 als auch des qualitativen Vorgehens in Schritt 2 abhängt. Ein zentraler Aspekt bei der Bewertung eines Qualitätsindikators pro Leistungserbringer ist dabei, inwieweit die Operationalisierung des Qualitätsaspekts durch den Indikator mit den vorliegenden Daten und Mengendefinitionen geeignet ist. Da jedoch sowohl der genaue Charakter der vorliegenden Informationen für die Entscheidung der Qualitätsbewertung als auch die entsprechende Verlustfunktion für die Entscheidung in Schritt 2 sich nur vage konkretisieren lassen, konzentriert sich die statistische Arbeit zur „Optimierung der Effizienz des Verfahrens“ auf den quantitativen Schritt 1 – wohlwissend, dass die Entscheidung in diesem Schritt große Auswirkungen auf das Vorgehen in Schritt 2 hat. Diese Auswirkungen können jedoch nur qualitativ bzw. nur sehr indirekt quantitativ beschrieben werden. Da sich die formale Betrachtung, wie oben begründet, allein auf Schritt 1 der in Abbildung 11 dargestellten Entscheidungssequenz konzentriert, ist das Ziel, eine Strategie für die Entscheidung der quantitativen Auffälligkeit zu entwickeln, indem der erwartete Verlust bzgl. der Verlustfunktion  $L_1$  minimiert wird.<sup>36</sup> Das heißt, das Interesse liegt in der Formulierung einer entsprechenden Verlustfunktion für Schritt 1,  $L_1$ , die das Entscheidungsproblem isoliert für Stufe 1 betrachtet und Informationen, Entscheidung und Verlustfunktion aus Schritt 2 werden im Folgenden nicht weiter betrachtet. Abbildung 12 zeigt das vereinfachte Entscheidungsdiagramm, welches nur die relevanten Teile für Schritt 1 enthält.

---

<sup>36</sup> Mit „erwartetem Verlust“ ist der Erwartungswert der Verlustfunktion unter Verwendung einer spezifischen Entscheidungsstrategie gemeint. Der Erwartungswert entspricht einer gewichteten Summe über die Verluste der jeweiligen Entscheidung summiert über die verschiedenen Zustände des Entscheidungsproblems. Die Gewichte bei der Summierung entsprechen der jeweiligen Wahrscheinlichkeit für den Zustand – vergleiche z. B. Abschnitt 9.2 in Jensen und Nielsen (2007) oder gängige ökonomische Literatur.

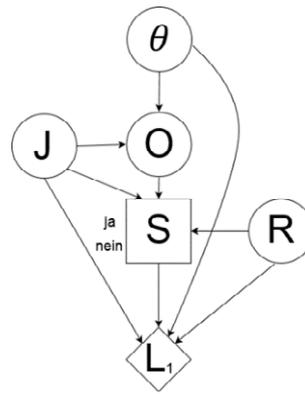


Abbildung 12: Vereinfachtes Entscheidungsdiagramm für die quantitative Auffälligkeitseinstufungsmethode bei einem Ratenindikator.

Tabelle 4: Vier-Felder-Schema zur Darstellung der Aufwände jeder Entscheidung bei der Klassifikation eines Leistungserbringers anhand dessen beobachteter Ergebnisse des Leistungserbringers, d. h.  $(o, J)$ . Exemplarisch dargestellt ist die Situation, wenn nur Fehlklassifikationen zu Aufwänden führen.

		Wahrer Wert des LE	
		$\theta \leq R$	$\theta > R$
Ergebnis der quantitativen Auffälligkeitseinstufung	$s(o, J, R) = \text{nein}$	0	$fn$
	$s(o, J, R) = \text{ja}$	$fp$	0

Vereinfachend kann die Verlustfunktion auch als Vier-Felder-Schema dargestellt werden (siehe Tabelle 4). In Tabelle 4 wird die Situation dargestellt, wenn nur Fehlklassifikationen in Schritt 1 zu Aufwänden<sup>37</sup> führen. Dabei bezeichnet  $fp > 0$  den Aufwand einer falsch-positiven Klassifikation (d. h. ein Leistungserbringer erfüllt den erwartbaren Qualitätsstandard und wird dennoch fälschlicherweise als quantitativ auffällig eingestuft, d. h. es liegt ein hinreichender Hinweis für ein Qualitätsdefizit vor, welcher dann mit entsprechendem Aufwand im Stellungnahmeverfahren korrigiert werden muss),  $fn > 0$  die Aufwände einer falsch-negativen Klassifikation (d. h. ein Leistungserbringer erfüllt nicht den erwartbaren Qualitätsstandard, wird jedoch nicht als quantitativ auffällig eingestuft). Die Frage, in welcher Einheit dieser Aufwand zu messen ist, ist dabei nebensächlich; entscheidend ist dagegen das relative Verhältnis zwischen diesen Aufwänden. Jedoch ist zu beachten, dass Verlustfunktionen, die nur die Fehlklassifikationsaufwände berücksichtigen, den Aufwand für die Durchführung eines Stellungnahmeverfahrens ignorieren, weil diese immer bei der Entscheidung  $s(o, J, R) = \text{ja}$  anfallen, d. h. auch wenn  $\theta > R$ . Des Weiteren berücksichtigt diese Art von Aufwandsbetrachtung auch nicht die Aufwände, welche durch ein mögliches Qualitätsdefizit verursacht werden, z. B. durch Schäden an Patienten und Patientinnen. Nur die Fehlklassifikationsaufwände zu berücksichtigen stellt also im Sinne von Abschnitt

<sup>37</sup> In diesem Kapitel werden die Begriffe „Aufwand“ bzw. „Aufwände“ verwendet, anstelle des gängigeren wirtschaftsmathematischen Begriffs „Kosten“, um zu verdeutlichen, dass die Betrachtungen weiter fassen, als monetäre Kosten.

5.2.3 eine starke Vereinfachung der Entscheidungssituation dar, erlaubt dagegen aber eine explizite Darstellung und Effizienzdiskussion.

### 5.2.3 Diskussion des entscheidungstheoretischen Ansatzes

Der in den vorigen Abschnitten vorgestellte Entscheidungstheoretische Ansatz formuliert explizit, dass es sich bei der quantitativen Auffälligkeitseinstufung um eine Klassifikation handelt. Die Wahl einer konkreten Vorgehensweise für die Auffälligkeitseinstufung fordert eine Abwägung zwischen Vor- und Nachteilen, die teilweise nur theoretisch betrachtet werden können, weil es keinen Goldstandard für die Klassifikation gibt. Um trotzdem die Abwägungen expliziter zu gestalten, werden Verlustfunktionen definiert, die dann unter bestimmten Annahmen zu optimieren sind, um z. B. kosteneffiziente Schwellenwerte für die Klassifikation von QI-Ergebnissen herzuleiten und somit auf theoretischer Grundlage eine Optimierung des Klassifikationsverfahrens herzuleiten. Die Schwierigkeit beim entscheidungstheoretischen Ansatz liegt darin, die Verlustfunktionen so zu formulieren, dass sie adäquat die vielen Aspekte der Entscheidung reflektieren. Dies erscheint im Rahmen der externen Qualitätssicherung extrem schwierig, da sehr viele unterschiedliche Interessen zu berücksichtigen sind. Als Beispiel: eine falsch-negative Klassifikation (d. h. die fälschliche Einstufung eines Leistungserbringer-Ergebnisses als „nicht auffällig“) hat beispielsweise zur Folge, dass kein Stellungnahmeverfahren stattfindet, was sich negativ auf die Behandlungsqualität zukünftiger Patientinnen und Patienten auswirken kann. Eine falsch-positive Klassifikation führt dagegen zu Aufwand auf Seiten des Leistungserbringers, der LAG bzw. Bundesstelle sowie möglicherweise der Patientinnen und Patienten, deren Entscheidung für einen Leistungserbringer durch Fehlinformationen beeinflusst wird. Diesen Aufwand überhaupt nur ansatzweise quantifizieren und gegeneinander abzuwägen ist keine mathematische, sondern vielmehr eine politische Entscheidung, ebenso wie jener, ob eine solche Aufwand-Nutzen-Abwägung überhaupt Grundlage für gesundheitspolitische Entscheidungen sein kann (Gerber-Grote et al. 2014).

Um trotzdem mithilfe des entscheidungstheoretischen Ansatzes zu einer formaleren Diskussion über Klassifikationsgüte und somit über Effizienzoptimierung zu kommen ist es hilfreich, lediglich auf eine wenige Aspekte der Entscheidung bei der Konstruktion der Verlustfunktion zu fokussieren. Dies könnte z. B. sein, nur auf die korrekte Klassifikation bzgl. des Kompetenzparameters zu fokussieren – egal, wie viele Patienten oder Patientinnen der Leistungserbringer behandelt und egal, wie groß die Abweichung vom Referenzwert ist. Alternativ können auch nur die patientenseitigen Nachteile einer falsch-negativen Klassifikationsentscheidung betrachtet und Entscheidungsregeln für die Klassifikation nur anhand von Grenzwerten für ein noch tolerierbares Ausmaß dieser Nachteile abgeleitet werden. Dies erlaubt zwar keine Abwägung der im Prozess entstehenden Gesamtaufwände für alle Parteien, hat jedoch den Vorteil, dass keine Aufwände für falsch-positive und falsch-negative Entscheidungen gegeneinander abgewogen werden müssen.

## 5.3 Methoden für die quantitative 1-Jahres-Auffälligkeitseinstufung

Ausgehend vom Rahmenkonzept in Abschnitt 5.1 und einer expliziten Formulierung der im vorherigen Abschnitt 5.2 vorgestellten Verlustfunktion  $L_1$  werden im Folgenden drei Methoden der

quantitativen Auffälligkeitseinstufung vorgestellt, die unter unterschiedlichen Annahmen über die Verlustfunktion und die Herangehensweise im entscheidungstheoretischen Kontext jeweils die optimale Vorgehensweise zur quantitativen Auffälligkeitseinstufung darstellen. Zwei dieser Methoden finden im Rahmen der externen Qualitätssicherung bereits Anwendung: die rechnerische Auffälligkeit nach § 10 QSKH-RL sowie die statistische Auffälligkeit für die Verfahren im Rahmen der plan. QI-RL (IQTIG 2016), die im Folgenden *statistisch signifikante Auffälligkeit* genannt wird. Für die statistische signifikante Auffälligkeit wird eine bayesianische Alternative vorgeschlagen, die quasi identisch zur Methodik des Verfahrens für planungsrelevante Qualitätsindikatoren ist (siehe Abschnitt 5.3.1.2). Die bayesianische Variante bietet einen flexibleren Rahmen für quantitative Auffälligkeitseinstufungen und führt dabei zu ähnlichen Klassifikationsentscheidungen, wie die in der plan. QI-RL verwendete frequentistische Variante. Zusätzlich zur rechnerischen und statistisch signifikanten Auffälligkeit wird anschließend eine dritte Einstufungsmethodik, die *statistisch relevante Auffälligkeit*, vorgestellt, welche von anderen Aufwandsannahmen ausgeht als die statistisch signifikante Auffälligkeit und die rechnerische Auffälligkeit.

Im Folgenden wird von einem ratenbasierten Qualitätsindikator mit festem Referenzwert ausgegangen, der anhand der Daten eines Erfassungsjahres ausgewertet werden soll. Erweiterungen für risikoadjustierte Indikatoren und verteilungsabhängige Referenzwerte werden in den Abschnitten 5.3.2.2 und 5.3.4 diskutiert.

### 5.3.1 Lösung des Entscheidungsdiagramms

Die rechnerische, die statistisch signifikante und die statistisch relevante Auffälligkeitseinstufung werden im Folgenden definiert und die entsprechenden Annahmen erläutert, in der die jeweilige Auffälligkeitseinstufungs-Methode die optimale Lösung des Entscheidungsdiagramms darstellt. Dazu werden der Aufwand für jede Entscheidung als Funktion des Kompetenzparameters, des Referenzwertes und der binären Klassifikation wie in Abschnitt 5.2 beschrieben analysiert. Um ein Entscheidungsdiagramm zu lösen, d. h. um die Strategie zu finden, die entsprechend der Verlustfunktion optimale Entscheidungen trifft, muss sequentiell von hinten nach vorne jeweils die optimale Entscheidungsstrategie gefunden werden (Jensen und Nielsen 2007). Dabei wird die zugehörige Verlustfunktion nach dem vorgegebenen Zielkriterium minimiert. Im Folgenden wird als Zielkriterium die Minimierung des erwarteten Verlusts verwendet. Es können aber auch andere Zielkriterien wie z. B. Minimierung des Worst-Case-Ergebnisses o. Ä. von Relevanz sein.

Im Rahmen eines Ratenindikators ist die optimale Strategie für die Einstufung in Schritt 1 als folgendes Problem zu lösen

$$s(o, J, R) = \operatorname{argmin}_{s \in S} E(L_1(s, J, R, \theta)) = \operatorname{argmin}_{s \in S} \int_0^1 L_1(s, J, R, \theta) f(\theta|o, J) d\theta,$$

wobei  $S = \{\text{ja, nein}\}$  und der Erwartungswert über alle Zufallsgrößen in der Verlustfunktion gebildet wird, d. h. über alle zum Zeitpunkt der Entscheidung unbekanntes Größen in der Verlustfunktion. Im konkreten Fall von Abbildung 11 wird nur über den Kompetenzparameter  $\theta$  integriert, für welchen zur expliziten Berechnung des Integrals eine Wahrscheinlichkeits-Verteilung

$f(\theta|o, J)$  spezifiziert werden muss (siehe Abschnitt 5.3.2.1 bzw. Abschnitt 5.3.2.2). Da für die Wahl der optimalen Entscheidung bei der quantitativen Auffälligkeitseinstufung genau zwei Entscheidungsalternativen miteinander kontrastiert werden, gilt entsprechend, dass

$$s(o, J, R) = I\{E[L_1(s = ja, J, R, \theta)] < E[L_1(s = nein, J, R, \theta)]\},$$

wobei  $I(\cdot)$  die Indikatorfunktion bezeichnet. Das heißt, dass die quantitative Auffälligkeitseinstufung so getroffen wird, dass der erwartete Verlust minimal ist. Hier wurde die Konvention gewählt, dass bei gleichem erwartetem Verlust der beiden Entscheidungsalternativen die Entscheidung – zugunsten des Leistungserbringers – für die Alternative „nein“ getroffen wird.

Um die Fallzahlabhängigkeit der gefundenen Entscheidungsregel explizit darzustellen, wird bei vorgegebenen  $J$  und Referenzwert  $R$  das kleinste  $o$  berichtet, welches zu einer Einstufung „quantitativ auffällig“ führt, d. h.

$$o_{\min}(J, R) = \operatorname{argmin}_{0 \leq o \leq J} \{s(o, J, R) = ja\}.$$

Eine Erkenntnis dabei ist, dass es in Abhängigkeit der Entscheidungsregel Fallzahlen geben kann, die so klein sind, dass es selbst bei der Beobachtung  $o = J$  nicht zu einer quantitativen Auffälligkeit kommt. Dies bedeutet, dass es bei dieser Fallzahl *unabhängig vom Ergebnis* nicht genügend statistische Evidenz für ein mögliches Qualitätsdefizit geben kann. Die kleinste Fallzahl, ab der ein Leistungserbringer überhaupt als quantitativ Auffällig eingestuft werden kann, wird mit  $J_{\min}(R) = \min\{J \geq 0 \mid s(o = J, J, R) = ja\}$  gekennzeichnet. Eine wichtige Frage für die Qualitätssicherung ist, wie mit Leistungserbringern umzugehen ist, die eine Fallzahl kleiner als  $J_{\min}(R)$  haben. Dies wird in Abschnitt 5.3.1.5 diskutiert.

Ausgangspunkt für alle weiteren Betrachtungen ist die Formulierung unterschiedlicher Verlustannahmen analog zur im letzten Abschnitt vorgestellten Vier-Felder-Tafel (vgl. Tabelle 4). Während in der Literatur zu einschrittigen Klassifikationsverfahren üblicherweise kein Aufwand für richtig-positive Klassifikationen (d. h. in diesem Fall korrekt klassifizierten Qualitätsdefiziten) angenommen werden (vgl. u. a. Berger (2010)), ist diese Annahme für ein zweistufiges Klassifikationsverfahren, bestehend aus quantitativer Auffälligkeitseinstufung und fachlicher Klärung, nicht sinnvoll. Vielmehr ist es sinnvoll, sowohl im Falle einer richtig-positiven als auch einer falsch-positiven Klassifikation feste Aufwände  $\zeta$  eines Stellungnahmeverfahrens anzunehmen. Betrachtet wird daher in Tabelle 5 eine im vgl. zu Tabelle 4 modifizierte Vier-Felder-Tafel.

Tabelle 5: Vier-Felder-Tafel der Aufwandsannahmen des zweischrittigen Klassifikationsproblems

		Wahrer Wert des LE	
		$\theta \leq R$	$\theta > R$
Ergebnis der quantitativen Auffälligkeitseinstufung	$s(o, J, R) = \text{nein}$	0	$fn$
	$s(o, J, R) = \text{ja}$	$\zeta$	$\zeta$

Unterschiedliche Aufwandsannahmen für falsch-negative Klassifikationen (d. h. nicht entdeckte Qualitätsdefizite) führen zu unterschiedlichen Lösungen des Entscheidungsdiagramms bzw. zu unterschiedlichen optimalen Einstufungsverfahren. In Abschnitt 5.3.1.1 wird gezeigt, dass die

aus der QSKH-Richtlinie als rechnerische Auffälligkeit bekannte Einstufungsstrategie entscheidungstheoretisch optimal ist, unter der Annahme, dass die Schätzung des Kompetenzparameters  $\theta$  keiner Unsicherheit unterliegt (enumerative Herangehensweise, also  $\theta = o/J$ ), und dass der Aufwand  $fn$  unabhängig von der Fallzahl  $J$  des Leistungserbringers und dem Ausmaß der Überschreitung  $(\theta - R)$  des Referenzbereiches ist. Unter den gleichen Aufwandsannahmen, aber mit Berücksichtigung statistischer Unsicherheit (analytische Herangehensweise) wird in Abschnitt 5.3.1.2 gezeigt, dass die statistisch signifikante Einstufung die optimale Einstufungsstrategie darstellt. Schließlich wird in Abschnitt 5.3.1.3 betrachtet, welches die optimale Einstufungsstrategie ist, unter der Annahme, dass der Aufwand  $fn$  sowohl vom Ausmaß der Überschreitung des Referenzbereiches als auch von der Fallzahl des Leistungserbringers abhängt.

### 5.3.1.1 Rechnerische Auffälligkeit

Betrachtet wird nun ein ratenbasierter Qualitätsindikator mit  $o$  als beobachtete Anzahl an interessierenden Ereignissen,  $J$  als Fallzahl an behandelten Patientinnen und Patienten im betrachteten Zeitraum und  $R$  als Referenzwert des Indikators, d. h. mit Referenzbereich  $[0, R]$ . Die rechnerische Auffälligkeit nach § 10 QSKH-RL ist definiert als Entscheidungsstrategie

$$s_{\text{rech}}(o, J, R) = I\left(\frac{o}{J} > R\right),$$

wobei  $I(A)$  die Indikatorfunktion darstellt, welche 1 (bzw. „ja“) ist, falls das Boolesche Argument  $A$  erfüllt ist<sup>38</sup> und 0 (bzw. „nein“) ist, falls  $A$  nicht erfüllt ist.

Aus den in Tabelle 5 beschriebenen Aufwandsannahmen ergibt sich der erwartete Verlust  $l_{\text{rech}}(\cdot)$  für die quantitative Einstufung eines Leistungserbringerergebnisses für die beiden Entscheidungsmöglichkeiten  $s_{\text{rech}} \in \{\text{ja}, \text{nein}\}$  als jeweils

$$E\left[l_{\text{rech}}\left(o, J, R \mid \theta = \frac{o}{J}, s = \text{ja}\right)\right] = \zeta$$

für  $s = \text{ja}$  und

$$E\left[l_{\text{rech}}\left(o, J, R \mid \theta = \frac{o}{J}, s = \text{nein}\right)\right] = fn \cdot P\left(\theta > R \mid \theta = \frac{o}{J}\right)$$

für  $s = \text{nein}$ . Da in der enumerativen Herangehensweise keine Unsicherheit berücksichtigt wird, gilt

$$P\left(\theta > R \mid \theta = \frac{o}{J}\right) = I\left(\frac{o}{J} > R\right),$$

d. h. die Kompetenz des Leistungserbringers ist entweder sicher über dem Referenzwert  $I\left(\frac{o}{J} > R\right) = 1$  oder sicher unterhalb des Referenzwertes  $I\left(\frac{o}{J} > R\right) = 0$ .

<sup>38</sup> im obigen Fall ist das Boolesche Argument  $A$  gleich  $o/J > R$ .

Die optimale Entscheidungsregel, welche den erwarteten Verlust minimiert, ist in diesem Fall daher:

$$s_{\text{rech}}\left(o, J, R \mid \theta = \frac{o}{J}\right) = \mathbb{I}\left(\zeta < fn \cdot \mathbb{I}\left(\frac{o}{J} > R\right)\right),$$

d. h.

$$s_{\text{rech}}\left(o, J, R \mid \theta = \frac{o}{J}\right) = \mathbb{I}\left(\frac{\zeta}{fn} < \mathbb{I}\left(\frac{o}{J} > R\right)\right).$$

Überwiegt der Aufwand eines Stellungnahmeverfahrens  $\zeta$  dem Aufwand eines nicht verfolgten Qualitätsdefizites  $fn$ , d. h. im Fall

$$\frac{\zeta}{fn} \geq 1,$$

würde die optimale Einstufungsstrategie keine Leistungserbringer als quantitativ auffällig einstufen. Umgekehrt, falls

$$\frac{\zeta}{fn} < 1,$$

wird ein Leistungserbringer genau dann quantitativ auffällig wenn gilt

$$\frac{o}{J} > R.$$

Dies entspricht der rechnerischen Auffälligkeitseinstufung der QSKH-RL. Die rechnerische Auffälligkeit ist daher die optimale Entscheidungsstrategie zur Minimierung des erwarteten Verlusts, sofern angenommen wird, dass  $\theta = o/J$  ist,  $fn > \zeta$  ist und der Aufwand  $fn$  als unabhängig von Fallzahl und Ausmaß der Überschreitung des Referenzbereiches angenommen werden. Mit anderen Worten ist die rechnerische Auffälligkeitseinstufung nur optimal, wenn es keine Unsicherheit bei der Bestimmung des Kompetenzparameters gibt, d. h. im Rahmen der vorgestellten Taxonomie nur bei enumerativer Herangehensweise mit Vollerhebung. In allen anderen Situationen ist die Verwendung der rechnerischen Auffälligkeitseinstufung suboptimal.

### 5.3.1.2 Statistisch signifikante Auffälligkeit

Bei der statistisch signifikanten Auffälligkeit wird, anders als bei der rechnerischen Auffälligkeit, davon ausgegangen, dass die beobachtete Anzahl an interessierenden Ereignissen eine Zufallsvariable und der zugrunde liegende Parameter  $\theta$  eines Leistungserbringers nicht direkt beobachtbar ist, sondern anhand der beobachteten Daten lediglich im inferenz-statistischen Sinne geschätzt werden muss. In der frequentistischen Version der statistischen Auffälligkeit wird anhand eines Hypothesentests zwischen der Nullhypothese  $H_0: \theta \leq R$  und der Alternativhypothese  $H_1: \theta > R$  entschieden, ob ein Leistungserbringerergebnis quantitativ auffällig oder unauffällig ist. In der bayesianischen Version wird diese Entscheidung über die A-posteriori-Wahrscheinlichkeit für die Nullhypothese unter zusätzlichen Verteilungsannahmen getroffen (siehe Abschnitte 5.3.2.1 und 5.3.2.2). Die frequentistische Version findet bereits Anwendung

im Verfahren nach plan. QI-RL (IQTIG 2016). Da die bayesianische Version jedoch zu fast identischen Klassifikationsstrategien und somit auch fast identischen Ergebnissen führt und gleichzeitig einen flexibleren Rahmen bietet, um auch Auffälligkeitseinstufungen bei komplexeren Indikatorarten wie z. B. Patientenbefragungen zu berücksichtigen, wird im Folgenden die bayesianische Version dargestellt und diskutiert.

### Bayesianische Version der statistisch signifikanten Auffälligkeit

Ausgehend von den Aufwandsannahmen in der Vier-Felder-Tafel aus Tabelle 5 werden die gleichen Aufwandsannahmen wie für die rechnerische Auffälligkeitseinstufung gestellt. Dabei ist  $fn > 0$  wieder der Aufwand einer falsch-negativen Klassifikation,  $\zeta > 0$  der Aufwand eines Stellungnahmeverfahrens. Wieder wird davon ausgegangen, dass der Aufwand einer Fehlklassifikation des Kompetenzparameters immer gleich ist, d. h. nicht von Eigenschaften des Leistungserbringers, wie z. B. der Fallzahl und dem Ausmaß der Überschreitung des Referenzbereichs, abhängen. Zusätzlich wird, anders als bei der rechnerischen Auffälligkeit, davon ausgegangen, dass der Parameter  $\theta$  nicht direkt beobachtbar ist, sondern nur aus den beobachteten Daten (und möglichen Vorinformationen) geschätzt werden kann. Diese Schätzung ist mit Unsicherheit behaftet, weshalb die Verteilung der Wahrscheinlichkeiten für den Parameter  $\theta$  von Interesse ist. Für die Klassifikationsentscheidung ist dabei die Wahrscheinlichkeit der Nullhypothese  $P(\theta \leq R)$ , d. h. das Vorliegen erwartbarer Qualität, von Interesse. Aus dieser und den oben angenommenen Aufwänden für Fehlklassifikationen ergeben sich die beiden erwarteten Verluste als:

$$E[l_{\text{stat.sig.bayes}}(o, J, R, \theta, s = \text{ja})] = \zeta$$

und

$$E[l_{\text{stat.sig.bayes}}(o, J, R, \theta, s = \text{nein})] = fn \cdot P(\theta > R).$$

Daraus ergibt sich

$$s_{\text{stat.sig.bayes}}(o, J, R, \alpha) = I(\zeta < fn \cdot (1 - P(\theta \leq R))) = I\left(P(\theta \leq R) < \frac{fn - \zeta}{fn}\right).$$

Bei der bayesianischen Version der statistisch signifikanten Auffälligkeit wird also die Wahrscheinlichkeit der Nullhypothese  $P(\theta \leq R)$  mit dem Aufwandsverhältnis  $\frac{fn - \zeta}{fn}$  verglichen. Wird also die Entscheidungsregel  $I(P(\theta \leq R) < \alpha)$  verwendet, dann ist diese Regel für die oben angenommene Verlustfunktion optimal, wenn der Schwellenwert  $\alpha = (fn - \zeta)/fn$  gewählt wird. Die Wahl von  $\alpha$  im Rahmen der (bayesianischen) statistisch signifikanten Auffälligkeit kann daher auch so interpretiert werden, dass über eine Festlegung des Aufwandsverhältnisses zwischen dem Aufwand eines Stellungnahmeverfahrens und dem Aufwand einer falsch-negativen Einstufung ein  $\alpha$  festgelegt wird. Der Fall  $\alpha = 0,05$  impliziert daher beispielsweise, dass das Ver-

hältnis des Aufwands eines Stellungnahmeverfahrens gegenüber einer falsch-negativen Klassifikation genau 19 zu 20 ist.<sup>39</sup> Mit dieser Wahl von  $\alpha = 0,05$  werden somit dem Übersehen eines Qualitätsdefizites nur wenig mehr Aufwände zugeschrieben, als dem Aufwand eines Stellungnahmeverfahrens.<sup>40</sup> Wird der Aufwand für das Übersehen eines Qualitätsdefizits dagegen beispielsweise als doppelt so hoch bewertet wie der Aufwand eines Stellungnahmeverfahrens, so führt ein  $\alpha$  von  $(2\zeta - \zeta)/(2\zeta) = 0,5$  zur optimalen Entscheidungsregel, welches zu einer strenger Bewertung als die rechnerische Auffälligkeit führen würde.

Die konkrete Berechnung der Wahrscheinlichkeit von  $P(\theta \leq R)$  im Rahmen von ratenbasierten und risikoadjustierten Indikatoren wird in Abschnitt 5.3.2 erläutert. Insgesamt ist die Interpretation des Schwellenwertes  $\alpha$  bei der Klassifikation im bayesianischen Rahmen sehr verwandt mit dem Signifikanzniveau  $\alpha$  des einseitigen frequentistischen Tests – oft wird der Schwellenwert im obigen bayesianischen Verfahren daher auch einfach Signifikanzniveau genannt.

Im bayesianischen Kontext gilt  $P(\theta \leq R) = 1 - P(\theta > R)$ , d. h. obiges Vorgehen ist äquivalent zum Vorgehen, die Alternativhypothese  $\theta > R$  anzunehmen, falls die Wahrscheinlichkeit dafür größer als  $1 - \alpha$ , d. h. z. B. größer als 95 % ist. Ist dies nicht der Fall, d. h. kommt es nicht zu einer Entscheidung für die Hypothese  $\theta > R$ , dann gilt im Umkehrschluss nicht unbedingt, dass ein hinreichender Hinweis für  $\theta \leq R$  vorliegt (vgl. u. a. Altman und Bland (1995)). Soll zusätzlich entschieden werden, ob ein hinreichender Hinweis dafür vorliegt, dass ein Leistungserbringergebnis im Referenzbereich liegt, könnte dies z. B. anhand der zusätzlichen Entscheidungsregel  $I(P(\theta \leq R) > 1 - \alpha)$  geschehen.

### 5.3.1.3 Statistisch relevante Auffälligkeit

Bisher wurde angenommen, dass der Aufwand für eine falsch-negative Einstufung eines Leistungserbringers fix ist und nicht von weiteren Faktoren abhängt, wie z. B. der behandelten Anzahl an Patienten und Patientinnen oder der Größe der Abweichung vom Referenzwert. Alternativ kann angenommen werden, dass die falsch-negative Klassifikation eines Leistungserbringers, der im nächsten Beobachtungszeitraum viele Fälle behandeln wird, aus Sicht der Patienten und Patientinnen höhere Aufwände verursacht, als jene eines Leistungserbringers mit wenigen Fällen. Gleiches gilt auch für die Größe der Abweichung des Kompetenzparameters vom Referenzwert bei zwei Leistungserbringern mit gleicher Fallzahl: Ist die Abweichung vom Referenzbereich sehr groß, dann ist die Anzahl an betroffenen Patienten und Patientinnen größer, als bei einem Leistungserbringer mit gleicher Fallzahl, aber mit geringer Abweichung vom Referenzbereich. In beiden Fällen liegt die Annahme zugrunde, dass das Stellungnahmeverfahren zu einer Aufdeckung von möglichen Qualitätsdefiziten führt und damit einer Verbesserung der Behandlungsqualität für alle nachfolgend behandelten Patientinnen und Patienten nach sich zieht. Die „Aufdeckung“ eines Qualitätsdefizits bei Leistungserbringern mit

<sup>39</sup> Verwendet man stattdessen die Verlustfunktion aus Tabelle 4 in Abschnitt 5.2.2, welche nur die Fehlklassifikationsaufwände betrachtet und die Aufwände des Stellungnahmeverfahrens ignoriert, dann ist das optimale  $\alpha$  gleich  $fn/(fn + fp)$ .

<sup>40</sup> Siehe auch Abschnitt 5.3.1.5 zu Limitationen der entscheidungstheoretischen Verlustbetrachtung.

großer Fallzahl wäre in dieser Betrachtungsweise wichtiger als die „Aufdeckung“ eines Qualitätsdefizits bei Leistungserbringer mit kleiner Fallzahl.

Eine alternative Operationalisierung des *hinreichenden Hinweises auf ein Qualitätsdefizit* in dieser Betrachtungsweise beruht nicht allein auf der statistische Evidenz für die Nullhypothese bzgl. des Kompetenzparameters des Leistungserbringers, d. h.  $H_0: \theta \leq R$ , sondern auf der *Relevanz* dieser Abweichung für Patienten und Patientinnen. Speziell hängt dies von der Größe der Abweichung vom Referenzwert sowie der Anzahl an betroffenen Patienten und Patientinnen beim Leistungserbringer in der Zukunft ab. Selbst wenn als Referenzrahmen nur die Fallzahl im *nächsten* Bewertungszeitraum, d. h. die Fallzahl im nächsten Erfassungsjahr als Metrik genommen wird, ist diese zum Zeitpunkt der Auffälligkeitsentscheidung unbekannt. Vereinfachend wird daher im Folgenden angenommen, dass die Fallzahl  $J$  im aktuellen Beobachtungszeitraum ein guter Schätzer für die Fallzahl im nächsten Beobachtungszeitraum ist. Die mit einer Fehlklassifikation verbundenen Aufwände aus patientenzentrierter Sichtweise lassen sich dann wie in untenstehender Matrix als abgewandelte Version der Vier-Felder-Tafel in Tabelle 5 quantifizieren:

Tabelle 6: Aufwandsannahmen für die statistisch relevante Einstufungsmethode

		Wahrer Wert des LE	
		$\theta \leq R$	$\theta > R$
Ergebnis der quantitativen Auffälligkeitseinstufung	$s_{\text{stat.rel}}(o, J, R) = \text{nein}$	0	$fn \cdot (\theta - R)_+ \cdot J$
	$s_{\text{stat.rel}}(o, J, R) = \text{ja}$	$\zeta$	$\zeta$

In dieser Darstellung ist dabei  $(\theta - R)_+ \cdot J$  die über die von Referenzwert  $R$  hinaus tolerierte Anzahl an Fällen mit unerwünschtem Ereignis, mit  $(\theta - R)_+ := \max(0, \theta - R)$ . Des Weiteren ist  $fn > 0$  der mit jedem dieser Fälle assoziierte Aufwand, und  $\zeta > 0$  wiederum der Aufwand eines Stellungnahmeverfahrens. Vereinfachend kann die Verlustfunktion auch als  $fn \cdot (\theta - R)_+ \cdot J$  bei  $s_{\text{stat.rel}} = \text{nein}$  und  $\zeta$  bei  $s_{\text{stat.rel}} = \text{ja}$  dargestellt werden.

Beispiel: Hat ein Leistungserbringer in einem Indikator mit  $R = 15\%$  im betrachteten Erfassungsjahr  $J = 40$  Fälle bei einer zugrunde liegenden Rate von  $\theta = 20\%$  behandelt, d. h. der Kompetenzparameter ist 20 %, so liegt die erwartete Anzahl an Fällen mit unerwünschtem Qualitätsergebnis genau  $0,05 \cdot 40 = 2$  über der vom Referenzwert tolerierten Anzahl (in diesem Fall  $0,15 \cdot 40 = 6$ ). Die konkrete Entscheidung, ob dieser Leistungserbringer als quantitativ auffällig eingestuft werden soll, hängt dann wiederum vom Aufwandsverhältnis zwischen  $\zeta$  und  $fn$  ab.

Die optimale Entscheidungsregel lässt sich aus den beiden Verlustfunktionen

$$E[l_{\text{stat.rel}}(o, J, R, \theta, s = \text{ja})] = \zeta$$

und

$$E[l_{\text{stat.rel}}(o, J, R, \theta, s = \text{nein})] = \int fn \cdot (\theta - R)_+ \cdot J \cdot f(\theta|o, J) d\theta$$

herleiten als Entscheidungsregel

$$s_{\text{stat.rel}}(o, J, R, \theta) = I \left( \zeta < \int f_n \cdot (\theta - R)_+ \cdot J \cdot f(\theta|o, J) d\theta \right),$$

mit  $f(\theta|o, J)$  als Dichtefunktion der Wahrscheinlichkeitsverteilung von  $\theta$  (gegeben die beobachteten Daten  $o, J$ ). Das heißt, es wird der Aufwand für die Durchführung eines Stellungnahmeverfahrens  $\zeta$  dem erwarteten Aufwand für die über die vom Referenzwert hinaus tolerierte Anzahl an Patienten mit unerwünschtem Qualitätsergebnis gegenübergestellt. Im Folgenden wird dabei vereinfachend  $f_n = 1$  gesetzt, weil die optimale Auswertungsstrategie lediglich vom Aufwandsverhältnis von  $\zeta/f_n$  abhängt. Als Beispiel:  $\zeta = 1$  bedeutet, dass der Aufwand eines Stellungnahmeverfahrens gleichzusetzen sind mit dem Aufwand für die vom Referenzwert tolerierte Anzahl an Fällen mit unerwünschtem Behandlungsergebnis plus genau einem weiteren Fall. Wird die Grenze für quantitative Auffälligkeit als  $\zeta = 1$  gesetzt, werden daher alle Leistungserbringer quantitativ auffällig, bei denen mindestens ein Fall mit unerwünschtem Behandlungsergebnis über die vom Referenzwert tolerierte Anzahl hinaus erwartet wird.

Zur expliziten Berechnung der statistisch relevanten Auffälligkeitseinstufung muss die Wahrscheinlichkeitsverteilung von  $\theta$  (gegeben die beobachteten Daten  $o, J$ ) spezifiziert werden. Dies erfolgt in einem bayesianischen Rahmen und wird am Beispiel des Beta-Binomial-Modelles in Abschnitt 5.3.2 für einen Ratenindikator diskutiert.

#### 5.3.1.4 Zusammenfassung: Annahmen zu Aufwänden und Herangehensweise der drei Einstufungsverfahren

Die in Abschnitt 5.3.1.1 bis 5.3.1.3 vorgestellten Einstufungsverfahren werden tabellarisch hinsichtlich der zu Grunde liegenden Annahmen gegenübergestellt:

Tabelle 7: Übersichtstabelle über die Annahmen und Entscheidungsregeln der drei vorgestellten quantitativen Auffälligkeitseinstufungsmethoden

Auffälligkeitseinstufungsmethode	rechnerisch	statistisch signifikant	statistisch relevant
Herangehensweise	enumerativ	analytisch	analytisch
Aufwand für Stellungnahmeverfahren	$\zeta$	$\zeta$	$\zeta$
Aufwand für falschnegative Klassifikation	$f_n$	$f_n$	$f_n \cdot (\theta - R)_+ \cdot J$
Entscheidungsregel	$I \left( \frac{o}{J} > R \right)$ , (sofern $\frac{f_n}{\zeta} < 1$ )	$I \left( P(\theta \leq R) < \frac{f_n - \zeta}{f_n} \right)$	$I \left( \zeta < \int f_n \cdot (\theta - R)_+ \cdot J \cdot f(\theta o, J) d\theta \right)$

### 5.3.1.5 Limitationen der entscheidungstheoretischen Verlustfunktionenbetrachtung

Die Messung und Steigerung der Effizienz des Strukturierten Dialogs, wie im Auftragstext dieses Berichtes formuliert (G-BA 2018a), erfordert es, Aufwand und Nutzen des Strukturierten Dialoges bzw. des Stellungnahmeverfahrens in expliziter Form miteinander ins Verhältnis zu setzen. Die Vorgehensweise, Strategien zur quantitativen Auffälligkeitseinstufung aus verschiedenen expliziten Aufwandsannahmen herzuleiten, bietet eine strukturierte Möglichkeit, die Annahmen des jeweiligen Einstufungsverfahrens transparent zu machen. Dabei ist zu berücksichtigen, dass bestimmte Vereinfachungen für die Formulierung der Verlustfunktionen getroffen wurden: Beispielsweise hängt das beobachtete Indikatorergebnis eines Leistungserbringers in der Regel nicht nur von der zugrunde liegenden Kompetenz ab, sondern auch von weiteren Faktoren wie beispielsweise der Qualität der Datenerfassung in den QI-relevanten Feldern oder unbeobachteten patientenseitigen Risikofaktoren (vgl. das Entscheidungsdiagramm in Abbildung 11). Wenn davon auszugehen ist, dass diese Faktoren nicht vollständig bei der Konstruktion und Messung eines Indikators berücksichtigt werden (können), kann dies dazu führen, dass ein strengerer oder weniger strenger Schwellenwert zu einer besseren Entscheidungsregel führt als die oben hergeleitete, welche diese Störfaktoren nicht berücksichtigt. Gibt es in einem Indikator beispielsweise unbeobachtete patientenseitige Risikofaktoren, die zu einer erhöhten Wahrscheinlichkeit der im Indikator betrachteten interessierenden Ereignisse führen, und werden diese nicht über eine Risikoadjustierung oder in der Festlegung des Referenzbereichs berücksichtigt, so führt dies in den Beobachtungen der Zufallsvariable  $O$  zu einer sogenannten Überdispersion, vgl. z. B. Spiegelhalter (2005). Damit wird die Entscheidung ob  $\theta > R$  vorliegt unsicherer, d. h. beispielsweise, dass die Wahrscheinlichkeit der Nullhypothese  $P(\theta \leq R)$  ohne Berücksichtigung der Störfaktoren tendenziell überschätzt wird. Somit führt bei der statistisch signifikanten Auffälligkeitseinstufungsmethode ein kleineres  $\alpha$  als jenes aus der Aufwandsabwägung zur optimalen Entscheidungsregel. Dies gilt unter der Annahme, dass diese unbeobachteten patientenseitigen Faktoren bei der fachlichen Bewertung bewertet und zu einer Entlastung des Leistungserbringers führen. Analoge Überlegungen treffen auf die statistisch relevante Auffälligkeitseinstufung zu. Dies sollte bei der Festlegung des Signifikanzniveaus bzw. des Schwellenwertes  $\zeta$  für die statistisch relevante Einstufung immer Berücksichtigung finden.

Eine weitere Vereinfachung ist, bei der Auslösung eines Stellungnahmeverfahrens die gleichen Aufwände anzunehmen, unabhängig davon, ob sich im Stellungnahmeverfahren ein Qualitätsdefizit bestätigt oder nicht. Prinzipiell könnten auch hier die Aufwände einer falsch-positiven Klassifikation höher sein, als bei einer richtig-positiven Klassifikation, z. B.  $\zeta + fp$  vs.  $\zeta$  mit einem weiteren Aufwandparameter  $fp$ . Jedoch ist das gewählte Vorgehen  $fp = 0$  zu setzen auch dadurch zu begründen, dass das anschließende Stellungnahmeverfahren im 2. Schritt eine falsch-positive Klassifikation verhindern würde. Eine weitere zusätzliche Vereinfachung, die getroffen wurde, ist es, die Sensitivität und Spezifität des Stellungnahmeverfahrens im 2. Schritt des Entscheidungsproblems unberücksichtigt zu lassen: Auch im Stellungnahmeverfahren können Fehler bei der Einstufung passieren, die wiederum unterschiedliche Aufwände verursachen können.

Grundsätzlich fordert die Aufwand-Nutzen-Betrachtung der vorangegangenen Abschnitte eine explizite Festlegung des Aufwandsverhältnisses von Stellungsverfahren gegenüber nicht entdeckten Qualitätsdefiziten. Dabei ist zu beachten, dass verschiedene Akteure sehr unterschiedliche Aufwandsperspektiven auf den Bewertungsprozess haben. Diese Komplexität der Aufwand-Nutzen-Betrachtung ist mit den vereinfachenden Annahmen der vorangegangenen mathematischen Herleitungen nur bedingt vereinbar. Trotzdem bietet der entscheidungstheoretische Rahmen eine Offenlegung der Abwägungen und kann eine Stütze bieten, um die Tuning-Parameter<sup>41</sup> der jeweiligen Einstufungsverfahren zu interpretieren. Prinzipiell legen die Schwellenwerte fest, ab wann ein hinreichender Hinweis auf ein Qualitätsdefizit vorliegt. Eine zu klärende Frage ist, ob dieses Niveau einheitlich für alle QS-Verfahren, Qualitätsindikatoren und Verwendungszwecke festgelegt werden soll, oder situativ anzupassen ist. Beispielsweise kann es Gründe geben, dass man für manche Qualitätsindikatoren (z. B. Sterblichkeitsindikatoren) weniger statistische Evidenz verlangt, bevor man im Rahmen eines Stellungsverfahrens den Kompetenzparameter des Leistungserbringers klären möchte, als bei anderen Qualitätsindikatoren. Auch können die Konsequenzen des Verwendungszweckes (Bsp. plan. QI-RL) dazu führen, dass man ein sehr hohes Niveau ansetzen möchte, dann aber im zweiten Schritt eine sehr umfangreiche fachliche Bewertung ansetzt, die entsprechend transparent dokumentiert ist, um den gravierenden Konsequenzen der Bewertung gerecht zu werden. Alle diese Gründe lassen sich im Rahmen der Verlustfunktion begründen. Eine weitere Diskussion zu diesem Thema und konkrete Empfehlungen findet sich in Abschnitt 5.6.

### 5.3.2 Bayesianische Modellierung

Die im vorangegangenen Abschnitt vorgestellten Ansätze zur Lösung des Entscheidungsdiagramms (Abbildung 11), müssen je nach Indikatorart mit konkreten Wahrscheinlichkeitsmodellen zur Modellierung der Verteilung des Kompetenzparameters  $\theta$  und der beobachteten Behandlungsergebnisse  $O_j$  ausgeführt werden. In der Literatur vorherrschend sind dabei bayesianische Modelle (vgl. u. a. Liu et al. (2003), George et al. (2017), Ash et al. (2012), Christiansen und Morris (1996)), welche die Wahrscheinlichkeitsverteilung der beobachteten Behandlungsergebnisse  $O_j$  für den  $j$ 'ten Patient oder Patientin bei gegebener Kompetenz  $\theta$  auf oberster Ebene (Patientenebene) und die Wahrscheinlichkeitsverteilung von  $\theta$  in einer zweiten Ebene (Leistungserbringerebene) modellieren. Diese Art der hierarchischen Modellierung berücksichtigt die Cluster-Struktur der erhobenen Daten, in der mehrere Patientinnen und Patienten jeweils von einem Leistungserbringer behandelt werden. Kernidee dieser Modelle ist, dass man durch die Modellierung einer Verteilung von  $\theta$  Vorwissen modelliert, welches zusammen mit den beobachteten Daten in die Schätzung der Kompetenz des Leistungserbringers einfließt. Vorwissen kann dabei aus verschiedenen Quellen kommen:

- der Historie eines Leistungserbringers
- Expertenwissen
- Daten anderer Leistungserbringer

---

<sup>41</sup> Als Tuning-Parameter sind gemeint: das Signifikanzniveau  $\alpha$  in der statistisch signifikanten Einstufungsmethode bzw. der Schwellenwert  $\zeta$  in der statistisch relevanten Einstufungsmethode

Dieses Vorwissen wird in der bayesianischen Statistik als sog. *A-priori-Verteilung* bezeichnet. Liegt kein oder nur sehr wenig Vorwissen vor, so spricht man von einer sog. *vagen A-priori-Verteilung*. Unter Beobachtung der Daten eines Leistungserbringers wird dieses Vorwissen mit der durch die Daten gewonnen Evidenz aufdatiert, und man erhält die sog. *A-posteriori-Verteilung*. Das Verwenden von bayesianischen Methoden erlaubt es, Vorwissen und Informationen transparent und zu unterschiedlichem Grad in die Einstufung von Leistungserbringern einzubeziehen, und somit je nach Zweck zu optimieren. Bei sogenannten *Empirical-Bayes-Ansätzen* werden A-priori-Informationen aus den Ergebnissen der Leistungserbringer geschätzt, was insbesondere bei verteilungsabhängigen Einstufungen (d. h. mit Perzentilreferenzbereich) sinnvoll ist. Je nach Indikatorart können A-priori-Verteilung und Annahmen über das Zustandekommen von Behandlungsergebnissen flexibel modelliert und, beispielsweise im Falle von Patientenbefragungen, sogar über mehr als zwei zu berücksichtigende Ebenen modelliert werden. Erweiterungen von George et al. (2017) beschreiben, wie durch explizite Modellierung der Fallzahlabhängigkeit Zusammenhänge zwischen Fallzahl und Ergebnisqualität bei der Einstufung berücksichtigt werden können. Dies führt zu sog. fallzahlabhängigen *Shrinkage Targets*, die es erlauben, je nach Fallzahl eine andere A-priori-Verteilung anzunehmen.

Im Folgenden werden zwei gängige Modelle für anteilsbasierte Indikatoren und für risikoadjustierte Indikatoren ( $\frac{O}{E}$ ) vorgestellt.

### 5.3.2.1 Anteilsbasierte Indikatoren

Im Falle von anteilsbasierten Indikatoren werden patientenseitige Einflüsse  $x_j$  (beobachtete) und  $z_j$  (unbeobachtete) auf das Behandlungsergebnis  $O_j$  einer Patientin oder eines Patienten  $j = 1, \dots, J$  nicht bei der Inferenz des Kompetenzparameters  $\theta$  berücksichtigt. Die Wahrscheinlichkeit

$$\pi_j = P(O_j = 1|\theta)$$

für das Auftreten des Behandlungsergebnisses  $O_j = 1$  wird daher innerhalb eines Leistungserbringers als für alle Patientinnen und Patienten gleich angenommen und kann direkt mit dem Kompetenzparameter  $\theta$  gleichgesetzt werden:  $\pi_j = \theta$  für alle  $j = 1, \dots, J$ . Die Summe  $O = \sum_{j=1}^J O_j$  folgt somit einer Binomial-Verteilung mit Erfolgswahrscheinlichkeit  $\theta$  und  $J$  Wiederholungen.

#### Das Beta-Binomial-Modell

Die Festlegung der A-priori-Verteilung von  $\theta$  als Beta-Verteilung  $\text{Beta}(\alpha, \beta)$  mit Parametern  $\alpha, \beta > 0$  führt zum sog. Beta-Binomial-Modell (vgl. z. B. (Carlin und Louis)):

$$O \sim \text{Bin}(J, \theta),$$

und

$$\theta \sim \text{Beta}(\alpha, \beta).$$

Die Beta-Verteilung kann je nach Wahl der sog. A-priori-Parameter  $\alpha$  und  $\beta$  sehr verschiedene Formen annehmen, und bietet somit einen sehr flexiblen Rahmen, um Vorwissen wiederzugeben. Der A-priori-Erwartungswert der Beta-Verteilung lautet

$$\mathbf{E}(\theta|\alpha, \beta) = \frac{\alpha}{\alpha + \beta}.$$

Die A-priori-Varianz der Beta-Verteilung

$$\mathbf{Var}(\theta|\alpha, \beta) = \frac{\alpha \cdot \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Die Familie der Beta-Verteilungen wird dabei gewählt, weil es sich um eine sog. *konjugierte A-priori-Verteilung* zur Binomial-Verteilung handelt. Diese besitzt die besondere Eigenschaft, dass unter Beobachtung von  $J$  Behandlungsfällen mit  $O = o$  interessierenden Ereignissen, die A-posteriori-Verteilung von  $\theta$  erneut einer Beta-Verteilung folgt. Dabei lässt sich zeigen (Carlin und Louis 2008), dass

$$\theta|o, J \sim \text{Beta}(\alpha + o, \beta + J - o).$$

Der A-posteriori-Erwartungswert dieser Verteilung lautet

$$\mathbf{E}(\theta|o, J, \alpha, \beta) = \frac{\alpha + o}{\alpha + \beta + J}.$$

Diese Formulierung bietet eine Interpretation der Parameter  $\alpha$  und  $\beta$  als a priori angenommene Anzahl von Fällen mit interessierendem Ereignis ( $\alpha$ ) und Fällen ohne interessierendem Ereignis ( $\beta$ ) an.

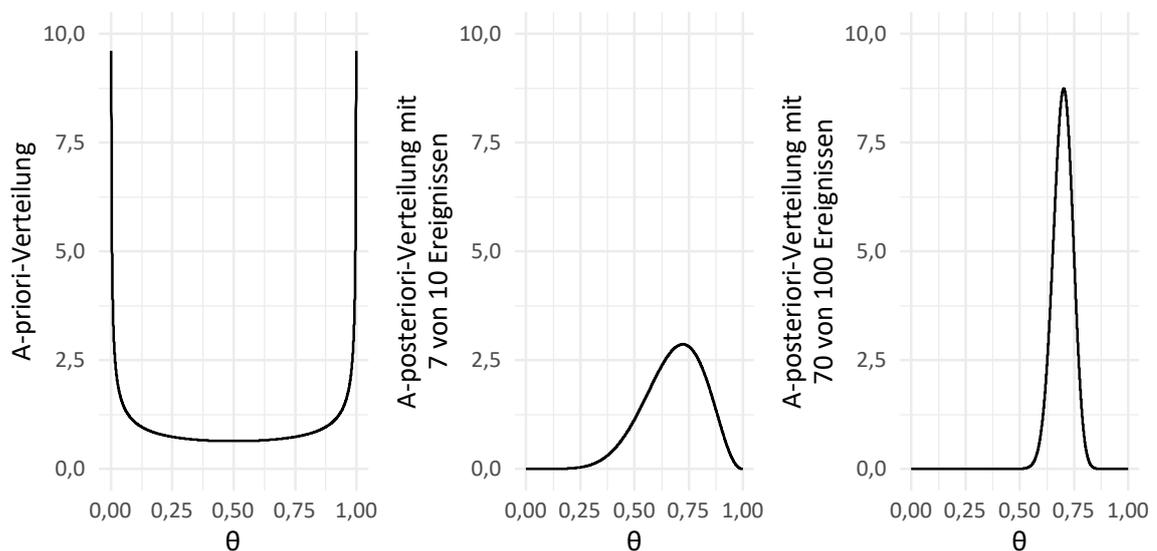


Abbildung 13: links: Beta-Verteilung mit Parametern  $\alpha = \beta = \frac{1}{2}$ ; Mitte: resultierende A-posteriori-Verteilung bei 7 von 10 Ereignissen; rechts: resultierende A-posteriori-Verteilung bei 70 von 100 Ereignissen

Abbildung 13 veranschaulicht die A-priori- und A-posteriori-Verteilung von  $\theta$ . Die linke Grafik zeigt dabei die A-priori-Verteilung  $\text{Beta}(\frac{1}{2}, \frac{1}{2})$ . Die mittlere und die rechte Grafik zeigen, wie sich

das Aufdatieren der A-priori-Verteilung mittels der Informationen aus den beobachteten Daten auswirkt. Beide Grafiken zeigen die A-posteriori-Verteilungen im Fall einer beobachteten Ereigniswahrscheinlichkeit von  $o/J = 0,7$ . Der Unterschied zwischen der mittleren und der rechten Grafik entsteht durch die Anzahl an beobachteten Fällen: In die mittlere Grafik gehen nur 10 Fälle ein, in die rechte Grafik dagegen 100. Dadurch erhalten im ersten Fall die Daten mehr und die A-priori-Informationen über den Parameter entsprechend weniger Gewicht als im zweiten Fall.

Alternativ lässt sich der A-posteriori-Erwartungswert auch als gewichtetes Mittel von A-priori-Erwartungswert und roher Rate  $\frac{o}{J}$  ausdrücken:

$$\mathbf{E}(\theta|o, J, \alpha, \beta) = \rho \cdot \frac{o}{J} + (1 - \rho) \cdot \frac{\alpha}{\alpha + \beta}.$$

Dabei ist das Gewicht  $\rho$ ,  $0 \leq \rho \leq 1$ , für die Mittelung wie folgt

$$\rho = \frac{J}{J + \alpha + \beta}.$$

Je kleiner  $J$  gegenüber  $\alpha + \beta$  ist, desto geringer wird die Evidenz der beobachteten Rate  $\frac{o}{J}$  gegenüber dem A-priori-Erwartungswert gewichtet. Dieses Phänomen wird auch als *Shrinkage* (dt. „Schrumpfung“) bezeichnet, weil die beobachtete Rate zum A-priori-Erwartungswert schrumpft.  $\rho$  wird auch als *Shrinkage-Gewicht* bezeichnet (vgl. u. a. Morris (1983)).

### Beispiel für Einstufungsverfahren mit Beta-Binomial-Annahme

Der Referenzwert und die Ausrichtung eines Indikators definieren einen fallzahlunabhängigen Referenzbereich, bei ratenbasierten Indikatoren je nach Ausrichtung entweder  $[0, R]$  oder  $[R, 1]$ , der angibt, welcher Wertebereich für den zugrunde liegenden Parameter  $\theta$  erwartbare Qualität widerspiegelt. Der Schwellenwert  $o_{\min}(J, R)$  einer vorgegebenen Entscheidungsregel  $s(o, J, R)$  ist dagegen definiert als der Wert, ab dem ein beobachtetes Leistungserbringerergebnis mit der Entscheidungsregel als quantitativ auffällig eingestuft wird. Während der Referenzbereich von der quantitativen Auswertungsmethodik unabhängig ist, hängen die Schwellenwerte für quantitative Auffälligkeit direkt von der verwendeten Klassifikationsmethodik für die Bewertung und deren Tuning-Parameter ab und sind, wie oben hergeleitet, in der Regel fallzahlabhängig. Die Fallzahlabhängigkeit der Schwellenwerte ist eine natürliche Konsequenz der Unsicherheit, die entsteht, wenn anhand der beobachteten QI-Ergebnisse Rückschlüsse auf den Kompetenzparameter gezogen werden. Aufgrund der Fallzahlabhängigkeit werden Schwellenwerte oft in sogenannten Funnelplots (IQTIG 2016, Spiegelhalter et al. 2012) dargestellt. Über diese Darstellungsweise werden im Folgenden die Schwellenwerte  $o_{\min}(J, R)$  für die drei oben hergeleiteten Klassifikationsmethoden verglichen, d. h. für die rechnerische, statistisch signifikante und statistisch relevante quantitative Auffälligkeitseinstufung. Der Vergleich wird anhand eines Qualitätsindikators mit Referenzbereich  $[0, 10\%]$  durchgeführt. Eine interaktive Version dieser Darstellung, bei der sowohl der Referenzbereich als auch die Tuning-Parameter der Klassifikationsmethoden (wie z. B.  $\alpha$  oder  $\zeta$ ) variiert werden können, ist als Shiny-App (Chang et al. 2019) unter <https://iqtig.shinyapps.io/funnelplot> zu finden.

Abbildung 14 zeigt einen Funnelplot für den oben beschriebenen Ratenindikator mit Referenzbereich  $[0, 10\%]$  und drei verschiedenen Methoden zur quantitativen Auffälligkeitseinstufung: rechnerische Auffälligkeit, statistisch signifikante Auffälligkeit (mit  $\alpha = 5\%$ ) und statistisch relevante Auffälligkeit (mit  $\zeta = 2$ ).

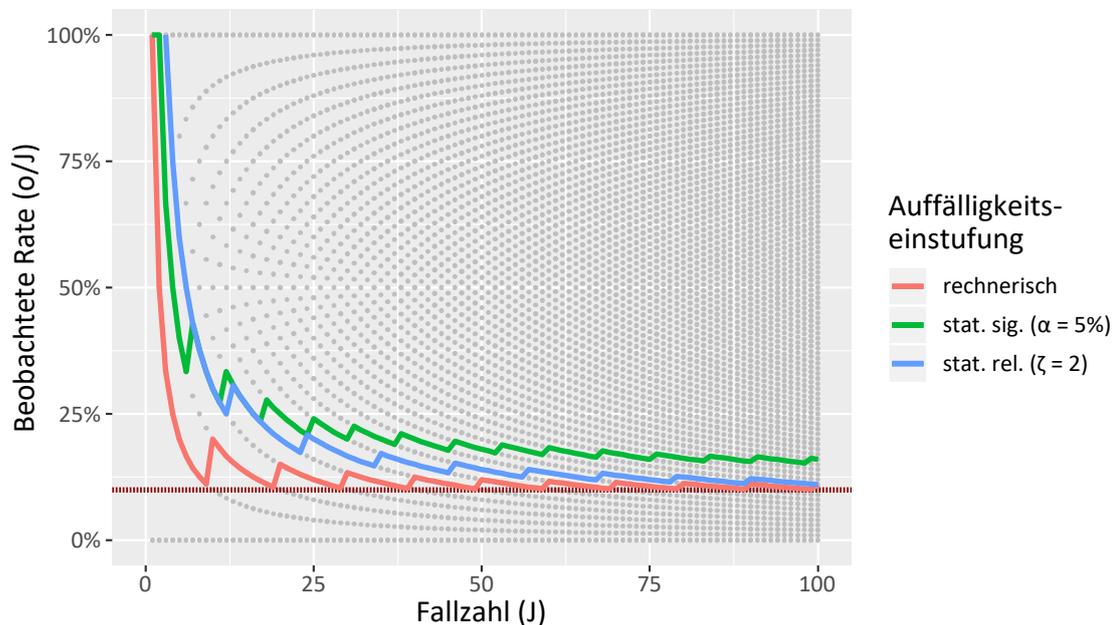


Abbildung 14: Funnelplot für die quantitative Auffälligkeitseinstufung der drei Klassifikationsmethoden für einen Indikator mit Referenzbereich  $[0-10\%]$ .

Für eine gegebene Fallzahl  $J$  sind dabei nur beobachtete Raten  $0, 1/J, 2/J, \dots, 1$  möglich, die als Punkte in der Grafik dargestellt werden. Liegt ein beobachtetes Leistungserbringerergebnis auf oder oberhalb der entsprechenden Linie für den Schwellenwert der Methode, wird dieser Leistungserbringer als quantitativ auffällig in dem entsprechenden Indikator klassifiziert. Aus der Abbildung wird deutlich, dass sowohl die rechnerische als auch die statistisch signifikante Auffälligkeit mit steigender Fallzahl gegen den Referenzwert  $R$  konvergieren. Die Schwellenwerte der rechnerischen Auffälligkeit liegen dabei durchgängig unter jenen der statistisch signifikanten Auffälligkeit; dies ist aber nicht grundsätzlich so, sondern von der Wahl von  $\alpha$  abhängig: wie oben beschrieben sind die Schwellenwerte bei einem  $\alpha$  von 0,465 nahezu identisch zu denen der rechnerischen Auffälligkeit.

Die Schwellenwerte der statistisch relevanten Auffälligkeit konvergieren scheinbar auch gegen den Referenzwert  $R$ ; dies ist aber nur scheinbar der Fall. Für Fallzahlen  $J > 240$  (in der Abbildung nicht dargestellt) liegen die Schwellenwerte dieser Einstufungsmethodik knapp unter dem Referenzwert, z. B. wird ein Ergebnis bei  $J = 241$  mit  $o = 24$  d. h. mit  $o/J = 9,9585\%$  bereits auffällig, bei  $J = 5000$  ab  $9,6\%$ . Das bedeutet, dass bei dieser Klassifikationsmethodik grundsätzlich auch Leistungserbringerergebnisse als quantitativ auffällig eingestuft werden können, die prinzipiell nicht rechnerisch auffällig sind. Dies liegt darin begründet, dass über die A-posteriori-Dichte integriert wird und daher der Fall eintreten kann, dass der Erwartungswert der Verlustfunktion für ein  $o$  größer ist als  $\zeta$ , obwohl für dieses  $o$  der Punktschätzer  $o/J$  innerhalb des

Referenzbereichs liegt. Wenn dies keine gewünschte Eigenschaft der Klassifikationsmethodik für die quantitative Auffälligkeit ist, könnte beispielsweise die rechnerische Auffälligkeit zu einer zusätzlichen Bedingung für die quantitative Auffälligkeit gemacht werden.

Die statistisch relevante Auffälligkeit ist bei dem betrachteten Indikator mit  $R = 10\%$  bis zu einer Fallzahl von  $J = 11$  gleich oder weniger streng als die statistisch signifikante Auffälligkeit. Ab dieser Fallzahl sinken die Schwellenwerte der statistisch relevanten Auffälligkeit jedoch schneller als jene der statistisch signifikanten Auffälligkeit, d. h. für diese Fallzahlen führt die statistisch relevante Auffälligkeit zu strengeren Schwellenwerten. Der Grund hierfür liegt in der Wahl der Verlustfunktion: Für die statistisch signifikante Auffälligkeitseinstufung relevant ist lediglich eine möglichst sichere Klassifikation des Kompetenzparameters des Leistungserbringers, d. h. ob  $\theta \leq R$  oder  $\theta > R$ , egal, wie groß eine potentielle Abweichung vom Referenzwert ist bzw. wie viele Patienten und Patientinnen davon betroffen sind. Die statistisch relevante Einstufung gewichtet dagegen die Relevanz der Abweichung für Patienten und Patientinnen höher. Beide Verfahren nehmen an, dass die Qualität der Leistungserbringer nicht fallzahlabhängig ist.

Tabelle 8 verdeutlicht die unterschiedlichen Schwellenwerte für ausgewählte Fallzahlen.

Tabelle 8: Kritische Zahl interessierender Ereignisse, ab der Leistungserbringer in den drei Auffälligkeitseinstufungsmethoden als quantitativ auffällig bewertet werden.

Fallzahl $J$	Schwellenwert $\sigma_{\min}(J, R)$ der		
	$S_{\text{rech}}$	$S_{\text{stat.sig.bayes}} (\alpha = 5\%)$	$S_{\text{stat.rel}} (\zeta = 2)$
1	1/1 = 100 %	1/1 = 100 %	-
2	1/2 = 50 %	2/2 = 100 %	-
3	1/3 = 33 %	2/3 = 67 %	3/3 = 100 %
5	1/5 = 20 %	2/5 = 40 %	3/5 = 60 %
10	2/10 = 20 %	3/10 = 30 %	3/10 = 30 %
20	3/20 = 15 %	5/20 = 25 %	4/20 = 20 %
50	6/50 = 12 %	9/50 = 18 %	7/50 = 50 %
100	11/100 = 11 %	16/100 = 16 %	11/100 = 11 %

Tabelle 8 verdeutlicht, dass die statistisch relevante Auffälligkeitseinstufung für Leistungserbringer mit kleinen Fallzahlen am liberalsten ist; Leistungserbringer mit weniger als 3 behandelten Fällen können unabhängig von ihrem beobachteten Ergebnis nicht quantitativ auffällig werden. Ab einer Fallzahl von  $J = 10$  ist die statistisch relevante Auffälligkeit dann gleich streng oder strenger als die statistisch signifikante Auffälligkeit. Für eine Abwägung der Vor- und Nachteile der verschiedenen Methoden siehe Abschnitt 5.6.

### 5.3.2.2 Risikoadjustierte Indikatoren

Ausgehend vom Entscheidungsdiagramm in Abbildung 11 sind in der quantitativen Auffälligkeitseinstufung risikoadjustierter Qualitätsindikatoren bei der Inferenz des Kompetenzparameters  $\theta$  auch explizit gemessene Einflüsse  $x_j$  der Patientinnen und Patienten  $j = 1, \dots, J$  auf die jeweiligen Behandlungsergebnisse  $O_j$  zu berücksichtigen. Anders als im Falle anteilsbasierter (nicht risikoadjustierter) Indikatoren muss die Wahrscheinlichkeit

$$\pi_j = P(O_j = 1 | x_j, \theta)$$

für das Eintreten des Behandlungsergebnisses  $O_j = 1$  als einerseits abhängig von der Kompetenz des Leistungserbringers und andererseits als abhängig von patientenseitigen Einflussgrößen  $x_j$  modelliert werden. Somit ist die Summe  $O = \sum_{j=1}^J O_j$  anders als im Falle anteilsbasierter Indikatoren, wegen der unterschiedlichen Wahrscheinlichkeiten, nicht mehr binomial-verteilt, sondern als generalisiert-binomial-verteilte Zufallsvariable

$$O \sim \text{GBin}(\pi_1, \dots, \pi_L) \quad (1)$$

zu betrachten. Das IQTIG verwendet zur Risikoadjustierung in der Regel das in der Epidemiologie unter der Bezeichnung „standardisierte Mortalitätsrate“ bzw. „Morbiditätsrate (SMR)“ bekannte Verhältnis von beobachteter zu erwarteter Anzahl interessierender Ereignisse ( $o/e$ ) (im Folgenden immer als *die SMR* bezeichnet).<sup>42</sup> Dies entspricht einer indirekten Standardisierung (Keiding und Clayton 2014), bei der die beobachtete Zahl an interessierenden Ereignissen  $O = o$  einer erwarteten Zahl an interessierenden Ereignissen  $e$  gegenübergestellt wird. Dabei wird die erwartete Zahl an interessierenden Ereignissen  $e$  in der Regel als Summe von Wahrscheinlichkeiten

$$e = \sum_{j=1}^J e_j$$

mit  $e_j = P(O_j = 1 | x_j)$  berechnet, welche meist mithilfe logistischer Regressionsmodelle prognostiziert werden. Dabei fließen nur patientenseitige Einflussgrößen  $x_j$ , nicht aber die Kompetenz des Leistungserbringers  $\theta$  in die Modellierung von  $e_j$  bzw.  $e$  ein. Im Allgemeinen gilt also  $e_j \neq \pi_j$ . Im Weiteren werden die Wahrscheinlichkeiten  $e_j$  vereinfachend als feste gegebene Größen betrachtet, deren Schätzunsicherheit vernachlässigt werden kann.<sup>43</sup>

<sup>42</sup> Äquivalent ist es das Verhältnis von beobachteter zu erwarteter *Rate* an interessierenden Ereignissen zu betrachten, d. h.  $\frac{o/J}{e/J}$ .

<sup>43</sup> Die Validität dieser Vereinfachung ist auf Ebene einzelner Qualitätsindikatoren zu prüfen. Insbesondere für Risikoadjustierungsmodelle mit kleiner Datenbasis kann die Schätzunsicherheit im Allgemeinen nicht vernachlässigt werden, und erfordert Erweiterungen der hier vorgestellten Methodik. Desweiteren ist davon auszugehen, dass nicht erhobene bzw. nicht in der Modellierung berücksichtigte Risikofaktoren  $z_j$  zur Unsicherheit der Schätzung von  $e_j$  beitragen.

### Das Poisson-Gamma-Modell

Das Poisson-Gamma-Modell (vgl. u. a. Liu et al. (2003), Christiansen und Morris (1996)) modelliert die Verteilung der beobachteten Anzahl an interessierenden Ereignissen  $O$  als Poisson-verteilte Zufallsvariable

$$O \sim \text{Po}(\theta \cdot e).$$

Dabei ist  $\theta$  die für einen Leistungserbringer zu schätzende SMR und  $e$  die erwartete Anzahl an interessierenden Ereignissen, basierend auf dem Risikoadjustierungsmodell für die Patientenrisiken (also ohne Berücksichtigung des Leistungserbringereinflusses). Die Annahme, dass man  $O$  als poisson-verteilt modellieren kann, und nicht generalisiert binomial-verteilt wie in Gleichung (1), wird durch die sog. Poisson-Approximation begründet: Ist das Verhältnis von interessierenden Ereignissen und Fallzahl klein, so kann die Verteilung einer generalisiert binomial-verteilten Zufallsvariable durch eine Poisson-Verteilung approximiert werden.<sup>44</sup>

Im Poisson-Gamma-Modell wird für  $\theta$  dabei eine Gamma-A-priori-Verteilung angenommen

$$\theta \sim \text{Gamma}\left(\alpha, \frac{\alpha}{\mu}\right),$$

mit  $\mu > 0, \alpha > 0$ . Der Parameter  $\mu$  ist als A-priori-Erwartungswert von  $\theta$  zu interpretieren, d. h.  $E(\theta) = \mu$ . Die A-priori-Varianz von  $\theta$  ist

$$\text{Var}(\theta) = \frac{\mu^2}{\alpha},$$

wodurch der Parameter  $\alpha$  als Präzisionsparameter interpretiert werden kann: Je kleiner  $\alpha$ , desto größer ist die A-priori-Unsicherheit in der Verteilung von  $\theta$ . Abbildung REF zeigt die Wahrscheinlichkeitsdichte einer Gamma-Verteilung für verschiedene Werte des Parameters  $\alpha$  und  $\mu = 1$ .

---

<sup>44</sup> Maßgeblich für die Validität dieser Approximation ist dabei, dass die für den Approximationsfehler geltende obere Schranke  $2 \sum_{j=1}^J \pi_j^2$  klein ist. Diese Annahme kann auch in der Praxis bei Indikatoren mit häufigen Ereignissen und kleinen Fallzahlen verletzt sein, weshalb im Einzelfall komplexere Modelle nötig werden können.

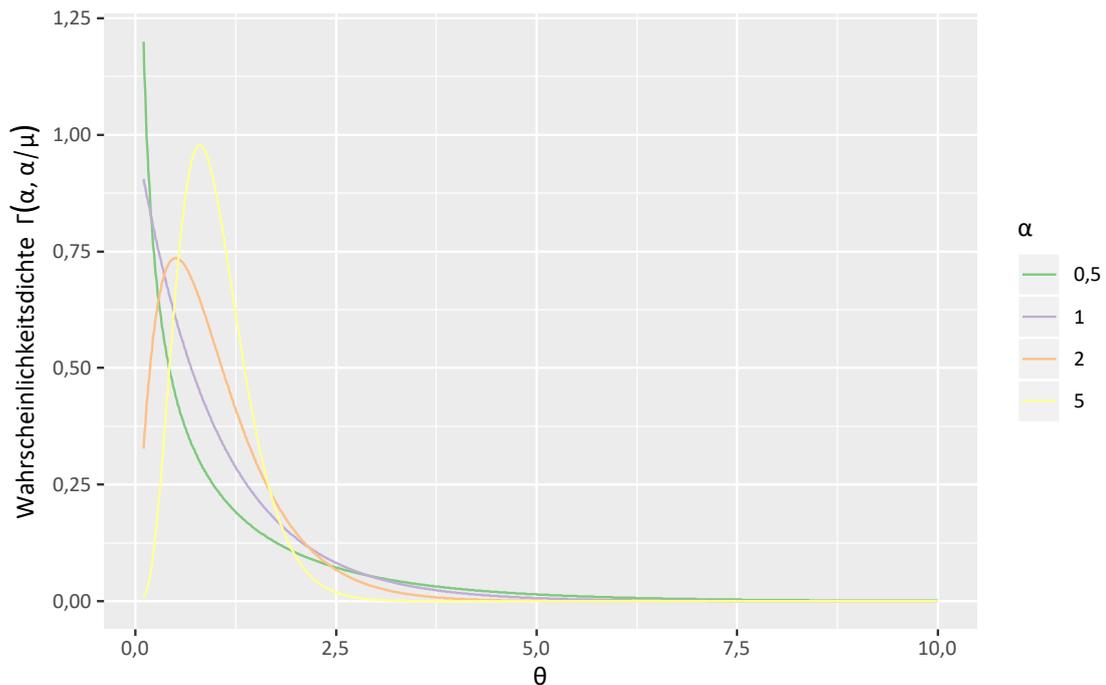


Abbildung 15: Wahrscheinlichkeitsdichte einer Gamma-Verteilung mit  $\vartheta = 1$  bei verschiedenen A-priori-Werten von  $\alpha$

Zu erkennen ist, dass sich die meiste Wahrscheinlichkeitsmasse im Bereich von ca. 0 bis 2 konzentriert. Während die Verteilung für  $\alpha = 1/2$  eine Polstelle bei 0 besitzt, wird mit  $\alpha = 1$  dem Wert  $\theta = 0$  die höchste Wahrscheinlichkeit beigemessen. Für  $\alpha = 2$  liegt der Modus der Verteilung bei  $\theta = 1/2$ . Mit wachsendem  $\alpha$  konzentriert sich die Wahrscheinlichkeitsverteilung beim Modus  $\mu \cdot \frac{\alpha-1}{\alpha}$ , der für große Werte von  $\alpha$  nahe am Erwartungswert  $\mu$  liegt.

Unter Beobachtung einer Anzahl  $O = o$  interessierender Ereignisse und erwarteter Anzahl  $e$  ist die A-posteriori-Verteilung von  $\theta$  (Christiansen und Morris 1996):

$$\theta|o, e \sim \text{Gamma}\left(\alpha + o, e + \frac{\alpha}{\mu}\right).$$

Der Erwartungswert dieser A-posteriori-Verteilung ist:

$$\mathbf{E}(\theta|o, e) = \frac{o + \alpha}{e + \alpha/\mu}.$$

Dies legt eine weitere Interpretation der Parameter  $\alpha$  und  $\mu$  nahe:  $\alpha$  entspricht einer beobachteten Anzahl von interessierenden Fällen, die man a priori annimmt und zu den tatsächlichen beobachteten interessierenden Fällen hinzuaddiert. Analog ist  $\frac{\alpha}{\mu}$  die a priori erwartete Anzahl von interessierenden Ereignissen, die man zur erwarteten Anzahl  $e$  hinzuaddiert.

Um den Einfluss der A-priori-Verteilung auf das Shrinkage-Verhalten zu verstehen, kann der A-posteriori-Erwartungswert wiederum als gewichtetes Mittel des SMR-Schätzers  $o/e$  und des A-priori-Erwartungswerts  $\mu$  ausgedrückt werden:

$$\mu' := \mathbf{E}(\theta|o, e) = \rho \cdot \frac{o}{e} + (1 - \rho) \cdot \mu,$$

mit  $\rho = \frac{e \cdot \mu}{\alpha + e \cdot \mu}$ . Je größer  $\alpha$  im Verhältnis zu  $e \cdot \mu$  gewählt wird, desto stärker ist der Einfluss der A-priori-Verteilung, d. h. der Shrinkage-Effekt wächst mit  $\alpha$ . Bei fester Wahl von  $\alpha$  und  $\mu$  steigt der Einfluss der beobachteten Daten mit steigendem  $e$  und der Shrinkage-Effekt nimmt ab.

Die A-posteriori-Varianz lautet

$$\text{Var}(\theta|o, e) = \mu' \frac{\rho}{e}$$

Im Kontext von Empirical-Bayes-Methoden, welche die A-priori-Parameter anhand empirischer Daten aller Leistungserbringer bestimmen, wird das Shrinkage-Gewicht  $\rho$  auch als Reliabilität der Messung interpretiert (Adams 2009), denn es gilt:

$$\rho = \frac{\text{Var}(\theta|\alpha, \mu)}{\text{Var}(\theta|\alpha, \mu) + \text{Var}\left(\frac{O}{e}|\alpha, \theta = \mu\right)}$$

Somit kann der Shrinkage-Faktor  $\rho$  auch als Varianzen-Verhältnis der sog. *between-provider-variance*  $\text{Var}(\theta|\alpha, \mu)$  und der Summe der Varianzen von  $\text{Var}(\theta|\alpha, \mu)$  und  $\text{Var}\left(\frac{O}{e}|\alpha, \theta = \mu\right)$  (*within-provider-variance*) interpretiert werden. Der A-posteriori-Erwartungswert  $\mu'$  wird aus diesem Grund auch als *reliability adjusted estimate* bezeichnet (vgl. u. a. Dimick et al. (2010), Krell et al. (2014b), Krell et al. (2014a)).

### Beispiel für Einstufungsverfahren mit Poisson-Gamma-Annahme

In Abbildung 16 werden verschiedene Einstufungsstrategien anhand eines Funnelplots gegenüber gestellt.

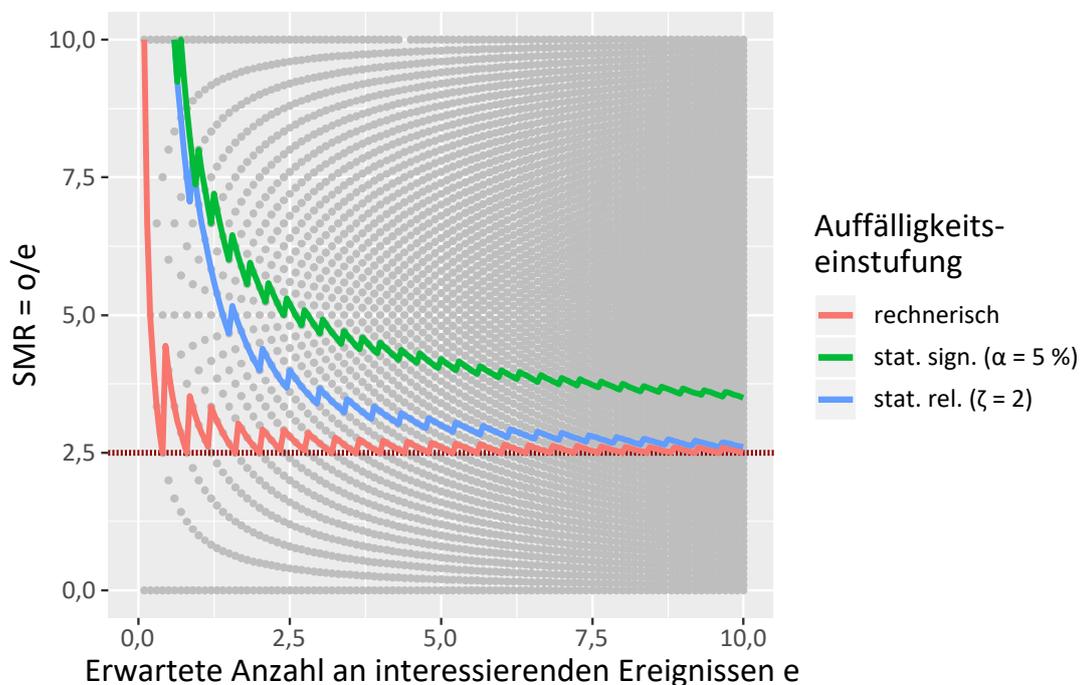


Abbildung 16: Funnelplot-Darstellung verschiedener Auffälligkeitseinstufungsmethoden für risiko-adjustierte Qualitätsindikatoren.

Auf der x-Achse wird die Anzahl interessierender Ereignisse  $e$  dargestellt. Auf der y-Achse werden SMR-Werte dargestellt. Graue Punkte in der Grafik stellen hypothetische Ergebnisse dar, die von Leistungserbringern angenommen werden können. Diese folgen je nach beobachteter Anzahl an interessierenden Ereignissen  $o$  den Hyperbel-Ästen  $\frac{1}{e}, \frac{2}{e}, \frac{3}{e}, \dots$  usw. Jede der farbigen Linien entspricht einer Einstufungsstrategie. Die dunkelrote horizontale Linie stellt einen Referenzwert von  $R = 2,5$  dar, welcher der rechnerischen Auffälligkeitseinstufung entspricht. Blau sind die Schwellenwerte eingezeichnet, ab welcher Leistungserbringer bei der Einstufung nach statistischer Relevanz für  $\zeta = 2$  quantitativ auffällig werden würden. Dabei wurde eine vage A-priori-Verteilung mit  $\alpha = 1/2$  und  $\mu = 1$  gewählt; zum Vergleich in Grün der Schwellenwert aus der statistisch signifikanten Auffälligkeitseinstufung ( $\alpha = 5\%$ ). Im Vergleich zur rechnerischen Auffälligkeit führen sowohl die Einstufung nach statistischer Relevanz als auch die Einstufung nach statistischer Signifikanz zu weniger strengen Schwellenwerten, vor allem im Bereich kleiner Fallzahlen. Je größer der Wert  $\zeta$  gewählt wird, desto großzügiger fällt die Einstufung insbesondere bei niedrigen Werten von  $e$  aus. Gegenüber der statistisch signifikanten Auffälligkeitseinstufung fällt die Einstufung für große Werte von  $e$  bei der statistisch relevanten Auffälligkeitseinstufung strenger aus, während für kleine Werte von  $e$  die statistisch signifikante Auffälligkeitseinstufung strenger ist (vgl. auch Tabelle 9).

Die folgende Tabelle zeigt die Anzahl interessierender Ereignisse  $o$ , die je nach Einstufungsstrategie und erwarteter Zahl  $e$  bei einer A-priori-Verteilung von  $\alpha = 1/2$  und  $\mu = 1$  nötig ist.

Tabelle 9: Kritische Anzahl an interessierenden Ereignissen, die zur quantitativen Auffälligkeit führt, je nach erwarteter Anzahl an interessierenden Ereignissen und Einstufungsstrategie

erwartete Anzahl an interessierenden Ereignissen $e$	kritische Anzahl beobachteter Ereignisse			
	rechnerische Auffälligkeit	statistisch relevant ( $\zeta = 1$ )	statistisch relevant ( $\zeta = 2$ )	statistisch signifikant ( $\alpha = 0,05$ )
0,1	1	7	13	4
0,5	2	4	6	6
1,0	3	5	7	8
2,0	5	7	8	11
5,0	13	13	15	21
10,0	25	24	26	35

### 5.3.2.3 Weitere Indikatorarten

In der externen Qualitätssicherung kommen über die ratenbasierten Indikatoren und die risikoadjustierten Indikatoren hinaus eine Reihe weiterer Indikatorarten bei der Berechnungsart zum Einsatz. Beispiele sind der Geburtshilfe-Index, der e/N-Indikator in HTXM-MKU, die Follow-up-

Indikatoren in den Transplantationsbereichen sowie die Follow-up-Indikatoren bei HEP, KEP und HSM (Anlage 3 QSKH-RL Follow-up-Indikatoren). Auch diese lassen sich einfach im bayesianischen Rahmen darstellen, falls nicht – wie beim e/N-Indikator – bereits bayesianisch formuliert. Auch die statistische Auswertungsmethodik für Qualitätsindikatoren bei Patientenbefragungen ist bereits mittels bayesianischen Modellen formuliert (IQTIG 2018c, IQTIG 2018d) und passt so nahtlos in das statistische Rahmenkonzept, sobald diese im Rahmen der DeQS-RL auszuwerten sind.

Die in IQTIG (2017c) und IQTIG (2018b) sowie Hengelbrock und Höhle (2019) beschriebene statistische Auswertungsmethodik für Follow-up-Indikatoren besteht für Ereigniszeit-adjustierte Ratenindikatoren bereits aus einer bayesianischen Vorgehensweise, welche Zensierung und Trunkierung bei der Datenerhebung berücksichtigt. Die Berechnung der risikoadjustierten, Ereigniszeit-adjustierten Indikatoren, welche als Kompetenzparameter das standardisierten Inzidenzratenverhältnis haben, lässt sich bayesianisch im Rahmen des oben beschriebenen Poisson-Gamma-Modells interpretieren. Insofern sollte die entwickelte Follow-up-Methodik prinzipiell auch für Sozialdaten-gestützte Follow-up-Indikatoren vorbereitet sein, wobei es sich erst in der Praxis und bei Vorliegen von Echtdateien zeigen wird, ob die längeren Zeitverzögerungen durch die Lieferung der Sozialdaten bei diesen Indikatoren weitere Modifikationen bei der statistischen Auswertungsmethodik erfordern.

#### 5.3.2.4 Wahl der A-priori-Verteilung in bayesianischen Modellen

Der bayesianische Ansatz zur Modellierung der Wahrscheinlichkeitsverteilungen für die Kompetenzparameter bietet einen flexiblen Weg, unterschiedliche Indikatorarten und Verlustannahmen in einer einheitlichen Systematik zu behandeln. Gleichzeitig erzwingt der bayesianische Ansatz mit der expliziten Modellierung der A-priori-Verteilung Transparenz über Vorwissen und Annahmen herzustellen. Die A-priori-Verteilung kann dabei wesentlichen Einfluss auf die Einstufungsmethodik haben. Im Beta-Binomial-Modell ist die Summe  $\alpha + \beta$  (d. h. die Anzahl virtuell angenommener A-priori-Beobachtungen) entscheidend dafür, wie stark die A-priori-Verteilung gegenüber den beobachteten Daten ist. Im Poisson-Gamma-Modell ist es der Parameter  $\alpha$ , welcher die Stärke des A-priori-Einflusses misst.

Wenn Shrinkage-Effekte unerwünscht sind, so müssen vage A-priori-Verteilungen mit niedrigen Werten von  $\alpha + \beta$  (Beta-Binomial-Fall) bzw.  $\alpha$  (Poisson-Gamma-Fall) gewählt werden. Ein Spezialfall der vagen A-priori-Verteilung ist mit  $\alpha = \beta = 1/2$  (Beta-Binomialfall) der sog. Jeffreys-Prior. Dieser spiegelt statistisch gesehen maximale Unwissenheit wieder. Wegen  $\alpha + \beta = 1$  ist gewährleistet, dass das Gewicht des Vorwissens nie stärkeres Gewicht als das der beobachteten Daten besitzt. Man kann zeigen, dass die statistisch signifikante Einstufung mit Jeffreys-Prior nahezu äquivalent zur statistisch signifikanten Einstufung im Verfahren *Planungsrelevante Qualitätsindikatoren* mit mid-P-Werten ist (Brown et al. 2001)

Neben der Stärke der A-priori-Verteilung muss auch das Shrinkage-Target, d. h. der A-priori-Erwartungswert so modelliert werden, dass er den erwarteten Anforderungen entspricht. Beispielsweise könnte für Indikatoren mit festem Referenzbereich der Referenzwert R das Shrinkage-Target sein. Geht man davon aus, dass es einen starken Zusammenhang zwischen

Fallzahl und Ergebnisqualität gibt, so können Shrinkage-Targets auch fallzahlabhängig gewählt werden (Ash et al. 2012, George et al. 2017).

Im Abschnitt 5.3.2 wurde bereits der Empirical-Bayes-Ansatz vorgestellt, bei dem die A-priori-Verteilung aus Daten mehrerer Leistungserbringer geschätzt wird. Da bei vielen QIs der Großteil der Leistungserbringerergebnisse im Referenzbereich liegt, führt dieser Ansatz tendenziell zu einem sehr wohlwollenden Shrinkage-Target. Da bei den meisten Leistungserbringern die Behandlungsqualität in Ordnung scheint, findet hier also die Qualitätssicherung unter der Grundannahme statt, dass auch beim nächsten betrachteten Leistungserbringer die Qualität vermutlich in Ordnung sein. Eine solche Annahme ist jedoch (anders als beim Public Reporting) in der Qualitätssicherung nicht unbedingt gerechtfertigt, wenn man sensitiv auf mögliche Qualitätsmängel reagieren will. Eine A-priori-Verteilung, bei der das Shrinkage-Target außerhalb des Referenzbereichs liegt, wie beispielsweise der Jeffreys-Prior, drückt dahingegen eine skeptische Grundhaltung aus. So eine A-priori-Verteilung basiert nicht auf empirischem Vorwissen, sondern auf Annahmen, mit denen man die quantitative Analyse der QS-Daten beginnen möchte. Daher lässt sich aus der Verwendung von Jeffreys-Prior, welche einen A-priori-Erwartungswert für den Kompetenzparameter von  $\frac{1}{2}$  annimmt und der Nullhypothese  $H_0: \theta \leq R$  eine A-priori-Wahrscheinlichkeit von  $F_{\text{Beta}}(R|\frac{1}{2}, \frac{1}{2})$  zuordnet<sup>45</sup> nicht die Behauptung ableiten, dass bei der Mehrheit der Leistungserbringer ein Qualitätsmangel vorliegt.

### 5.3.3 Sensitivität und Spezifität der Einstufungsmethoden

Die Entscheidung, ob ein Leistungserbringerergebnis als quantitativ auffällig oder unauffällig eingestuft werden soll, ist immer eine Abwägung von Sensitivität und Spezifität bei der Klassifikation von  $\theta$ : je strenger die Schwellenwerte einer Einstufungsmethode sind, desto höher ist die Sensitivität und desto niedriger die Spezifität, und vice versa. Für einen detaillierteren modellbasierten Vergleich von Sensitivität und Spezifität der Methoden werden im Folgenden vier Szenarien angenommen, bei denen ein Leistungserbringer jeweils die zugrunde liegende Rate von  $\theta = 5\%$ ,  $\theta = 10\%$ ,  $\theta = 15\%$ , und  $\theta = 20\%$  hat, wiederum für einen Qualitätsindikator mit Referenzbereich  $[0, 10\%]$ . Für diese vier Szenarien wird die Wahrscheinlichkeit einer quantitativ auffälligen Einstufung für die oben diskutierten Einstufungsmethoden berechnet, jeweils für Fallzahlen zwischen  $J = 1, \dots, 1000$ . Die oberen beiden Darstellungen in Abbildung 17 zeigen dabei die Wahrscheinlichkeit falsch-positiver Einstufungen ( $1 - \text{Spezifität}$ ) der Methoden, da die wahren zugrunde liegenden Raten innerhalb des Referenzbereichs liegen. Die unteren beiden Abbildungen zeigen die Wahrscheinlichkeit richtig-positiver Einstufungen (Sensitivität). Verglichen werden rechnerische Auffälligkeit, statistisch signifikante Auffälligkeit mit  $\alpha = 5\%$  sowie die statistisch relevante Auffälligkeit mit  $\zeta = 2$ . Eine interaktive Version der Abbildung, bei der sowohl  $\alpha$  als auch  $\zeta$  variiert werden können, ist als Shiny-App unter [https://iqtig.shinyapps.io/sensitivitaet\\_spezifitaet/](https://iqtig.shinyapps.io/sensitivitaet_spezifitaet/) zu finden.

<sup>45</sup> Ist der Referenzwert z. B.  $R = 10\%$ , dann ist die A-priori-Wahrscheinlichkeit für die Nullhypothese z. B.  $20,5\%$ .

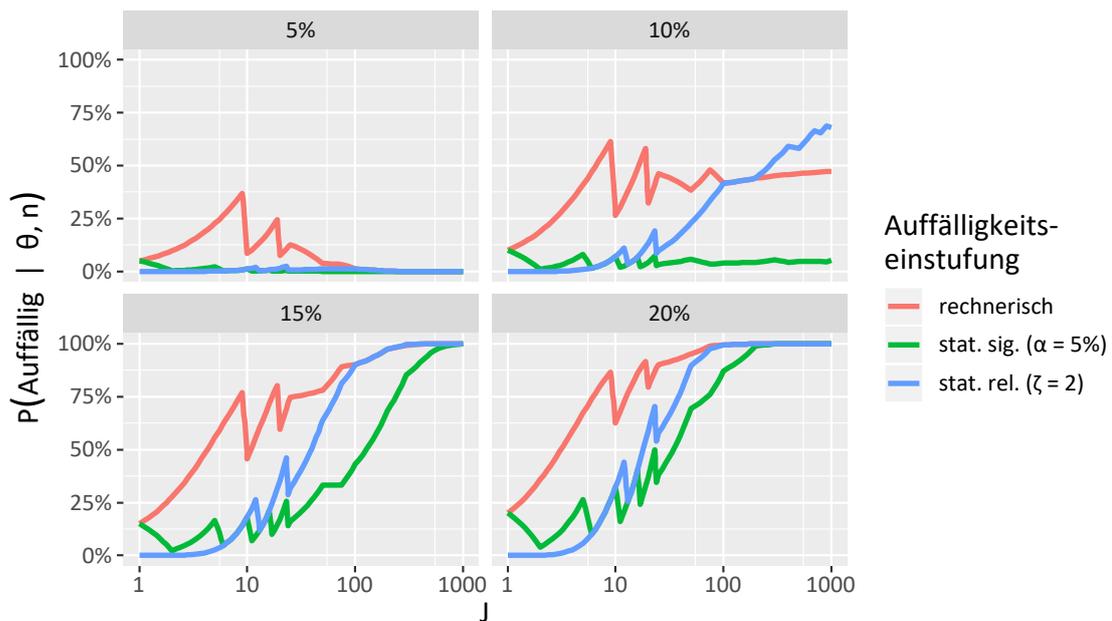


Abbildung 17: Vergleich von Sensitivität und 1-Spezifität für die drei Methoden der quantitativen Auffälligkeitseinstufung. Die Fallzahl auf der x-Achse ist dabei auf einer logarithmischen Skala dargestellt.

Der Fall, in dem die wahre zugrunde liegende Rate genau auf dem Referenzwert liegt, d. h.  $\theta = 10\%$ , ist insofern besonders, als die Methodik der statistisch signifikanten Auffälligkeit so ausgelegt ist, dass in diesem Fall (im Mittel über alle Fallzahlen  $J$ ) die Wahrscheinlichkeit einer auffälligen Einstufung genau  $\alpha$  entspricht, in diesem Fall  $5\%$ . Die Spezifität ist dabei wie auch im Fall  $\theta = 5\%$  im Mittel über alle Fallzahlen bei der statistisch signifikanten Auffälligkeit sehr groß. Bei der statistisch relevanten Auffälligkeit dagegen ist die Spezifität bei  $\theta = 5\%$  ebenfalls groß, bei  $\theta = 10\%$  jedoch besonders bei großen Fallzahlen sehr klein und konvergiert mit steigender Fallzahl gegen 0. Umgekehrt hat die statistisch signifikante Auffälligkeit bei großen Fallzahlen die geringste Sensitivität der hier verglichenen Methoden. Der Grund für die schlechte Spezifität der statistischen Relevanz bei  $\theta = 10\%$  für große Leistungserbringer ist, dass große Leistungserbringer in diesem Grenzfall bei der statistischen relevanten Auffälligkeit schnell eine relevante Anzahl an Patienten und Patientinnen haben können, die über den Referenzwert hinaus ein unerwünschtes Qualitätsergebnis hatten. Hat ein großer Leistungserbringer jedoch einen Kompetenzparameter, der nur leicht unter  $R$  liegt, dann ist auch hier die Spezifität gut. Mit anderen Worten: Es spielt bei der statistischen Relevanz weniger eine Rolle, ob nun genau  $\theta \leq R$  oder nicht, sondern der mit der Fallzahl gewichtete Abstand zum Referenzwert ist von Interesse.

Grundsätzlich bieten die statistisch signifikante und statistisch relevante Auffälligkeitseinstufung gegenüber der rechnerischen Auffälligkeit den Vorteil, dass der Trade-Off zwischen Sensitivität und Spezifität über einen Parameter ( $\alpha$ , bzw.  $\zeta$ ) gesteuert und damit beispielsweise an QI-spezifische Besonderheiten angepasst werden kann, ohne den Referenzbereich verändern zu müssen. Diese Flexibilität erscheint vor dem Hintergrund der unterschiedlichen QS-Verfahren mit ihren unterschiedlichen Fallzahlen wünschenswert.

Die statistisch relevante Auffälligkeitseinstufung basierend auf  $\zeta$  sowie die statistisch signifikante Auffälligkeit haben in den analysierten Szenarien beide die wünschenswerte Eigenschaft, dass ihre Spezifität bei wahrer zugrunde liegender Rate innerhalb des Referenzbereichs (aber nicht auf dem Referenzwert) mit steigender Fallzahl gegen 100 % und ihre Sensitivität bei zugrunde liegender Rate außerhalb des Referenzwerts gegen 100 % konvergieren. Wie bei der Betrachtung der Schwellenwerte im Funnelplot schon deutlich wurde, ist die statistisch signifikante Auffälligkeit mit  $\alpha = 5\%$  dabei bei kleinen Fallzahlen strenger als die statistisch relevante mit  $\zeta = 2$ , bei großen Fallzahlen jedoch liberaler. Diese Abwägung zwischen strengeren Schwellenwerten für Leistungserbringer mit kleiner Fallzahl (statistisch signifikante Auffälligkeit) gegenüber strengeren Schwellenwerten für solche mit großer Fallzahl (statistisch relevante Auffälligkeit) leitet sich direkt aus den zugrunde liegenden Verlustfunktionen der beiden Methoden ab und ist zentral bei der Gegenüberstellung dieser beiden Klassifikationsmethoden.

### 5.3.4 Umgang mit verteilungsabhängigen Referenzwerten

Im Gegensatz zu fixen Referenzwerten definieren perzentilbasierte Referenzwerte nicht erwartbare Qualität (IQTIG 2019a). Sie erlauben lediglich einen Vergleich der Leistungserbringer anhand der beobachteten Leistungserbringerergebnisse. Für einen festen Referenzbereich  $R$  ist die interessierende Wahrscheinlichkeit  $P(\theta_i > R | o_i, J_i, \dots)$  für den Leistungserbringer  $i$ . Wird anstelle eines fixen Referenzwertes lediglich ein Perzentil  $\xi$  vorgegeben, so interessiert die Wahrscheinlichkeit, dass der Kompetenzparameter des Leistungserbringers  $i$  zu den schlechtesten  $\xi \times 100\%$  der Kompetenzparameter gehört, d. h.

$$P(\text{Perc}(\theta_i) > \xi),$$

wobei

$$\text{Perc}(\theta_i) = \frac{1}{I} \sum_{l=1}^I I(\theta_i \geq \theta_l)$$

der Perzentilrang<sup>46</sup> des Leistungserbringers  $i$  bzgl. seines Kompetenzparameters in der Population aller  $I$  Leistungserbringer ist. Dies kann auch als empirischen Quantil von  $\theta_i$  in der Population der Werte  $\theta_1, \dots, \theta_I$  aufgefasst werden. Wichtig ist auch die Erkenntnis, dass bei der analytischen Herangehensweise der Kompetenzparameter nicht direkt beobachtet wird, sondern nur anhand der beobachteten Daten im inferenz-statistischen Sinne geschätzt werden kann.

Während die Wahrscheinlichkeit  $P(\theta_i > R | o_i, J_i, \dots)$  nur vom Leistungserbringer  $i$  und dem Referenzbereich  $R$  abhängt, fließen in die Wahrscheinlichkeit  $P(\text{Perc}(\theta_i) > \xi | o_1, J_1, \dots, o_I, J_I, \dots)$  auch die Ergebnisse anderer Leistungserbringer ein. Liegt beispielsweise in einem Indikator eine geringe Streuung der Ergebnisse vor, so ist die Unsicherheit bei der Bestimmung des Perzentil-Ranges größer als für Qualitätsindikatoren, die eine große Streuung der Indikatorergebnisse aufweisen. In der Praxis vieler gegenwärtiger Qualitätsindikatoren nach QSKH-RL wird häufig das 95. Perzentil als Referenzbereich gewählt. Unter Berücksichtigung von statistischer Unsicherheit

<sup>46</sup> Der Perzentilrang des Wertes 9 in der Menge  $\{1, 2, \dots, 9, 10\}$  ist z. B.  $2/10 = 20\%$ .

ist eine treffsichere Einstufung, ob der Leistungserbringer unter die schlechtesten 5 % der Leistungserbringer gehört, sehr schwer und setzt voraus, dass sich diese 5 % der Leistungserbringer stark von den restlichen 95 % absetzen, um diese sicher als „schlechteste 5 %“ zu klassifizieren. In den meisten Anwendungsfällen lässt sich aufgrund der beobachteten Daten nur in sehr wenigen Fällen sagen, dass ein Leistungserbringer mit hoher Sicherheit zu den schlechtesten 5 % gehört. Die Berücksichtigung von statistischer Unsicherheit mit festem Konfidenzniveau  $\alpha$  bei gleichzeitiger Festlegung des 95. Perzentils als Referenzwert würde somit zu einem doppelten Toleranzmechanismus führen, der nur extrem ausreißende Leistungserbringerergebnisse als quantitativ auffällig klassifiziert.

Daher wird im Folgenden ein Vorgehen für die Bestimmung eines perzentilbasierten Referenzwerts vorgeschlagen, bei welchem ein vorher definierter Anteil an Leistungserbringerergebnissen quantitativ auffällig wird. Dies entspricht in etwa dem aktuellen Vorgehen bei der Bestimmung von verteilungsabhängigen Referenzwerten in der QSKH-RL, da auch dort prinzipiell auch ein bestimmter Anteil an Leistungserbringern quantitativ auffällig werden soll. Sei dafür  $\{1, \dots, I\}$  die Menge der Leistungserbringer mit zugrunde liegendem Kompetenzparameter  $\{\theta_1, \dots, \theta_I\}$  und  $q$  der Anteil an Leistungserbringerergebnissen, welche in einem Indikator quantitativ auffällig werden sollen. Da es sich immer um eine diskrete Menge an Leistungserbringern handelt, kann es vorkommen, dass der Anteil an quantitativen Auffälligkeiten nicht genau  $q$  sein kann, weshalb im Folgenden davon ausgegangen wird, dass *mindestens*  $q \cdot I$  Leistungserbringerergebnisse quantitativ auffällig werden sollen. Weiterhin wird davon ausgegangen, dass es sich um einen Qualitätsindikator handelt, bei dem ein niedriges  $\theta$  gute zugrunde liegende Qualität bedeutet.

Für die rechnerische Auffälligkeit wurde im Rahmen von (analytischen) Auswertungen<sup>47</sup> der Referenzwert nach der RAW20-Methode (Paddock 2014) bestimmt, indem das  $q \cdot 100$ . Perzentil der beobachteten Ergebnisse unter jenen Leistungserbringern ermittelt wird, die eine Grundgesamtheit von mindestens 20 Fällen haben:

$$R = Q\left(q \cdot 100, \left\{\frac{o_i}{J_i}\right\}_{i \in H}\right),$$

mit  $Q$  als empirischer Quantilsfunktion und  $H$  als Menge aller Leistungserbringer mit  $J_i \geq 20$ . Die Grenze von 20 Fällen wird verwendet, damit die verwendeten beobachteten Leistungserbringerergebnisse halbwegs präzise Schätzungen der zugrunde liegenden Kompetenzparameter der jeweiligen Leistungserbringer sind. Diese Vorgehensweise kann jedoch, wie in Abschnitt 2.6 gezeigt, zu größeren Abweichungen zwischen der nominell festgelegten Anzahl an Auffälligkeiten  $\lceil q \cdot I \rceil$  und der tatsächlichen Anzahl an Auffälligkeiten führen. Für enumerative Auswertungen ist die 20'er Fallzahl-Grenze nicht notwendig.

Für die statistisch signifikante Auffälligkeit im Rahmen von analytischen Auswertungen schlägt das IQTIG vor, dass der Referenzwert so gewählt wird, dass die A-posteriori-Wahrscheinlichkeit für die Nullhypothese für mindestens  $q \cdot I$  Leistungserbringer kleiner gleich dem gewählten Schwellenwert  $\alpha$  ist:

<sup>47</sup> Zum Beispiel Auswertungen nach QSKH-RL.

$$\min \left\{ R : \sum_{i=1}^I [P(\theta_i \leq R | o_i, J_i) \leq \alpha] \geq q \cdot I \right\}.$$

Analog ist dazu für die statistische relevante Auffälligkeitseinstufung im Rahmen von Ratenindikatoren:

$$\min \left\{ R : \sum_{i=1}^I \left[ \int_R^1 f n \cdot (\theta_i - R) \cdot J_i \cdot f_{\text{Beta}}(\theta | \dots) d\theta \geq \zeta \right] \geq q \cdot I \right\}.$$

In beiden Fällen kann das Minimum über eine numerische Lösung des Optimierungsproblems bestimmt werden. Im Fall der statistisch signifikanten Auffälligkeit ist dies analog zur Bestimmung des  $q \cdot 100$ . Perzentils der unteren Grenze der zweiseitigen  $(1-2\alpha) \times 100$  % Unsicherheitsintervalle aller  $I$  Leistungserbringer, d. h. zum Beispiel im Rahmen von Ratenindikatoren

$$R = Q(q \cdot 100, F_{\text{Beta}}(\alpha; \dots)),$$

mit  $F_{\text{Beta}}$  als Verteilungsfunktion der A-posteriori-Wahrscheinlichkeit.

Gerade bei Qualitätsindikatoren, bei denen die beobachteten Ergebnisse der Leistungserbringer eine geringe Streuung aufweisen, kann es dabei bei der Bestimmung dieser Referenzwerte vorkommen, dass der verteilungsabhängige Referenzwert nah am oder sogar unterhalb des Bundesdurchschnitts liegt. Dies ist bei der statistisch signifikanten und statistischen relevanten Auffälligkeit eher der Fall als bei der rechnerischen Auffälligkeit, da bei letzterer keine Stochastizität der beobachteten Daten (zugunsten der Leistungserbringer) berücksichtigt wird.

### 5.3.5 Vor- und Nachteile der Klassifikationsmethoden

Die rechnerische Auffälligkeit entspricht in enumerativen Studiendesigns ohne Unsicherheit bei der Bestimmung von  $\theta$  der optimalen Entscheidungsstrategie. Die Annahme, dass der zugrunde liegende Parameter eines Leistungserbringers ohne Unsicherheit über die beobachteten Daten bestimmt werden kann, trifft jedoch im Regelfall nicht zu. Da die rechnerische Auffälligkeit diese Unsicherheit nicht berücksichtigt, führt sie – im Vergleich zur statistisch signifikanten bzw. statistisch relevanten Einstufung – zu verhältnismäßig vielen falsch-positiven Klassifikationen (unter der Annahme, dass  $\alpha$  bzw.  $z$  in der Regel so gewählt werden, dass die Methoden der statistischen Auffälligkeit zu höheren Schwellenwerten führt, als die rechnerische Auffälligkeit). Dies wird auch durch die empirischen Ergebnisse der Hintergrundanalyse in Abschnitt 2.6 untermauert. Dieser niedrigen Spezifität der rechnerischen Auffälligkeit bzgl. der Klassifikation des Kompetenzparameters  $\theta$  steht eine hohe Sensitivität gegenüber.

Die statistisch signifikante Auffälligkeit ist wiederum in analytischen Studiendesigns, bei denen nur Aufwände für falsch-positive und falsch-negative Klassifikationen angenommen werden, die optimale Entscheidungsstrategie. Sie ist außerdem ein in der QS etabliertes Klassifikationsverfahren (Spiegelhalter et al. 2012) und die frequentistische Version findet bereits in den Verfahren nach plan. QI-RL Anwendung. Darüber hinaus lassen sich Sensitivität und Spezifität über den Parameter  $\alpha$  steuern und ggf. pro Qualitätsindikator oder QS-Verfahren individuell festsetzen. Die Wahl dieses Parameters stellt jedoch gleichzeitig eine Herausforderung dar, da hier anders

als bei der rechnerischen Auffälligkeit – implizit oder explizit – die Aufwände für falsch-positive gegenüber falsch-negativen Klassifikationen abgewogen werden müssen.

Dies trifft auch auf die statistisch relevante Klassifikationseinstufung zu, allerdings wird dort der Aufwand eines Stellungnahmeverfahrens abgewogen mit dem Aufwand eines Falls mit unerwünschtem Ereignis (über die vom Referenzbereich hinaus tolerierte Anzahl an Fällen). Diese Aufwandsabwägung berücksichtigt, dass die Aufwände einer falsch-negativen Klassifikation unter Umständen von der Fallzahl des zu klassifizierenden Leistungserbringers abhängen; sofern dies der Fall ist, ist die statistisch relevante Auffälligkeitsklassifikation die optimale Entscheidungsstrategie, da sie die erwarteten Aufwände unter diesen Annahmen minimiert. Ihre Anwendung führt jedoch auch dazu, dass im Vergleich zur rechnerischen oder statistisch signifikanten Auffälligkeitseinstufung die Sensitivität bzgl. der Klassifikation des Kompetenzparameters bei kleinen Leistungserbringern vergleichsweise geringer ist, da diese aufgrund ihrer Fallzahl von geringer Relevanz sind. Umgekehrt ist die Spezifität bei Leistungserbringern mit großer Fallzahl gering, da deren Relevanz stark ins Gewicht fällt. Demgegenüber stehen eine hohe Spezifität bei kleinen Fallzahlen und eine hohe Sensitivität bei großen Fallzahlen, was sich aus der zugrunde liegenden Verlustfunktion ableitet.

Ein kritischer Punkt bei sowohl der statistisch signifikanten als auch der statistisch relevanten Auffälligkeitseinstufung ist, dass es, je nach Wahl der A-priori-Verteilung und der Tuning-Parameter der Methoden, Fallzahlen geben kann, bei denen Leistungserbringer unabhängig vom Ergebnis nicht auffällig werden können. Als Beispiel: bei der statistisch relevanten Auffälligkeitseinstufung mit  $\zeta = 2$  kann ein Leistungserbringer erst bei einer Fallzahl von 3 überhaupt auffällig werden (3/3). Die Frage ist, wie man mit den Leistungserbringern unter diese Fallzahl umgehen möchte: akzeptiert man, dass hier im Rahmen eines Erfassungsjahres nicht genügend Hinweise für ein Qualitätsdefizit gesammelt werden können? Muss man stattdessen die Datenbasis erweitern? Oder versucht man für diese Leistungserbringer über eine andere Vorgehensweise, z. B. anhand einer qualitativen Prüfung, ggf. nur für eine Stichprobe der Leistungserbringer, zu einer Bewertung zu kommen? Eine weitere Alternative ist, anstelle die Qualität bei den Leistungserbringern im Rahmen eines Zweifachstichproben-basierten Audit-Verfahren zu untersuchen (Tenenbein 1970, Raats und Moors 2003). Die genaue Wahl der Verfahrensweise hängt hier sehr von den Zielen der Qualitätsbewertung ab und wird weiter in Abschnitt 5.6 diskutiert.

#### **5.4 Berücksichtigung der Daten mehrerer Erfassungsjahre**

Die Güte der im Abschnitt 5.3 dargestellten Methoden zur quantitativen Einstufung von Leistungserbringern lässt sich verbessern, indem zusätzlich Informationen über die vergangenen Erfassungsjahre mit betrachtet werden. Speziell bei Leistungserbringern mit kleinen Fallzahlen ist die Hoffnung, dass eine solche Vergrößerung der Datengrundlage zu einer sichereren quantitativen Bewertung führt. Betrachtet wird als Beispiel ein Leistungserbringer mit QI-Ergebnissen für einen Ratenindikator mit Referenzbereich  $[0 - 10 \text{ \%}]$  wie in Abbildung 18.

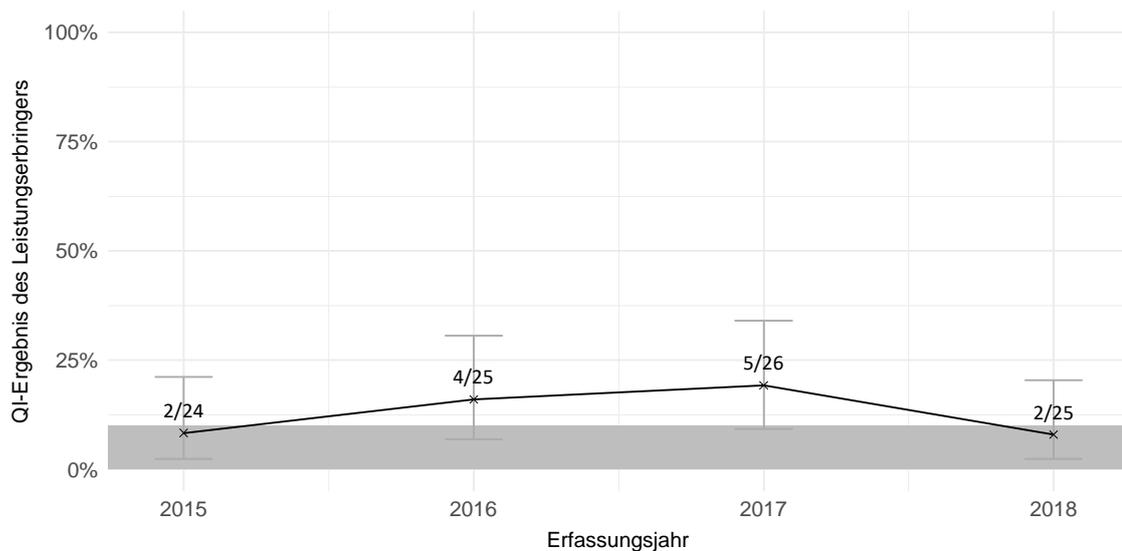


Abbildung 18: Beispielhafter Verlauf des QI-Ergebnis pro Erfassungsjahr eines hypothetischen Leistungserbringers für einen Ratenindikator. Der Referenzbereich  $\leq 10\%$  ist grau schattiert. Um die Ergebnisse sind bayesianische 90 %-Unsicherheitsintervalle eingezeichnet.

Im Erfassungsjahr 2016 ist das QI-Ergebnis  $4/25 = 16\%$  rechnerisch auffällig, jedoch ist das Ergebnis bei einem Schwellenwert von  $\alpha = 0,05$  nicht statistisch signifikant auffällig (Wahrscheinlichkeit für die Nullhypothese nach der in Abschnitt 5.3.1.2 vorgestellten bayesianischen Methodik 0,1561). Im Erfassungsjahr 2017 ist das QI-Ergebnis  $5/26 = 19\%$  erneut rechnerisch auffällig, jedoch ebenfalls nicht statistisch signifikant auffällig (Wahrscheinlichkeit für die Nullhypothese 0,0683). Im Erfassungsjahr 2017 ist der Leistungserbringer jedoch schon das zweite Mal in Folge rechnerisch auffällig. Außerdem ist das Gesamtergebnis  $9/51 = 18\%$  für den 2-Jahres-Zeitraum 2016 bis 2017 statistisch auffällig (Wahrscheinlichkeit für die Nullhypothese 0,0429). Die Daten der 1-Jahres-Zeiträume liefern also jeweils wenig statistische Evidenz auf einen möglichen Qualitätsmangel, doch in den vereinigten Daten des 2-Jahres-Zeitraums ist die statistische Evidenz dafür größer, d. h. hier besteht ein Hinweis auf ein Qualitätsdefizit.

Bei der Bewertung ist zu berücksichtigen, dass die Daten aus größeren Erfassungszeiträumen nicht automatisch dazu führen, dass Qualitätsdefizite schneller entdeckt werden. Wird z. B. ein gleitendes Fenster von mehreren Erfassungsjahren für die Auffälligkeitseinstufung verwendet, würde eine große Verschlechterung im aktuellen Erfassungsjahr möglicherweise von guten Ergebnissen aus den früheren Jahren ausgeglichen. Gleichmaßen können alte, bereits behobene Qualitätsprobleme dazu führen, dass aktuell gute Ergebnisse in den schlechten Ergebnissen der Vergangenheit untergehen. Eine Erweiterung der Datenbasis für die quantitative Auffälligkeitseinstufung um mehrere Erfassungsjahre bedingt daher auch eine gemeinsame Betrachtung der jährlichen Entscheidungen zur quantitativen Auffälligkeit und deren Auswirkungen.

#### 5.4.1 Zusammenhang mit Strukturbruchproblemen, statistischer Prozesskontrolle und sequentiellen Entscheidungsproblemen

Es wird die folgende Notation verwendet: Die Zählvariable  $t$  läuft über die Erfassungsjahre. Die Anzahl an interessierenden Ereignissen im Erfassungsjahr  $t$  wird mit  $o_t$  bezeichnet, die Kardinalität der Grundgesamtheit mit  $J_t$ . Wie bereits im Abschnitt 5.2 wird die Modellannahme verwendet, dass  $o_t$  die Realisierung einer binomial  $\text{Bin}(J_t, \theta_t)$ -verteilten Zufallsvariablen ist, wobei der Kompetenzparameter  $\theta_t$  des Leistungserbringers ebenfalls vom Erfassungsjahr abhängen kann.

Das Entscheidungsproblem lässt sich formal als Strukturbruchproblem (*change point problem*) betrachten (Montgomery 2013, Tartakovsky et al. 2015). Verlässt  $\theta_t$  den Referenzbereich, so wird dies als Strukturbruch interpretiert.<sup>48</sup> Der Strukturbruch soll möglichst zeitnah detektiert werden, weswegen sequentielle Methoden zur Anwendung kommen müssen (im Gegensatz zu retrospektiven Methoden). Anders als bei klassischen Strukturbruchproblemen sind nicht alle Parameteränderungen von Interesse, sondern vorwiegend solche, an denen sich  $\theta_t$  von einem Wert im Referenzbereich zu einem Wert außerhalb des Referenzbereichs ändert.

Die Detektion solcher Strukturbrüche ist auch das Ziel der statistischen Prozesskontrolle (*statistical process control, SPC*) (Montgomery 2013). Bei gutartigen Parameterwerten (d. h.  $\theta_t$  liegt im Referenzbereich) sagt man in der statistischen Prozesskontrolle auch, der Prozess sei „unter Kontrolle“, während bei ungünstigen Parameterwerten (d. h.  $\theta_t$  liegt außerhalb des Referenzbereichs) der Prozess „außer Kontrolle“ ist.

Das Einflussdiagramm bietet wieder einen formalen Rahmen, um das sequentielle Entscheidungsproblem zu visualisieren und optimale Strategien zu identifizieren (Jensen und Nielsen 2007). Abbildung 19 zeigt die Erweiterung des in Abbildung 11 dargestellten Einflussdiagramms, in dem die jährliche quantitative Auffälligkeitseinstufung für einen analytisch auszuwertenden Ratenindikator als sequentielles Entscheidungsproblem dargestellt wird.

---

<sup>48</sup> Es kann natürlich sein, dass der Kompetenzparameter bereits beim ersten Zeitpunkt, an dem ein Leistungserbringer Daten für einen QI liefert, außerhalb des Referenzbereichs liegt. In diesem Fall wird formal dieser Startzeitpunkt des Verfahrens als Strukturbruch angesehen.

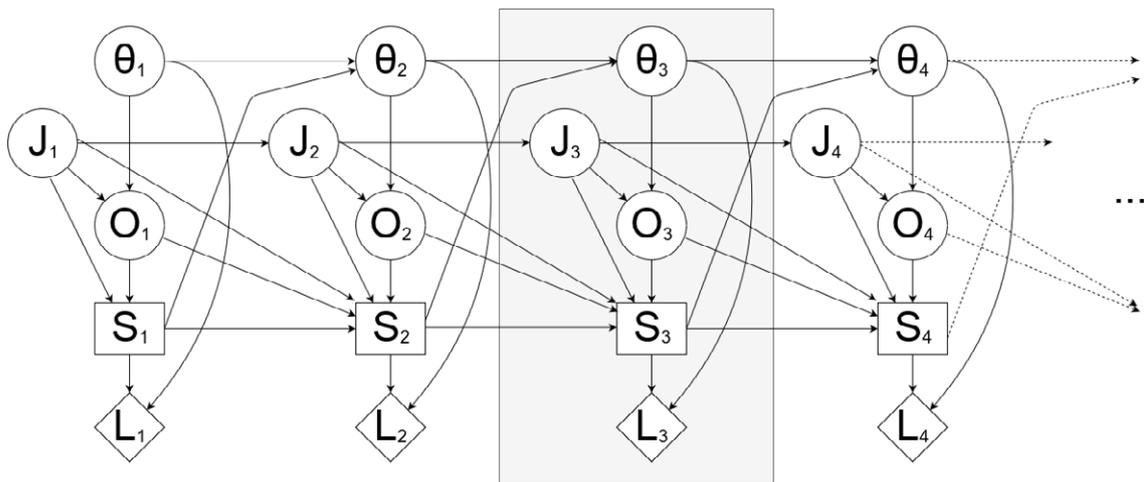


Abbildung 19: Entscheidungsdiagramm für das sequentielle Entscheidungsproblem der quantitativen Auffälligkeitseinstufung. Zeitpunkt 1 stellt das erste Erfassungsjahr mit Datenerfassung und Bewertung da. Die schattierte Fläche stellt die Knoten für ein ausgewähltes Erfassungsjahr ( $t = 3$ ) dar.

In Abbildung 19 ist Zeitpunkt  $t = 1$  das erste Jahr mit Datenerfassung und Bewertung für den Qualitätsindikator. Die quantitative Auffälligkeitseinstufung für dieses Jahr,  $S_1$ , wird anhand der beobachteten Anzahl an Fällen mit unerwünschtem Qualitätsergebnis,  $O_1$  und der Gesamtanzahl an Fällen,  $J_1$ , getroffen. Die vorhandene beobachtete Evidenz ( $O_1, J_1$ ) erlaubt somit Rückschlüsse auf  $\theta_1$ , den Kompetenzparameter des Leistungserbringers in diesem Jahr. Ähnlich der Vorgehensweise in Abschnitt 5.2 wird basierend auf einer Abwägung verschiedener Aufwände eine quantitative Bewertung des Leistungserbringers vorgenommen, indem der Leistungserbringer für das Jahr als quantitativ auffällig ( $S_1 = \text{ja}$ ) oder unauffällig ( $S_1 = \text{nein}$ ) klassifiziert wird. Die unmittelbaren Aufwände dieser Klassifikation für das Erfassungsjahr sind in der Verlustfunktion  $L_1$  festgehalten. Neu ist jedoch, dass die optimale Entscheidung nicht alleine auf der Abwägung der unmittelbaren Aufwände  $L_1$  geschieht, sondern eine Betrachtung der Auswirkungen der Entscheidung und deren zukünftigen Aufwände notwendig sind. Dafür ist zum einen die zeitliche Dynamik des Kompetenzparameters  $\theta_t$  des Leistungserbringers von Interesse, d. h. der Zeitreihe  $\theta_1, \theta_2, \theta_3, \dots$ , und welche Entscheidungen in der Zukunft getroffen werden, d. h. nach welcher Strategie die Entscheidungen  $S_2, S_3, \dots$  getroffen werden. Um entsprechende optimale Entscheidungen treffen zu können ist es relevant zu wissen, wie stark sich der Kompetenzparameter von Jahr zu Jahr ändert; z. B. wie wahrscheinlich ist es, dass ein Leistungserbringer mit zureichender Qualität im nächsten Jahr sich verschlechtert bzw. ob es Leistungserbringer aus Eigeninitiative schaffen, Qualitätsdefizite zu beheben.

Wie aus dem Diagramm hervorgeht kann  $J_t$  prinzipiell über die Erfassungsjahre fluktuieren. Zur Vereinfachung wird in den folgenden Betrachtungen jedoch  $J_t$  als fest und gegeben angenommen. Eine andere wichtige Komponente sind die Auswirkungen der quantitativen Auffälligkeitseinstufung (und der anschließenden fachlichen Bewertung) auf den Kompetenzparameter. Wie wahrscheinlich sind beispielsweise Verbesserungen nach einer quantitativen Auffälligkeit, wenn diese berechtigt ist? Alle diese Abhängigkeiten sind konzeptionell durch Pfeile im Diagramm illustriert. Eine konkrete Operationalisierung des Entscheidungsproblems durch das Dia-

gramm bedarf jedoch zusätzlich einer konkreten Festlegung der betrachteten Zustände der Variablen im Diagramm, den konkreten Übergangswahrscheinlichkeiten sowie den einzelnen Verlustfunktionen. Für ein Beispiel einer solchen Operationalisierung vgl. Abschnitt 5.4.4.

Prinzipiell liegen für jede Entscheidung im Diagramm immer alle Informationen aus der Vergangenheit vor, d. h. die Informationsgrundlage für  $S_1$  besteht aus  $(O_1, J_1)$ , für  $S_2$  ist sie  $(O_1, J_1, S_1, O_2, J_2)$ , für  $S_3$  ist die Grundlage  $(O_1, J_1, S_1, O_2, J_2, S_2, O_3, J_3)$  etc. Daraus wird klar, dass die Strategien mit der Zeit immer komplexer werden können, weil die Strategie von immer mehr Variablen abhängt.

Zur Vereinfachung bei der Handhabung und somit gleichzeitig zur besseren Kommunizierbarkeit der Strategien wird im folgenden Text vereinfachend eine obere Grenze für den Einbezug der vergangenen Information verwendet. In der Abbildung 19 geht z. B. für die Entscheidung immer nur die Information aus dem aktuellen Jahr sowie die beobachtete Information aus dem Vorjahr ein, d. h. die Informationsgrundlage für  $S_3$  ist nur  $(O_2, J_2, S_2, O_3, J_3)$  – die frühere Vergangenheit wird vergessen. Dieses Vorgehen entspricht dem *Limited Memory Influence Diagram* (Lauritzen und Nilsson 2001) bzw. der Vorgehensweise bei *Partially Observed Markov Decision Processes* (Kaelbling et al. 1998, Åström 1965). Die Einschränkung auf zwei aufeinanderfolgender Erfassungsjahre soll nicht bedeuten, dass zu einem späteren Zeitpunkt (oder in der fachlichen Bewertung) nicht noch weitere Erfassungsjahre betrachtet werden könnten. Je größer jedoch die Anzahl an betrachteten Jahren ist, desto größer ist die Rolle, die Trends in der medizinischen Versorgung, Veränderungen beim Leistungserbringer bzw. Veränderungen wegen früheren Qualitätsinterventionen spielen, sodass die Daten nur bedingt aussagekräftig für die aktuelle Kompetenz eines Leistungserbringers sind.

Für die Beschreibung von Entscheidungsstrategien, die auf den Daten zweier Erfassungsjahre beruhen, wird im Folgenden das zum Zeitpunkt der Entscheidung aktuelle Erfassungsjahr mit  $t = 0$  bezeichnet, das Vorjahr entsprechend mit  $t = -1$ .

Eine grundlegende Idee für die Formulierung solcher 2-Jahres-Strategien ist nun die Kombination aus zwei Sachverhalten: Zum einen möchte man möglichst frühzeitig große Veränderungen im Kompetenzparameter detektieren können. Andererseits kann es kleinere Änderungen geben, die nur schwierig anhand der Daten eines Erfassungsjahres detektiert werden können. In dieser Situation ist es notwendig, die Daten  $o_0$  und  $o_{-1}$  bzw.  $J_0$  und  $J_{-1}$  gemeinsam zu betrachten. Die gepoolte Betrachtungsweise birgt jedoch die Gefahr, dass gute Ergebnisse im ersten Jahr schlechte Ergebnisse im zweiten Jahr ausgleichen können. Damit die gepoolte Betrachtungsweise der Daten tatsächlich zusätzliche Informationen liefert, muss allgemein davon ausgegangen werden, dass  $\theta_{-1} \approx \theta_0$ . Daher lohnt es sich, die Daten der beiden Erfassungsjahre sowohl separat als auch gemeinsam zu betrachten.<sup>49</sup> Vor allem aufgrund der zeitlichen Verzögerung, mit der die Ergebnisse der quantitativen Einstufung an die Leistungserbringer übermittelt und

---

<sup>49</sup> Eine ähnliche Situation tritt bei der Strukturbruchdetektion auf, wo große Parameteränderungen ebenfalls schneller zu detektieren sind. In der Strukturbruchdetektion werden ebenfalls gleitende Zeitfenster verwendet. Der Vorschlag, gleitende Zeitfenster unterschiedlicher Größe (1- und 2-Jahres-Zeiträume) parallel zu betrachten, ergibt sich daraus, dass man zumindest bei der Betrachtung des 1-Jahres-Zeitraums ohne die Hilfsannahme  $\theta_{-1} \approx \theta_0$  auskommt und diese Hilfsannahme daher erst so spät wie möglich gemacht wird.

Stellungnahmeverfahren geführt werden (im Sommer des Folgejahres), ist zu vermuten, dass eine quantitative Auffälligkeit nur wenig Einfluss auf den zugrunde liegenden Kompetenzparameter im direkt anschließenden Erfassungsjahr hat. Daher wird davon abgesehen, das Ergebnis  $s_{-1}$  in der Strategie für  $s_0$  zu verwenden.

Die Strategien werden nach folgendem Prinzip entworfen: Im ersten Schritt werden die QS-Daten des aktuellen Erfassungsjahres für den Leistungserbringer betrachtet. Liefern die Daten einen hinreichenden Hinweis auf ein mögliches Qualitätsdefizit des Leistungserbringers im aktuellen Erfassungsjahr, so wird der Leistungserbringer im ersten Schritt als „quantitativ auffällig“ eingestuft. Liefern die Daten keinen hinreichenden Hinweis auf ein mögliches Qualitätsdefizit, so wird der Leistungserbringer als „quantitativ unauffällig“ eingestuft. Legen die Daten hingegen den Verdacht nahe, dass ein Qualitätsdefizit bestehen könnte, ohne dass der Hinweis als ausreichend erachtet wird, um im Sinne der Effizienz des Verfahrens ein Stellungnahmeverfahren auszulösen (d. h. die Abweichung vom Referenzwert ist vermutlich gering), so werden die QS-Daten des Vorjahres hinzugezogen. Nun werden im zweiten Schritt für den Leistungserbringer die QS-Daten des Vorjahres und die QS-Daten des aktuellen Jahres gemeinsam betrachtet. Liefern die QS-Daten beider Jahre zusammen genug statistische Evidenz für ein mögliches Qualitätsdefizit, wird ein Stellungnahmeverfahren ausgelöst. Lassen die QS-Daten beider Jahre keinen Verdacht auf ein mögliches Qualitätsdefizit aufkommen, so wird der Leistungserbringer in Schritt 1 als „quantitativ unauffällig“ eingestuft, d. h. es besteht kein Hinweis auf ein Qualitätsdefizit. Bei dieser Vorgehensweise spielen die QS-Daten des aktuellen Erfassungsjahres daher auch stets bei der Bewertung die wichtigste Rolle.

Insgesamt ist das Vorgehen motiviert durch Ideen aus der statistischen Prozesskontrolle sowie durch sequentielle Hypothesentests. Das zweistufige Vorgehen ähnelt einem zweifachen Prüfplan (vgl. Abschnitt 15.3.1 in Montgomery (2013)), wobei jedoch im Unterschied zu diesem keine Stichproben gezogen, sondern komplette Erfassungsjahre betrachtet werden.

Um das Verfahren zu konkretisieren, muss man festlegen, wie in jedem Schritt die betrachteten QS-Daten bewertet werden. Dazu sind zwei Festlegungen nötig:

- Unter welchen Bedingungen wird der Hinweis auf ein mögliches Qualitätsdefizit als ausreichend gesehen, um ein Stellungnahmeverfahren zu starten?
- Unter welchen Bedingungen liegt kein Verdacht auf ein Qualitätsdefizit vor?

Beispielsweise könnte man für den ersten Schritt des soeben beschriebenen Verfahrens, in dem lediglich die aktuellen QS-Daten betrachtet werden, folgende Festlegung treffen: Ist das QI-Ergebnis des aktuellen Erfassungsjahres statistisch signifikant auffällig beim vorgegebenen Signifikanzniveau, so gibt es statistische Evidenz für ein mögliches Qualitätsdefizit im aktuellen Erfassungsjahr. Ist das QI-Ergebnis dagegen rechnerisch unauffällig, so kann man definieren, dass wenig statistische Evidenz für ein mögliches Qualitätsdefizit im aktuellen Erfassungsjahr vorliegt.<sup>50</sup> Unabhängig von der Festlegung wird jedes Verfahren hier wegen stochastischer Einflüsse auch Fehler begehen: Leistungserbringer mit Qualitätsdefiziten könnten per Zufall trotzdem ein

---

<sup>50</sup> Der fehlende hinreichende Hinweis auf ein Qualitätsdefizit bedeutet natürlich keineswegs im Umkehrschluss, dass ein Hinweis für zureichende Qualität vorliegt.

gutes Ergebnis vorweisen und umgekehrt, vgl. Abschnitt 5.3.3. Die Frage ist, wie es zu einer guten Abwägung der verschiedenen Perspektiven kommen kann.

In den späteren Schritten muss man festlegen, wie man die QS-Daten aus mehreren Erfassungsjahren gemeinsam betrachtet. Dazu gibt es zwei prinzipielle Möglichkeiten:

- Man kann Bewertungen der QS-Daten der einzelnen Erfassungsjahre kombinieren. Beispielsweise könnte man im zweiten Schritt festlegen, dass ein hinreichender Hinweis auf ein Qualitätsdefizit vorliegt, falls das QI-Ergebnis des Leistungserbringers sowohl im aktuellen Erfassungsjahr als auch im Vorjahr rechnerisch auffällig ist.<sup>51</sup>
- Man kann die QS-Daten vereinigen und gemeinsam bewerten. Beispielsweise könnte man im zweiten Schritt festlegen, dass ein hinreichender Hinweis auf ein Qualitätsdefizit gegeben ist, wenn das über den gesamten 2-Jahres-Zeitraum berechnete QI-Ergebnis statistisch signifikant auffällig ist.

In beiden Fällen kann dabei auf verschiedene bekannte Bewertungsmethoden für QS-Daten zurückgegriffen werden, wie rechnerische Auffälligkeit, statistische signifikante Auffälligkeit oder die im Abschnitt 5.3.1.3 eingeführte Einstufung nach der statistischen Relevanz. Bei der statistisch signifikanten Einstufung ist in diesem Abschnitt stets die in Abschnitt 5.3.1.2 vorgestellte bayesianische Methodik gemeint. Somit können die Mehrjahreerweiterungen in diesem Abschnitt als Erweiterungen nach dem Baukastenprinzip gesehen werden, bei denen die Verfahren aus Abschnitt 5.3 kombiniert werden.

#### 5.4.2 Untersuchte Methoden

In diesem Abschnitt werden einige konkrete Verfahren beschrieben. Ein Vergleich der Methoden findet in den Abschnitten 5.4.3 bis 5.4.4 statt.

##### 5.4.2.1 Prüfung, ob kein Verdacht auf ein mögliches Qualitätsdefizit vorliegt

In allen Methoden wird die Einschätzung, ob anhand der vorliegenden Daten ein Verdacht auf ein Qualitätsdefizit im aktuellen Erfassungsjahr besteht, mithilfe der rechnerischen Auffälligkeitseinstufung getroffen. Alternativ könnte man auch hier stochastische Einflüsse berücksichtigen und beispielsweise analog zur Methodik der statistisch signifikanten Auffälligkeit überprüfen, ob es statistische Evidenz (zu einem festzulegenden Schwellenwert  $\alpha$ ) dafür gibt, dass es kein Qualitätsdefizit gibt, d. h. im frequentistischen Rahmen ein Test der Art  $\theta \geq R$  vs.  $\theta < R$  bzw. im bayesianischen Rahmen eine Entscheidung für die Nullhypothese  $\theta \leq R$ . Ähnliche Verfahren werden beim sequentiellen Testen betrachtet. Die Verwendung der rechnerischen Auffälligkeit entspricht in gewisser Weise einem sehr schwachen Signifikanzniveau bzw. einem großen Schwellenwert für die A-posteriori-Wahrscheinlichkeit für die Nullhypothese.

<sup>51</sup> Hier ist zu beachten, dass die rechnerische Auffälligkeit an sich ein recht schwaches Kriterium ist, welches für sich alleine genommen zu zu vielen Auffälligkeiten führt. Dadurch, dass man rechnerische Auffälligkeit in mehreren aufeinanderfolgenden Jahren verlangt, verschärft man das Kriterium; vgl. die Diskussion in den Abschnitten 5.4.2.2 und 5.4.3.

Üblicherweise nimmt man jedoch beim sequentiellen Testen an, dass der betrachtete Prozess stationär ist. Bei den betrachteten Leistungserbringern ist dies, wie bereits erwähnt, nicht unbedingt der Fall. Ein Leistungserbringer, der im aktuellen Erfassungsjahr nicht rechnerisch auffällig ist, bei dem dennoch ein Stellungnahmeverfahren gestartet wird, könnte stets als Erklärung vorbringen, dass es eine Qualitätsverbesserung gegeben habe und dass es keinen aktuellen Qualitätsmangel gebe. Außerdem sollte ein konsistentes Verfahren, welches im aktuellen Jahr trotz rechnerischer Unauffälligkeit ein Stellungnahmeverfahren auslöst, mit hoher Wahrscheinlichkeit bereits in einem der Vorjahre ein Stellungnahmeverfahren ausgelöst haben, sodass die interessierenden Fälle der Vergangenheit bereits in einem Stellungnahmeverfahren besprochen wurden (siehe auch Abschnitt 5.4.5.1).

Soll die Prüfung, ob kein Verdacht auf ein Qualitätsdefizit vorliegt, strenger gestaltet werden, so ist zu beachten, dass dadurch eine größere Datengrundlage benötigt wird, um zum Schluss zu kommen, dass kein Qualitätsdefizit vorliegt. Dies liegt daran, dass die Referenzbereiche in der Regel klein sind (d. h. der Referenzwert ist bei Ratenindikatoren nahe bei 0 bzw. 1). Dadurch ist es schwieriger, Evidenz dafür zu sammeln, dass  $\theta$  im Referenzbereich liegt, als Evidenz dafür zu sammeln, dass  $\theta$  außerhalb des Referenzbereichs liegt. Als Beispiel wird ein Referenzbereich von  $[0,10\%]$  betrachtet. Bereits bei einer Grundgesamtheit von lediglich  $J = 1$  Fall, welcher interessierend ist, liegt die Wahrscheinlichkeit für die  $H_0: \theta \leq 10\%$  bei 0,01. Hingegen liegt die Wahrscheinlichkeit für die umgedrehte Nullhypothese  $H_0: \theta \geq 10\%$  immer über 0,05, solange die Grundgesamtheit kleiner ist als 18. Während also ein einzelner interessierender Fall bereits einen Hinweis auf ein mögliches Qualitätsdefizit liefern kann, benötigt der Leistungserbringer mindestens 18 Fälle ohne interessierendes Ereignis, um zum gleichen Schwellenwert statistische Evidenz für das Einhalten des Referenzbereichs zu liefern.

#### 5.4.2.2 Laufregel

Betrachte die folgende Entscheidungsregel, die die QS-Daten eines Leistungserbringers aus zwei Erfassungsjahren verwendet:

- Ist das QI-Ergebnis des Leistungserbringers im aktuellen Erfassungsjahr statistisch auffällig, so wird der Leistungserbringer als „quantitativ auffällig“ eingestuft.
- Sind die QI-Ergebnisse des Leistungserbringers sowohl im aktuellen Erfassungsjahr als auch im Vorjahr rechnerisch auffällig, so wird der Leistungserbringer als „quantitativ auffällig“ eingestuft.
- Andernfalls wird der Leistungserbringer als „quantitativ unauffällig“ eingestuft.

Diese Entscheidungsregel ist von der Form, wie sie in der Einleitung vorgestellt wurde. Sie wird im Folgenden *Laufregel* genannt, da ähnliche Regeln unter diesem Namen z. B. in der statistischen Prozesskontrolle Anwendung finden, vgl. Abschnitt 5.3.5 in Montgomery (2013).

Formal lässt sich die Laufregel darstellen als

$$s_{\text{lauf2}}(o_0, o_{-1}, J_0, J_{-1}, R, \alpha) = s_{\text{stat.sig.Bayes}}(o_0, J_0, R, \alpha) \vee (s_{\text{rech}}(o_0, J_0, R) \wedge s_{\text{rech}}(o_{-1}, J_{-1}, R)).$$

Im Beispiel aus Abbildung 18 wäre der Leistungserbringer also im Erfassungsjahr 2017 quantitativ auffällig, da im zweiten Jahr in Folge eine rechnerische Auffälligkeit besteht.

Die Abbildung 20 illustriert, welche Ergebnisse zur Auslösung führen. Dabei wird ein Leistungserbringer mit  $J = 25$  Fällen im Indikator und ein Referenzbereich  $\leq 10\%$  angenommen. Die Punkte in den Abbildungen entsprechen möglichen Ergebniskombinationen von Vorjahr und aktuellem Jahr. Sie sind nach der quantitativen Einstufung nach der Laufregel eingefärbt. Die durchgezogene Linie ist die Grenze zur rechnerischen Auffälligkeit, die gestrichelte Linie die Grenze zur statistischen Auffälligkeit bei dem derzeit verwendeten Signifikanzniveau von  $\alpha = 5\%$ . Da die Laufregel die QS-Daten jedes Jahres isoliert betrachtet, setzt sich die Entscheidungsgrenze zwischen „quantitativ unauffällig“ und „quantitativ auffällig“ aus stückweise horizontalen und vertikalen Linien zusammen.

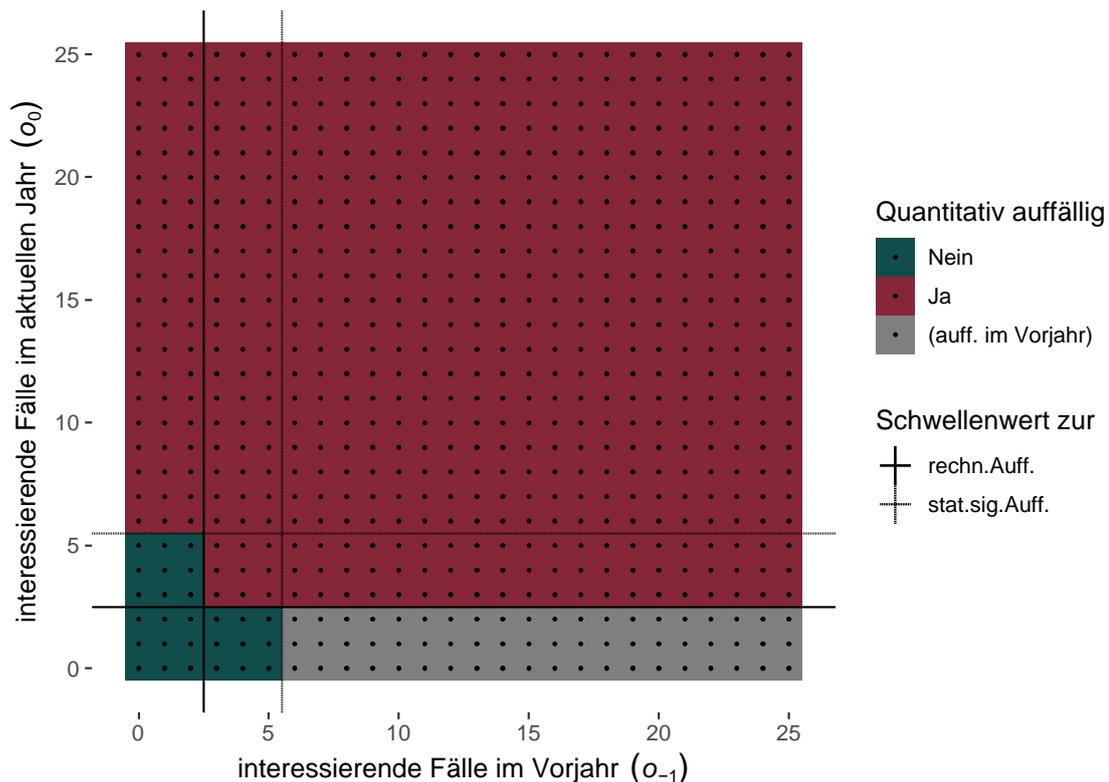


Abbildung 20: Verhalten der Laufregel bei verschiedenen Ergebnissen aus zwei Jahren für einen Leistungserbringer mit 25 Fällen in jedem Jahr

Beispielsweise wird ein Leistungserbringer, der im aktuellen Erfassungsjahr und im Vorjahr jeweils  $J = 25$  Fälle in der Grundgesamtheit des QI hatte, als „quantitativ unauffällig“ klassifiziert, wenn es im Vorjahr zwei interessierende Ereignisse gab und im aktuellen Erfassungsjahr fünf (also insgesamt sieben interessierende Ereignisse in beiden Jahren). Gab es hingegen in beiden Jahren jeweils mindestens drei interessierende Ereignisse (d. h. insgesamt sechs), so wird der Leistungserbringer wegen der wiederholten rechnerischen Auffälligkeit als „quantitativ auffällig“ klassifiziert.

Bei den klassischen Laufregeln in der statistischen Prozesskontrolle (z. B. die Western Electric-Regeln, vgl. Abschnitt 5.3.5 in Montgomery (2013)) werden in der Regel Häufungen von Ereig-

nissen gezählt, die wesentlich seltener sind als rechnerische Auffälligkeiten bei Leistungserbringern mit Kompetenz im Referenzbereich. Wie später im Abschnitt 5.4.3 gezeigt, führt die Laufregel zu einer vergleichsweise hohen Auslösewahrscheinlichkeit für ein Stellungnahmeverfahren. Eine einfache Art, die Auslösewahrscheinlichkeit zu verringern, erhält man, wenn man die Bedingung, dass die QI-Ergebnisse des Leistungserbringers in den zwei aufeinanderfolgenden Jahren größer als der Referenzwert  $R$  sein sollen, durch die Bedingung, dass die QI-Ergebnisse des Leistungserbringers in den zwei aufeinanderfolgenden Jahren größer als ein verschärfter Referenzwert  $R'$  (mit  $R' > R$ ) sein sollen, ersetzt. Diese Methode soll im Folgenden nicht weiter verfolgt werden, da die Laufregel nicht nur den Nachteil einer großen Auslösewahrscheinlichkeit besitzt, sondern daneben auch eine starke Fallzahlabhängigkeit aufweist, wie im Abschnitt 5.4.3 ausgeführt wird.

#### 5.4.2.3 Statistische Auffälligkeit auf mehreren Jahren

Eine weitere einfache Regel zur quantitativen Auffälligkeit verwendet zur Beurteilung der Evidenzlage bzgl. eines möglichen Qualitätsdefizits beim Leistungserbringer das Konzept der statistisch signifikanten Auffälligkeit (dabei ist hier wie auch im Rest dieses Abschnittes stets die Abschnitt 5.3.1.2 vorgestellte bayesianische Methodik gemeint) in den QS-Daten mehrerer Erfassungsjahre. Zur Beurteilung der statistischen Evidenz für einen möglichen Qualitätsmangel wird also die A-posteriori-Wahrscheinlichkeit dafür, dass der Kompetenzparameter im Referenzbereich liegt, verwendet. Für die Erweiterung auf Daten mehrerer Jahre soll entsprechend die A-posteriori-Wahrscheinlichkeit auf den vereinigten QS-Daten des Leistungserbringers berechnet werden. Als Kriterium, ob in den betrachteten QS-Daten überhaupt ein Verdacht auf ein mögliches Qualitätsdefizit besteht, wird die rechnerische Auffälligkeit, berechnet basierend auf der Vereinigung der betrachteten QS-Daten, verwendet.

Bei maximal zwei betrachteten Jahren lautet die Regel also:

- Ist das QI-Ergebnis des Leistungserbringers im aktuellen Erfassungsjahr statistisch signifikant auffällig, so wird der Leistungserbringer als „quantitativ auffällig“ eingestuft.
- Ist das QI-Ergebnis für die Daten im aktuellen Erfassungsjahr rechnerisch auffällig und über die Daten beider Erfassungsjahre zusammen statistisch auffällig, so wird der Leistungserbringer als „quantitativ auffällig“ eingestuft.
- Andernfalls wird der Leistungserbringer als „quantitativ unauffällig“ eingestuft.

Formal lässt sich diese Entscheidungsregel darstellen als

$$s_{\text{stat.sig.Bayes2}}(o_0, o_{-1}, J_0, J_{-1}, R, \alpha) = (s_{\text{stat.sig.Bayes}}(o_0, J_0, R, \alpha) \vee s_{\text{stat.sig.Bayes}}(o_0 + o_{-1}, J_0 + J_{-1}, R, \alpha)) \wedge s_{\text{rech}}(o_0, J_0, R).$$

Im Beispiel aus Abbildung 18 wäre der Leistungserbringer also im Erfassungsjahr 2017 quantitativ auffällig, da im zweiten Jahr in Folge eine rechnerische Auffälligkeit besteht.

Die Abbildung 21 zeigt wieder für einen Ratenindikator mit Referenzbereich  $\leq 10\%$ , welche Ergebnisse für einen Leistungserbringer mit  $J = 25$  Fälle pro Erfassungsjahr nach diesem Verfahren zur Auslösung von Stellungnahmen führen. Es sind die statistischen Einstufungen in den beiden Jahren sowie die statistische Einstufung über beide Jahre wichtig. Daher setzt sich die

Entscheidungsgrenze aus horizontalen und vertikalen Linien sowie aus einer Gegendiagonale ( $o_0 + o_{-1} = \text{konst.}$ ) zusammen.

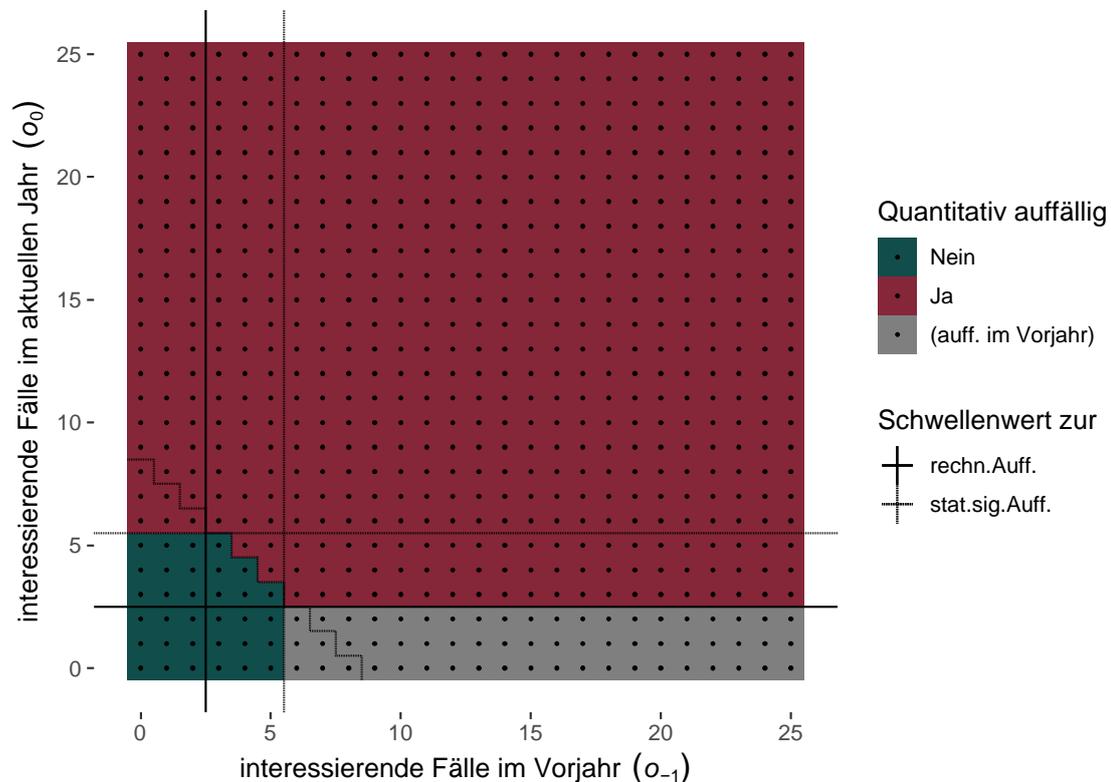


Abbildung 21: Verhalten der statistischen Auffälligkeit auf bis zu zwei Jahren bei verschiedenen Ergebnissen für einen Leistungserbringer mit 25 Fällen in jedem Jahr

Beispielsweise wird ein Leistungserbringer, der im aktuellen Erfassungsjahr und im Vorjahr jeweils  $J = 25$  Fällen in der Grundgesamtheit des QI hatte, als „quantitativ unauffällig“ klassifiziert, wenn es im Vorjahr und im aktuellen Erfassungsjahr jeweils vier interessierende Ereignisse gab. Gab es hingegen im aktuellen Jahr vier interessierende Ereignisse und im Vorjahr fünf, so wird der Leistungserbringer als „quantitativ auffällig“ klassifiziert.

#### 5.4.2.4 GLR-Binom-Verfahren

In der Statistischen Prozesskontrolle werden auch gerne CUSUM-Verfahren (Englisch: *cumulative sum*, Deutsch: kumulierte Summe) verwendet (Montgomery 2013, Abschnitt 9.1). In CUSUM-Verfahren wird für jeden einzelnen Zeitschritt eine Kennzahl berechnet, die ein Maß für die Evidenz ist, dass der Prozess „außer Kontrolle“ ist. Aus diesen Kennzahlen, die positiv oder negativ sein können, werden kumulative Summen gebildet. Dabei werden negative Zwischensummen auf 0 zurückgesetzt, um zu vermeiden, dass aus alten Daten zu viel „Guthaben“ angesammelt werden kann. Falls die kumulierte Summe einen vorher festgelegten Schwellenwert überschreitet, wird ein Signal gegeben.

Um solch ein CUSUM-Verfahren klassischer Weise einzustellen, muss man sich im Vorneherein auf zwei Hypothesen an die zugrunde liegende Rate  $\theta$  an interessierenden Ereignissen festlegen:

Die Nullhypothese  $H_0: \theta = \theta_{uk}$  modelliert die Annahme, dass der Prozess „unter Kontrolle“ ist, während die Alternativhypothese  $H_1: \theta = \theta_{ak}$  (mit  $\theta_{ak} > \theta_{uk}$ ) modelliert, dass der Prozess „außer Kontrolle“ ist. Dadurch ist implizit festgelegt, welche Größenordnung von Änderung im Parameter  $\theta$  man detektieren will. Die Güte der Detektion hängt dabei stark von den getroffenen Festlegungen für  $\theta_{uk}$  und  $\theta_{ak}$  ab. Diese Formulierung der Nullhypothese und Alternativhypothese vereinfacht die Vorgehensweise der statistischen Auffälligkeit, bei der sowohl die Nullhypothese  $\theta \leq R$  als auch die Alternativhypothese  $\theta > R$  zusammengesetzter Natur ist, d. h. nicht einzelnen Werten, sondern Intervallen entspricht.

Im Folgenden soll stattdessen das GLR-Binom-Verfahren betrachtet werden, bei dem die erwartete Rate unter der Alternativhypothese nicht vorher festgesetzt werden muss, sondern für jeden Zeitpunkt aus den vorhandenen Daten geschätzt wird. Das GLR-Binom-Verfahren kann auch als eine sequentielle Aneinanderreihung von Likelihood-Quotiententests gesehen werden. In das oben eingeführte Verfahrensschema passt es wie folgt: Der Hinweis auf einen Qualitätsmangel in den QS-Daten eines Leistungserbringers in einem oder mehreren Erfassungsjahren wird als hinreichend erachtet, wenn ein Likelihood-Quotiententest die Nullhypothese  $H_0: \theta \leq R$  zugunsten der Alternativhypothese  $H_1: \theta > R$  ablehnt. Der Likelihood-Quotiententest hat als Parameter einen Schwellenwert  $c > 0$ . Die Nullhypothese wird abgelehnt, wenn der Logarithmus des Likelihood-Quotienten der beiden Hypothesen größer ist als  $c$ .

Insgesamt kann die Strategie auf folgende algorithmische Art dargestellt werden: Sei

$$s_{Irt}(o, J, R, c) = I \left( \log \frac{f_{Bin}(o; J, \frac{o}{J})}{f_{Bin}(o; J, R)} \geq c \right)$$

die Entscheidungsfunktion eines Likelihood-Quotiententests für die Nullhypothese  $\theta \leq R$  mit Schwellenwert  $c$ . Damit ist

$$s_{glrbin2}(o_0, o_{-1}, J_0, J_{-1}, R, c) = (s_{Irt}(o_0, J_0, R, c) \vee s_{Irt}(o_0 + o_{-1}, J_0 + J_{-1}, R, c)) \wedge s_{rech}(o_0, J_0, R),$$

vgl. z. B. Huang et al. (2012) bzw. Höhle und Paul (2008).

Abbildung 22 zeigt wieder, welche Ergebnisse zur Auslösung führen (mit  $J = 25$  und Referenzbereich  $\leq 10\%$ ). Verwendet wird als Schwellenwert  $c = 1,5$ . Wie bei der statistischen Auffälligkeit auf zwei Jahren setzt sich die Entscheidungsgrenze aus horizontalen und vertikalen Linien sowie aus einer Gegendiagonale ( $o_1 + o_2 = konst.$ ) zusammen.

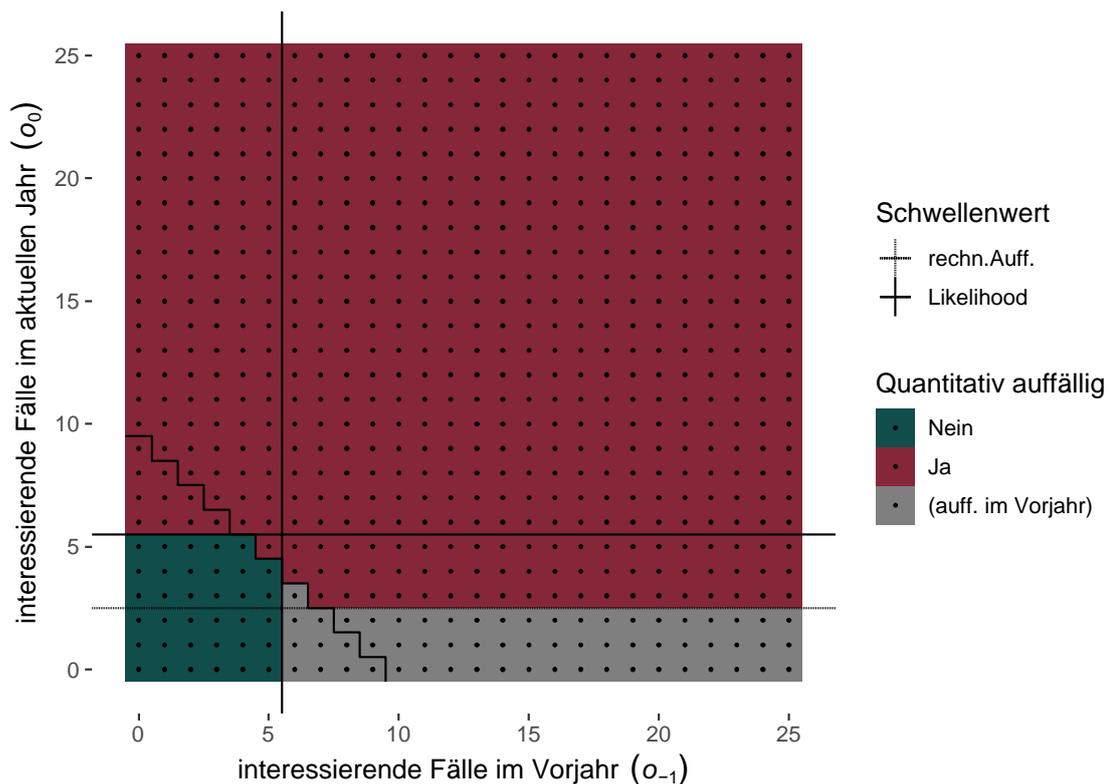


Abbildung 22: Verhalten des GLR-Binom-Verfahrens bei verschiedenen Ergebnissen aus zwei Jahren für einen Leistungserbringer mit 25 Fällen in jedem Jahr

Beispielsweise wird ein Leistungserbringer, der im aktuellen Erfassungsjahr und im Vorjahr jeweils  $J = 25$  Fälle in der Grundgesamtheit des QI hatte, als „quantitativ unauffällig“ klassifiziert, wenn es im Vorjahr und im aktuellen Erfassungsjahr jeweils vier interessierende Ereignisse gab. Gab es hingegen im aktuellen Jahr und im Vorjahr jeweils fünf interessierende Ereignisse, so wird der Leistungserbringer als „quantitativ auffällig“ klassifiziert.

#### Vergleich von statistisch signifikanter Auffälligkeit und GLR-Binom-Verfahren

Das GLR-Binom-Verfahren ähnelt der statistisch signifikanten Auffälligkeit, die auf Basis mehrerer Erfassungsjahre berechnet wird: Während das GLR-Binom-Verfahren auf dem Likelihood-Quotienten-Test aufbaut, verwendet die frequentistische statistisch signifikante Auffälligkeit stattdessen einen bayesianischen Hypothesentest. In beiden Verfahren sind daher die Entscheidungsgrenzen ähnlich. Beide Verfahren besitzen einen Parameter ( $\alpha$  bzw.  $c$ ), über den sich Sensitivität bzw. Spezifität (pro Durchlauf des Verfahrens, also im esQS-Kontext pro Jahr) einstellen lassen. Bei der statistisch signifikanten Auffälligkeit besitzt der Parameter  $\alpha$  eine einfache Interpretation als Signifikanzniveau. Diese Interpretation gilt jedoch nur für einen einzelnen Hypothesentest und beschreibt nicht das Gesamtverfahren, welches multiple Tests verwendet. Diese Interpretation kann auch ein Nachteil sein und unter Umständen eine freie Wahl des Parameters erschweren, zugunsten des aus plan. QI bekannten Parameterwerts  $\alpha = 0,05$ . Im vorliegenden Bericht wird daher der Parameter  $\alpha$  auch nicht variiert.

#### 5.4.2.5 Statistisch relevante Auffälligkeitseinstufung

Zur Beurteilung der statistischen Evidenz für einen möglichen Qualitätsmangel lässt sich auch die in Abschnitt 5.3.1.3 (Statistisch relevante Auffälligkeit) eingeführte Methodik verwenden, welche den Begriff „hinreichender Hinweis“ im Sinne der „Relevanz“ für Patienten und Patientinnen operationalisiert. Dazu wird aus den betrachteten QS-Daten der Erwartungswert  $E[l(\cdot, y)]$  der entsprechenden Verlustfunktion geschätzt und anschließend mit einem vorgegebenen Schwellenwert  $\zeta$  verglichen.

Bei der Berechnung von  $E[l(\cdot, y)]$  wird angenommen, dass die Anzahl der Fälle, für die der Verlust geschätzt wird, gleich der Anzahl der Fälle im aktuellen Erfassungsjahr ist. Will man hingegen den erwarteten Verlust für ein Jahr abschätzen und verwendet man dafür die QS-Daten aus  $k > 1$  Erfassungsjahren, so ist das Ergebnis  $E[l(\cdot, y)]$  in diesem Fall durch  $k$  zu teilen. Alternativ kann man auch  $E[l(\cdot, y)]$  mit  $k \cdot \zeta$  vergleichen um festzustellen, ob der erwartete Verlust pro Erfassungsjahr größer gleich  $\zeta$  ist.

Formal lautet die betrachtete 2-Jahres-Strategie also:

$$s_{\text{stat.rel2}}(o_{-1}, o_0, J_{-1}, J_0, \zeta, R) \\ = (s_{\text{stat.rel}}(o_{-1} + o_0, J_{-1} + J_0, 2\zeta, R) \vee s_{\text{stat.rel}}(o_0, J_0, \zeta, R)) \wedge s_{\text{rech}}(o_0, J_0, R)$$

Die Abbildung 23 zeigt wieder, welche Ergebnisse zur Auslösung führen (mit  $J = 25$  und Referenzbereich  $\leq 10\%$ ). Verwendet wird als Schwellenwert  $\zeta = 2$ . Die Entscheidungsgrenze setzt sich aus einer horizontalen Linie sowie aus einer Gegendiagonale ( $o_1 + o_2 = \text{konst.}$ ) zusammen.

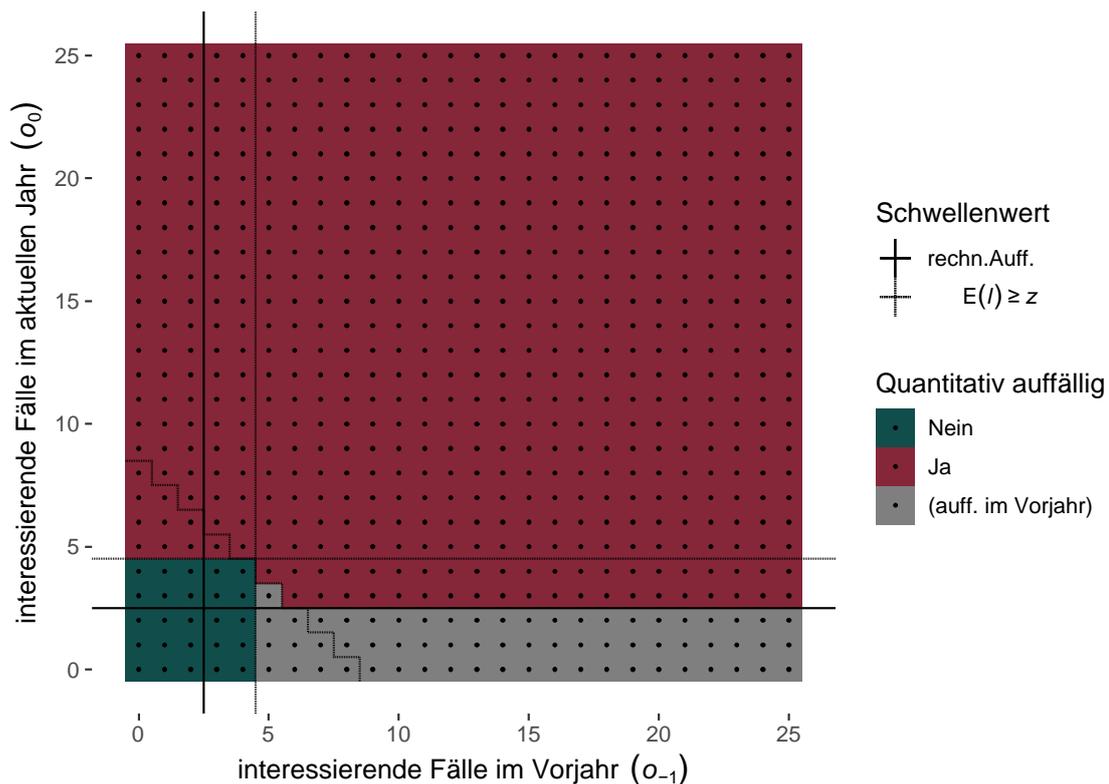


Abbildung 23: Verhalten der statistisch relevanten Auffälligkeitseinstufung auf den Daten zweier Jahre bei verschiedenen Ergebnissen für einen Leistungserbringer mit 25 Fällen in jedem Jahr

Beispielsweise wird ein Leistungserbringer, der im aktuellen Erfassungsjahr und im Vorjahr jeweils  $J = 25$  Fälle in der Grundgesamtheit des QI hatte, als „quantitativ unauffällig“ klassifiziert, wenn es im Vorjahr und im aktuellen Erfassungsjahr jeweils vier interessierende Ereignisse gab. Gab es hingegen im aktuellen Jahr vier interessierende Ereignisse und im Vorjahr fünf oder mehr interessierende Ereignisse, so wird der Leistungserbringer als „quantitativ auffällig“ klassifiziert.

### 5.4.3 Vergleich anhand von Sensitivität und Spezifität

In diesem Abschnitt sollen die vorgestellten Methoden zur 2-Jahres-Einstufung mittels Sensitivität und Spezifität verglichen werden. Der nächste Abschnitt 5.4.4 kombiniert dann die beiden Dimensionen im Rahmen einer konkreten entscheidungstheoretischen Modellierung. Alternativ wäre es möglich, die Methoden an der Auslösewahrscheinlichkeit nach mehreren Jahren zu vergleichen. Man könnte z. B. fragen: Wie wahrscheinlich ist es, dass die Methode für einen gegebenen Beispielleistungserbringer innerhalb von 5 Jahren ein Stellungnahmeverfahren auslöst? Eine weitere Kennzahl, die in der statistischen Prozesskontrolle oft hinzugezogen wird, ist die erwartete Lauflänge (*average run length, ARL*), also die mittlere Anzahl von Jahren, nach denen ein Stellungnahmeverfahren ausgelöst wird. Hier wird lediglich die Auslösewahrscheinlichkeit pro Jahr betrachtet, da das Ziel der esQS die jährliche Bewertung von Leistungserbringern ist.

Die vorgeschlagenen Methoden mit ihren zugehörigen Tuning-Parametern unterscheiden sich in Sensitivität und Spezifität (in egal welchem Zeitraum). Im Folgenden soll in Anlehnung an den Vergleich der 1-Jahres-Methoden in Abschnitt 5.3.3 die Wahrscheinlichkeit für das Auslösen

eines Stellungnahmeverfahrens (pro Durchlauf des Verfahrens, also pro Erfassungsjahr) für Beispielleistungserbringer anhand dreier Szenarien verglichen werden. Den Vergleich mittels Sensitivität und Spezifität pro Erfassungsjahr durchzuführen begründet sich darin, dass jährlich eine Entscheidung getroffen werden muss. Bei Beispielleistungserbringern, bei denen angenommen wird, dass der Kompetenzparameter im Referenzbereich liegt, d. h.  $\theta \leq R$ , entspricht diese Auslöswahrscheinlichkeit einer Fehlerwahrscheinlichkeit 1. Art und sollte möglichst klein sein. Bei Beispielleistungserbringern mit angenommener mangelhafter Kompetenz entspricht diese Auslöswahrscheinlichkeit der Power des Verfahrens und sollte möglichst groß sein. Die Parameter der Verfahren müssen so gewählt werden, dass ein guter Kompromiss zwischen geringer Fehlerwahrscheinlichkeit und großer Power erreicht wird.

Die Auslöswahrscheinlichkeit hängt neben der angenommenen Kompetenz des Leistungserbringers auch von der angenommenen Fallzahl des Leistungserbringers ab. Aufgrund der Diskretheit der Fallzahlen oszilliert die Auslöswahrscheinlichkeit mit der Fallzahl: Leistungserbringer mit gleicher Kompetenz und ähnlicher Fallzahl können sehr unterschiedliche Auslöswahrscheinlichkeit haben. Da dieser Effekt unerwünscht ist, ist ein weiteres Bewertungskriterium der verschiedenen Methoden, wie glatt die Abhängigkeit der Auslöswahrscheinlichkeit von der Fallzahl ist.

Im Folgenden wird lediglich der Parameter  $c$  der GLR-Binom-Methode variiert. Der Schwellenwert bei statistisch signifikanten Auffälligkeiten wird auf den auch im Verfahren plan. QI verwendeten Wert  $\alpha = 0,05$  fixiert, und bei rechnerischen Bewertungen wird stets mit dem Referenzwert  $R$  verglichen. Der Parameter  $\zeta$  der statistisch relevanten Auffälligkeiten wird auf den Wert  $\zeta = 2$  fixiert. Über den Parameter  $c$  lassen sich die Fehlerwahrscheinlichkeiten 1. Art der GLR-Binom-Methode für einen Beispielleistungserbringer einstellen. Damit lässt sich diese Methoden mit den anderen Methoden anhand der Power vergleichen.

Als zweiter Vergleichsansatz, welche Sensitivität und Spezifität der Klassifizierung gegeneinander abwägt, wird in Abschnitt 5.4.4 eine zeitlich reduzierte Version des Entscheidungsdiagramms aus Abschnitt 5.4 mit nur zwei Zeitpunkten verwendet.

Betrachtet wird ein Ratenindikator mit Referenzbereich  $\leq 10\%$ . Wie in der Einleitung erwähnt, ist eines der Ziele der gemeinsamen Betrachtung mehrerer Erfassungsjahre die Verbesserung der Aussagekraft für kleine Leistungserbringer. Daher werden Fallzahlen bis maximal  $J = 30$  betrachtet. Der Einfachheit halber wird angenommen, dass die Leistungserbringer in beiden aufeinanderfolgenden Jahren die gleiche Fallzahl aufweisen. Unterschiedliche Fallzahlen in beiden Jahren führen vermutlich zu etwas glatteren Kurven.

#### 5.4.3.1 Abhängigkeit von der Fallzahl bei fester Kompetenz

##### Szenario: Kompetenz gleich Referenzwert

Für den Fehler erster Art wird zunächst beispielhaft der Fall eines Leistungserbringers mit  $\theta = R = 10\%$  betrachtet. Es handelt sich also um einen Leistungserbringer mit gerade noch akzeptablem Kompetenzparameter. Die Auslöswahrscheinlichkeit ist hier also die Wahrscheinlich-

keit für den Fehler 1. Art im aktuellen Erfassungsjahr. Die Auslösewahrscheinlichkeit in Abhängigkeit der Fallzahl pro Jahr ist in Abbildung 24 dargestellt. Zusätzlich zu den in Abschnitt 5.4.2 eingeführten Methoden wird als Vergleich noch die statistisch signifikante Auffälligkeitseinstufung auf der Basis des aktuellen Erfassungsjahres sowie die statistisch relevante Auffälligkeitseinstufung auf der Basis des aktuellen Erfassungsjahres betrachtet.

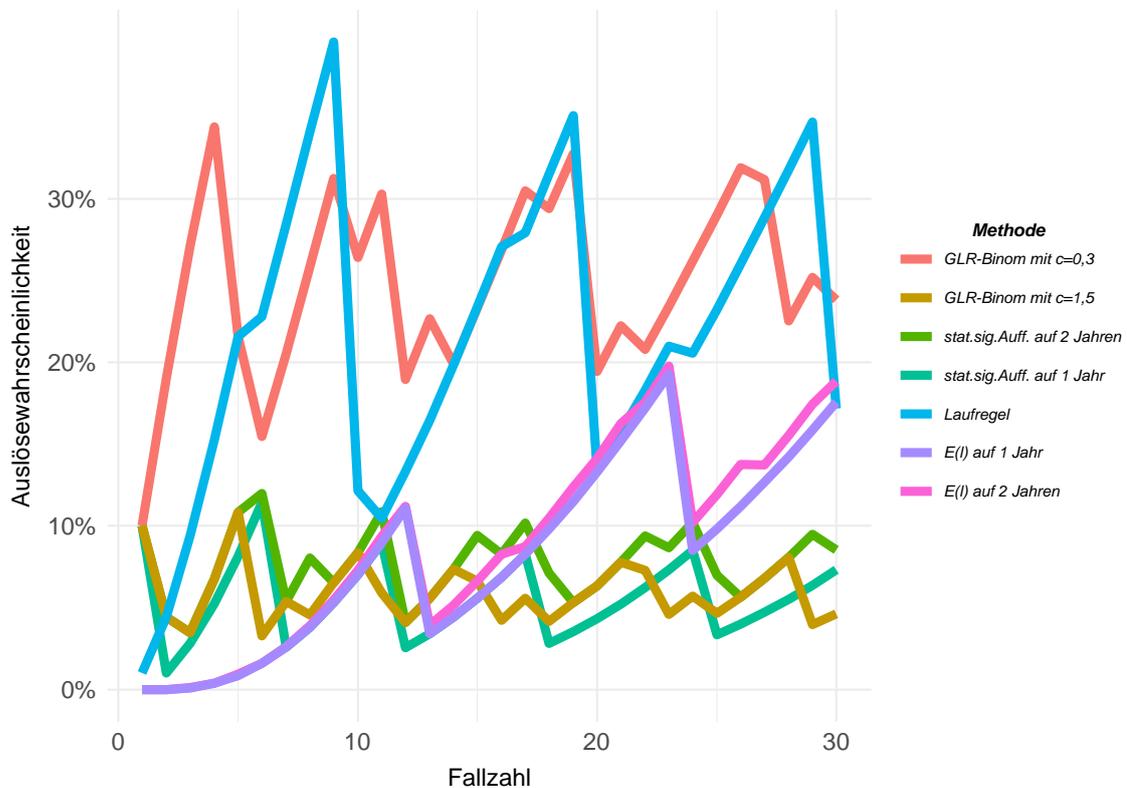


Abbildung 24: Die Auslösewahrscheinlichkeit verschiedener Methoden bei  $\vartheta=R$  als Funktion von  $J$ .

Man sieht, dass die Wahrscheinlichkeit für einen Fehler 1. Art für die Laufregel am größten ist und außerdem die stärksten Oszillationen vorweist. Zum Teil hängt beides zusammen, d. h. die Oszillationen der Wahrscheinlichkeit sind größer, weil auch die Wahrscheinlichkeit absolut größer ist als die Wahrscheinlichkeit der anderen Methoden. Darüber hinaus kommen die großen Oszillationen auch durch die Form der Menge an Ereignissen, die zur Auslösung führen: Bei der quadratischen Form in Abbildung 20 können kleine Änderungen in der Fallzahl zu großen Änderungen in der berechneten Fehlerwahrscheinlichkeit führen. Zum Vergleich wird auch ein GLR-Binom-Verfahren mit recht niedrigem  $c = 0,3$  betrachtet, welches eine vergleichbare durchschnittliche Fehlerwahrscheinlichkeit wie die Laufregel bietet, allerdings bei deutlich geringeren Oszillationen. Wie im Abschnitt 5.4.2.2 besprochen, lässt sich die Auslösewahrscheinlichkeit der Laufregel durch die Einführung eines weiteren Parameters  $R' > R$  verringern. Dies würde jedoch nicht das Problem der Oszillationen lösen.

Die Fehlerwahrscheinlichkeit 1. Art der statistisch signifikanten Auffälligkeit über zwei Jahre,  $S_{\text{stat.sig.Bayes2}}$ , liegt leicht über der einfachen statistisch signifikanten Auffälligkeit. Beide Methoden haben ähnliche Eigenschaften wie die GLR-Binom-Methode mit  $c = 1,5$ . Tendentiell sind die Oszillationen für das GLR-Binom-Verfahren am geringsten.

Für die Einstufung nach der statistischen Relevanz steigt die Auslösewahrscheinlichkeit mit wachsender Fallzahl und ist für kleine Leistungserbringer sehr gering. Beide Effekte wurden bereits im Abschnitt 5.3.3 besprochen. Man sieht auch, dass es bei dieser Methodik vor allem bei kleinen Leistungserbringern keinen großen Unterschied gibt, ob man ein oder zwei Jahre betrachtet.

### Szenario: $\theta = 2R$

Im Folgenden wird die Auslösewahrscheinlichkeit für Leistungserbringer mit unzureichendem Kompetenzparameter  $\theta = 2R = 20\%$  geplottet (also die „Power“ für diesen Wert von  $\theta$ ):

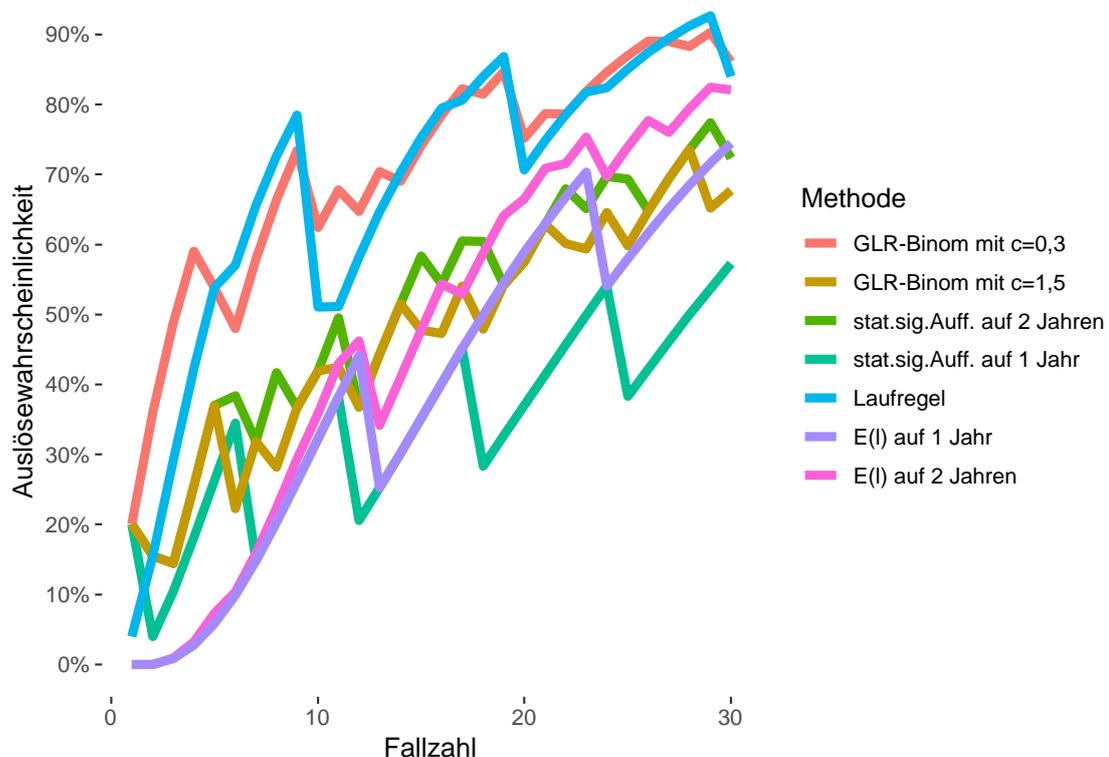


Abbildung 25: Die Auslösewahrscheinlichkeit verschiedener Methoden bei  $\vartheta=2R$  als Funktion von  $J$ .

Die Power der GLR-Binom-Methode mit  $c = 1,5$  ist vergleichbar mit der Power der einfachen statistisch signifikanten Auffälligkeit und der statistisch signifikanten Auffälligkeit über zwei Jahre.

**Szenario:  $\theta = R/2$** 

Leistungserbringer mit  $\theta = R$  haben „akzeptable“ Qualität; liegt der Bundesdurchschnitt im Referenzbereich, so bedeutet dies in der Regel, dass die meisten Leistungserbringer besser als der Referenzbereich sind. Daher wird als drittes ein Leistungserbringer mit  $\theta = \frac{R}{2} = 0,05$  betrachtet.

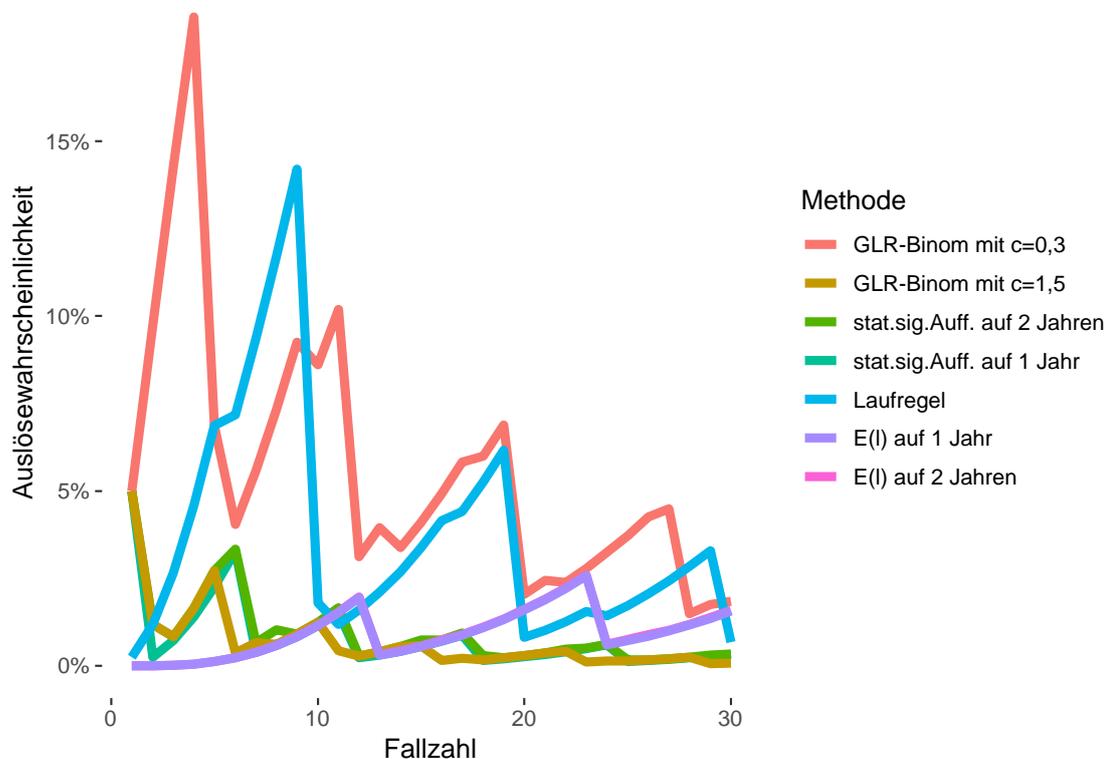


Abbildung 26: Die Auslösewahrscheinlichkeit verschiedener Methoden bei  $\vartheta=R/2$  als Funktion von  $J$ .

Bei den meisten Methoden geht die Auslösewahrscheinlichkeit mit steigender Fallzahl gegen Null. Die Laufregel und die Vergleichsmethode GLR-Binom mit  $c = 0,3$  erreichen dabei jedoch relativ hohe Auslösewahrscheinlichkeiten (teilweise über 10 %). Die Auslösewahrscheinlichkeit der auf der statistischen Relevanz basierenden Methoden besitzen im betrachteten Fallzahlenbereich eine steigende Tendenz, bleiben jedoch klein (unter 3 %).

#### 5.4.3.2 Abhängigkeit vom Kompetenzparameter bei fester Fallzahl

In diesem Abschnitt wird die Abhängigkeit der Auslösewahrscheinlichkeit vom Kompetenzparameter dargestellt, wobei die Fallzahl fixiert wird. Um die Oszillationen der Auslösewahrscheinlichkeit mit der Fallzahl etwas auszugleichen wird dabei über einen Bereich von Fallzahlen gemittelt. Wie gehabt wird ein Referenzwert von 10 % angenommen.

Abbildung 27 zeigt die Auslösewahrscheinlichkeit gemittelt für  $J = 3, \dots, 7$ . Abbildung 28 zeigt die Auslösewahrscheinlichkeit gemittelt für  $J = 11, \dots, 20$ .

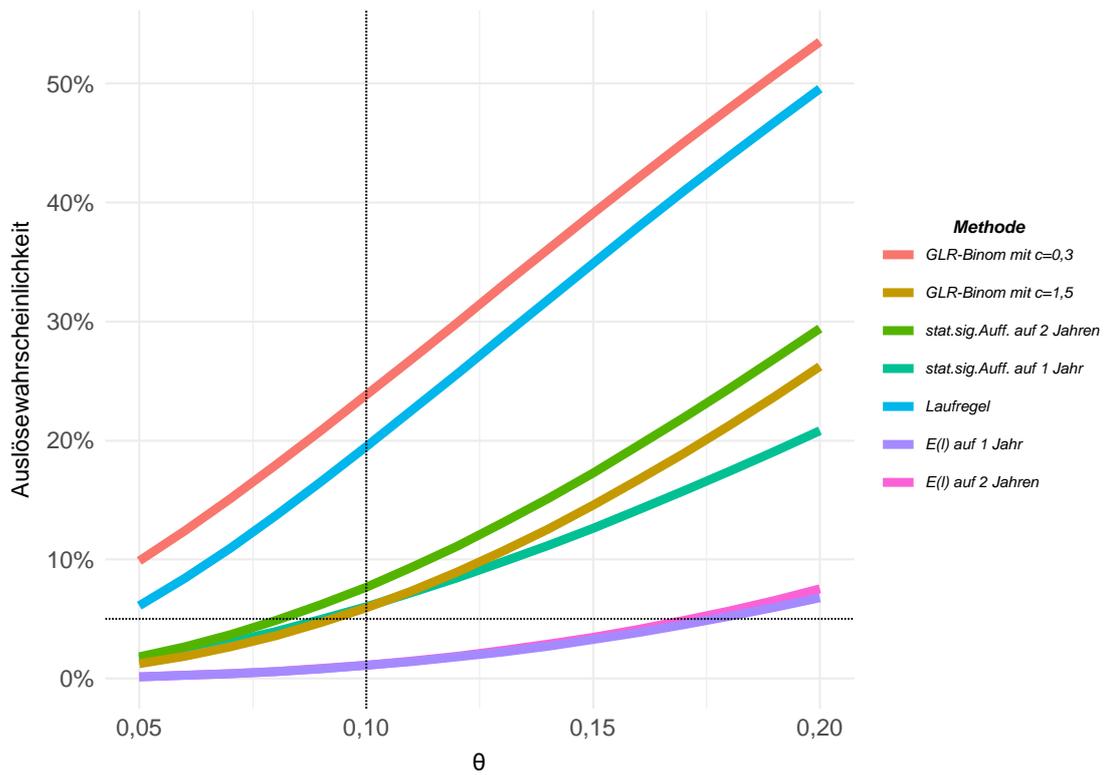


Abbildung 27: Die Auslösewahrscheinlichkeit verschiedener Methoden als Funktion von  $\vartheta$  gemittelt über  $J=3, \dots, 7$ . Die gestrichelten Linien markieren den Referenzwert  $R$  sowie den Schwellenwert  $\alpha=0,05$ .

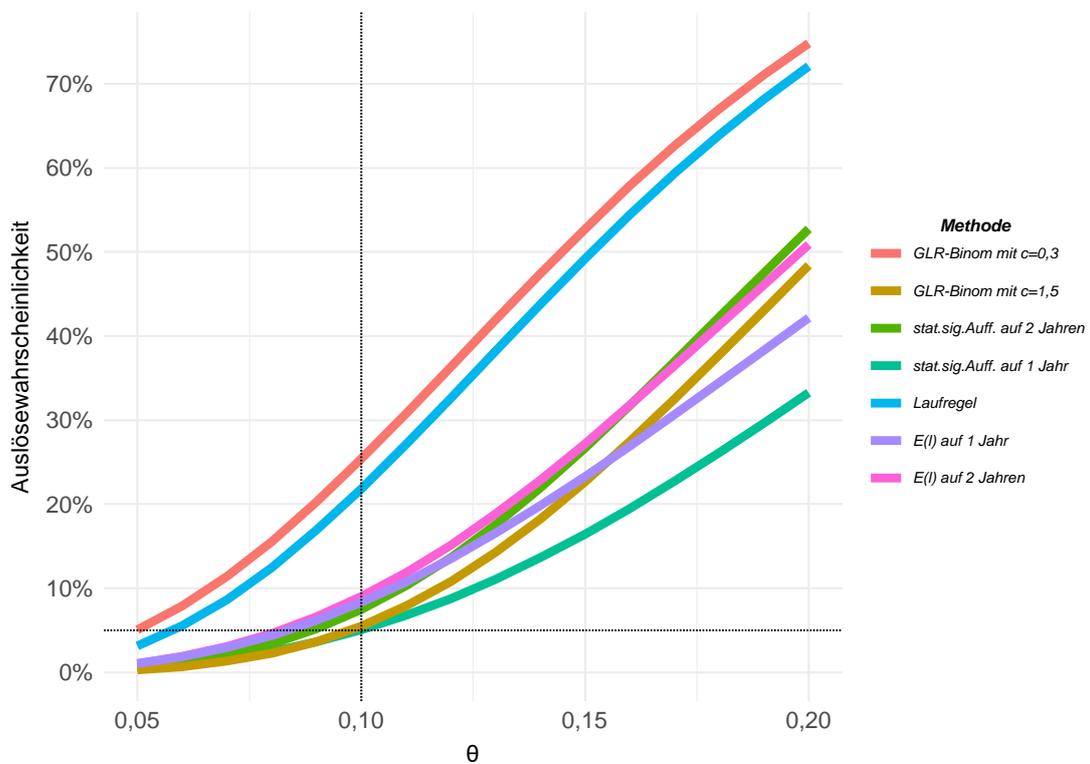


Abbildung 28: Die Auslösewahrscheinlichkeit verschiedener Methoden als Funktion von  $\vartheta$  gemittelt über  $J=11, \dots, 20$ . Die gestrichelten Linien markieren den Referenzwert  $R$  sowie den Schwellenwert  $\alpha=0,05$ .

Man sieht, dass die Kurven bei den Verfahren, die nur ein einzelnes Erfassungsjahr berücksichtigen, jeweils langsamer ansteigen, als die Kurven der entsprechenden Verfahren, die zwei Erfassungsjahre berücksichtigen. Dies illustriert den Vorteil der Vergrößerung der Datengrundlage. Wie schon im vorangegangenen Abschnitt sichtbar ist die große Auslösewahrscheinlichkeit der Laufregel und der Vergleichsmethode GLR-Binom mit  $c = 0,3$ . Die Kurven für die statistisch signifikante Auffälligkeitseinstufung und das GLR-Binom-Verfahren mit  $c = 1,5$  laufen parallel, was auf ähnliche Eigenschaften der Verfahren hindeutet.<sup>52</sup>

#### 5.4.4 Vergleich durch Entscheidungsdiagramm mit zwei Zeitpunkten

Im vorherigen Abschnitt wurden jeweils Sensitivität und Spezifität der verschiedenen 2-Jahres-Auffälligkeitseinstufungsmethoden als separate Dimensionen betrachtet, was hilfreiche Einblicke liefert, aber es dem Entscheidungsträger überlässt, implizit zwischen diesen beiden Dimensionen zu gewichten. Für eine konkrete Empfehlung einer Methodik bedarf es, ähnlich der Vorgehensweise zur quantitativen Auffälligkeitseinstufung für die Daten eines Erfassungsjahres, der Kombination der Dimensionen im Rahmen einer entscheidungstheoretischen Modellierung.

Da generelle Betrachtungen zur optimalen Vorgehensweise im sequentiellen Kontext stark von den Annahmen zur z. B. zeitlichen Dynamik abhängen, sind generelle Aussagen zur optimalen Vorgehensweise deutlich schwieriger als für die 1-Jahres-Auffälligkeitseinstufung. Stattdessen werden die vorgeschlagenen Methoden im Rahmen eines konkreten Einflussdiagramms mit nur zwei Entscheidungen verglichen. In kalendarischer Zeitangabe sind die Entscheidungsvariablen im Einflussdiagramm unter dieser Annahme  $S_1$  und  $S_2$ , die jeweils nach den vorgeschlagenen Strategien  $S_{\text{lauf}2}$ ,  $S_{\text{stat.sig.Bayes}2}$ ,  $S_{\text{stat.rel}2}$ , und  $S_{\text{glrbin}2}$  aus Abschnitt 5.4.2 getroffen werden sollen. Für die Entscheidung  $S_1$  stehen die Informationen bzgl.  $O_1$  und  $J_1$  zur Verfügung; für die Entscheidung  $S_2$  stehen die Informationen bzgl.  $O_1$ ,  $J_1$ ,  $S_1$ ,  $O_2$  und  $J_2$  zur Verfügung. Der Träger des Kompetenzparameters  $\theta_t$  wird vereinfachend als die zwei möglichen Werte  $\theta_{\text{unter-kontrolle}}$  bzw.  $\theta_{\text{außer-kontrolle}}$  angenommen, wobei die Begriffe „unter-Kontrolle“ und „außer-Kontrolle“ für diese beiden Werte aus der statistischen Prozesskontrolle inspiriert sind (Montgomery 2013). Das heißt, das Ziel ist es, optimal für einen Leistungserbringer für jedes Erfassungsjahr differenzieren zu können, ob dieser die zugrunde liegenden Rate von  $\theta_{\text{unter-kontrolle}}$  oder die zugrunde liegenden Rate  $\theta_{\text{außer-kontrolle}}$  hat. Für die konkrete Berechnung wird  $\theta_{\text{unter-kontrolle}} = 5\%$  und  $\theta_{\text{außer-kontrolle}} = 20\%$ , gewählt. Als A-priori-Verteilung wird für  $\theta_1$  die folgende Verteilung gewählt,

$$P(\theta_1 = \text{unter-kontrolle}) = 90\%, \quad P(\theta_1 = \text{außer-kontrolle}) = 10\%.$$

Das heißt man geht davon aus, das 90 % der Leistungserbringer eine Kompetenz entsprechend dem Kompetenzparameter  $\theta_{\text{unter-kontrolle}}$  haben und 10 % eine Kompetenzparameter von  $\theta_{\text{außer-kontrolle}}$  haben.

<sup>52</sup> Für einen noch faireren Vergleich der Verfahren anhand des Verlaufs der Auslösewahrscheinlichkeiten müsste man die Parameter der Verfahren so wählen, dass die Auslösewahrscheinlichkeiten bei einem Vergleichswert  $\theta = R$  nahezu identisch sind. Die Schlussfolgerungen bleiben jedoch gleich. Die entsprechenden Abbildungen werden hier nicht gezeigt, da, wie oben besprochen, in diesem Abschnitt lediglich der Parameter  $c$  variiert wird.

Für die Übergangswahrscheinlichkeiten  $P(\theta_t | \theta_{t-1}, S_{t-1})$  des Kompetenzparameters werden folgende bedingte Wahrscheinlichkeiten gewählt:

Tabelle 10: Wahl der Übergangswahrscheinlichkeiten für den Kompetenzparameter

$\theta_t \setminus \theta_{t-1}, S_{t-1}$	Unter-Kontrolle		Außer-Kontrolle	
	$S_{t-1}=\text{nein}$	$S_{t-1}=\text{ja}$	$S_{t-1}=\text{nein}$	$S_{t-1}=\text{ja}$
Unter-Kontrolle	0,95	0,95	0	0,2
Außer-Kontrolle	0,95	0,95	1	0,8

Es wird also angenommen, dass ein Leistungserbringer, der im letzten Erfassungsjahr mit einem Kompetenzparameter entsprechend „unter-kontrolle“ gearbeitet hat, dies mit 95 % Wahrscheinlichkeit auch wieder im aktuellen Erfassungsjahr tut. Diese Wahrscheinlichkeit ist unabhängig davon, ob es im letzten Erfassungsjahr zu einer quantitativen Auffälligkeit kam oder nicht. War der Leistungserbringer dagegen im letzten Erfassungsjahr „außer-Kontrolle“, hängt die Situation im aktuellen Erfassungsjahr von der Entscheidung im letzten Erfassungsjahr ab: Hatte man sich damals gegen ein Signal entschieden ( $S_1 = \text{nein}$ ), ist der Leistungserbringer weiterhin „außer-Kontrolle“. War die Entscheidung im letzten Jahr anstelle, ( $S_1 = \text{ja}$ ), dann wird angenommen, dass mit einer Wahrscheinlichkeit von 20 % der Leistungserbringer im aktuellen Erfassungsjahr wieder „unter-Kontrolle“ ist.

Die Wahrscheinlichkeitsverteilung von  $O_t | \theta_t, J_t$  entspricht wie immer einer Binomialverteilung, d. h.

$$P(O_t = o_t | \theta_t, J_t) = \binom{J_t}{o_t} \theta_t^{o_t} (1 - \theta_t)^{J_t - o_t},$$

welches sowohl für  $\theta_{\text{unter-kontrolle}}$  als auch  $\theta_{\text{außer-kontrolle}}$  für den Träger von  $O_t$ , d. h. für 0, 1, ...,  $J_t$ , berechnet werden muss. Die Verlustfunktionen sind in Analogie zu Abschnitt 5.3.1 jeweils zunächst entsprechend der Verlustfunktion beim statistisch-signifikanten Ansatz, welcher nur die Aufwände für Fehlklassifikationen betrachtet und in Tabelle 11 dargestellt ist.

Tabelle 11: Wahl der mit verschiedenen Fehlklassifikationen assoziierten Aufwände

Unter-Kontrolle	Außer-Kontrolle	Unter-Kontrolle	Außer-Kontrolle
$S_{t-1}=\text{nein}$	$S_{t-1}=\text{ja}$	$S_{t-1}=\text{nein}$	$S_{t-1}=\text{ja}$
0	$fp$	$fn$	0

mit  $fp = 19$  und  $fn = 1$ .

Durch Lösen des Einflussdiagramms nach dem Algorithmus von Shenoy (1992) für einen Leistungserbringer mit  $J_1 = J_2 = 25$  erhält man als optimale Strategie  $S_1$ , dass ein Signal generiert wird wenn  $o_1 \geq 5$ . Anhand des gleichen Lösungsalgorithmus findet man, dass das die optimale Entscheidung für  $S_2$  ist, welche auch die Information zu  $S_1$  (nein oder ja) mit einbezieht, die in Abbildung 29 dargestellte Form hat.

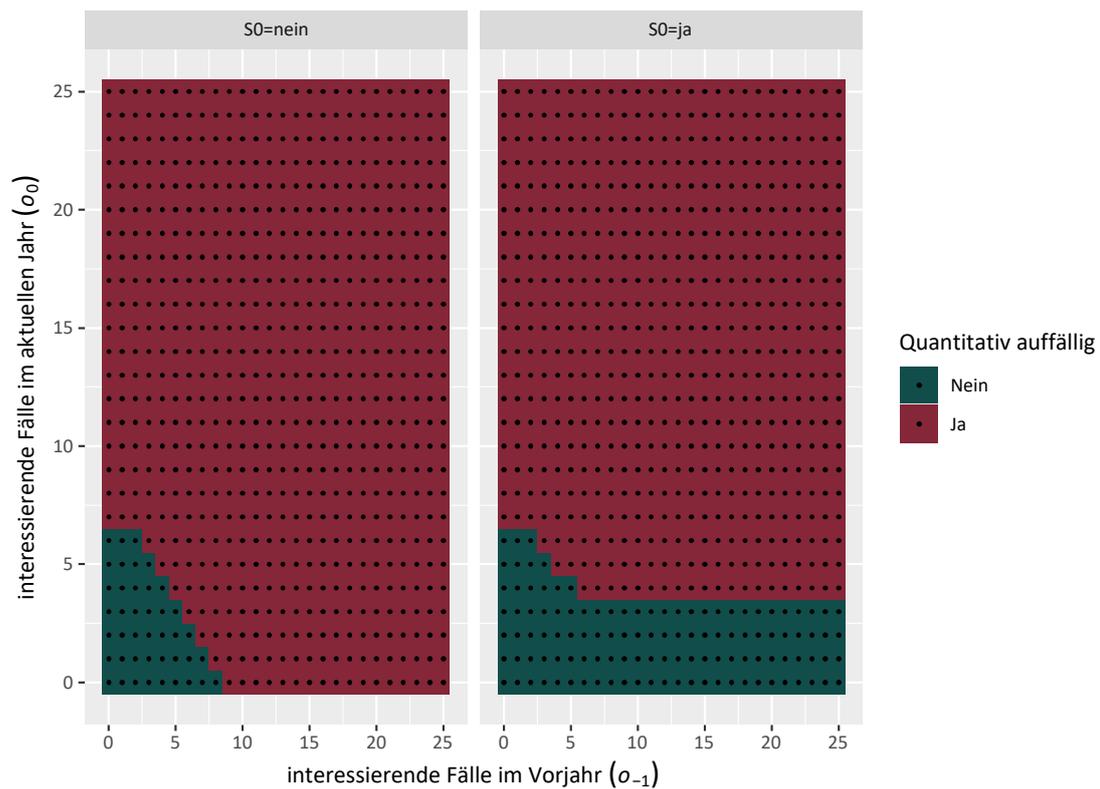


Abbildung 29: Optimale Strategie im gewählten Entscheidungsdiagramm in Abhängigkeit von  $o_0$ ,  $o_{-1}$  und  $S_0$ . Links die Entscheidung in Abhängigkeit von  $o_0$  und  $o_{-1}$ , falls  $S_0 = \text{nein}$  und rechts die Entscheidung für den Fall, dass  $S_0 = \text{ja}$ .

Man sieht, dass die optimale Strategie für  $S_2$  tatsächlich von  $S_1$  abhängt. Hat man sich für eine quantitative Auffälligkeit bei  $S_1$  entschieden und weil die Annahme ist, dass dies die zugrunde liegende Rate verbessert, würde man sich anschließend etwas zurückhaltender für eine quantitative Auffälligkeit bei  $S_2$  entscheiden.

Innerhalb des gewählten Einflussdiagramms können die verschiedenen Strategien mittels des jeweiligen erwarteten Verlustes verglichen werden. Die oben erwähnte optimale Entscheidungsstrategie sowie die sehr vereinfachten Strategien immer bzw. niemals eine Auffälligkeit für sowohl  $S_1$  als auch  $S_2$  auszulösen, liefern entsprechende Baseline-Niveaus für den Vergleich. Die Ergebnisse finden sich in Tabelle 4.

Tabelle 12: Erwarteter Verlust der verschiedenen 2-Jahres-Methoden für das konkrete Einflussdiagramm mit zwei Zeitpunkten.

Methode	erwarteter Verlust
optimal	0,165
GLR-Binom ( $c=1,5$ )	0,183
stat.sig.bayes2 ( $\alpha=0,005$ )	0,186
stat.sig.bayes2 ( $\alpha=0,05$ )	0,186

Methode	erwarteter Verlust
nie auffällig	0,245
stat.rel2 ( $\zeta=2$ )	0,350
lauf2	0,420
QSKH	4,304
immer auffällig	33,725

Per Konstruktion liefert die optimale Strategie den niedrigsten erwarteten Verlust. Aus der Tabelle wird klar, dass die Methode  $s_{\text{stat}2}$  mit  $\alpha = 0,05$  die besten Ergebnisse unter den in Abschnitt 5.4.2 vorgestellten Baukasten-Methoden für das betrachtete Setting liefert. Die Verwendung der rechnerischen Auffälligkeit in jedem Erfassungsjahr einzeln (die QSKH-RL-Methode) liefert wie erwartet schlechte Ergebnisse, weil es hier zu sehr vielen unnötigen Auffälligkeitseinstufungen kommt.

Insgesamt zeigen die obigen Ergebnisse, dass es prinzipiell möglich ist sehr komplexe Strategien zu verwenden, um im konkret angenommenen Einflussdiagramm optimale Entscheidungsstrategien zu erhalten. Jedoch sind diese stark abhängig von den verwendeten Annahmen, wie z. B. der A-priori-Verteilung für  $\theta_1$  und der Wahrscheinlichkeit, dass eine quantitative Auffälligkeit auch zu einer Qualitätsverbesserung führt. Auch das von der statistischen Prozesskontrolle inspirierte GLR-Binom-Verfahren liefert hervorragende Ergebnisse, jedoch ist das Verfahren bereits relativ komplex und es gibt keine gute Methodik zur Wahl des Schwellenwertes. Somit erscheinen die etwas einfacheren 2-Jahres-Einstufungen nach dem Baukasten-Prinzip anhand von statistischen Verfahren und ohne Berücksichtigung früherer Entscheidungen als robuste Alternativen, die auch gut in anderen Settings als dem betrachteten Einflussdiagramm funktionieren. Des Weiteren sind diese Strategien auch einfacher zu kommunizieren und liefern fast so gute Ergebnisse wie die optimale Strategie oder das GLR-Binom-Verfahren im konkreten Setting.

Bei der Interpretation der obigen Ergebnisse ist zu beachten, dass das konkrete Entscheidungsdiagramm für einen Leistungserbringer mit  $J = 25$  Fällen und konkreten Festlegungen für die Wahrscheinlichkeitstabellen nur eine von vielen möglichen Einflussdiagrammen für einen Vergleich der Methoden ist. Trotzdem ist diese Art der Bewertung hilfreich, um zu einer kombinierten Sichtweise von Sensitivität und Spezifität zu kommen und liefert wichtige Einblicke in die Verhaltensweisen der verschiedenen Methoden.

#### 5.4.5 Weitere Aspekte

##### 5.4.5.1 Berücksichtigung vorangegangener Stellungnahmeverfahren

Bei der Verwendung von QS-Daten aus Vorjahren will man vermeiden, dass schlechte Ergebnisse in der Vergangenheit zu lange einen Einfluss auf die Gesamteinstufung haben. Einerseits ist dies dadurch gewährleistet, dass vor Erweiterung der Datengrundlage stets geprüft wird, ob die aktuell betrachteten Daten überhaupt einen Verdacht auf einen möglichen Qualitätsmangel nahelegen. Dadurch wird verhindert, dass ein Leistungserbringer mehrfach aus identischen Gründen

in ein Stellungnahmeverfahren gelangt. Daneben könnte man auch die Ergebnisse vorangegangener Stellungnahmeverfahren berücksichtigen.

Wenn im Vorjahr bei einem Leistungserbringer ein Stellungnahmeverfahren ausgelöst wurde, bei dem der Leistungserbringer bei der Bewertung dann abschließend als „unauffällig“ eingestuft wurde, so könnte man dies als Anlass sehen, zur Bewertung des aktuellen Jahres keine Vorjahresdaten hinzuzunehmen. Da im Vorjahr ein Stellungnahmeverfahren ausgelöst wurde, enthalten die Vorjahresdaten scheinbare statistische Evidenz für einen möglichen Qualitätsmangel. Diese statistische Evidenz hat sich jedoch im Stellungnahmeverfahren nicht erhärtet. Daher sollte man diese statistische Evidenz nicht unbedingt zur Bewertung der aktuellen QI-Ergebnisse heranziehen. Dabei muss man jedoch vorsichtig die Gründe analysieren, die zu der Bewertung des Leistungserbringers geführt haben. Zum Beispiel könnte man bei der Bewertung, dass es sich bei den interessierenden Fällen im Vorjahr vermutlich um „tragische Einzelfälle“ gehandelt haben könnte, durchaus die Evidenz für einen möglichen Qualitätsmangel in den QS-Daten des Vorjahres erneut berücksichtigen, da die Bewertung, ob es sich tatsächlich um Einzelfälle gehandelt hatte wiederum im Licht der neueren QS-Daten überprüft werden müsste.

Oft ist es wünschenswert, bei Leistungserbringern, bei denen im Vorjahr ein Stellungnahmeverfahren stattfand, ein genaues Augenmerk auf das Behandlungsergebnis zu haben, vor allem, wenn im Stellungnahmeverfahren ein Qualitätsdefizit festgestellt wurde. Bei Betrachtung der Vorjahresdaten ist dies in gewisser Weise gegeben: Das Stellungnahmeverfahren im Vorjahr wurde ja unter anderem aufgrund des Hinweises in den Vorjahresdaten auf ein mögliches Qualitätsdefizit gestartet, und dieser Hinweis wird nun auch im aktuellen Jahr berücksichtigt. Entschieden man sich alternativ dazu, die Vorjahresdaten im Falle eines Stellungnahmeverfahrens im Vorjahr aus der Betrachtung auszuschließen, so kann man die Wachsamkeit anderweitig erhöhen, indem man im Falle eines Stellungnahmeverfahrens im Vorjahr die Parameter der Entscheidungsregeln (also z. B.  $\alpha$ ,  $\zeta$  oder  $c$ ) verändert.

So eine Anpassung der Parameter der Entscheidungsregeln an das Ergebnis eines Stellungnahmeverfahrens erfordert jedoch weitere technische Festlegungen. Eine einfachere, zielführendere Alternative ist es, den Wunsch, einen Leistungserbringer im Folgejahr nach einem Stellungnahmeverfahren erneut gezielt zu betrachten, während des Stellungnahmeverfahrens separat festzuhalten. Die erneute Betrachtung könnte dann losgekoppelt von quantitativen Überlegungen ausgelöst werden.

Die verschärfte Betrachtung von Leistungserbringern, bei denen im Vorjahr ein Stellungnahmeverfahren stattfand, erweist sich auch im Widerspruch zur optimalen Strategie im Sinne der Entscheidungstheorie (vgl. Abschnitt 5.4.4). Dies bedeutet nicht, dass so eine verschärfte Betrachtung nicht sinnvoll sein kann, sondern liegt darin begründet, dass diese verschärfte Betrachtung (wo es darum geht, wie Leistungserbringer auf die Ergebnisse des Stellungnahmeverfahrens reagieren) ein anderes Ziel hat als die quantitative Auffälligkeitseinstufung im Allgemeinen (wo es darum geht, unbekannte Qualitätsdefizite zu detektieren). Dies ist ein weiteres Argument dafür, erneute Betrachtungen in Folgejahren nach Stellungnahmeverfahren unabhängig von der allgemeinen quantitativen Einstufungsmethodik zu betrachten.

Im Folgenden wird stets davon ausgegangen, dass es für die quantitative Einstufung unerheblich ist, ob im Vorjahr bereits ein Stellungnahmeverfahren stattgefunden hat.

#### 5.4.5.2 Probleme bei zeitlichen Veränderungen in Daten und Indikatoren

Bei Änderungen in der Spezifikation oder in der Definition von Indikatoren muss man prüfen, ob die Berücksichtigung von Vorjahresergebnissen bei der Bestimmung des Hinweises auf ein Qualitätsdefizit Sinn macht. Die meisten der in diesem Abschnitt vorgeschlagenen Methoden nehmen an, dass sich QI-Ergebnisse auch über Mehrjahreszeiträume berechnen lassen.

Außerdem sollte möglichst ein konstanter Referenzbereich gegeben sein. Ändert sich der Referenzbereich, so muss man überlegen, ob man die Vorjahresdaten auch mit dem neuen Referenzbereich auswerten will. Unter Umständen kann dies durchaus legitim sein, sollte aber begründet werden: Die Idee ist ja, die aktuellen QS-Daten im Lichte der Vorjahresdaten zu bewerten. Die Vorjahresdaten können natürlich Evidenz dafür enthalten, dass der Parameter  $\theta$  außerhalb des neuen Referenzbereichs liegt, auch wenn man dem Leistungserbringer nicht vorwerfen kann, dass der Parameter im letzten Jahr noch nicht im neuen Referenzbereich lag. Schwierig zu interpretieren sind in diesem Kontext allerdings perzentilbasierte Referenzbereiche, die sich jedes Jahr ändern.

#### 5.4.5.3 Perzentilbasierte Referenzwerte

Für sequentielle Verfahren, bei denen Daten von mehr als einem Erfassungsjahr in die Auffälligkeitsklassifikation einfließen, wird vorgeschlagen, den für die Einstufung relevanten Referenzbereich anhand der Daten des aktuellsten Erfassungsjahres – wie oben beschrieben – zu bestimmen und diesen Referenzbereich für die gesamte sequentielle Entscheidungsregel anzuwenden. Dies kann dazu führen, dass unter Umständen Vorjahresergebnisse quantitativ auffällig sind, die zum Referenzbereich des Vorjahres unauffällig waren, und dass in der Regel mehr als die durch das Perzentil festgelegte Anzahl an Leistungserbringerergebnissen auffällig wird. Dies folgt daraus, dass die Vorjahresdaten dazu genutzt werden, stärkere Evidenz dafür zu generieren, ob ein Leistungserbringerergebnis innerhalb des *aktuellen* Referenzbereichs liegt oder nicht.

#### 5.4.5.4 Risikoadjustierung

Die generellen Ideen dieses Abschnitts lassen sich auch auf risikoadjustierte Indikatoren anwenden. Bei der Untersuchung der Auslösewahrscheinlichkeiten ist jedoch das Risikoprofil des betrachteten Leistungserbringers ein weiterer Parameter, der modelliert werden muss.

Viele risikoadjustierte Indikatoren besitzen einen perzentilbasierten Referenzbereich. Außerdem wird oft die Risikoadjustierung regelmäßig überarbeitet, wodurch sich die Rechenregeln ändern. In diesen Fällen sind die Bemerkungen aus dem vorangegangenen Abschnitt 5.4.5.2 zu beachten.

#### 5.4.6 Fazit

In diesem Abschnitt wurden verschiedene Methoden vorgestellt, wie man auf der Basis zweier aufeinanderfolgender Erfassungsjahre zu einer quantitativen Auffälligkeitseinstufung eines Leistungserbringers kommen kann. Die Methoden wurden anhand des Verhaltens der Auslösewahrscheinlichkeit als Funktion von Fallzahl und Kompetenzparameter miteinander verglichen. Unter den vorgestellten Methoden haben die statistisch signifikante Auffälligkeit auf zwei Jahre und die GLR-Binom-Methode die besten Eigenschaften. Die Auslösewahrscheinlichkeit oszilliert schwächer als bei den anderen Methoden als Funktion von der Fallzahl, und Auslösewahrscheinlichkeit steigt stärker mit dem Kompetenzparameter  $\theta$  an. Ein weiterer Vergleich fand im Rahmen eines konkret operationalisierten Entscheidungsdiagramms anhand des erwarteten Verlusts statt, hier konnten die Methoden mit der optimalen Entscheidungsstrategie verglichen werden: wiederum schnitten die GLR-Binom-Methode und die statistisch signifikante Auffälligkeit auf zwei Jahren am besten ab.

Die statistisch signifikante Auffälligkeit über zwei Jahre hat den zusätzlichen Vorteil, dass die verwendeten Konzepte so oder so ähnlich bereits in der externen stationären Qualitätssicherung etabliert sind. Eine Prüfung der statistischen Signifikanz findet beispielsweise bereits im Verfahren plan. QI statt (IQTIG 2016). Dadurch kann man erwarten, dass bei Landesgeschäftsstellen für Qualitätssicherung und Leistungserbringern ein Grundverständnis über die Rolle des Parameters  $\alpha$  vorhanden ist (auch wenn sich die Interpretation im Kontext der erweiterten Datengrundlage ändert, wie im Abschnitt 5.4.2.4 erläutert). Hingegen ist das GLR-Binom-Verfahren relativ technisch, und der Parameter  $c$  ist schwer interpretierbar.

### 5.5 Illustration der Auswertungsmethodik

Die vorgeschlagene statistische Auswertungsmethodik wird im Folgenden beispielhaft anhand eines an den QI 54003 aus dem QS-Verfahren *Hüftendoprothesenversorgung* angelehnten Raten-indikator für einen Leistungserbringer dargestellt.

Zähler, Nenner und Referenzbereich des Qualitätsindikators 54003 sind in Tabelle 13 dargestellt.

*Tabelle 13: Rechenregeln eines Beispielindikators, angelehnt an den Qualitätsindikator 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“*

<b>Zähler</b>	Eingriffe, bei denen die Operation später als 48 Stunden nach der Aufnahme erfolgte
<b>Nenner</b>	Alle Eingriffe zur endoprothetischen Versorgung einer hüftgelenknahen Femurfraktur
<b>Referenzbereich</b>	$\leq 15 \%$

Gegenüber der bisherigen Methode zur Feststellung einer sog. rechnerischen Auffälligkeit soll zukünftig auf Basis der Beobachtungen von „Anzahl Fälle im Zähler“ und „Anzahl Fälle im Nenner“ eine *Wahrscheinlichkeit* dafür ermittelt werden, dass die wahre, nicht direkt beobachtbare Rate eines Leistungserbringers innerhalb des Referenzbereichs liegt.

Damit soll dem Sachverhalt Rechnung getragen werden, dass gleiche rechnerische Ergebnisse „Anzahl Fälle in Zähler durch Anzahl Fälle in Nenner“ von Leistungserbringern mit unterschiedlicher Nenner-Fallzahl nicht gleichbedeutend sind; Ergebnisse mit größerer Nenner-Fallzahl liefern nämlich stärkere statistische Evidenz dafür, dass die zugrunde liegende Rate eines Leistungserbringers im Qualitätsindikator inner- bzw. außerhalb des Referenzbereichs liegt. Beispielsweise überschreiten die Ergebnisse 50/100 und 1/2 beide den Referenzbereich  $\leq 15\%$  um 35 %-Punkte. Die Evidenz für ein Qualitätsdefizit bei einer Fallzahl von 100 ist jedoch größer, als bei einer Fallzahl von 2, bei der das Ergebnis auch allein aufgrund zufälliger Variation zustande gekommen sein kann.

Daher bezieht sich die vorgeschlagene Einstufungsmethodik nicht auf der beobachteten Rate, welche auch aufgrund stochastischer Einflüsse variiert, sondern bezieht sich auf die wahre, zugrunde liegende Rate. Anhand des beobachteten Indikatorergebnisses kann, unter Berücksichtigung stochastischer Einflüsse, die Wahrscheinlichkeit dafür berechnet werden, dass diese zugrunde liegende Rate innerhalb des Referenzbereichs liegt.

Nach der vorgeschlagenen Auswertungsmethodik wird für einen Leistungserbringer im ersten Schritt die Wahrscheinlichkeit dafür berechnet, dass die zugrunde liegende Rate im betrachteten Indikator innerhalb des Referenzbereichs liegt. Illustriert wird dies im Folgenden anhand eines hypothetischen Leistungserbringers und dem oben genannten Qualitätsindikator mit:

$$J = 20; o = 5$$

Bei 5 von insgesamt 20 behandelten Fällen im aktuellen Erfassungsjahr erfolgte demnach die Operation später als 48 Stunden nach Aufnahme. Die beobachtete Rate des Leistungserbringers liegt bei  $o/J = 25\%$ . Um Rückschlüsse auf die zugrunde liegende, nicht direkt beobachtbare Rate des Leistungserbringers zu ziehen, wird auf Basis der beobachteten Daten und mittels des oben beschriebenen bayesianischen Modells die sogenannte A-Posteriori-Wahrscheinlichkeit der zugrundeliegenden Rate  $\theta$  berechnet. Die Verteilung der Wahrscheinlichkeit für den oben betrachteten Fall in Abbildung 30 dargestellt.

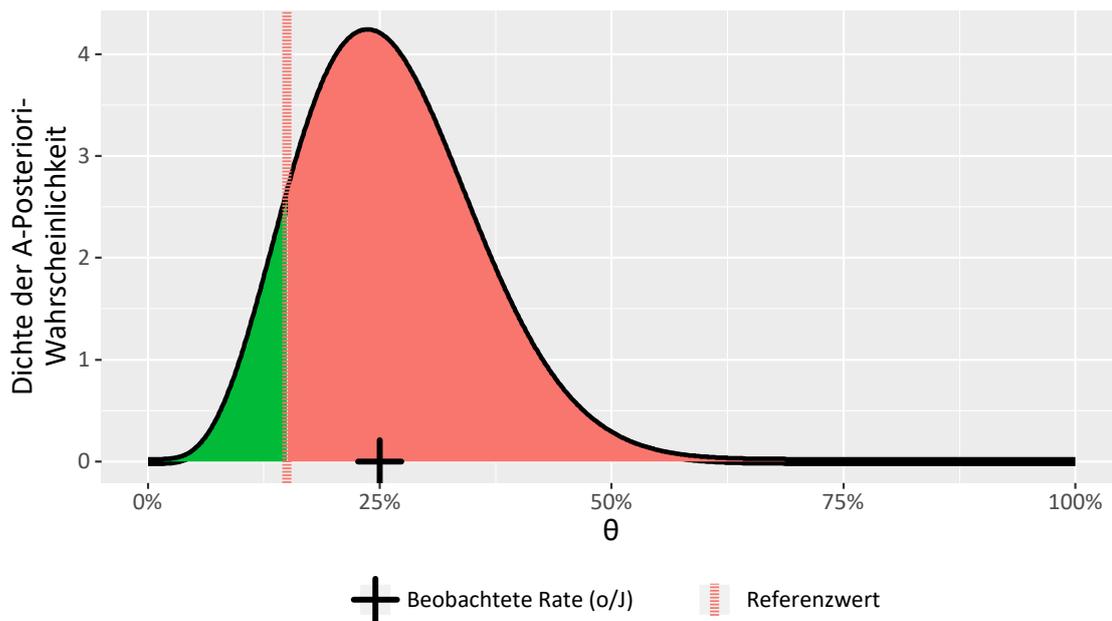


Abbildung 30: Dichte der A-posteriori-Wahrscheinlichkeit für die zugrunde liegende Rate  $\vartheta$  bei einem Leistungserbringer mit  $J = 20$  Fällen im Nenner und  $o = 5$  Fällen im Zähler. Anhand des Referenzwerts von 15 % wird diese Dichte in Dichte für Werte innerhalb und außerhalb des Referenzbereiches unterteilt.

Die grüne Fläche kennzeichnet dabei die Dichte der Verteilung innerhalb des Referenzbereichs, die rote Fläche jene außerhalb des Referenzbereichs. Insgesamt summieren die beiden Flächen zu dem Wert 1. Die A-posteriori-Wahrscheinlichkeit dafür, dass die zugrunde liegende Rate  $\theta$  innerhalb des Referenzbereichs liegt, kann als Fläche der Dichte innerhalb des Referenzbereichs ausgedrückt werden und liegt in diesem Fall bei 11 %. Unterschreitet diese Wahrscheinlichkeit den vorgegebenen Schwellenwert  $\alpha$  (Signifikanzniveau), wird das Leistungserbringerergebnis als quantitativ auffällig eingestuft, weil die Wahrscheinlichkeit dafür, dass die zugrunde liegende Rate des Leistungserbringers im Referenzbereich liegt, unwahrscheinlicher ist als  $\alpha$ . Ist dies nicht der Fall, wird das Indikatorergebnis als quantitativ unauffällig eingestuft.

Sofern die Einstufungsmethodik die Betrachtung von zwei Erfassungsjahren einschließt, wird in einem zweiten Schritt auf Basis der Daten des aktuellen Erfassungsjahres und des Vorjahres die Wahrscheinlichkeit dafür berechnet, dass die zugrunde liegende Rate *innerhalb* des Referenzbereichs liegt, da der Punktschätzer für das aktuelle Erfassungsjahr außerhalb des Referenzbereichs liegt. Unter der Annahme, dass sich die wahre Rate des Leistungserbringers zwischen den Jahren nicht ändert, erlaubt die Vergrößerung der Datenbasis eine präzisere statistische Einschätzung, ob die wahre Rate innerhalb des Referenzbereichs liegt. Wurde für den Leistungserbringer im Vorjahr dasselbe Ergebnis wie im aktuellen Erfassungsjahr, d. h.  $5/20 = 25\%$  beobachtet, liegt die auf Basis der gepoolten Daten, d. h.  $J = 20 + 20 = 40$  und  $o = 5 + 5 = 10$  berechnete Wahrscheinlichkeit dafür, dass die zugrunde liegende Rate im Referenzbereich liegt, bei 0,9 % und ist damit deutlich niedriger, als die nur auf Basis der Daten des aktuellen Erfassungsjahres berechnete Wahrscheinlichkeit. Unterschreitet diese Wahrscheinlichkeit das Signifikanzniveau  $\alpha$ , wird das Leistungserbringerergebnis für den QI als quantitativ auffällig eingestuft. Eine interaktive Version der Abbildung, in der sowohl  $o, J$  als auch der Referenzwert sowie

das Signifikanzniveau verändert werden können, ist unter [https://iqtig.shinyapps.io/illustration\\_einstufungsmethodik/](https://iqtig.shinyapps.io/illustration_einstufungsmethodik/) zu finden.

## 5.6 Zusammenfassung und Empfehlungen

In diesem Kapitel wurde ein Rahmenkonzept für die statistische Auswertungsmethodik von Qualitätsindikatoren präsentiert. Ein solches statistisches Rahmenkonzept existiert bisher nicht für die DeQS-RL, ist jedoch eine notwendige Voraussetzung für eine differenzierte Diskussion über die Auswertung von Qualitätsindikatoren. Ein wichtiger Bestandteil des Rahmenkonzepts ist, dass für jeden Indikator bei der Entwicklung entschieden wird, ob der Indikator nach einer analytischen oder enumerativen Herangehensweise ausgewertet werden soll. Der Großteil der Indikatoren in der DeQS-RL ist aus methodischer Sicht des IQTIG nach der analytischen Herangehensweise auszuwerten. Das bedeutet, dass bei der Berechnung und Bewertung der Leistungserbringerergebnisse dieser Indikatoren statistische Unsicherheit zu berücksichtigen ist, indem zwischen dem beobachteten Indikatorwert des Leistungserbringers und dem zugrunde liegenden Parameter des Leistungserbringers (in Abschnitt 5.1.1 auch als Kompetenzparameter bezeichnet) unterschieden wird. Dies ist ein klarer Unterschied zu den Festlegungen in § 10 QSKH-RL, welche die Berücksichtigung statistischer Unsicherheit beim rechnerischen Teil des Bewertungsprozesses (die rechnerische Auffälligkeitseinstufung) explizit ausschließen. Insbesondere auch für die Effizienzsteigerung des Verfahrens ist jedoch die Berücksichtigung statistischer Unsicherheit bei analytisch auszuwertenden Indikatoren notwendig, wie sowohl empirisch in Abschnitt 2.6 (Aufwandsanalyse/Hintergrund) als auch theoretisch in Abschnitt 5.3 gezeigt wird. Eine zentrale Frage, der in diesem Kapitel nachgegangen wird, ist, wie genau statistische Unsicherheit bei der Bewertung der Leistungserbringerergebnisse Berücksichtigung finden soll.

Abschnitt 5.2 stellt aus statistischer Sicht dar, dass der Bewertungsprozess aus zwei Schritten besteht, die für analytisch auszuwertende Indikatoren im Rahmen eines *Entscheidungsproblems unter Unsicherheit* aufgefasst werden können. Der erste Schritt besteht aus einem rein empirischen Vorgehen, bei dem der Begriff „hinreichender Hinweis für ein Qualitätsdefizit“ aus Kapitel 4 als binäre Klassifikation der Leistungserbringerergebnisse operationalisiert wird. Der zweite Schritt besteht in einer fachlichen Bewertung dieser Klassifikation in Form eines Stellungnahmeverfahrens. Für die Klassifikation im ersten Schritt wird in diesem Kapitel der Überbegriff der „quantitativen Auffälligkeitseinstufung“ eingeführt, der konkrete Vorgehensweisen wie die Einstufung in „rechnerische Auffälligkeit“ nach QSKH-RL oder die Einstufung in „statistische Auffälligkeit“ der plan. QI-RL umfasst. Im Rahmen des Entscheidungsproblems müssen zunächst geeignete Kriterien aufgestellt werden, welche die Klassifikation optimieren soll, um festzulegen, welche konkrete Vorgehensweise für die quantitative Auffälligkeitseinstufung zu wählen ist. Die Festlegung dieser Kriterien entspricht der expliziten Festlegung, wie man – aus statistischer Sicht – zu einer Steigerung der Effizienz des gesamten Bewertungsprozesses kommen kann. Im Kapitel werden zwei Vorgehensweisen unterschieden:

- *statistisch signifikante Auffälligkeitseinstufung*: Betrachtet wird die statistisch signifikante Abweichung des Kompetenzparameters des Leistungserbringers vom Referenzbereich. Ziel ist,

anhand der beobachteten Ergebnisse korrekt zu klassifizieren, ob der zugrunde liegende Kompetenzparameter innerhalb oder außerhalb des Referenzbereiches liegt. Dabei spielen die Größe der Abweichung oder die Konsequenzen der Abweichung des Kompetenzparameters keine Rolle – selbst minimale Abweichungen vom Referenzbereich führen zur quantitativen Auffälligkeit.

- *statistisch relevante Auffälligkeitseinstufung*: Das Ausmaß der Abweichung des Kompetenzparameters des Leistungserbringers vom Referenzbereich wird in Bezug zu der Anzahl an betroffenen Patientinnen und Patienten gesetzt. Somit werden die statistischen Hinweise auf ein Qualitätsdefizit in betroffenen Patientinnen und Patienten gemessen.

Jede dieser beiden Vorgehensweisen wird in Abschnitt 5.3.1 theoretisch als Optimierer einer entsprechenden Verlustfunktion in einem bayesianischen Rahmen hergeleitet. Eine Konsequenz der Berücksichtigung statistischer Unsicherheit ist, dass belastbare Aussagen über Ergebnisse von Leistungserbringern mit sehr kleinen Fallzahlen – unabhängig davon, auf welche Art die Unsicherheit berücksichtigt wird (d. h. bei beiden Vorgehensweisen) – schwierig sind, da selten genug statistische Evidenz generiert wird, um von einem hinreichenden Hinweis auf ein Qualitätsdefizit zu sprechen. Während es bei der statistisch signifikanten Einstufung – je nach Wahl des Referenzwerts und des Signifikanzniveaus – für Leistungserbringer meist auch bei sehr kleinen Fallzahlen möglich ist, quantitativ auffällig zu werden,<sup>53</sup> ist eine quantitative Auffälligkeit bei der statistisch relevanten Einstufung erst ab einer gewissen Fallzahl möglich, da diese Vorgehensweise danach geht, wie viele Patientinnen und Patienten von einem möglichen Qualitätsdefizit betroffen sind. Zwar mag dies richtig im vereinfachten Rahmen einer Aufwand-Nutzen-Abwägung des Bewertungsprozesses sein, ist aber kaum vereinbar mit den Zielen der externen Qualitätssicherung nach § 135a Abs. 1 SGB V und könnte unerwünschte, nicht in der Verlustfunktion abgebildete Nebeneffekte haben. Somit müsste man für diese Gruppe zu anderen Mitteln der Bewertung greifen, wie z. B. Peer-Review-Verfahren für eine Stichprobe der nicht berücksichtigten Leistungserbringer. Dies wiederum führt zu einer neuen strukturellen Heterogenität des Bewertungsverfahrens, weil nicht mehr die gleiche Evidenzgrundlage für die Bewertung der Leistungserbringer gesichert ist.

Untersuchungen im Rahmen einer Sensitivitäts- und Spezifitätsanalyse für die Klassifikation des Kompetenzparameters im Abschnitt 5.3.3 zeigen, dass die statistisch signifikante Auffälligkeitseinstufung bei einem zugrunde liegenden Kompetenzparameter, der dem Referenzwert entspricht (die schwierigste Situation für die Klassifikation), einen konstanten Fehler 1. Art besitzt. Dagegen hat die statistisch relevante Auffälligkeitseinstufung in dieser Situation einen mit der Fallzahl ansteigenden Fehler 1. Art. Das bedeutet, dass die Schwellenwerte für eine quantitative Auffälligkeit für größere Leistungserbringer sehr nahe am Referenzwert liegen können und es, bei gleichem zugrunde liegendem Kompetenzparameter, tendenziell schneller zu einer fachlichen Bewertung kommt, als bei kleinen Leistungserbringern. Liegt der Kompetenzparameter etwas weiter vom Referenzwert entfernt, haben beide Methoden geringe Fehler 1. Art. Die

---

<sup>53</sup> Bei der Wahl von sehr kleinen Signifikanzniveaus kann es aber selbst bei der statistisch signifikanten Auffälligkeitseinstufung passieren, dass ein Leistungserbringer mit einer sehr kleinen Fallzahl nicht auffällig werden kann. Als Beispiel: 1/1 ist nicht auffällig, wenn das Signifikanzniveau  $\alpha = 0,01$  ist.

Power beider Methoden, d. h. ihre Eigenschaft, Abweichungen des zugrunde liegenden Kompetenzparameters vom Referenzbereich zu detektieren, ist stark fallzahlabhängig und unterscheidet sich je nach Größe der Abweichung. Die bisher in der QSKH-RL verwendete „rechnerische Auffälligkeitseinstufung“ kann im analytischen Kontext als eine hoch-sensitive, aber nur sehr niedrig-spezifische Methode angesehen werden, was auch durch die Analysen im Abschnitt 2.6 (Aufwandsanalyse) empirisch belegt wird. Dies führt dann zu einer hohen Menge an Auffälligkeiten, die sich im Stellungnahmeverfahren oft als unbegründet herausstellen.

Jede Entscheidung für eine Methodik zur quantitativen Auffälligkeitseinstufung enthält daher eine Abwägung zwischen Sensitivität und Spezifität der Methode für die Klassifikation des Kompetenzparameters anhand der Konsequenzen. Um diese Abwägung zu systematisieren, ist es notwendig, sowohl Sensitivität als auch Spezifität in einer einheitlichen Metrik zu kombinieren. Hierfür wurde im Abschnitt 5.2 ein Verlustfunktionsansatz im Rahmen des vorliegenden Entscheidungsproblems präsentiert, der in einem vereinfachten Aufwand-Nutzen-Rahmen erlaubt, strukturiert über diese Abwägungen zu diskutieren und Annahmen transparent zu machen. Grundsätzlich fordert der Verlustfunktionsansatz der vorangegangenen Abschnitte eine explizite Festlegung des Aufwandsverhältnisses von Stellungnahmeverfahren gegenüber nicht entdeckten Qualitätsdefiziten. Die Komplexität des Bewertungsprozesses und die vielen Akteure mit unterschiedlichen Aufwandsperspektiven stellen hier eine fast unüberwindbare Herausforderung dar. Trotzdem bietet der entscheidungstheoretische Rahmen mit seinen Vereinfachungen eine Offenlegung der Abwägungen und kann eine Stütze bieten, um z. B. die Schwellenwerte der jeweiligen Einstufungsverfahren zu interpretieren, welche festlegen, ab wann von einem hinreichenden Hinweis auf ein Qualitätsdefizit ausgegangen werden kann. Der vorgestellte entscheidungstheoretische Rahmen erlaubt auch weiterführende Analysen, z. B. zu den Konsequenzen von Fehlklassifikationen<sup>54</sup> bei den interessierenden Ereignissen. Dadurch bietet er prinzipiell auch die Möglichkeit, bei der quantitativen Auffälligkeitseinstufung bereits die fachliche Prüfung im zweiten Schritt des Bewertungsprozesses zu berücksichtigen (Raats und Moors 2003). Diese Überlegungen sind eng mit der Methodik für die Festlegung von Referenzwerten bei Qualitätsindikatoren verknüpft, weil sich u. a. die Frage stellt, ob die jetzigen Referenzwerte bereits Toleranz für besondere Versorgungssituationen enthalten, die nun zusätzlich als stochastische Einflüsse berücksichtigt werden.

Insgesamt zeigen die Analysen, dass die korrekte Klassifikation kleiner Leistungserbringer eine besondere Herausforderung darstellt, die entweder nur über eine Erweiterung der Datenbasis oder mittels nicht statistischer Methoden zu lösen ist. Hier zeigt der in Abschnitt 5.4.3 gebrachte theoriegeleitete Vergleich verschiedener, aus der statistischen Prozesskontrolle inspirierter 2-Jahres-Einstufungsmethoden, dass diese Methoden tatsächlich bessere Klassifikationsergebnisse erreichen, als die 1-Jahres-Einstufungsmethoden.

---

<sup>54</sup> Mit „Fehlklassifikation“ ist hier die folgende Situation gemeint: nach den Rechenregeln des Indikators wird das durch den Indikator betrachtete Ereignis bei einem Fall als unerwünscht kategorisiert, dieses stellt sich jedoch nach der fachlichen Prüfung als besondere Versorgungssituation heraus. Auch die umgekehrte Situation, d. h. ein Ereignis wird fälschlicherweise nicht als unerwünscht kategorisiert, führt zu einer Fehlklassifikation. Prinzipiell können Fehler bei der Datenerhebung des Falls zu Fehlklassifikationen führen.

### Konkrete Vorschläge für die statistische Methodik des Bewertungskonzepts: Auslösung von Stellungnahmen

- Die Kriterien für den Bewertungsprozess in Kapitel 4 legen nahe, dass es bei der Bewertung der Leistungserbringerergebnisse bei einem Qualitätsindikator um eine Klassifikation des Kompetenzparameters des Leistungserbringers im Vergleich zum Referenzbereich geht, welche in der Natur des Vergleichs unabhängig von Fallzahl und der Größe der Abweichung vom Referenzwert ist. Aus diesem Grund wird empfohlen, die Einstufung nach der statistisch signifikanten Auffälligkeitseinstufung vorzunehmen. Beim inferenzstatistischen Vorgehen dieser Auffälligkeitseinstufung manifestiert sich dann Fallzahl und Größe der Abweichung vom Referenzwert in den beobachteten Daten als unterschiedliche Grade von empirischer Evidenz für eine Abweichung vom Referenzwert.
- Da bayesianische Inferenz im Gegensatz zu frequentistischer Inferenz einen deutlich flexibleren Rahmen bildet, um Auffälligkeitseinstufungen auch für komplexe Indikatortypen (wie z. B. bei Patientenbefragungen) vorzunehmen, empfiehlt das IQTIG für die DeQS-RL die bayesianische Version der statistisch signifikanten Auffälligkeit (d. h.  $S_{\text{stat.sig.bayes}}$ , siehe Abschnitt 5.3.1.2) zu verwenden, um einen einheitlichen statistischen Inferenzrahmen für die Auswertungen zu ermöglichen. Dieser Vorschlag entspricht der Empfehlung der Präsidiumskommission der Statistischen Gesellschaften des amerikanischen *Centers for Medicare and Medicaid Services* (CMS) (Ash et al. 2012), hierarchische bayesianische Modelle zu verwenden. Mit der Wahl der Jeffrey-Prior sind die in diesem Kapitel erarbeiteten Vorschläge analytisch kongruent mit der in der plan. QI-RL verwendeten Methode zur Auffälligkeitseinstufung bei Ratenindikatoren und risikoadjustierten Indikatoren. Komplexere Überlegungen zur Risikoadjustierung, Einfluss von Fehlklassifikationen, Unsicherheitsfortpflanzung bei der Verwendung von mehreren Datenquellen<sup>55</sup> können nur kohärent im bayesianischen Rahmen vorgenommen werden. Auch bietet der bayesianische Ansatz, wie in Abschnitt 5.3.1 gezeigt, einen Rahmen, um optimale Entscheidungsstrategien für die Auffälligkeitseinstufungsmethode in diesen komplexen Situationen zu analysieren.
- Für belastbare Aussagen bei Leistungserbringern mit kleinen Fallzahlen ist die Datengrundlage eines Erfassungsjahres häufig nicht ausreichend. Es besteht somit die Gefahr, dass mögliche Qualitätsdefizite nicht detektiert werden können. In diesen Fällen empfiehlt das IQTIG, die Datengrundlage zu erweitern und beispielsweise zwei Erfassungsjahre gemeinsam für die Bewertung heranzuziehen. Konkret empfiehlt das IQTIG die bayesianische Version der statistisch signifikanten Auffälligkeitseinstufung basierend auf den Daten zweier Erfassungsjahre ( $S_{\text{stat.sig.bayes2}}$ , siehe Abschnitt 5.4.2.3), unter der Einschränkung, dass die Methodik zur fachlichen Prüfung und Neuberechnung diese Erweiterung der Datenbasis auf zwei Erfassungsjahre erlaubt oder sie entsprechend erweitert werden kann (vergleiche Empfehlungen in Kapitel 11). Um Erfahrungen mit dem Verfahren zu bekommen und um die in Abschnitt 5.4.5 erwähnten Probleme zu lösen, empfiehlt das IQTIG vor einer flächendeckenden Umsetzung, diese Methodik zunächst in einem direkten QS-Verfahren einzusetzen.

<sup>55</sup> Wie z. B. bei Risikoadjustierungsmodellen, Risikostatistik bzw. Sozialdaten.

- Sowohl  $S_{\text{stat.sig.bayes}}$  als auch  $S_{\text{stat.sig.bayes2}}$  hängen vom Schwellenwert  $\alpha$  ab, welcher als Signifikanzniveau interpretiert werden kann. Dieser Tuning-Parameter stellt zusammen mit dem Referenzwert des Qualitätsindikators die Stellschrauben der Operationalisierung des hinreichenden Hinweises auf ein Qualitätsdefizit im Indikator dar. Diese neue Stellschraube ist eine Konsequenz davon, dass im Rahmen einer analytischen Herangehensweise Unsicherheit bei der Klassifikation der Leistungserbringerergebnisse berücksichtigt wird, und legt fest, ab wann unter Berücksichtigung von Unsicherheit von hinreichender statistischen Evidenz für eine Abweichung vom Referenzbereich ausgegangen werden kann. Somit ist  $\alpha$  sorgfältig und behutsam zu wählen: Möchte man eine einheitliche Grundlage für die Operationalisierung des Begriffs „hinreichender Hinweis auf ein Qualitätsdefizit“ über alle QS-Verfahren haben, dann ist *ein* Wert von  $\alpha$  für alle QS-Verfahren festzulegen. Möchte man dagegen berücksichtigen, dass unterschiedliche Konsequenzen, Ressourcen und Verfahrenszwecke eine Rolle spielen, kann es ggf. eine unterschiedliche Wahl von  $\alpha$  für die quantitative Auffälligkeitseinstufungsmethode geben. Dabei ist zu beachten, dass die 2-Jahres-Einstufung einen leicht höheren Fehler 1. Art hat, als das nominal vorgegebene  $\alpha$ , sodass das Erreichen des nominellen Niveaus nur durch die Verwendung eines  $\alpha^* < \alpha$  geschieht.
- Das IQTIG empfiehlt für die Wahl des Signifikanzniveaus für die statistische signifikante Auffälligkeitseinstufung einen der folgenden Werte, die auch über eine Abwägung zwischen Aufwänden für ein Stellungnahmeverfahren und den Aufwänden für ein nicht entdecktes Qualitätsdefizit hergeleitet werden können, d. h. es besteht die Möglichkeit, die Relevanz und Auswirkungen der im Indikator betrachteten interessierenden Ereignisse mit zu berücksichtigen:
  - Ein nach Aufwand-Nutzen-Überlegungen festgelegtes  $\alpha$ , welches die Methodik zur Wahl des Referenzwerts und quantitative Rückmeldungen zur Treffsicherheit des Qualitätsindikators aus den Stellungnahmeverfahren mit berücksichtigt. Beispielsweise enthalten einige der jetzigen Referenzwerte einen gewissen Puffer für Fehlklassifikationen, die später im Stellungnahmeverfahren geklärt werden sollen. Dieser Puffer kompliziert die Aufwand-Nutzen-Überlegungen. Um daher solche Festlegungen adäquat treffen zu können, sind weitere methodische Arbeiten bzgl. der Wahl der Referenzwerte sowie Korrekturen für Fehlklassifikationen bei den entscheidungstheoretischen Überlegungen notwendig.
  - $\alpha = 0,05$  – dies entspricht der Auffälligkeitseinstufungsmethode der plan. QI-RL und würde zu einer konsistenten Vorgehensweise mit dieser Richtlinie führen.
  - Ein  $\alpha$ -Wert, welcher in etwa die gleiche Anzahl an Stellungnahmen produziert, wie sie zzt. beobachtet wird – vgl. Abschnitt 10.4 für eine exemplarische Betrachtung bei den QS-Verfahren *Hüftendoprothesenversorgung* und *Nierentransplantation*. Für QS-Verfahren, bei denen jede rechnerische Auffälligkeit zu einer Stellungnahme geführt hat, würde dies einem  $\alpha$  von 0,465 (entsprechend der rechnerischen Auffälligkeitseinstufung) bedeuten. Die konkrete Wahl würde somit vom betrachteten QS-Verfahren abhängen oder müsste als Mittelung verschiedener Verfahren vorgenommen werden.
- Perspektivisch sollte geprüft werden, inwieweit es sinnvoll ist, bei der Wahl des  $\alpha$ -Werts zu berücksichtigen, wie oft es bei einem Qualitätsindikator zu Unterschieden zwischen der quantitativen Auffälligkeitseinstufung und der fachlichen Prüfung kommt. Beispielsweise könnte

es sinnvoll sein, die notwendige statistische Evidenz für einen hinreichenden Hinweis auf ein Qualitätsdefizit umso geringer anzusetzen, je größer die Übereinstimmung beider Verfahren ist. Dafür sollen im Rahmen des vorgeschlagenen Stellungnahmeverfahrens Daten dazu gesammelt werden, wie groß diese Übereinstimmung bzw. Abweichung auf QI-Ebene ist. Anschließend sollte geprüft werden, ob die oben dargelegten Vorschläge für die Wahl von  $\alpha$  um diese Komponente auf QI- oder QS-Verfahrenebene erweitert werden sollen (siehe Kapitel 11).

- Obwohl für die quantitative Auffälligkeitseinstufung unterschiedliche Signifikanzniveaus und Datengrundlagen gewählt werden können, sollte das Public Reporting der Ergebnisse eines Erfassungsjahres bei den Unsicherheitsintervallen immer ein konsistentes Signifikanzniveau festlegen. Im Rahmen der Qualitätsberichte der Krankenhäuser wäre dies zzt.  $\alpha = 0,025$ , um damit zweiseitige 95 %-Unsicherheitsintervalle für die Ergebnisse eines Erfassungsjahres zu bekommen.<sup>56</sup>

Eine umsetzungsorientierte Evaluation zu Konsequenzen zur Wahl des Signifikanzniveaus findet sich in Abschnitt 10.3 und daraus abgeleitete Empfehlungen in Abschnitt 11.5.1.

Zusammenfassend ist die Berücksichtigung von Unsicherheit einer der großen Gewinne der vorgestellten statistischen Auswertungsmethodik, weil es ein fachlich fundiertes Konzept darstellt, um das zurzeit verwendete Instrument der Hinweise als Teil des Bewertungsprozesses zu vermeiden. Unsicherheit ist inhärenter Bestandteil des Bewertungsprozesses, jedoch können statistische Methoden helfen, die Konsequenzen von Unsicherheit bei der Bewertung zu standardisieren und zu reduzieren.

---

<sup>56</sup> Siehe in den Regelungen des Gemeinsamen Bundesausschusses gemäß § 136b Absatz 1 Satz 1 Nummer 3 SGB V über Inhalt, Umfang und Datenformat eines strukturierten Qualitätsberichts für nach § 108 SGB V zugelassene Krankenhäuser (Regelungen zum Qualitätsbericht der Krankenhäuser, Qb-R), Anlage 1: Inhalt, Umfang und Datenformat eines strukturierten Qualitätsberichts für das Berichtsjahr 2018, Abschnitt C-1.2.2 Ergebnisse für Qualitätsindikatoren.

## 6 Fachliche Bewertung

Durch die Abgabe von Stellungnahmen erhalten Leistungserbringer, deren Indikatorergebnis einen hinreichenden Hinweis auf ein Qualitätsdefizit geliefert hat, die Möglichkeit, besondere Konstellationen geltend zu machen, die das Verfehlen des Referenzbereichs erklären und von ihnen nicht zu verantworten sind, aber in der indikatorbasierten Bewertung nicht berücksichtigt sind (siehe Abschnitt 3.1). Die fachliche Bewertung einer Stellungnahme hat die Funktion, die in der Stellungnahme angeführten Gründe für das Verfehlen des Referenzbereichs zu prüfen und zu beurteilen, ob tatsächlich eine solche besondere, nicht vom Leistungserbringer zu verantwortende Konstellation vorliegt, die den Hinweis aus dem Indikatorergebnis entkräftet. Auf Grundlage der fachlichen Bewertung wird eine abschließende Einstufung des Indikatorergebnisses vorgenommen.

In den folgenden Abschnitten wird zunächst hergeleitet, in welche Kategorien diese Einstufung erfolgen sollte. Das daraus folgende Bewertungsschema ist in Abschnitt 6.1.4 dargestellt. In den Abschnitten 6.2 bis 6.5 werden Methodik, Ablauf und Rahmenbedingungen der fachlichen Bewertung erläutert, mit denen die Einstufung in das Bewertungsschema vorgenommen werden soll. In Abschnitt 6.6 werden praktische Umsetzungsfragen und der erwartete Aufwand für die Umsetzung der Empfehlungen diskutiert.

### 6.1 Bewertungsschema zur Einstufung der Indikatorergebnisse

Die Ergebnisse der Bewertung im Stellungnahmenverfahren sind der Ausgangspunkt für verschiedene Handlungsanschlüsse. Zum einen werden die Ergebnisse für den Zweck eingesetzt, Rechenschaft über die erreichte Versorgungsqualität abzugeben, im Sinne von *accountability* (siehe Abschnitt 2.1). Zum anderen dienen die Bewertungsergebnisse dem Zweck, die Qualitätsverbesserung der Leistungserbringer anzustoßen, unterstützt durch die Qualitätsförderung externer Stellen.

Das Schema, nach dem die Ergebnisse der fachlichen Bewertung kategorisiert werden, soll diese unterschiedlichen Zwecke und damit einhergehenden unterschiedlichen Adressaten berücksichtigen. Während für den Zweck der *accountability* eine eindeutige, für die Allgemeinheit verständliche Aussage über die erreichte Qualität im Vordergrund steht, ist für den Zweck der Qualitätsförderung eine differenziertere Darstellung anzustreben, die die Auswahl angemessener Fördermaßnahmen erlaubt, indem sie die Gründe für vom Referenzbereich abweichende Indikatorergebnisse benennt, also insbesondere zwischen Datenfehlern und fachlich-inhaltlichen Gründen unterscheidet. Diese Anforderungen an den Detailgrad des Bewertungsschemas sind zunächst gegensätzlich.

Das vom IQTIG im Folgenden vorgeschlagene Bewertungsschema stellt daher einen Kompromiss dar, der die oben genannten Anforderungen berücksichtigt, indem zwei Ebenen verwendet werden. In diesem Schema soll durch die Kategorien auf der ersten Ebene eine klare und allgemeinverständliche Aussage getroffen werden, ob ein Leistungserbringer die an ihn gestellten Anforderungen erfüllt oder nicht. Eine weitere Differenzierung dieser Aussage z. B.

für Public Reporting wird auf dieser Darstellungsebene als nicht sinnvoll beurteilt. So ist es für Patientinnen und Patientinnen wenig hilfreich, zu unterscheiden, ob ein Qualitätsdefizit auf fachliche Probleme, Organisationsprobleme oder Dokumentationsprobleme zurückzuführen ist und mit welcher Methode (Vergleich mit dem Referenzbereich versus fachliche Bewertung) die Aussage getroffen wurde. Daher wird empfohlen, diese Details auf einer zweiten Ebene des Schemas zu dokumentieren.

Außerdem beschränkt sich das Bewertungsschema im Einklang mit der empfohlenen Trennung zwischen Qualitätsbewertung und Qualitätsförderung auf das Ergebnis der abschließenden Qualitätsbewertung<sup>57</sup>. Eine Dokumentation der eingeleiteten Fördermaßnahmen wird zwar ebenfalls empfohlen, soll jedoch separat von der Qualitätsbewertung erfolgen (vgl. Kapitel 7).

### 6.1.1 Kategorien des Bewertungsschemas

Es wird vorgeschlagen, die bisherigen Bezeichnungen „qualitativ auffällig“ und „qualitativ unauffällig“ durch eindeutigeren Begrifflichkeiten zu ersetzen. Gerade das Adjektiv „qualitativ“ birgt ein Risiko für Missverständnisse, da es sich sowohl auf den Prozess der Bewertung als auch auf das Ergebnis beziehen kann: Das Ergebnis des qualitativen (im Gegensatz zum quantitativen) Bewertungsprozesses könnte „unauffällig“ lauten oder die Qualität der Versorgung könnte als Ergebnis des gesamten Bewertungsprozesses (quantitativ und qualitativ) als „unauffällig“ bewertet worden sein. Auch bleibt bei dieser Bezeichnung offen, ob „auffällige“ Qualität positiv oder negativ ist. Eine eindeutige Interpretation der Bewertungsergebnisse ist jedoch eine wichtige Voraussetzung für die angemessene Verwendung der Qualitätsergebnisse z. B. für die öffentliche Berichterstattung oder für Handlungsentscheidungen von Institutionen des Gesundheitswesens (siehe Abschnitt 2.2 und 3.7).

Für die Verwendung im Rahmen von *accountability* soll daher am Ende der fachlichen Bewertung eine der folgenden Bewertungskategorien für das Indikatorergebnis eines Leistungserbringers vergeben werden:

- Kein Hinweis auf Qualitätsdefizit
- Qualitätsdefizit
- Sonstiges

Die Bezeichnung „Qualitätsdefizit“ kennzeichnet dabei allgemeinverständlich das Nicht-Erreichen der Anforderungen, die durch den Referenzbereich des Indikators für erwartbare Qualität repräsentiert werden. Die Bezeichnung „kein Hinweis auf ein Qualitätsdefizit“ soll dementsprechend verwendet werden, wenn nach Abschluss des Stellungnahmeverfahrens kein Anhalt besteht, dass das vom Referenzbereich repräsentierte Soll nicht erreicht wurde.

Aus dem zweischrittigen Vorgehen aus quantitativer Bewertung mittels Qualitätsindikator und Prüfung mittels fachlicher Bewertung können sich für einen Leistungserbringer am Ende des Moduls „Qualitätsbewertung“ die in Abbildung 31 dargestellten Kombinationen von Indikatorergebnis und Ergebnis des Stellungnahmeverfahrens ergeben. Bestätigt sich der

<sup>57</sup> Eine Bewertung der Daten- und Dokumentationsqualität, die im Rahmen von Verfahren zur Datenvalidierung untersucht wurde, sollte anhand eines eigenen Bewertungsschemas erfolgen.

indikatorbasierte hinreichende Hinweis auf ein Qualitätsdefizit im Stellungnahmeverfahren, folgt aus dem Verständnis von Qualitätsindikatoren (vgl. Kapitel 3), dass das mittels des Referenzbereichs für erwartbare Qualität geforderte Soll nicht erreicht wurde. Diese Bestätigung kann entweder durch den Leistungserbringer erfolgen, indem dieser den Verzicht auf eine Stellungnahme erklärt und damit die Gültigkeit der Indikатораussage anerkennt, oder dadurch, dass in der fachlichen Bewertung einer Stellungnahme keine ausreichenden Gründe anerkannt werden, weshalb der Referenzbereich des Indikators ohne Verantwortung des Leistungserbringers verfehlt wurde. Das bestätigte Unterschreiten des Solls in einem Indikator wird mit der Bezeichnung „Qualitätsdefizit“ gekennzeichnet.

Liegt auf Grundlage des Indikatorergebnisses kein hinreichender Hinweis auf ein Qualitätsdefizit vor oder kann dieser Hinweis im Rahmen des Stellungnahmeverfahrens entkräftet werden, soll die Bewertung „Kein Hinweis auf ein Qualitätsdefizit“ erfolgen.

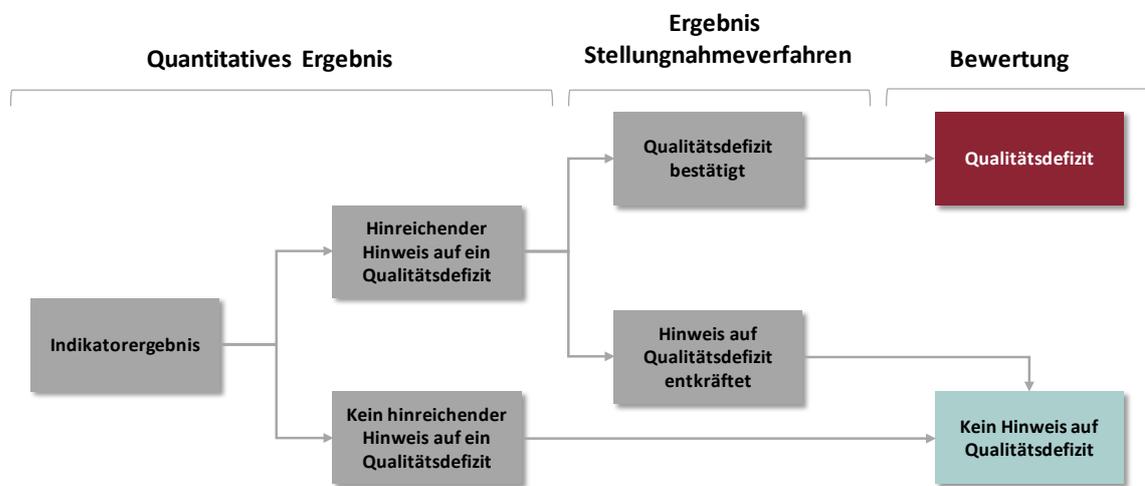


Abbildung 31: Mögliche Kombinationen von quantitativem Indikatorergebnis und Ergebnis des Stellungnahmeverfahrens

Die Kategorie „Sonstiges“ sollte nur in Ausnahmefällen verwendet werden und ist daher immer von der LAG bzw. der Bundesstelle zu begründen, beispielsweise wenn aus organisatorischen Gründen, die nicht vom Leistungserbringer zu verantworten sind, kein Stellungnahmeverfahren durchgeführt werden konnte.

Differenziertere Informationen über die Gründe für die Einstufung in die jeweilige Bewertungskategorie sollen auf der zweiten Darstellungsebene abgebildet werden. Im Folgenden ist erläutert, welche Unterkategorien dazu unterschieden werden sollten.

### 6.1.2 Differenzierung der Kategorie „kein Hinweis auf Qualitätsdefizit“

Zeigt sich in der auf Indikatorergebnis und ggf. auf dem Stellungnahmeverfahren basierenden Bewertung kein Qualitätsdefizit, so ist zunächst keine weitere Differenzierung erforderlich, da sich keine Maßnahmen der Qualitätsförderung anschließen. Eine weitere Differenzierung dieser Kategorie kann allerdings die Transparenz für die Leistungserbringer erhöhen. Diese können anhand einer genaueren Aufschlüsselung des Bewertungsergebnis auf einen Blick erkennen, wie

ihre Stellungnahme bewertet wurde. Ein weiterer Vorteil besteht darin, dass in der leistungserbringer-übergreifenden Betrachtung deutlich wird, ob es beim betreffenden Qualitätsindikator beispielsweise bei vielen Leistungserbringern zur Entkräftung der Indikатораussage gekommen ist und ob dies Datenfehlern oder einer besonderen Versorgungskonstellation geschuldet ist. Diese Informationen können zur zielgerichteten Weiterentwicklung der Indikatoren genutzt werden, indem z. B. häufig auftretende Versorgungskonstellationen in der Berechnungsvorschrift des Indikators berücksichtigt werden. Folgende Unterkategorien zur Differenzierung der Kategorie „kein Hinweis auf ein Qualitätsdefizit“ werden daher vorgeschlagen:

- „kein hinreichender Hinweis auf ein Qualitätsdefizit“: Mit dieser Unterkategorie soll dokumentiert werden, dass die quantitative Bewertung anhand des Qualitätsindikators keinen schlüssigen Hinweis auf ein Qualitätsdefizit geliefert hat, da das Indikatorergebnis des Leistungserbringers nicht mit der geforderten statistischen Sicherheit außerhalb des Referenzbereichs lag<sup>58</sup>. Gleichzeitig wird durch Einstufung in diese Unterkategorie dokumentiert, dass kein Stellungnahmeverfahren durchgeführt wurde.

Wurde ein Stellungnahmeverfahren durchgeführt und in diesem die Indikатораussage entkräftet, sollen die Gründe für die Entkräftung im differenzierten Einstufungsergebnis ablesbar sein (U1 bis U3; siehe Abschnitt 6.4.5):

- „Hinweis auf Qualitätsdefizit entkräftet wegen vom Leistungserbringer nicht zu verantwortender Datenfehler“: Diese Unterkategorie soll anzeigen, dass die Geltendmachung von nicht vom Leistungserbringer zu verantwortenden Datenfehlern<sup>59</sup> für die Entkräftung maßgeblich war.
- „Hinweis auf Qualitätsdefizit entkräftet wegen besonderer Versorgungskonstellation“: Diese Unterkategorie soll dokumentieren, dass die Entkräftung wegen nicht vom Leistungserbringer zu verantwortender, fachlich-inhaltlicher Einflussfaktoren erfolgte.
- „Hinweis auf Qualitätsdefizit entkräftet wegen vom Leistungserbringer nicht zu verantwortender Datenfehler und besonderer Versorgungskonstellation“: Eine Einstufung in diese Unterkategorie soll erfolgen, wenn sowohl Datenfehler als auch fachlich-inhaltliche Einflussfaktoren zur Entkräftung geführt haben. Diese Einstufung soll dabei nicht für jede Konstellation vergeben werden, in der beide Arten von Einflussfaktoren vorgelegen haben, sondern nur für Konstellationen, in denen beide Arten von Einflussfaktoren für die Einstufung (in die Hauptkategorie) *entscheidend* waren. Dies ist dann gegeben, wenn entweder erst die Kombination aus beiden Arten von Gründen zur Entkräftung des Indikatorergebnisses geführt

---

<sup>58</sup> Aus Gründen der besseren Verständlichkeit werden in diesem Kapitel die Bezeichnungen „hinreichender Hinweis auf ein Qualitätsdefizit“ und „Verfehlen des Referenzbereichs“ synonym verwendet. In beiden Fällen ist gemeint, dass das Indikatorergebnis unter Berücksichtigung von stochastischen Einflüssen den Referenzbereich nicht erreicht, d. h. es ist kein einfacher Größer-kleiner-Vergleich des Indikatorwerts mit dem Referenzbereich gemeint. Details zur Berechnung des indikatorbasierten hinreichenden Hinweises auf ein Qualitätsdefizit sind in Kapitel 5 beschrieben.

<sup>59</sup> Im Bericht wird der Begriff „Datenfehler“ statt „Dokumentationsfehler“ verwendet, um deutlich zu machen, dass in seltenen Fällen auch Fehler in den Daten vorliegen können, die nicht vom Leistungserbringer zu verantworten sind, also keine Fehler des Dokumentationsprozesses darstellen (siehe Abschnitt 6.4.1).

hat oder wenn sowohl Datenfehler alleine als auch fachlich-inhaltliche Faktoren alleine für die Entkräftung hinreichend gewesen wären.

### 6.1.3 Differenzierung der Kategorie „Qualitätsdefizit“

Wird ein Indikatorergebnis nach Abschluss des Stellungnahmeverfahrens mit „Qualitätsdefizit“ bewertet, sollen die Gründe für die Bewertung transparent sein. Sowohl für den Leistungserbringer als auch für die Stelle, die das Stellungnahmeverfahren durchführt, kann durch die transparente Darstellung der Gründe für die Bewertung deutlich gemacht werden, wie eine Stellungnahme gewertet wurde und mit welchem Fokus Maßnahmen des internen Qualitätsmanagement bzw. externe Maßnahmen der Qualitätsförderung eingesetzt werden sollten. Zudem können in einer leistungserbringerübergreifenden Betrachtung Hinweise zur Datenqualität und Validität des Qualitätsindikators gewonnen werden (siehe oben). Die Dokumentation, welche der Stellungnahmen wegen Nichteinhaltung der formalen Vorgaben nicht fachlich-inhaltlich bewertet werden konnten, liefert darüber hinaus wichtige Hinweise zur Durchführengsevaluation des Stellungnahmeverfahrens (s. a. Kapitel 10).

#### Differenzierung zwischen Datenfehlern und Defizit in der Versorgungsqualität

In seiner Stellungnahme zum Indikatorergebnis kann ein Leistungserbringer sowohl Datenfehler als auch fachlich-inhaltliche Einflussfaktoren geltend machen. Beide Komponenten können zum Verfehlen des Referenzbereichs des Indikators beigetragen haben. Für die Ableitung geeigneter Verbesserungsmaßnahmen ist es wichtig, die Gründe für das Verfehlen korrekt zu identifizieren. Außerdem sollen bei der fachlichen Bewertung Defizite der Versorgungsqualität nicht durch angegebene oder tatsächliche Datenfehler verdeckt werden, damit Anreize vermieden werden, die Qualitätsbewertung zu umgehen. Für die Differenzierung zwischen Datenfehlern und Defiziten der Versorgungsqualität werden daher folgende Unterkategorien empfohlen:

- Wenn Dokumentationsfehler, d. h. vom Leistungserbringer zu verantwortende Datenfehler, entscheidend für das Verfehlen des Referenzbereichs des Qualitätsindikators waren, es also in der fachlichen Bewertung ohne diese Dokumentationsfehler nicht zu einer Bewertung als „Qualitätsdefizit“ gekommen wäre, soll die Einstufung „Vom Leistungserbringer zu verantwortende Datenfehler haben eine Fehlmessung verursacht“ (Unterkategorie A1) vergeben werden.

Ist das Verfehlen des Referenzbereichs durch den Leistungserbringer nicht allein durch Dokumentationsfehler zu erklären, so ist zumindest ein Teil dieser Abweichung durch ein Defizit der Versorgungsqualität bedingt. In diesem Fall soll das Bewertungsergebnis widerspiegeln, in welchem Ausmaß die beiden Komponenten Dokumentationsfehler bzw. Versorgungsdefizit jeweils *entscheidend* für die Bewertung als „Qualitätsdefizit“ waren:

- Werden in der fachlichen Bewertung sowohl Dokumentationsfehler als auch Defizite der Versorgungsqualität als bedeutsame Gründe identifiziert, soll dies durch die Unterkategorie „Qualitätsdefizit, teilweise durch vom Leistungserbringer zu verantwortende Datenfehler bedingt“ gekennzeichnet werden (Unterkategorie A2). Diese Unterkategorie soll also nicht jede

Konstellation kennzeichnen, bei der sowohl Datenfehler als auch fachlich-inhaltliche Faktoren zum Verfehlen des Referenzbereiches geführt haben<sup>60</sup>, sondern vielmehr solche Konstellationen, bei denen beide Komponenten in relevantem Ausmaß zur Abweichung beigetragen haben.

- Die Unterkategorie „Hinweis auf Qualitätsdefizit im Stellungnahmeverfahren bestätigt oder Verzicht auf Einreichung einer Stellungnahme“ (A3) soll dagegen darstellen, dass für die Bewertung als „Qualitätsdefizit“ fachlich-inhaltliche Gründe führend waren. Dies ist insbesondere dann gegeben, wenn der Leistungserbringer keine Datenfehler geltend gemacht hat oder durch Verzicht auf eine Stellungnahme die Validität der Indikатораussage anerkannt hat.

### **Differenzierung zwischen inhaltlichen und formalen Defiziten**

Ist wegen der Nichteinhaltung formaler Vorgaben durch einen Leistungserbringer keine fachliche-inhaltliche Bewertung einer Stellungnahme möglich, kann die oben erläuterte Differenzierung der Gründe für ein Qualitätsdefizit nicht vorgenommen werden. Für diesen Fall wird im Bewertungsschema die zusätzliche Unterkategorie „Stellungnahme nicht fristgerecht eingereicht oder Stellungnahme entsprach nicht den formalen Anforderungen“ benötigt. Eine Einstufung in diese Unterkategorie sollte aus den in Abschnitt 6.5.1 erläuterten Gründen als Qualitätsdefizit gewertet werden (Unterkategorie A0). Sie bringt zum Ausdruck, dass der indikatorbasierte Hinweis auf ein Qualitätsdefizit nicht entkräftet wurde, da vom Leistungserbringer keine adäquaten Informationen für eine Entkräftung zur Verfügung gestellt wurden.

#### **6.1.4 Bewertungsschema im Überblick**

Tabelle 14 zeigt das vollständige Bewertungsschema im Überblick. Wie bisher sollen auch zukünftig das Bewertungsergebnis auf der oberen Ebene durch Buchstaben gekennzeichnet und die Binnendifferenzierung der Bewertungskategorien durch Ziffern kodiert werden. In Anlehnung an den bisherigen Schlüssel werden für ein Qualitätsdefizit der Buchstabe A, für fehlende Hinweise auf ein Qualitätsdefizit der Buchstabe U und für die Kategorie „Sonstiges“ der Buchstabe S vorgeschlagen. In den Abschnitten 6.2 bis 6.4 dieses Kapitels wird beschrieben, mit welchen Methoden im Rahmen der fachlichen Bewertung eine angemessene Einstufung in diese Kategorien und Ziffern erfolgen kann.

---

<sup>60</sup> Dies würde andernfalls dazu führen, dass z. B. bereits bei Feststellung eines einzelnen Behandlungsfalls mit Datenfehler die Mischkategorie vergeben würden, obwohl der überwiegende Teil der Abweichung vom Qualitätsziel inhaltlich bedingt war.

Tabelle 14: Bewertungsschema

	Kategorie	Einstufung	Ziffer	Erläuterung
Ergebnis der Qualitätsbewertung	U	Kein Hinweis auf Qualitätsdefizit	0	Kein hinreichender Hinweis auf ein Qualitätsdefizit
			1	Hinweis auf Qualitätsdefizit entkräftet wegen vom Leistungserbringer nicht zu verantwortender Datenfehler
			2	Hinweis auf Qualitätsdefizit entkräftet wegen vom Leistungserbringer nicht zu verantwortender Datenfehler und besonderer Versorgungskonstellation
			3	Hinweis auf Qualitätsdefizit entkräftet wegen besonderer Versorgungskonstellation
	A	Qualitätsdefizit	0	Stellungnahme nicht fristgerecht eingereicht oder Stellungnahme entsprach nicht den formalen Anforderungen
			1	Vom Leistungserbringer zu verantwortende Datenfehler haben eine Fehlmessung verursacht
			2	Qualitätsdefizit, teilweise durch vom Leistungserbringer zu verantwortende Datenfehler bedingt
			3	Hinweis auf Qualitätsdefizit im Stellungnahmeverfahren bestätigt <i>oder</i> Verzicht auf Einreichung einer Stellungnahme
keine Qualitätsbewertung	S	Sonstiges – ohne Bewertung	-	<i>[genauer Sachverhalt durch LAG/Bundesstelle anzugeben]</i> <sup>61</sup>

### 6.1.5 Empfehlung für die zukünftige Einstufung der Indikatorergebnisse

Im Bewertungsschema gemäß Tabelle 14 wird zwischen „Qualitätsdefizit“ auf der einen Seite und „kein Hinweis auf Qualitätsdefizit“ auf der anderen Seite unterschieden. In diesen Bezeichnungen spiegelt sich wider, dass die Abwesenheit eines hinreichenden Hinweises auf ein Qualitätsdefizit nicht gleichzusetzen ist mit einem Hinweis auf zureichende Qualität. Aufgrund des Einflusses unbekannter oder nicht gemessener Faktoren auf den beobachteten

<sup>61</sup> Der genaue Sachverhalt wird in Textform verpflichtend im Qualitätsbericht des Krankenhauses veröffentlicht.

Indikatorwert ist dieser mit statistischer Unsicherheit verbunden<sup>62</sup> und kann von dem „wahren“, zugrunde liegenden Indikatorwert des Leistungserbringers abweichen (IQTIG 2019a: 163 f.). Unter Berücksichtigung dieser Unsicherheit ergibt sich somit noch eine dritte Situation. So kann beispielsweise bei sehr kleinen Fallzahlen und gleichzeitig Indikatorergebnissen nahe der Referenzbereichsgrenze die statistische Unsicherheit so groß sein, dass keine Aussage darüber möglich ist, ob der zugrunde liegende Indikatorwert innerhalb oder außerhalb des Referenzbereichs liegt. In dieser Situation kann weder die Aussage getroffen werden, dass ein hinreichender Hinweis auf zureichende Qualität vorliegt, noch dass ein hinreichender Hinweis auf ein Qualitätsdefizit existiert.

Während die Bezeichnung „Qualitätsdefizit“ einen Qualitätsmangel deutlich benennt, bleibt bei der Bezeichnung „kein Hinweis auf Qualitätsdefizit“ also offen, ob zureichende Qualität vorliegt oder nur Unsicherheit über die Qualitätsbewertung besteht. Dies ist nicht nur aus Patientensicht, sondern auch aus Sicht der Leistungserbringer verbesserungsbedürftig. Aber auch vor dem Hintergrund der unterschiedlichen Verwendungszwecke der Qualitätsergebnisse, wie dem G-BA-Qualitätsportal oder für die Versorgungssteuerung, ist aus Sicht des IQTIG eine methodisch fundierte Feststellung zureichender Versorgungsqualität unabdingbar. Im Hinblick auf qualitätsverbessernde Maßnahmen nach dem Leitprinzip „Lernen von den Besseren“ (siehe Kapitel 7) ist eine Identifizierung von Leistungserbringern mit zureichender Qualität und ggf. zukünftig mit besonders guter Qualität ebenfalls wichtig. Daher wird empfohlen, die Methodik dahingehend weiterzuentwickeln, dass hinreichende Sicherheit über das Erreichen des Solls in einem Indikator (z. B. des Referenzbereichs für erwartbare Qualität) mit einer entsprechenden Bewertungskategorie (z. B. „zureichende Qualität“ ) bezeichnet werden kann.

Die in Kapitel 2 empfohlene biometrische Operationalisierung des hinreichenden Hinweises auf ein Qualitätsdefizit berücksichtigt zunächst nur die Differenzierung zwischen einem hinreichenden Hinweis auf ein Qualitätsdefizit und der Abwesenheit eines solchen Hinweises. Daher ist eine Unterscheidung zwischen Leistungserbringern, deren Indikatorergebnis einen Hinweis auf zureichende Qualität aufweist und Leistungserbringern, deren Indikatorergebnis weder einen Hinweis auf zureichende Qualität noch auf ein Qualitätsdefizit aufweist, mit dieser Methodik zunächst nicht möglich. Um Leistungserbringern, die zureichende Qualität erbringen, dies auch attestieren zu können, wird im Sinne des Rahmenkonzepts für die weitere Verwendung der Ergebnisse z. B. im G-BA-Qualitätsportal oder für Qualitätszu- und -abschläge auch die Weiterentwicklung der biometrischen Methodik um diese Komponente empfohlen. Ansätze zu einer solchen Erweiterung der Methodik werden in Kapitel 2 skizziert.

## 6.2 Methodik der fachlichen Bewertung

Die fachliche Bewertung der Stellungnahmen soll gemäß den in Kapitel 3 dargestellten methodischen Grundsätzen nach einem möglichst transparenten und objektiven Verfahren erfolgen. Sie hat die Funktion, die in der Stellungnahme angeführten Gründe für das Verfehlen

---

<sup>62</sup> Wie in Kapitel 5 dargelegt, spielt statistische Unsicherheit bei manchen Indikatoren, wie beispielsweise Strukturindikatoren keine Rolle. Daher kann für Strukturindikatoren die Einhaltung des Referenzbereichs durch einfache Differenzbildung mit der Referenzbereichsgrenze ermittelt und in diesem Fall theoretisch schon jetzt ohne Weiteres zureichende Qualität attestiert werden.

des Referenzbereichs zu prüfen und zu beurteilen, ob eine besondere Konstellation vorliegt, die den Hinweis aus dem Indikatorergebnis entkräftet. Unter einer *besonderen Konstellation* wird in diesem Bericht das Vorliegen eines oder mehrerer Einflussfaktoren auf das Indikatorergebnis bei einem Leistungserbringer verstanden, die nicht von diesem zu verantworten sind und die in ihrer Gesamtheit das Verfehlen des Referenzbereichs bei diesem Qualitätsindikator erklären. Die in einer Stellungnahme genannten Gründe müssen das Vorliegen solcher Einflussfaktoren nachvollziehbar darstellen.

Bei der Beurteilung der Stellungnahmen ist zu berücksichtigen, dass manche Einflussfaktoren nur einen Teil der Behandlungsfälle des Leistungserbringers betreffen oder nur für einen Teil der Behandlungsfälle anerkannt werden können. Darüber hinaus kann die Wirkung eines Einflussfaktors auch bezogen auf einen einzelnen Behandlungsfall unterschiedlich groß sein (siehe Abschnitt 6.4.1). Für die fachliche Bewertung der Stellungnahmen reicht es daher nicht aus, zu bewerten, ob überhaupt ein nicht vom Leistungserbringer zu verantwortender Einflussfaktor bei der Messung vorlag. Es muss vielmehr beurteilt werden, ob das Ausmaß aller in der Stellungnahme zu einem Indikator angeführten und als nicht vom Leistungserbringer zu verantwortend anerkannten Einflussfaktoren in ihrer Gesamtheit groß genug war, um eine besondere Konstellation zu begründen. Dies erfordert eine Zusammenführung und Gewichtung der verschiedenen Einflussfaktoren, damit am Ende des Bewertungsprozesses eine eindeutige Qualitätsaussage je Indikatorergebnis eines Leistungserbringers getroffen werden kann. Gemäß den Ausführungen in Abschnitt 3.4 kann diese Gesamtbeurteilung der Einflussfaktoren und der zugehörigen Begründungen in der Stellungnahme grundsätzlich auf impliziten oder expliziten Regeln basieren.

Eine Beurteilung anhand impliziter Regeln würde bedeuten, dass durch die LAG bzw. die Bundesstelle und die beratenden Expertinnen und Experten durch individuelle Einschätzung abgewogen würde, wie nachvollziehbar die in der Stellungnahme vorgebrachten Gründe sind und wie relevant die dadurch anerkannten Einflussfaktoren für die Bewertung sind. Das Gesamturteil ergäbe sich bei diesem Vorgehen aus der Summe der individuellen Einschätzungen der Bewertenden auf Grundlage ihrer fachlichen Erfahrung und könnte in einem Abstimmungsverfahren ermittelt werden. Dies entspricht größtenteils dem bisherigen Vorgehen. Im Folgenden werden Beurteilungsschritte, die von den Beurteilenden auf Basis solcher impliziter Regeln vorgenommen werden, vereinfachend als „heuristische Beurteilung“ bezeichnet. Ein Vorgehen anhand expliziter Beurteilungsregeln erfordert dagegen eine möglichst eindeutige Beschreibung der Entscheidungssituation einschließlich der Informationsgrundlage sowie eindeutiger Entscheidungsregeln, welche Gründe zu bewerten sind und auf welche Art und Weise die Bewertung erfolgen muss. Um bei einem solchen Vorgehen eindeutige Regeln aufzustellen, nach welchen Kriterien in den Stellungnahmen vorgebrachte Gründe berücksichtigt werden sollen, bedarf es mindestens zum Teil einer expliziten und quantitativen Gewichtung der als Grund anerkannten Einflussfaktoren. Dazu würden für die Einflussfaktoren vorab ein Algorithmus und quantitative Gewichte festgelegt werden, anhand derer eine Neubewertung unter Berücksichtigung dieser Faktoren durchgeführt würde (z. B. in Form einer Nachberechnung des Indikatorwerts für diesen Leistungserbringer und eines erneuten Vergleichs mit einem ggf. angepassten Referenzbereich).

Wie in Abschnitt 3.4 ausgeführt, sind aus methodischer Sicht bei der fachlichen Bewertung explizite Beurteilungsregeln auf Basis standardisierter Informationen zu bevorzugen, da die Erweiterung der Verwendung von Qualitätsmessungen auf Zwecke der *accountability* eine möglichst hohe Objektivität und Reliabilität der Bewertung erfordert. Eine vollständig explizite und quantitative Bewertung der in den Stellungnahmen geltend gemachten Einflussfaktoren ist jedoch aus mehreren Gründen nicht möglich:

- Für die quantitative Neubewertung eines Indikatorergebnisses müsste für alle im Stellungnahmeverfahren vorgebrachten Einflussfaktoren bestimmbar sein, in welchem Ausmaß sie das Indikatorergebnis verändern. Wie in Abschnitt 6.4.1 erläutert, lässt sich jedoch für manche Faktoren nicht zuverlässig angeben, wie sie bei einer Nachberechnung berücksichtigt werden sollten.
- Um beim Stellungnahmeverfahren standardisiert und einheitlich bei den unterschiedlichen LAG angewendet zu werden, müsste die Gewichtung der möglichen Einflussfaktoren bereits vor Beginn des Stellungnahmeverfahrens festgelegt werden. Es handelt sich aber in der Regel um Einflussfaktoren, die in der Rechenvorschrift des Indikators nicht berücksichtigt sind, weil sie zu diesem Zeitpunkt noch nicht bekannt sind oder weil sie zu selten auftreten. In beiden Fällen kann ein Gewicht nicht oder nicht zuverlässig angegeben werden.
- In den Stellungnahmen werden bisher in der Regel nur Angaben zu den Fällen mit einem interessierenden Ereignis (z. B. Komplikationen oder die Nicht-Durchführung eines geforderten Prozesses) gemacht. Um einen Einflussfaktor bei einer Nachberechnung korrekt zu berücksichtigen, müsste die Ausprägung dieses Einflussfaktors für alle Fälle des Leistungserbringers bekannt sein.
- Die Ergebnisse risikoadjustierter Indikatoren basieren nicht nur auf Informationen eines Leistungserbringers, sondern auch auf dem Risikoprofil der Gesamtpopulation. Falls ein Einflussfaktor im Stellungnahmeverfahren als ein zu berücksichtigender Grund anerkannt wird, so müsste daraufhin für eine exakte Nachberechnung das Risikoadjustierungsmodell angepasst werden. Die Berücksichtigung des zusätzlichen Einflussfaktors kann außerdem dazu führen, dass sich auch die Effekte der bereits vorhandenen Einflussgrößen im neuen Risikoadjustierungsmodell ändern. Dies passiert insbesondere, wenn der zusätzliche Einflussfaktor mit den bereits berücksichtigten Risikofaktoren assoziiert ist (beispielsweise weil eine palliative Behandlungsintention bei Patientinnen und Patienten mit dem Vorliegen schwerer Begleiterkrankungen assoziiert ist). Um das Risikoadjustierungsmodell anzupassen, würden jedoch nicht nur für die Leistungserbringer, die im Stellungnahmeverfahren kontaktiert wurden, sondern für alle Leistungserbringer Daten über den neu zu berücksichtigenden Einflussfaktor benötigt.

Eine Einschätzung des Indikatorwertes einer Einrichtung, der vorgelegen hätte, wenn alle nicht vom Leistungserbringer beeinflussbaren Faktoren berücksichtigt worden wären, ist also nur teilweise möglich<sup>63</sup>. Diese Einschätzung ist entsprechend der in Abschnitt 3.1 hergeleiteten

---

<sup>63</sup> Eine exakte Nachberechnung wäre eine wichtige Voraussetzung für die Verwendung bereinigter Indikatorergebnisse in vergleichenden Darstellungen wie beispielsweise qualitätsbezogenen Rangreihenfolgen. Für diesen Zweck ist das Stellungnahmeverfahren allerdings ohnehin nicht geeignet, da es nicht alle Leistungserbringer umfasst. Vergleiche von nachberechneten Indikatorergebnissen der

Funktion von Stellungnahmen allerdings auch nur soweit notwendig, dass entschieden werden kann, ob der Indikatorwert die Erfüllung der Qualitätsanforderungen durch einen Leistungserbringer abbildet oder nicht. Das IQTIG empfiehlt daher ein kombiniertes Vorgehen, bei dem Einflussfaktoren, deren Beitrag zum Indikatorergebnis eindeutig angegeben werden kann, in Form einer partiellen Nachberechnung des Indikatorwertes und die übrigen angeführten Einflussfaktoren durch heuristische Beurteilung mit Unterstützung der Expertinnen und Experten berücksichtigt werden.

### 6.3 Ablauf der fachlichen Bewertung

Abbildung 32 gibt einen Überblick über den empfohlenen Ablauf der fachlichen Bewertung.

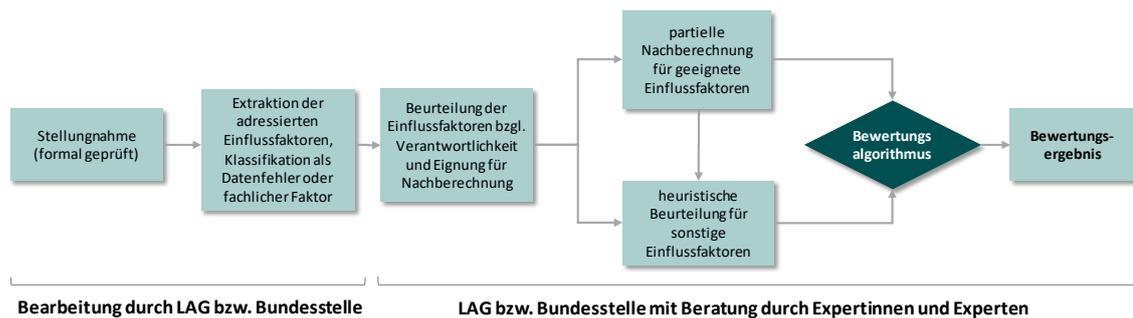


Abbildung 32: Ablauf der fachlichen Bewertung von Stellungnahmen im Überblick

Nach Abschluss der formalen Prüfung (siehe Abschnitt 6.5) werden anhand der in der jeweiligen Stellungnahme angeführten Gründe die adressierten Einflussfaktoren von der LAG bzw. Bundesstelle extrahiert<sup>64</sup> und ihre Art entweder als Datenfehler oder als fachlich-inhaltlicher Einflussfaktor klassifiziert. Im nächsten Schritt wird jeder der Datenfehler oder fachlich-inhaltlichen Einflussfaktoren dahingehend beurteilt, ob es sich um einen vom Leistungserbringer zu verantwortenden Faktor handelt, und ob sein Beitrag zum Indikatorergebnis so eindeutig ist, dass er bei der Indikatorberechnung berücksichtigt werden könnte (siehe Abschnitt 6.4.1).

Im darauffolgenden Schritt wird zum einen auf Grundlage der dafür als geeignet eingeschätzten Einflussfaktoren eine Nachberechnung des Indikatorwertes für den stellungnehmenden Leistungserbringer vorgenommen. Diese wird im folgenden als partielle Nachberechnung bezeichnet, da sie ggf. nicht alle Einflussfaktoren berücksichtigen kann. Zum anderen werden die weiteren, nicht für eine Nachberechnung geeigneten Einflussfaktoren mit Unterstützung der Expertinnen und Experten – und unter Berücksichtigung der Ergebnisse der partiellen Nachberechnung – heuristisch auf Grundlage fachlicher Erfahrung beurteilt. Mittels eines Bewertungsalgorithmus, der der unterschiedlichen Art der in den Stellungnahmen geltend gemachten Einflussfaktoren Rechnung trägt, werden die Ergebnisse aus partieller

Leistungserbringer mit Stellungnahme mit originalen Indikatorergebnissen der Leistungserbringer ohne Stellungnahme sind methodisch nicht angemessen.

<sup>64</sup> Damit ist nicht gemeint, dass jeder klinische Parameter, der möglicherweise in einer Stellungnahme erläutert wird, separat aufgeführt werden soll, sondern nur der Einflussfaktor oder die Konstellation von Einflussfaktoren, die vom Leistungserbringer als besonders angesehen wird (z. B. „Vorliegen einer Kontraindikation“ oder „Multimorbidität“).

Nachberechnung und heuristischer Beurteilung anschließend zu einem Bewertungsergebnis zusammengeführt.

Die Klassifikation der aus den Stellungnahmen extrahierten Einflussfaktoren hinsichtlich ihrer Eignung für eine Nachberechnung und insbesondere hinsichtlich der Verantwortlichkeit des Leistungserbringers erfordert fachliche Expertise. Das gleiche gilt für die Beurteilung der nicht für eine Nachberechnung geeigneten Einflussfaktoren. Die LAG bzw. Bundesstelle wird deshalb von Expertinnen und Experten in Form einer Fachkommission beraten (siehe Abschnitt 6.6). Dadurch soll eine angemessene Berücksichtigung der oft komplexen Sachverhalte in den Stellungnahmen gewährleistet werden. Die partielle Nachberechnung erfolgt also zwangsläufig im Austausch mit der Fachkommission, was eine angemessene technische Unterstützung erfordert (siehe Abschnitt 6.8). Dabei sollen die Leistungserbringer gegenüber den Mitgliedern der Fachkommission pseudonymisiert werden, um Verzerrungen der fachlichen Beurteilung zu vermeiden.

Die LAG bzw. Bundesstelle trifft die abschließende Entscheidung in Bezug auf die Einordnung des Indikatorergebnisses in das Bewertungsschema. Weicht die LAG bzw. Bundesstelle von den Empfehlungen der Fachkommission ab, begründet und dokumentiert sie diese Abweichung. Mit der abschließenden Entscheidung in Bezug auf die Einordnung des Indikatorergebnisses in das Bewertungsschema ist die Qualitätsbewertung abgeschlossen.

## 6.4 Bewertungsalgorithmus

In den folgenden Abschnitten werden die Details des empfohlenen Vorgehens bei der Bewertung von Stellungnahmen und der partiellen Nachberechnung von Indikatorwerten erläutert. Abbildung 33 zeigt den grundlegenden Bewertungsalgorithmus im Überblick. Ziel des Algorithmus ist, zu einer möglichst objektiven und einheitlichen Einstufung in die Kategorien des in Abschnitt 6.1 entwickelten Bewertungsschemas zu kommen (siehe Tabelle 14). Er ist daher so angelegt, dass die Beurteilungsschritte so weit wie möglich auf expliziten, quantitativen Regeln basieren. Dies ist insbesondere für die unter dem Gesichtspunkt der *accountability* maßgeblichen Ergebnisse auf erster Ebene des Bewertungsschemas, „Qualitätsdefizit“ versus „Hinweis auf Qualitätsdefizit entkräftet“, von Bedeutung. Verbleibende Unsicherheit wird in nachfolgenden, qualitativen Beurteilungsschritten adressiert.

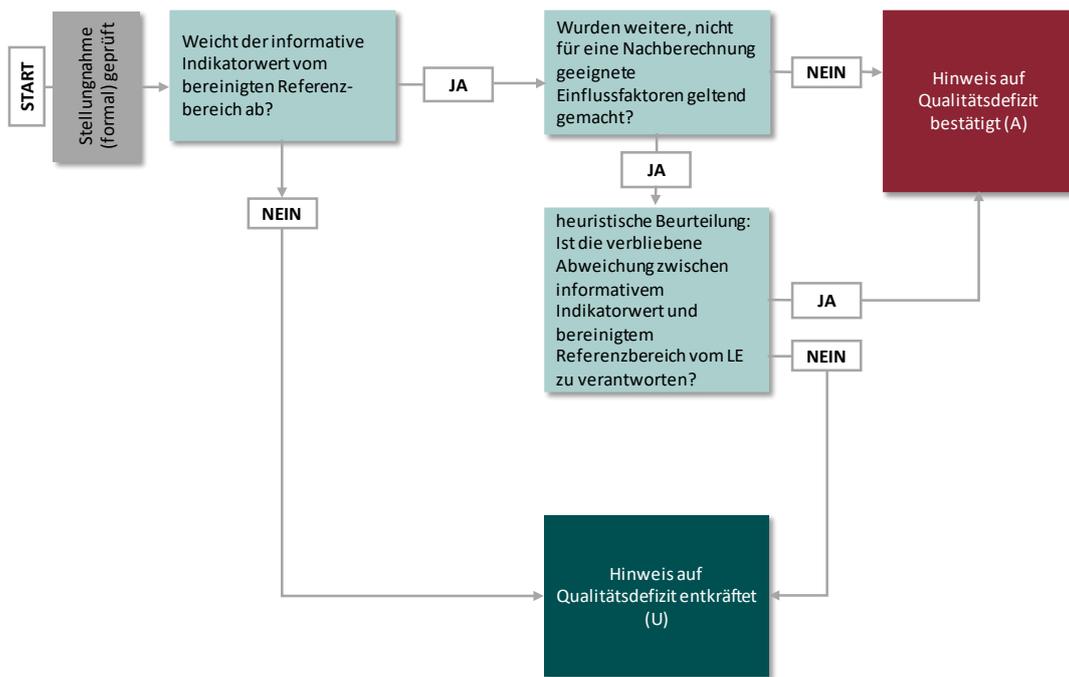


Abbildung 33: Grundlegender Algorithmus der fachlichen Bewertung

Im ersten Schritt wird eine (partielle) Nachberechnung des Indikatorwertes für den Leistungserbringer und ein Vergleich mit einem entsprechend modifizierten Referenzbereich vorgenommen. Dazu wird ein Indikatorwert berechnet, der um diejenigen nicht vom Leistungserbringer zu verantwortenden Datenfehler und fachlich-inhaltlichen Einflussfaktoren bereinigt ist, deren Auswirkung auf das Indikatorendergebnis eindeutig benannt werden kann. Diese Bereinigung erfolgt z. B. dadurch, dass Behandlungsfälle, die vom Indikator nicht hätten adressiert werden sollen, aus der Grundgesamtheit des Indikators ausgeschlossen werden (Details siehe Abschnitt 6.4.1). Der so bereinigte Wert wird im Folgenden als *informativer Indikatorwert* bezeichnet und dient ausschließlich als Information im Rahmen der fachlichen Bewertung; er hat nicht die Funktion, den originalen Indikatorwert<sup>65</sup> der Einrichtung zu ersetzen.

Desweiteren wird aus dem originalen Referenzbereich<sup>66</sup> des Indikators ein modifizierter Referenzbereich abgeleitet. Dieser ist – wie der informative Indikatorwert – um den Anteil bereinigt, der schon im originalen Referenzbereich für die gerichteten, nicht vom Leistungserbringer zu verantwortenden Einflussfaktoren einkalkuliert ist (siehe Abschnitt 6.4.2). Dieser Referenzbereich kann als der Referenzbereich verstanden werden, der festgelegt worden wäre, wenn die Berechnungsvorschrift des bereinigten Indikators von vorneherein die gerichteten Einflüsse berücksichtigt hätte, und wird im Folgenden als *bereinigter Referenzbereich* bezeichnet.

<sup>65</sup> Unter „originaler Indikatorwert“ wird in diesem Bericht der Indikatorwert verstanden, der anhand der gültigen Rechenregeln des Indikators im Regelbetriebs berechnet und in den Auswertungen der Leistungserbringer dargestellt wird.

<sup>66</sup> Mit „originaler Referenzbereich“ wird in diesem Bericht der vom G-BA beschlossene Referenzbereich eines Indikators bezeichnet, der bei der Erstellung der Auswertungen im Regelbetrieb angewendet wird.

Der informative Indikatorwert wird mit dem bereinigten Referenzbereich unter Anwendung der üblichen Vergleichsregel (statistisches Verfahren) verglichen. Wenn bei diesem Vergleich der informative Indikatorwert den bereinigten Referenzbereich nicht verfehlt, ist die ursprüngliche Abweichung (des originalen Indikatorwerts vom originalen Referenzbereich) hinreichend durch die besondere Konstellation erklärt, die vom Leistungserbringer geltend gemacht wurde. In diesem Fall lautet die Bewertung des Indikatorergebnisses „Hinweis auf Qualitätsdefizit entkräftet“. Soll genauer gekennzeichnet werden, ob Datenfehler, fachlich-inhaltliche Einflussfaktoren oder eine Kombination aus beiden Arten von Faktoren für das ursprüngliche Verfehlen des Referenzbereichs ursächlich waren, kann anschließend eine Binnendifferenzierung dieser Unterkategorien auf Grundlage weiterer Berechnungen erfolgen (siehe Abschnitt 6.4.5).

Erreicht dagegen der informative Indikatorwert den bereinigten Referenzbereich nicht, ist zu prüfen, ob vom Leistungserbringer zusätzliche Einflüsse geltend gemacht wurden, die in der partiellen Nachberechnung nicht berücksichtigt werden konnten. Ist dies nicht der Fall, liegen in der Stellungnahme keine Belege vor, die das ursprüngliche Indikatorergebnis entkräften. Die Qualitätsanforderung des Indikators gilt damit als nicht erfüllt und die Einstufung des Indikatorergebnisses lautet „Qualitätsdefizit“.

Wurden dagegen in der Stellungnahme des Leistungserbringers über die in der Nachberechnung berücksichtigten Faktoren hinaus zusätzliche Einflüsse auf den Indikatorwert angeführt, werden diese durch die Expertinnen und Experten der Fachkommission qualitativ beurteilt. Bei dieser Beurteilung wird – ausgehend vom informativen Indikatorwert und vom bereinigten Referenzbereich – eingeschätzt, ob das Ausmaß dieser verbliebenen, zusätzlichen Einflüsse das Verfehlen des bereinigten Referenzbereiches trotz partieller Nachberechnung erklärt. Sofern Datenfehler nicht vom Leistungserbringer zu verantworten sind und nicht bereits bei der partiellen Nachberechnung berücksichtigt werden konnten, sollen sie ebenfalls in diese Bewertung mit einbezogen werden. Die zu beurteilende Frage lautet:<sup>67</sup> *„Ist es schlüssig, dass die verbliebene Abweichung des Indikatorwerts des Leistungserbringers vom Referenzbereich nach partieller Nachberechnung vollständig durch die restlichen, nicht vom Leistungserbringer zu verantwortenden Faktoren und Datenfehler erklärt ist?“*. In Abhängigkeit von der Beantwortung dieser Frage erfolgt dann die Einstufung als „Hinweis auf Qualitätsdefizit entkräftet“ (wenn die Abweichung schlüssig nicht vom Leistungserbringer zu verantworten ist) bzw. als „Qualitätsdefizit“.

Um bei einer Einstufung als „Qualitätsdefizit“ genauer zwischen Datenfehlern und einem Defizit der Versorgungsqualität zu differenzieren, kann anschließend eine Binnendifferenzierung dieser Unterkategorien erfolgen (siehe Abschnitt 6.4.4).

Die Details dieses Bewertungsalgorithmus werden in den folgenden Abschnitten ausführlich erläutert.

---

<sup>67</sup> Siehe Abschnitt 6.4.3 für methodische Limitationen bei verteilungsbezogenen Referenzbereichen.

### 6.4.1 Analyse der in den Stellungnahmen genannten Gründe

Bevor der informative Indikatorwert im ersten Beurteilungsschritt berechnet werden kann, muss zunächst anhand der in einer Stellungnahme angeführten Gründe für das Verfehlen des Referenzbereichs analysiert werden, welche Einflussfaktoren auf das Indikatorergebnis vom Leistungserbringer geltend gemacht werden. Beispielsweise können sich die angeführten Gründe auf ein besonderes Risikoprofil der Patientinnen und Patienten, auf Notfallsituationen oder auf Datenfehler beziehen. Für jeden dieser Einflussfaktoren sind folgende Fragen zu klären:

- Handelt es sich bei dem Einflussfaktor um einen Fehler in den QS-Daten (Datenfehler) oder um einen fachlich-inhaltlichen Einflussfaktor (besondere Versorgungskonstellation)?
- Ist der Datenfehler oder der fachlich-inhaltliche Einflussfaktor vom Leistungserbringer zu verantworten oder hatte der Leistungserbringer keine Möglichkeit, diesen Faktor zu beeinflussen?
- Ist die Auswirkung des Datenfehlers oder des fachlich-inhaltlichen Einflussfaktors auf das Indikatorergebnis so eindeutig zu beziffern, dass dieser Einflussfaktor bei einer Nachberechnung berücksichtigt werden kann?

#### Datenfehler vs. fachlich-inhaltlicher Einflussfaktor

Neben fachlich-inhaltlichen Einflussfaktoren können auch Fehler in den übermittelten Daten dazu führen, dass ein Indikator die Versorgungsqualität eines Leistungserbringers nicht angemessen abbildet. Beispielsweise können Behandlungsfälle in die Indikatorberechnung für einen Leistungserbringer eingegangen sein, die bei korrekten Daten nicht hätten berücksichtigt werden dürfen. Um eine nachvollziehbare Einstufung des Indikatorergebnisses zu erreichen, muss daher die Beurteilung von in den Stellungnahmen angeführten Datenfehlern mit der Beurteilung angeführter fachlich-inhaltlicher Einflussfaktoren verknüpft werden. Die Klassifizierung, bei welchen der in der Stellungnahme eines Leistungserbringers angeführten Einflussfaktoren es sich um Datenfehler handelt, nimmt die LAG bzw. Bundesstelle vor.

Die fachliche Bewertung bezieht sich prinzipiell auf das Indikatorergebnis einer Einrichtung, das auf Grundlage der regulär dokumentierten Fälle berechnet wurde. Eine Nachdokumentation von Behandlungsfällen, die fälschlich nicht dokumentiert wurden, ist im Rahmen des Stellungnahmeverfahrens nicht vorgesehen. Solche zusätzlichen Behandlungsfälle können bei der Bewertung der Stellungnahmen nicht berücksichtigt und nicht als besondere Konstellation anerkannt werden, da sich ihre Auswirkung auf den Indikatorwert einer Einrichtung nicht angemessen beurteilen lässt.

#### Beurteilung der Verantwortlichkeit für einen Einflussfaktor

Sowohl bei Datenfehlern als auch bei fachlich-inhaltlichen Einflussfaktoren ist zwischen solchen Faktoren zu unterscheiden, die vom Leistungserbringer nicht zu verantworten sind, und solchen, die der Leistungserbringer zu verantworten hat (Beispiele siehe Anhang, Kapitel 5). Die Beurteilung erfolgt auf Basis der in den Stellungnahmen vorgebrachten Gründe und dazugehörigen Belege.

Die Beurteilung, welche der in der Stellungnahme eines Leistungserbringers angeführten Datenfehler vom Leistungserbringer zu verantworten sind, trifft die LAG bzw. Bundesstelle. Die Beurteilung, ob ein fachlich-inhaltlicher Einflussfaktor vom Leistungserbringer zu verantworten ist, erfolgt grundsätzlich unter Beratung der LAG bzw. Bundesstelle durch die Expertinnen und Experten der Fachkommission. In diese Beurteilung soll auch einfließen, ob vom Leistungserbringer mit der gebotenen Sorgfaltspflicht alles Zumutbare getan wurde, um den aus seiner Sicht nicht von ihm zu verantwortenden Einflussfaktor aufzufangen (z. B. durch die Einplanung von genügend Ressourcen und genügend Zeit).

Bei manchen Einflussfaktoren kann es erforderlich sein, die Verantwortlichkeit des Leistungserbringers je nach Konstellation des Behandlungsfalls unterschiedlich einzuschätzen. Beispielsweise könnte eine Wiederbelebensmaßnahme bei einem Patienten oder einer Patientin als vom Leistungserbringer nicht zu verantwortender Faktor eingeschätzt werden, wenn sie vor Eintreffen in der Einrichtung erfolgt ist, aber nicht, wenn sie erst im weiteren Behandlungsverlauf notwendig wurde. Für solche Einflussfaktoren muss entschieden werden, ob die Verantwortlichkeit für jeden Behandlungsfall eindeutig als gegeben oder nicht gegeben klassifiziert werden kann und sie sich für eine partielle Nachberechnung auf Grundlage der anerkannten Behandlungsfälle eignen (siehe unten), oder ob die Auswirkung auf das Indikatorergebnis für diesen Einflussfaktor nicht zuverlässig angegeben werden kann.

### **Auswirkung auf das Indikatorergebnis**

Ein wichtiges Merkmal der in den Stellungnahmen adressierten Einflussfaktoren ist das Ausmaß, in dem sie das Indikatorergebnis beeinflussen. Damit ist gemeint, dass sich für manche Einflussfaktoren zuverlässig angeben lässt, wie sie bei einer Nachberechnung des Indikatorergebnisses berücksichtigt werden sollten, während dies für andere Einflussfaktoren nicht möglich ist. Wie groß die Auswirkung eines geltend gemachten Einflussfaktors auf das Indikatorergebnis ist, hängt unmittelbar von seiner Beeinflussbarkeit durch den Leistungserbringer ab (siehe IQTIG 2019a). Die Beurteilung, ob ein Einflussfaktor durch partielle Nachberechnung berücksichtigt werden kann, erfolgt daher zusammen mit der Beurteilung der Verantwortlichkeit durch die LAG bzw. Bundesstelle unter Beratung durch die Expertinnen und Experten der Fachkommission.

Beschreibt ein in der Stellungnahme angeführter Grund einen Einflussfaktor, bei dessen Vorliegen die betreffenden Behandlungsfälle nicht vom Qualitätsindikator hätten adressiert werden sollen, d. h. nicht zur Zielpopulation gehören, so kann das Ausmaß auf den Indikatorwert eindeutig angegeben werden. Beispielsweise soll ein Indikator zur Sterblichkeit durch Pneumonie die Behandlungskompetenz der Leistungserbringer in Bezug auf die Verhinderung krankheitsbedingter Todesfälle messen. Das Qualitätsmerkmal bezieht sich auf Patientinnen und Patienten, die mit dem Ziel der Heilung behandelt werden und nicht auf solche, die mit palliativer Intention behandelt werden<sup>68</sup>. Für den Einflussfaktor palliative Behandlungsintention lässt sich

---

<sup>68</sup> Im Leistungsbereich „Ambulant erworbene Pneumonie“ wird zwar über ein Datenfeld erhoben, ob mit palliativer Intention behandelt wird. Derzeit wird dieses Datenfeld jedoch nicht als valide genug erachtet, um Fälle mit Therapieverzicht aus der Grundgesamtheit des Indikators zur Sterblichkeit auszuschließen. In diesem Fall erweist es sich schwierig, die Fälle, die man ausschließen möchte, durch ein Datenfeld zu identifizieren. Dieser Einflussfaktor muss daher im Stellungnahmeverfahren berücksichtigt werden.

bei diesem Indikator also klar entscheiden, wie er bei der Indikatorberechnung berücksichtigt werden sollte. Die partielle Nachberechnung erfolgt bei solchen Einflussfaktoren dadurch, dass die Behandlungsfälle, bei denen das Vorliegen des Einflussfaktors anerkannt wird, aus der Grundgesamtheit (Nenner) des Indikators ausgeschlossen werden (siehe Anhang, Kapitel 6). Bei den für eine Nachberechnung geeigneten Einflussfaktoren handelt es sich oft um besondere Patientencharakteristika wie Kontraindikationen für eine Behandlung oder das Auftreten unerwarteter Notfallsituationen. Auch andere Einflussfaktoren, bei denen keine Verantwortlichkeit des Leistungserbringers für diesen Faktor anzunehmen ist (z. B. ein extern bedingter Stromausfall), eignen sich für eine Berücksichtigung im Rahmen der partiellen Nachberechnung.

Andere Einflussfaktoren erhöhen hingegen lediglich die Wahrscheinlichkeit für ein ungünstigeres Indikatorergebnis. Ein Beispiel hierfür sind Begleiterkrankungen der Patientinnen und Patienten, die bisher in der Risikoadjustierung des Indikators nicht oder nicht adäquat berücksichtigt ist. So könnte etwa das Vorliegen von Übergewicht durch die Begünstigung lokaler Komplikationen bei Operationswunden das Ergebnis eines Qualitätsindikators beeinflussen, der Komplikationshäufigkeiten einer Behandlung misst. Werden Informationen zu solchen Einflussfaktoren nicht erfasst und im Risikoadjustierungsmodell des Indikators nicht berücksichtigt, lässt sich ihr Einfluss auf den Indikatorwert eines Leistungserbringers nicht eindeutig benennen. Kommen solche Faktoren bei einem Leistungserbringer häufiger vor als in der Gesamtpopulation der Leistungserbringer, so können sie bei der fachlichen Bewertung nicht im Rahmen der partiellen Nachberechnung, sondern nur im qualitativen Beurteilungsschritt berücksichtigt werden.

Je nach betroffenem Datenfeld können auch Datenfehler unterschiedlich gut durch eine Nachberechnung adressiert werden. Wurden durch Datenfehler Behandlungsfälle fälschlich in die Indikatorberechnung einbezogen, so lässt sich deren Auswirkung durch Ausschluss der Fälle aus der Grundgesamtheit des Indikators ermitteln. Hat ein Datenfehler dazu geführt, dass das interessierende Ereignis (z. B. Komplikationen oder die Durchführung eines Prozesses) fälschlich als gegeben bzw. als nicht gegeben gewertet wurde, so kann das Ausmaß des Datenfehlers auf den Indikatorwert ebenfalls einfach angegeben werden. Die Nachberechnung besteht dann darin, den Behandlungsfall als Behandlungsfall ohne interessierendes Ereignis zu werten, also ihn z. B. aus dem Zähler des Indikators auszuschließen (siehe Anhang, Kapitel 6). Für Datenfehler muss also entschieden werden, ob sie durch Ausschluss der Behandlungsfälle aus der Grundgesamtheit oder nur aus dem Zähler korrigiert werden können. Geht ein fehlerhafter Wert dagegen eher indirekt in das Indikatorergebnis ein, beispielsweise wenn Werte für Risikofaktoren in einer falschen Einheit dokumentiert wurden, so könnte die Auswirkung des Datenfehlers auf den Indikatorwert nur mit hohem Aufwand ermittelt werden.

Weitere Beispiele für die Beurteilung von Einflussfaktoren sind im Anhang, Kapitel 5 dargestellt. Eine abschließende Auflistung und Kategorisierung aller denkbaren Datenfehler und fachlich-inhaltlichen Einflussfaktoren mit Angabe von Verantwortlichkeit des Leistungserbringers und Eignung für eine partielle Nachberechnung ist nicht möglich, da immer wieder neue besondere Konstellationen auftreten können. Die Beurteilungsergebnisse von in früheren Stellungnahmen angegebenen Gründen können daher zwar als Orientierung genutzt werden, um eine

Einheitlichkeit der Beurteilung zu fördern, die abschließende Beurteilung der Gründe und Einflussfaktoren muss aber jeweils auf Grundlage der jeweils aktuellen Stellungnahme erfolgen.

Der informative Indikatorwert, der die Grundlage für den quantitativen Teil der fachlichen Bewertung ist, wird ausgehend vom originalen Indikatorwert berechnet, indem die auszuschließenden Behandlungsfälle und die auszuschließenden interessierenden Ereignisse berücksichtigt werden. Details zur Berechnung sind in Anhang, Kapitel 6 dargestellt. Alle anderen Einflussfaktoren, die sich nicht diesen beiden Kategorien zuordnen lassen, gehen nicht in die partielle Nachberechnung ein, sondern werden durch die LAG bzw. Bundesstelle und die Expertinnen und Experten der Fachkommission heuristisch beurteilt.

#### **6.4.2 Festlegung von Referenzbereichen für die partielle Nachberechnung**

Für die quantitative Beurteilung des Indikatorwerts eines Leistungserbringers, der um gerichtete, nicht vom Leistungserbringer zu verantwortende Einflüsse bereinigt ist, wird ein modifizierter Referenzbereich benötigt. Dies ist deshalb erforderlich, weil die Referenzbereiche der externen Qualitätssicherung eine Abweichung vom Idealwert erlauben, die aus zwei Anteilen besteht:

1. Zum einen sind gerichtete Einflüsse auf den Indikatorwert einkalkuliert, die nicht erfasst werden und die nicht vom Leistungserbringer zu verantworten sind (siehe Abschnitt 3.2.1). Beispielsweise könnte ein Ergebnisindikator die Komplikationshäufigkeit nach einer Operation messen, die im Idealfall 0 % betrage. Aufgrund systematischer, nicht erfasster Einflüsse wie pathophysiologischer Gesetzmäßigkeiten oder der Charakteristika der notwendigen Operationen sei angenommen, dass bei dem betreffenden Krankheitsbild und Operationsverfahren eine mittlere Komplikationshäufigkeit von 5 % selbst bei bester Versorgungsqualität zu erwarten ist. Würden die zugrunde liegenden Indikatorwerte nur um den Betrag dieser Einflussfaktoren (im Beispiel 5 %) vom Idealwert abweichen, läge optimale Qualität im Sinn des Qualitätsziels des jeweiligen Indikators vor.
2. Zum anderen berücksichtigen die Referenzbereiche, dass nicht jeder Leistungserbringer optimale Qualität erreichen kann. Das Erreichen des Referenzbereichs repräsentiert also nicht optimale Qualität, sondern die Erfüllung erwartbarer Standards (zureichende Qualität), die von jedem Leistungserbringer verlangt werden kann. Es wird damit anerkannt, dass es oberhalb der Schwelle zu zureichender Qualität weitere Qualitätsunterschiede zwischen den Leistungserbringern geben kann (z. B. besonders gute Erfüllung der Anforderungen). Würden im obigen Indikatorbeispiel 2 Prozentpunkte zusätzliche Abweichung toleriert, läge die Referenzbereichsgrenze bei  $5\% + 2\% = 7\%$ .

Für die Bewertung nach partieller Nachberechnung sind diese beiden Anteile des Referenzbereichs von großer Bedeutung: Es wird angenommen, dass durch die Stellungnahme gerichtete Einflussfaktoren, die zum Zeitpunkt der Indikatorberechnung noch nicht berücksichtigt waren, bekannt werden. Wird nun eine partielle Nachberechnung vorgenommen und würde der informative Indikatorwert mit dem originalen Referenzbereich verglichen werden, würden die gerichteten Einflussfaktoren auf den Indikatorwert doppelt berücksichtigt werden: Einmal im Referenzbereich, bei dessen Setzung sie einkalkuliert wurden, und einmal

durch den Vorgang der partiellen Nachberechnung des Indikatorwerts. Für einen angemessenen Vergleich mit dem informativen Indikatorwert muss der Referenzbereich also um den Anteil gerichteter Einflussfaktoren verringert werden (siehe Abbildung 34). Der Anteil des Referenzbereiches, der eine Toleranz für zulässige Qualitätsunterschiede ausdrückt (Nr. 2), bleibt dagegen auch nach partieller Nachberechnung erhalten.<sup>69</sup>

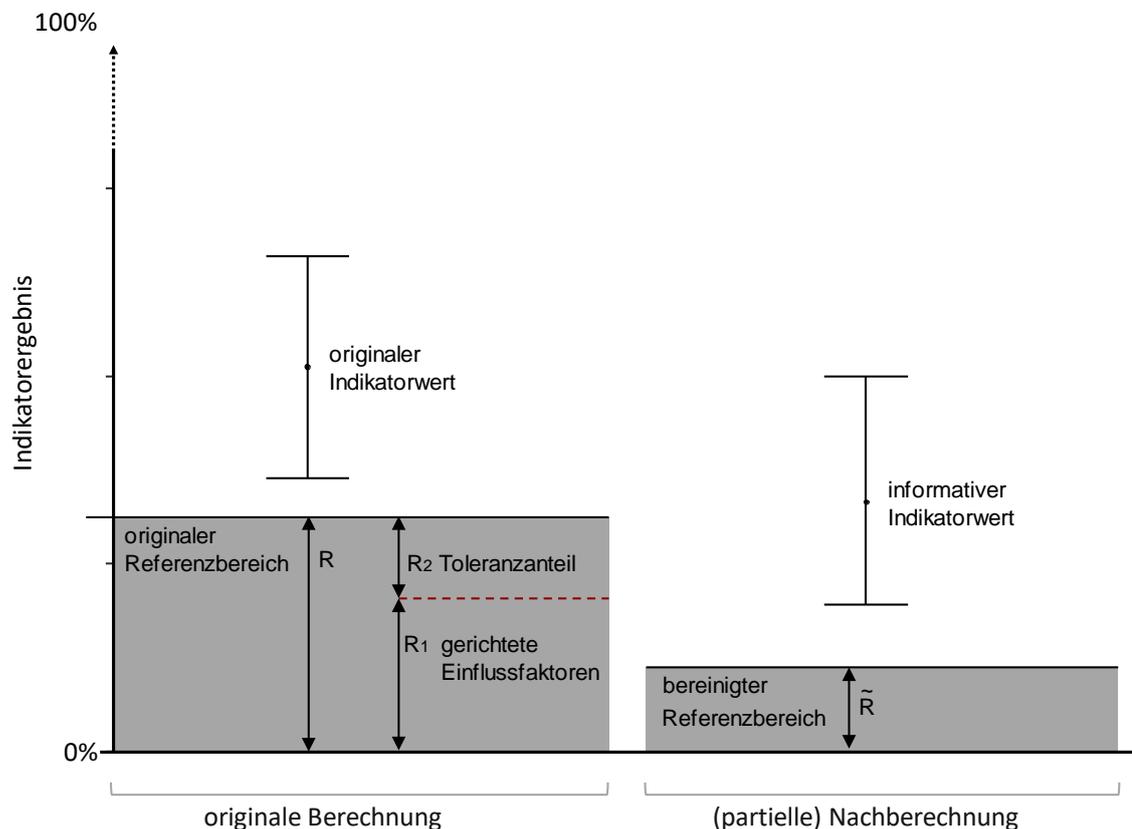


Abbildung 34: Beispiel für die Nachberechnung von Indikatorergebnis und Referenzbereich zur fachlichen Bewertung. Links: Das ursprüngliche Indikatorergebnis vor der fachlichen Bewertung. Rechts: Nachberechnung von Referenzbereich und Indikatorergebnis.

Der so „bereinigte“ Referenzbereich kann als der Referenzbereich verstanden werden, der festgelegt worden wäre, hätte die Rechenregel des Indikators von vorneherein die gerichteten Einflüsse berücksichtigt. Beim Vergleich von informativem Indikatorwert mit dem bereinigten Referenzbereich gibt der Referenzbereich also an, welche vom Leistungserbringer zu verantwortende Abweichung vom Idealwert des Indikators (entweder 0 % oder 100 %) noch toleriert wird, bevor ein Qualitätsdefizit attestiert wird (im obigen Beispiel eine Abweichung von ca. 2 %). Weitere Details zur empfohlenen Berechnungsvorschrift sind in Anhang, Kapitel 6 dargestellt.

<sup>69</sup> Für verteilungsbasierte Referenzbereiche (z. B. perzentilbasierte) gelten bestimmte Einschränkungen, die in Abschnitt 6.4.3 näher erläutert sind.

Zum Zeitpunkt der Berichtslegung entwickelt das IQTIG eine Methodik, wie bei Festlegung von Referenzbereichen die Größe der beiden Anteile des Referenzbereichs angemessen und nachvollziehbar festgelegt werden kann. Für die bisher in der externen Qualitätssicherung eingesetzten Referenzbereiche wurde jedoch die jeweilige Größe der beiden Anteile nicht gesondert ausgewiesen. Sie muss daher zunächst retrospektiv geschätzt werden, bevor auf dieser Basis ein bereinigter Referenzbereich berechnet werden kann. Ein mögliches Vorgehen für diese retrospektive Abschätzung ist ebenfalls in Anhang, Kapitel 6 dargestellt.

#### 6.4.3 Methodische Limitationen verteilungsbezogener Referenzbereiche

Verteilungsbezogene Referenzbereiche erlauben keine Aussagen über die Erfüllung definierter Standards, sondern lediglich vergleichende Aussagen relativ zu den Ergebnissen aller Leistungserbringer (IQTIG 2019a: 161 f.). Ob beispielsweise ein Leistungserbringer, der im 5. Perzentil der Verteilung aller Leistungserbringer in diesem Qualitätsindikator liegt, erwartbare Standards in der Versorgung erfüllt, kann nicht aus dem Indikatorergebnis erschlossen werden. Durch ein Stellungnahmeverfahren kann zwar die Validität der Messung des Indikators für einen Leistungserbringer geprüft werden, jedoch keine umfassende Qualitätsaussage zu einem Leistungserbringer (siehe Abschnitt 3.3 für eine ausführliche Begründung) getroffen werden. Daher kann die fachliche Bewertung bei verteilungsbezogenen Referenzbereichen zwar ergeben, dass der Vergleich des Indikatorwerts eines Leistungserbringers mit dem Referenzbereich nicht angemessen ist, aber nicht, ob der Leistungserbringer mit diesem Indikatorwert einen definierten Standard erreicht. Denn wenn ein Standard für die Versorgung nicht schon bei der Entwicklung des Indikators festgelegt werden konnte, kann dies nicht im Rahmen eines Stellungnahmeverfahrens geschehen.

Beispielsweise könnte sich im Stellungnahmeverfahren mit einem Leistungserbringer bestätigen, dass keine besonderen Konstellationen Einfluss auf den Indikatorwert genommen haben und dieser den Erfüllungsgrad des Qualitätsziels des Indikators angemessen darstellt. Dennoch könnte das Ergebnis des Leistungserbringers absolut gesehen keinen Qualitätsmangel darstellen, sondern lediglich schlechter sein als das Ergebnis von z. B. 90 % der Leistungserbringer (bei einem Referenzbereich auf Basis des 10. Perzentils).

Strenggenommen müsste daher aus methodischer Sicht für Indikatoren mit einem verteilungsbezogenen Referenzbereich das Einstufungsergebnis „keine Qualitätsbewertung“ (S) lauten.<sup>70</sup> Da derzeit für viele Qualitätsindikatoren der externen Qualitätssicherung noch keine festen Referenzbereiche definiert sind, hätte ein solches Vorgehen den Nachteil, dass trotz aufwändiger Datenerhebung keine Qualitätsaussage getroffen werden würde. Als Interimslösung empfiehlt das IQTIG daher, perzentilbasierte Referenzbereiche im

---

<sup>70</sup> Denkbar wäre grundsätzlich auch, in die Klassifikation des Bewertungsergebnisses weitere Unterkategorien für verteilungsbezogene Referenzbereiche einzufügen. Diese könnten dann deutlich machen, dass z. B. „Qualitätsdefizit“ in diesem Fall keinen definierten Qualitätsmangel darstellt, sondern „schlechter als andere Leistungserbringer“ bedeutet. Von einer solchen Aufteilung in zusätzliche Kategorien wird aber abgeraten, da sie für die Nutzung der Informationen im Rahmen von *accountability* (z. B. für die Qualitätsberichte der Krankenhäuser oder für Vergütungsentscheidungen; siehe Abschnitt 2.2) weniger verständlich sein dürfte, den Blick auf wichtige Informationen erschwert und zudem keine Konsequenzen für Handlungsentscheidungen hat.

Stellungnahmeverfahren wie feste Referenzbereiche zu behandeln. Perspektivisch sollte durch Festlegung fester, kriteriumsbezogener Referenzbereiche für diese Indikatoren Klarheit geschaffen werden, welche Qualitätsstandards von den Leistungserbringern gefordert werden.

#### 6.4.4 Vergabe der Unterkategorien für die Einstufung „Qualitätsdefizit“

Wurde der indikatorbasierte Hinweis auf ein Defizit der Versorgungsqualität nach Berücksichtigung aller vom Leistungserbringer nicht zu verantwortenden Einflussfaktoren nicht entkräftet, wird als Ergebnis der fachlichen Bewertung ein „Qualitätsdefizit“ festgestellt. Das Bewertungsschema zum Stellungnahmeverfahren sieht Unterkategorien (Ziffern) vor, die unterscheiden, ob Datenfehler oder fachlich-inhaltliche Faktoren oder eine Kombination aus beiden entscheidend für die Bewertung als „Qualitätsdefizit“ waren (siehe Abschnitt 6.1). Im Folgenden wird ein Vorgehen für einheitliche und nachvollziehbare Kategorisierungen in die Unterkategorien für ein Qualitätsdefizit beschrieben. Das Vorgehen dazu wird in Abbildung 35 visualisiert.

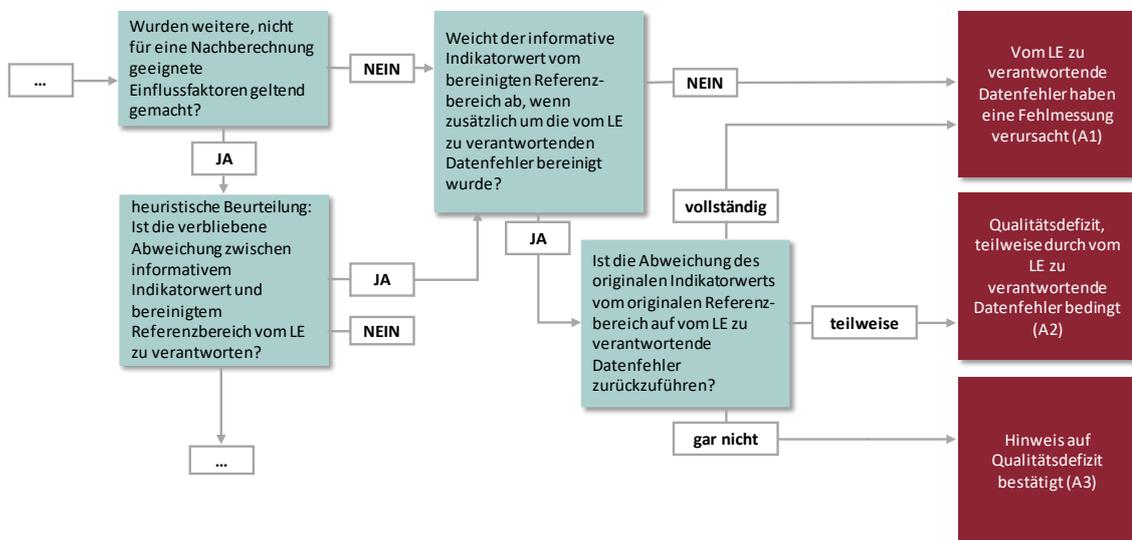


Abbildung 35: Algorithmus zur Differenzierung der Gründe für die Einstufung als Qualitätsdefizit

Der erste Schritt der Differenzierung besteht in einer expliziten Beurteilungsregel, um eine einheitliche Beurteilung sicherzustellen. Dazu wird eine Variante des informativen Indikatorwerts berechnet, bei der nicht nur Datenfehler berücksichtigt werden, die nicht vom Leistungserbringer zu verantworten sind, sondern alle Datenfehler, soweit das rechnerisch eindeutig möglich ist. Diese zusätzliche partielle Nachberechnung erfolgt analog zu dem in den vorherigen Abschnitten geschilderten Vorgehen. Führt diese modifizierte Berechnung des informativen Indikatorwerts dazu, dass der bereinigte Referenzbereich nicht mehr verfehlt wird, ist das ursprüngliche Indikatorergebnis vollständig durch die Kombination aus nicht vom Leistungserbringer zu verantwortenden Einflussfaktoren (falls vorhanden) sowie von ihm zu verantwortenden Datenfehlern erklärt. Das Einstufungsergebnis lautet in diesem Fall „Vom Leistungserbringer zu verantwortende Datenfehler haben eine Fehlmessung verursacht“ (A1). An den entsprechenden Leistungserbringer werden damit hinsichtlich der Beurteilung, inwieweit

(zusätzlich) ein Defizit der Versorgungsqualität vorgelegen hatte, die gleichen Anforderungen gestellt wie an andere Leistungserbringer ohne Datenfehler.

Erklärt die zusätzliche Berücksichtigung der vom Leistungserbringer zu verantwortenden Datenfehler das Qualitätsdefizit dagegen nicht, so ist in der Regel zumindest anteilig von Defiziten der Versorgungsqualität auszugehen. Die Einstufung in die passende erläuternde Unterkategorie A2 oder A3 soll in diesem Fall von der LAG bzw. Bundesstelle mit Unterstützung durch die beratenden Expertinnen und Experten der Fachkommission vorgenommen werden. Sie soll sich daran orientieren, in welchem Ausmaß die Datenfehler einerseits und die inhaltlichen Einflussfaktoren andererseits ausschlaggebend für die Bewertung als „Qualitätsdefizit“ waren. Eine exakte Quantifizierung des jeweiligen Anteils ist jedoch nicht möglich, da Datenfehler zum Verfehlen des Referenzbereiches beigetragen haben können, deren Auswirkung auf den Indikatorwert sich nicht über eine Nachberechnung exakt angeben lässt, und da für einzelne Behandlungsfälle, die in die Indikatorberechnung eingegangen sind, gleichzeitig Datenfehler und fachlich-inhaltliche Faktoren relevant sein können.

In seltenen Fällen könnte das ursprüngliche Abweichen vom Referenzbereich vollständig durch vom Leistungserbringer zu verantwortende Datenfehler verursacht sein, die nicht durch eine Nachberechnung adressiert werden können (denkbar wären beispielsweise gehäufte Eingabefehler bei einem für die Risikoadjustierung genutzten Datenfeld). Für solche Konstellationen sieht der Algorithmus die Möglichkeit vor, auf Grundlage der nicht-quantitativen Einschätzung die Unterkategorie A1 zu vergeben.

#### **6.4.5 Vergabe der Unterkategorien bei Entkräftung des Hinweises auf ein Qualitätsdefizit**

Falls auch Transparenz über die Gründe gewünscht wird, die dazu geführt haben, dass der indikatorbasierte Hinweis auf ein Qualitätsdefizit entkräftet wurde, können die entsprechenden Unterkategorien des Bewertungsschemas genutzt werden, die zwischen Datenfehlern und fachlich-inhaltlichen Faktoren unterscheiden (U1 bis U3; siehe Abschnitt 6.1). Wenn eine zuverlässige Differenzierung angestrebt wird, sollte diese entsprechend den in Kapitel 5 erläuterten methodischen Anforderungen möglichst anhand expliziter, quantitativer Kriterien erfolgen.

Gegenüber dem eingangs dargestellten grundlegenden Algorithmus (siehe Abbildung 33) wird der Einstufungsprozess nach Entkräftung des Hinweises auf ein Qualitätsdefizit an zwei Stellen weitergeführt, die im Folgenden beschrieben werden.

#### **Differenzierung nach Entkräftung durch partielle Nachberechnung**

Um eine Differenzierung vorzunehmen, wenn die partielle Nachberechnung des Indikatorergebnisses zur Entkräftung des Hinweises auf ein Qualitätsdefizit geführt hat, sollte der Bewertungsalgorithmus wie in Abbildung 36 dargestellt erweitert werden.

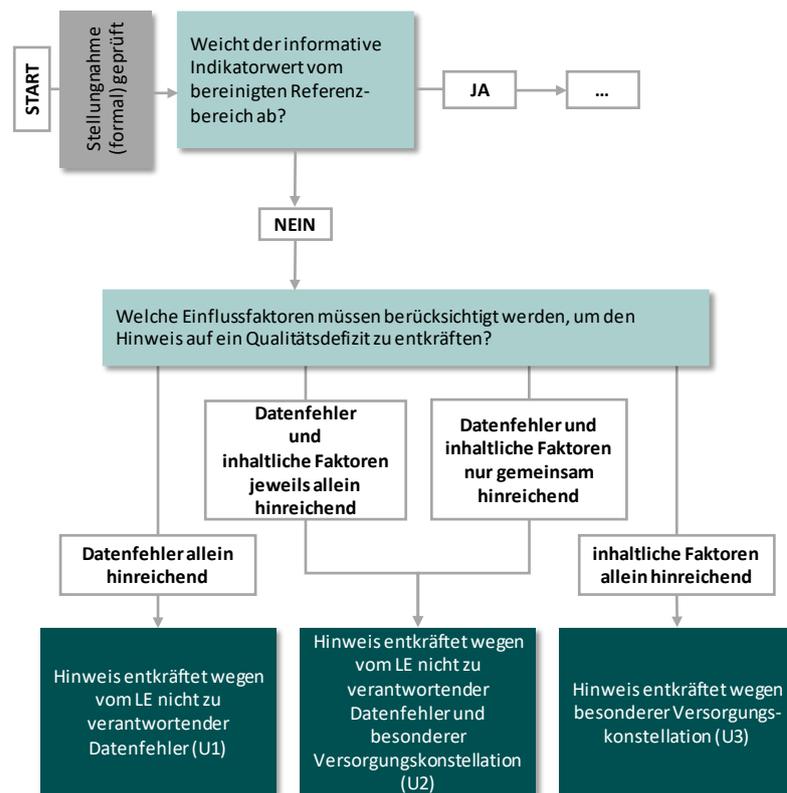


Abbildung 36: Algorithmus für die quantitative Differenzierung der Gründe für einen entkräfteten Hinweis auf ein Qualitätsdefizit

Die Differenzierung der Unterkategorien erfolgt in diesem Fall durch zwei weitere Varianten des informativen Indikatorwerts, bei denen der ursprüngliche Indikatorwert einmal nur um Datenfehler und einmal nur um fachlich-inhaltliche Faktoren bereinigt wurde. Die erforderlichen partiellen Nachberechnungen, die für die Prüfung der hier beschriebenen Bedingungen erforderlich sind, erfolgen analog zu dem in den vorangegangenen Abschnitten geschilderten Vorgehen.

Führt die partielle Nachberechnung bereits dann zu einem Indikatorergebnis im Referenzbereich, wenn ausschließlich Datenfehler, die nicht vom Leistungserbringer zu verantworten sind, bei der partiellen Nachberechnung berücksichtigt wurden, so ist die ursprüngliche Abweichung (des originalen Indikatorwerts vom Referenzbereich) hinreichend durch die Datenfehler erklärt. Führt zusätzlich die ausschließliche Berücksichtigung fachlich-inhaltlicher Einflussfaktoren, die nicht vom Leistungserbringer zu verantworten sind, bei der partiellen Nachberechnung ebenfalls zu einem Ergebnis innerhalb des Referenzbereichs, so ist die ursprüngliche Abweichung genauso hinreichend durch fachlich-inhaltliche Einflussfaktoren erklärt. Der ursprüngliche Hinweis wurde in diesem Fall also sowohl durch Datenfehler als auch durch eine besondere Versorgungskonstellation entkräftet (Kategorie U2). Andernfalls waren allein die Datenfehler für die Entkräftung des Hinweises entscheidend (Kategorie U1)<sup>71</sup>.

<sup>71</sup> Dabei ist es unerheblich, ob zusätzliche Dokumentationsfehler vorgelegen haben, die der Leistungserbringer beeinflussen konnte.

Reicht die alleinige Berücksichtigung der nicht vom Leistungserbringer zu verantwortenden Datenfehler dagegen nicht aus, um ein Indikatorergebnis im Referenzbereich bei der partiellen Nachberechnung zu erhalten, so war für die Entkräftung des Hinweises eine Berücksichtigung fachlich-inhaltlicher Einflussfaktoren notwendig. Führt in diesem Fall bereits die alleinige Berücksichtigung fachlich-inhaltlicher Einflussfaktoren, die nicht vom Leistungserbringer zu verantworten sind, zu einem Ergebnis innerhalb des Referenzbereichs, so waren diese inhaltlichen Faktoren entscheidend (Kategorie U3). Andernfalls wurde der ursprüngliche Hinweis auf ein Qualitätsdefizit erst durch die kombinierte Berücksichtigung von Datenfehlern und besonderer Versorgungskonstellation entkräftet (Kategorie U2).

### Differenzierung nach Entkräftung des Hinweises anhand der heuristischen Beurteilung

Wenn der indikatorbasierte Hinweis auf ein Qualitätsdefizit nicht durch die partielle Nachberechnung und den Vergleich mit dem bereinigten Referenzbereich, sondern erst im Rahmen der heuristischen Beurteilung entkräftet wurde, so kann auch der jeweilige Anteil von Datenfehlern und fachlich-inhaltlichen Faktoren am Einstufungsergebnis nicht zuverlässig quantifiziert werden. Die Einstufung in die passende Unterkategorie (U1, U2, U3) soll in diesem Fall auf Grundlage der Expertise von LAG bzw. Bundesstelle und beratenden Expertinnen und Experten (Fachkommission) vorgenommen werden (Abbildung 37).

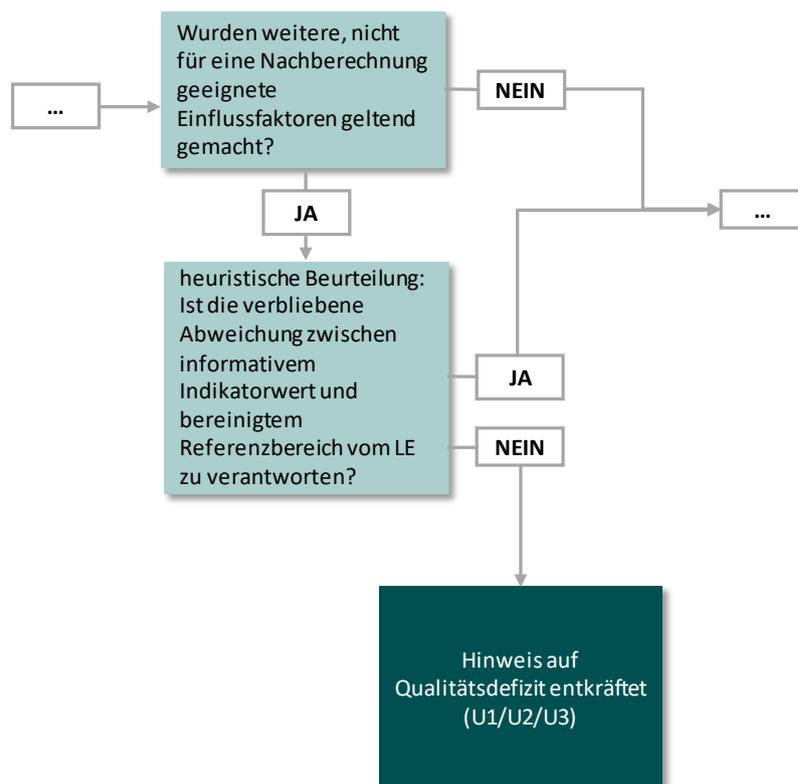


Abbildung 37: Heuristische Differenzierung der Gründe für einen entkräfteten Hinweis auf ein Qualitätsdefizits

## 6.5 Mindestanforderungen an Stellungnahmen

Weder in der QSKH-RL noch in der DeQS-RL finden sich konkrete Vorgaben für Form und Inhalt von Stellungnahmen sowie Angaben dazu, in welchem Zeitraum Stellungnahmen bei der LAG bzw. Bundesstelle abzugeben sind (vgl. Abschnitt 2.4). Dadurch ergibt sich ein großer Spielraum für die Ausgestaltung des Stellungnahmeverfahrens und damit eine mögliche Quelle für Heterogenität in den Ergebnissen, die nicht auf Qualitätsunterschiede zurückzuführen ist. Durch die Vorgabe von Kriterien für Stellungnahmen sollen diese formal und inhaltlich soweit wie möglich standardisiert werden und der Zeitraum für deren Erstellung beschränkt werden. Diese Standardisierung soll einerseits die Objektivität der Bewertung und andererseits eine Aufwandsreduktion bei der Bearbeitung der Stellungnahmen bewirken und bildet die Grundlage für einen einheitlichen Bewertungsprozess. Gerade das Einholen und die Bewertung von Präzisierungen zu Stellungnahmen (vgl. Abschnitt 2.4) stellen einen zusätzlichen Arbeitsaufwand dar. Mittels konkreter Kriterien für Stellungnahmen soll dieser Aufwand tendenziell verringert werden und eine weniger heterogene Bewertungsgrundlage (vgl. Abschnitt 3.3) erreicht werden. In die folgenden Empfehlungen sind u. a. die Rückmeldungen der Landesstellen im gemeinsamen Workshop mit dem IQTIG eingegangen (siehe Anhang, Kapitel 1).

### 6.5.1 Formale Kriterien

Die folgenden formalen Kriterien sollen bundeseinheitlich als Mindestanforderung für jede Stellungnahme gelten und von der LAG bzw. Bundesstelle geprüft werden. Es wird empfohlen, bundesweit einheitlich die Textvorschläge aus Anhang, Kapitel 7 für die Einholung von Stellungnahmen zu verwenden.

- Für jedes auffällige Indikatorergebnis ist eine Stellungnahme abzugeben.
- Die Stellungnahme ist innerhalb von 5 Wochen nach Anforderung der Stellungnahme abzugeben.
- Die Stellungnahme einschließlich eventuell notwendiger Anhänge ist in Schriftform zu erstellen und einzureichen.
- Belege, die ohne erläuternden Text übermittelt werden, sollen nicht als Stellungnahme anerkannt werden. Dazu zählen beispielsweise Einzelfalldokumente wie Arztbriefe, Epikrisen, Verlegungsberichte, Operationsberichte oder Befunde. Solche Belege können nur als Ergänzung einer Stellungnahme berücksichtigt werden und dazu in angemessenem Umfang vom Leistungserbringer eingereicht werden, um seine Argumentation zu unterstützen.
- Die fachlich verantwortliche Instanz des Leistungserbringers (z. B. die Chefarztin/der Chefarzt) ist über den Hinweis auf ein Qualitätsdefizit zu informieren und soll die angefertigte Stellungnahme autorisieren.<sup>72</sup>
- Die Anonymität sowohl der Patientinnen und Patienten (personenidentifizierende Daten) als auch ggf. weiterer Beteiligter (z. B. Zuweiser) ist zu wahren. Bei Nichteinhaltung der Vorgaben

---

<sup>72</sup> Anders als in der Richtlinie zu planungsrelevanten Indikatoren (plan. QI-RL), die festlegt, dass die juristisch verantwortliche Instanz die Stellungnahme zu autorisieren hat, wird die Autorisierung durch die juristisch verantwortliche Instanz für den Verwendungszweck der Qualitätsförderung oder für Public Reporting als zu aufwendig angesehen.

zur Wahrung der Anonymität soll die Stellungnahme aus Datenschutzgründen vernichtet werden und der Leistungserbringer über diesen Vorgang benachrichtigt werden.

- Der Leistungserbringer soll eine Erklärung darüber abgeben, ob die übermittelten QS-Daten, auf denen das Indikatorergebnis basiert, aus seiner Sicht korrekt sind.

### **Umgang mit der Nichteinhaltung formaler Vorgaben**

Jedes Bewertungsverfahren benötigt Rahmenbedingungen, die die zulässigen Umstände der Bewertung vorgeben, sowie Regelungen für den Umgang mit der Nichteinhaltung solcher Vorgaben. Für das Modul „Qualitätsbewertung“ sollen die obigen Kriterien Form und Inhalt von Stellungnahmen so weit wie möglich standardisieren und die empfohlenen Fristen für Datenlieferungen und für die Einreichung von Stellungnahmen etc. (siehe auch Abschnitt 8) sollen dem Modul einen zeitlichen Rahmen geben. Im Folgenden werden Empfehlungen für den Umgang mit der Nichteinhaltung dieser empfohlenen Vorgaben gegeben.

Gemäß dem Verständnis von Qualitätsindikatoren (vgl. Kapitel 3) gibt ein Indikatorergebnis außerhalb des Referenzbereichs einen Hinweis auf ein Qualitätsdefizit. Wird eine Stellungnahme nicht fristgerecht eingereicht oder erfüllt die eingereichte Stellungnahme die formalen Kriterien nicht (z. B. wegen Verstößen gegen den Datenschutz), kann sie nicht zur Entkräftung des Indikatorergebnisses herangezogen werden. Zwar ist eine Nicht-Einhaltung der formalen Vorgaben nicht gleichbedeutend mit einer Bestätigung der Indikатораussage. Jedoch sollte es – im Sinne der Fairness – nicht möglich sein, durch nicht kriterienkonforme oder nicht fristgerechte Stellungnahmen die Qualitätsbewertung zu umgehen oder den Bewertungsprozess zu verzögern. Um Anreize für Formfehler zu vermeiden, wird daher empfohlen, in Fällen, in denen die formalen Vorgaben oder die Fristen nicht eingehalten wurden, die Bewertung „Qualitätsdefizit“ zu vergeben. Dadurch wird deutlich, dass der betreffende Leistungserbringer keine adäquaten Informationen zur Verfügung gestellt hat, die die Indikатораussage entkräften. In Fällen, in denen die Frist für die Einreichung einer Stellungnahme noch nicht verstrichen ist, soll dem Leistungserbringer die Möglichkeit gegeben werden, eine den Kriterien entsprechende Stellungnahme einzureichen.

#### **6.5.2 Inhaltlicher Fokus**

In den Stellungnahmen können für das Verfehlen des Referenzbereichs Gründe angegeben werden, die nur auf einzelne oder wenige Behandlungsfälle zutreffen (z. B. dass ein Notfall ein Abweichen vom üblichen Vorgehen erforderlich machte) oder die sich auf alle Behandlungsfälle beziehen (z. B. dass wegen einer Spezialisierung der Einrichtung auf bestimmte Behandlungen die Patientinnen und Patienten ein besonderes Risikoprofil haben). Vor dem Hintergrund der folgenden Argumente sollte die Stellungnahme in beiden Fällen möglichst auf eine Behandlungsfall-übergreifende Analyse fokussieren:

- Qualitätsindikatoren treffen eine aggregierte Qualitätssaussage auf Ebene des Leistungserbringers und dienen nicht dazu, Einzelfälle zu bewerten. Eine zu stark an einzelnen Fällen orientierte Stellungnahme kann den Blick auf systemische Ursachen und notwendige

systemische Verbesserungsansätze, wie z. B. die Verbesserung in der Einrichtungsorganisation oder der Patientensicherheitskultur, versperren.

- Das Ergebnis der fachlichen Bewertung soll keine umfassende Qualitätsaussage treffen (siehe Abschnitt 3.3), sondern nur die Angemessenheit der Indikатораussage untersuchen. Ob eine besondere, nicht vom Leistungserbringer zu verantwortende Konstellation vorgelegen hat, lässt sich ggf. anhand weniger, auch zusammenfassender Informationen beurteilen, ohne alle betreffenden Behandlungsfälle im Detail zu prüfen. Beispielsweise ist es denkbar, dass ein Leistungserbringer, dessen In-Hospital-Sterblichkeit von Patientinnen und Patienten mit Pneumonie außerhalb des Referenzbereichs liegt, in seiner Stellungnahme die räumliche Nähe zu einer Hospiz-Einrichtung als unberücksichtigten Risikofaktor anführen möchte. Als Beleg könnte in diesem Fall z. B. eine Übersicht dienen, welche der Patientinnen und Patienten im Nenner des Indikators mit palliativer Behandlungsintention aufgenommen wurden, ggf. ergänzend belegt durch die jeweiligen Nebendiagnosen. Eine ausführliche Untersuchung aller Behandlungsverläufe anhand von Epikrisen wird für die Beurteilung, ob der Indikator das Qualitätsmerkmal (Sterblichkeit an Pneumonie bei kurativer Behandlungsintention) angemessen für diesen Leistungserbringer abbildet, voraussichtlich nicht benötigt.
- Umfangreiche Diskussionen von Einzelfällen in den Stellungnahmen gehen mit einem hohen Aufwand bei Stellungnehmenden und LAG bzw. Bundesstelle einher. Dazu trägt bei, dass schriftliche Epikrisen sehr detailliert sein müssen und potenzielle Fragen der beurteilenden Fachkommission bereits vorwegnehmen müssen, wenn der Leistungserbringer die Nachvollziehbarkeit seiner Argumentation sicherstellen möchte. Die Erkenntnisse, die sich aus detaillierten Einzelfallbetrachtungen ergeben, sind jedoch weniger für die Bewertung der Indikатораussage wichtig (s. o.), sondern vielmehr für die Identifizierung von Ansätzen und Maßnahmen zur Qualitätsverbesserung hilfreich. Detaillierte Einzelfallbetrachtungen sollten daher möglichst dem Modul „Qualitätsförderung“ vorbehalten bleiben. Dies führt auch zu einer Aufwandsreduktion, erstens, weil in den Fällen, in denen keine detaillierten Informationen für die Qualitätsbewertung benötigt werden und in denen letztendlich keine Qualitätsförderung notwendig ist, auf die Anfertigung ausführlicher Epikrisen für die externe QS und deren Diskussion verzichtet wird. Zweitens können im Rahmen der Qualitätsförderung die relevanten Fragestellungen durch den Einsatz echter dialogischer Elemente (z. B. Gespräche und Visitationen) zielgerichteter beantwortet werden als durch ein schriftliches Verfahren.

Bei kleinen Fallzahlen oder kleinen Zahlen von interessierenden Ereignissen, die zu einem Hinweis auf ein Qualitätsdefizit geführt haben, ist es einem Leistungserbringer oft nicht möglich, die dem Problem zugrunde liegenden Systemfaktoren zu identifizieren und eine Behandlungsfall-übergreifende Analyse vorzunehmen. Die unter Abschnitt 6.4 geschilderte Methodik erlaubt es daher, sowohl fallbezogene als auch fallübergreifende Gründe für das Verfehlen des Referenzbereichs bei der fachlichen Bewertung zu berücksichtigen.

Für die Prüfung der Indikатораussage in der fachlichen Bewertung ist es wichtig, dass der Leistungserbringer in seiner Stellungnahme für jeden Grund angibt, auf welche seiner Behandlungsfälle dieser Grund zutrifft bzw. ob der Grund für alle Behandlungsfälle zutrifft. Dazu

sind die Vorgangsnummern der entsprechenden Behandlungsfälle mit der Stellungnahme zu übermitteln. Diese Angabe ist auch für diejenigen Behandlungsfälle zu machen, bei denen kein interessierendes Ereignis aufgetreten ist. Für dieses Vorgehen gibt es zwei Gründe:

- Zum einen sollte ein Behandlungsfall, für den das Qualitätsziel des Indikators nicht anwendbar ist, da ein nicht vom Leistungserbringer zu verantwortender Einflussfaktor vorliegt, nicht zur Qualitätsbeurteilung herangezogen werden, unabhängig davon, ob dieser Behandlungsfall ein interessierendes Ereignis aufgewiesen hat oder nicht. Beispielsweise könnte ein Leistungserbringer eine Stellungnahme einreichen zu einem Qualitätsindikator zur Durchführung einer postoperativen Thromboseprophylaxe, der den Anteil der Patientinnen und Patienten mit erhaltener Prophylaxe misst und einen Referenzbereich von > 90 % hat. Wenn der Leistungserbringer für die 50 von ihm im Erfassungsjahr behandelten Patientinnen und Patienten einen Indikatorwert von 80 % aufweist (d. h. in 40 der 50 Fälle wurde eine Prophylaxe durchgeführt) und er geltend machen möchte, dass bei einem besonders hohen Anteil der Patientinnen und Patienten eine Kontraindikation zur blutverdünnenden Behandlung vorlag, reicht es nicht aus, das Vorliegen einer Kontraindikation in z. B. 6 der 10 Fälle zu belegen, die bei der Indikatorberechnung als interessierendes Ereignis (keine Thromboseprophylaxe) gewertet wurden. Es bliebe in diesem Fall unbekannt, ob alle verbleibenden 44 Fälle des Leistungserbringers für eine Thromboseprophylaxe in Frage kamen oder weniger als 44. Im ersten Fall betrüge der nachberechnete Indikatorwert 40 von 44 Fällen (90,9 %). Im anderen Fall wäre denkbar, dass die anerkannte Kontraindikation z. B. auch bei 15 der 40 Fälle ohne interessierendes Ereignis (d. h. mit durchgeführter Thromboseprophylaxe) vorlag. Die Zielpopulation des Indikators (Fälle, bei denen eine Thromboseprophylaxe durchgeführt werden soll) wäre dann nur 29 Fälle groß (50 Fälle abzgl. 21 Fällen mit Kontraindikation), bei denen der Leistungserbringer in 4 Fällen nicht begründen konnte, warum keine Thromboseprophylaxe durchgeführt wurde (nachberechneter Indikatorwert 86,2 %). Liegen Informationen über das Vorliegen eines Einflussfaktors nicht für alle Behandlungsfälle vor, kann dies also zu relevanten Unterschieden führen.
- Zum anderen kann es vorkommen, dass mehrere Einflussfaktoren auf das Indikatorergebnis von einem Leistungserbringer geltend gemacht und bei der fachlichen Bewertung als nicht von ihm zu verantworten anerkannt werden. In diesen Fall muss anhand der Vorgangsnummern erkennbar sein, in welcher Kombination die anerkannten Einflussfaktoren bei den Behandlungsfällen vorliegen, damit die Zielpopulation bei der Nachberechnung des Indikatorwert korrekt bestimmt werden kann.

In der Berechnung des informativen Indikatorergebnisses sind diese Überlegungen berücksichtigt (siehe Abschnitte 6.4.1 und Anhang, Kapitel 6). In Kapitel 7 des Anhangs findet sich ein Beispiel, auf welche Weise in einer Stellungnahme die Vorgangsnummern fallbezogen übermittelt werden sollten. Es wird empfohlen, eine digitale Übermittlung der Vorgangsnummern festzulegen, um Übertragungsfehler in die Software für die Nachberechnung des informativen Indikatorwerts (siehe Abschnitt 6.6) zu vermeiden. Auf die Angabe der Vorgangsnummern soll nur dann verzichtet werden, wenn dies nicht sinnvoll oder nicht möglich ist, beispielsweise weil ein struktureller Einflussfaktor (etwa die Anfahrtszeit des Rettungsdienstes in einer ländlichen Region, siehe Beispiel 3 in Kapitel 5 des Anhangs) sich auf

alle Behandlungsfälle gleichermaßen bezieht. Diese Einflussfaktoren können bei der fachlichen Bewertung im heuristischen Beurteilungsschritt (siehe Abschnitt 6.4) berücksichtigt werden. Ein Einflussfaktor, bei dem aus der Stellungnahme nicht hervorgeht, in welchem Maße er auch bei Fällen ohne interessierendes Ereignis vorgelegen hat, soll in der fachlichen Bewertung nicht berücksichtigt werden.

Werden in der Stellungnahme Gründe für das auffällige Indikatorergebnis angeführt, die aus Sicht des Leistungserbringers nicht von ihm zu verantworten waren, so sollen diese Gründe nicht nur benannt werden, sondern es soll auch nachgewiesen werden, dass mit der gebotenen Sorgfaltspflicht alles Zumutbare getan wurde, um diese aufzufangen (z. B. die Einplanung von genügend Ressourcen und genügend Zeit). Damit ist nicht gemeint, dass jeder Behandlungsfall in der Stellungnahme ausführlich diskutiert wird (s. o.), sondern dass z. B. anhand einer fallübergreifenden Erläuterung für die Fachkommission nachvollziehbar werden muss, warum der geltend gemachte Einflussfaktor tatsächlich nicht vom Leistungserbringer zu verantworten war. Das gleiche trifft sinngemäß auf Datenfehler zu: Hat ein Leistungserbringer in seiner Stellungnahme angegeben, dass Fehler in den übermittelten QS-Daten vorliegen, soll die Stellungnahme erklären, welche Fehler seiner Ansicht nach vorliegen und welche der Fehler in den übermittelten Daten von ihm nicht zu vertreten sind.

Analog zu den formalen Vorgaben für die Stellungnahmen (siehe Abschnitt 6.5.1) liegt auch die Verantwortung für die Vollständigkeit und Angemessenheit der Inhalte einer Stellungnahme beim Leistungserbringer. Aus Gründen der Fairness und der Verfahrenseffizienz sollte auf die Nachforderung von ergänzenden Informationen durch die LAG verzichtet werden (siehe Abschnitt 8.2.2). Kann innerhalb der Einreichungsfrist ein Sachverhalt, der mutmaßlich die Indikатораussage entkräftet, nicht hinreichend durch entsprechende Informationen und Belege plausibel gemacht werden, so stellt der indikatorbasierte Hinweis auf ein Qualitätsdefizit die bestverfügbare Evidenz bzgl. der Erfüllung des Qualitätsmerkmals dar. Die fachliche Bewertung soll dann auf Basis der bis zu diesem Zeitpunkt vorliegenden Informationen erfolgen.

## **6.6 Umsetzung und Aufwand**

Das in den vorangehenden Abschnitten empfohlene Konzept für die fachliche Bewertung zeichnet sich gegenüber dem bisherigen Vorgehen durch mehr explizite Verfahrensregeln und einem höheren Grad an Standardisierung aus. Dies kann zu Herausforderungen bei der Implementation und zu höherem Aufwand bei der Durchführung als bisher führen. Diese Aspekte werden im Folgenden diskutiert.

### **Klassifikation der Einflussfaktoren**

Die initialen Schritte bei der fachlichen Bewertung bestehen in der Klassifikation der in den Stellungnahmen angeführten Einflussfaktoren auf das Indikatorergebnis (siehe Abschnitte 6.3 und 6.4.1). Sie ist eine Voraussetzung für die Beurteilung, ob nicht berücksichtigte Einflussfaktoren die Indikатораussage entkräften. Wenn auch im bisherigen Strukturierten Dialog der Einfluss solcher Faktoren auf das Indikatorergebnis in die Beurteilung eingeht, führt das vorgeschlagene Vorgehen zu keinem höheren Aufwand, da in diesem Fall auch bisher schon eine solche

Klassifikation vorgenommen werden musste. Wenn im bisherigen Strukturierten Dialog ein anderes Ziel verfolgt wird (z. B. eine Gesamtschau der Behandlungsqualität eines Leistungserbringers), kann sich der Aufwand bei den LAG und den Fachkommissionen gegenüber dem Status quo unterscheiden. Die Erfahrungen mit dem Vorgehen im Verfahren zu planungsrelevanten Qualitätsindikatoren zeigen, dass eine explizite Klassifikation der in Stellungnahmen angeführten Einflussfaktoren auf das Indikatorergebnis grundsätzlich umsetzbar ist.

### **Partielle Nachberechnung**

Die Nachberechnung des informativen Indikatorwerts auf Grundlage der anerkannten Einflussfaktoren ist ein wesentliches Element des vorgeschlagenen Algorithmus für die fachliche Bewertung und im Vergleich zum bisherigen Vorgehen neu. Die grundsätzliche Umsetzbarkeit dieses Vorgehens zeigen die Erfahrungen aus dem Verfahren der planungsrelevanten Qualitätsindikatoren. Eine zuverlässige und aufwandsangemessene Umsetzung dieses Elements setzt geeignete technische Unterstützung voraus, beispielsweise in Form eines für alle Bewertungsstellen einheitlichen Web-basierten Portals. Eine einheitliche Softwarelösung für alle Bewertungsstellen würde zudem Heterogenität in der Bewertung aufgrund unterschiedlicher Softwarelösungen bei den Landesstellen vermeiden. Die Ergebnisse der Nachberechnungen sollten von der Software im Echtzeit-Betrieb bereitgestellt werden, sodass während der Beratungen in den Fachkommissionssitzungen keine Verzögerungen entstehen. Zu berücksichtigen ist der Aufwand für die Implementation und Pflege einer solchen Software, der bei einer zentralen technischen Lösung z. B. durch die Schaffung des erforderlichen Rechtessystems und von Berechnungsmöglichkeiten in Echtzeit entsteht. Diese notwendige IT-seitige Entwicklungsarbeit muss bei einer möglichen Beauftragung der Umsetzung des Konzeptes berücksichtigt werden.

Nach der Klassifikation der in einer Stellungnahme angeführten Einflussfaktoren (s. o.) besteht der zusätzliche Aufwand durch die hier empfohlene Methodik darin, der Software die Behandlungsfälle z. B. in Form einer Liste (siehe Kapitel 8 des Anhangs) anzugeben, die bei der Nachberechnung aus der Indikatorberechnung ausgeschlossen werden sollen. Dazu müssen diese Angaben vom Leistungserbringer zur Verfügung gestellt werden. Möchte ein Leistungserbringer einen bisher nicht erfassten Einflussfaktor geltend machen, muss er – sofern nicht alle Behandlungsfälle betroffen sind – die Vorgangsnummern aller Behandlungsfälle mit diesem Einflussfaktor übermitteln (siehe Abschnitt 6.5.2). Im Vergleich zum bisher üblichen Vorgehen, bei dem die Angabe üblicherweise nur für Behandlungsfälle mit interessierendem Ereignis erfolgt, ergibt sich für die Leistungserbringer ein Mehraufwand. Dieser Mehraufwand ist nicht vermeidbar, wenn eine Verzerrung durch die selektive Betrachtung von bestimmten Fällen vermieden werden soll.

### **Differenzierung mittels Ziffern für Qualitätsdefizit**

Zur Erhöhung der Transparenz über die Bewertungsergebnisse für die Verfahrensbeteiligten wurden im Bewertungsschema auf der zweiten Ebene Ziffern und ein aus zwei Schritten bestehender Bewertungsalgorithmus vorgeschlagen (siehe Abschnitte 6.1.3 und 6.4.4). Eine Software für die Nachberechnung des informativen Indikatorwertes kann den ersten, expliziten Beurtei-

lungsschritt für diese Differenzierung ohne zusätzlichen Aufwand für die LAG bzw. Fachkommission vornehmen, da er auf denselben Informationen basiert. Der zweite, heuristische Beurteilungsschritt bedeutet im Vergleich zum bisherigen Vorgehen ebenfalls keinen Zusatzaufwand, da eine Beurteilung von Dokumentationsproblemen auch im bisherigen Verfahren schon durchgeführt wird, wie sich an der Vergabe entsprechender Schlüsselwerte zeigt (z. B. in der Kategorie „D“ des bisherigen Bewertungsschemas, siehe Abschnitt 2.4.3).

#### **Differenzierung mittels Ziffern für „kein Hinweis auf Qualitätsdefizit“**

Für die Unterscheidung der Gründe, weshalb der Hinweis aus einem Indikatorergebnis als entkräftet angesehen wurde, wurden ebenfalls Ziffern und Bewertungsregeln vorgeschlagen (siehe Abschnitte 6.1.2 und 6.4.5). Erfolgte die Entkräftung auf Grundlage der Nachberechnung des informativen Indikatorwerts, so kann die passende Ziffer ohne zusätzlichen Aufwand für die LAG bzw. Fachkommission von einer entsprechenden Software mit ausgegeben werden, da alle notwendigen Informationen bereits angegeben wurden. Wurde der Hinweis auf ein Qualitätsdefizit durch heuristische Beurteilung entkräftet, so waren entweder nicht zu verantwortende Datenfehler, eine besondere Versorgungskonstellation oder eine Kombination aus beidem ausschlaggebend für die Entkräftung. Die Differenzierung der Ziffern ist dann lediglich eine Dokumentation der Entscheidungsgründe für die getroffene Beurteilung und bedeutet keinen Mehraufwand.

#### **Trennung von Qualitätsbewertung und -förderung**

Die Prozesse von Qualitätsförderung und Qualitätsbewertung sollen zukünftig getrennt werden, auch um die Objektivität der Bewertung zu erhöhen. Für die Qualitätsförderung können jedoch über das bloße Einstufungsergebnis hinaus weitere Informationen aus der fachlichen Bewertung relevant sein. Beispielsweise können sich aus Fallberichten, die vom Leistungserbringer als Beleg für eine besondere Versorgungskonstellation angeführt wurden, Hinweise ergeben, durch welche Maßnahmen der Leistungserbringer Qualitätsverbesserungen erreichen kann. Damit diese Informationen bei der Qualitätsförderung genutzt werden können, müssen diese bei der fachlichen Bewertung neben der Qualitätsbewertung im eigentlichen Sinne begleitend dokumentiert werden. Wenn die Qualitätsförderung zu einem späteren Zeitpunkt als die Qualitätsbewertung erfolgt (z. B. nicht in derselben Sitzung der Fachkommission), so bedeutet bereits die notwendige erneute Befassung einen gewissen Aufwand. Dieser entsteht allerdings ebenfalls im bisherigen Vorgehen, wenn Qualitätsverbesserungsmaßnahmen wie kollegiale Gespräche und Visitationen zu einem späteren Zeitpunkt als die initiale Beurteilung der schriftlichen Stellungnahme durchgeführt werden. Dies dürfte auch für die in § 17 Abs. 3 DeQS-RL genannten Maßnahmen der Qualitätsförderung zutreffen. Maßnahmen, die allein auf Basis der Stellungnahme erfolgen (beispielsweise die Zusendung eines indicatorspezifischen Fragebogens zur Selbsteinschätzung des Leistungserbringers), können dagegen noch innerhalb derselben Sitzung der Fachkommission veranlasst werden. Dies ist auch im vorgeschlagenen Konzept möglich.

## Zusammenfassung

Durch den vorgeschlagenen, auf möglichst expliziten Beurteilungsregeln basierenden Bewertungsprozess entstehen voraussichtlich folgende zusätzliche Aufwände gegenüber einem ausschließlich heuristischen, auf impliziten Beurteilungsregeln basierenden Vorgehen:

- Bereitstellung und Pflege einer technischen Unterstützung (Software) von Nachberechnungen
- Eingabe der relevanten Information in die Software bei der fachlichen Bewertung
- Nochmalige Befassung mit Informationen aus den Stellungnahmen bei von der Qualitätsbewertung zeitlich getrennten Befassung der Fachkommission mit Maßnahmen der Qualitätsverbesserung.

Eine Reihe weiterer Aufwände bei der Qualitätsbewertung sind dagegen nicht spezifisch für das vorgeschlagene Vorgehen, sondern sind in jedem Fall zu berücksichtigen, wenn fundierte Aussagen über die Erfüllung des Qualitätsmerkmals eines Indikators gemäß dem in Abschnitt 6.1 dargestellten Bewertungsschema getroffen werden sollen:

- Klassifikation der in den Stellungnahmen genannten Einflussfaktoren hinsichtlich Verantwortlichkeit, Art des Faktors und Auswirkung auf das Indikatorergebnis
- Angabe durch den Leistungserbringer, welche Behandlungsfälle von einem geltend gemachten Einflussfaktor betroffen sind
- Beurteilung, inwieweit die Abweichung eines Indikatorergebnisses vom Referenzbereich auf Datenfehler zurückzuführen ist, sofern sich dies nicht berechnen lässt
- Beurteilung, ob sich bereits auf Grundlage der schriftlichen Stellungnahme Hinweise auf Maßnahmen zur Qualitätsverbesserung ableiten lassen

Eine Reduktion des Aufwandes für die fachliche Bewertung könnte in geringem Maße durch Verzicht auf die Differenzierung der Ziffern des Bewertungsschemas erreicht werden. Allerdings ist die Vergabe dieser Ziffern sowohl bei einem expliziten als auch bei einem heuristischen Vorgehen im Wesentlichen ein Nebenprodukt des Beurteilungsprozesses und erzeugt wenig Zusatzaufwand (s. o.).

## 6.7 Zusammensetzung von Fachkommissionen

Die Fachkommissionen haben eine wichtige beratende Funktion für die LAG bzw. Bundesstelle bei der Qualitätsbewertung und Qualitätsförderung. Ihre Arbeit verfolgt im Rahmen des Stellungnahmeverfahrens zwei Ziele. Zum einen soll die Fachkommission die Angemessenheit der Abbildung der Versorgungsqualität anhand des Indikators aus medizinisch-fachlicher Sicht beurteilen. Dafür müssen die Mitglieder der Fachkommission die indikatorbezogenen Stellungnahmen lesen und eine fachliche Beurteilung vornehmen. Die fachliche Beurteilung beinhaltet die Prüfung, ob eine besondere Versorgungskonstellation vorliegt, die zur Fehlmessung im Indikator geführt hat, sowie eine Empfehlung an die LAG bzw. Bundesstelle, wie das Ergebnis in der abschließenden Qualitätsbewertung kategorisiert werden sollte (siehe Abschnitte 6.3 und 6.4). Zum anderen soll die Fachkommission nach dem Bewertungsprozess und bei einem bestätigten

Qualitätsdefizit Empfehlungen darüber abgeben, welche Maßnahmen der LAG bzw. Bundesstelle zur Qualitätsverbesserung an das interne Qualitätsmanagement adressiert werden.

Orientiert an den Zielen und Aufgaben der Fachkommission sind bestimmte Anforderungen sowohl an die einzelnen Expertinnen und Experten als auch an die Zusammensetzung der Gruppe als Ganzes zu formulieren. Diese sollten insbesondere die in Abschnitt 3.7 beschriebenen Erfolgsfaktoren (fachliche Kompetenz, Unabhängigkeit und Interdisziplinarität) bei der Zusammensetzung der Fachkommission berücksichtigen. Daher wird empfohlen, eine fachlich qualifizierte, unabhängige und interdisziplinäre Fachkommission zusammenzusetzen, um die Akzeptanz des Gesamtverfahrens sicherzustellen und Qualitätsdefizite erfolversprechend abzubauen. Die Verantwortung für die Zusammensetzung einer fachlich qualifizierten, unabhängigen und interdisziplinären Fachkommission soll bei der LAG bzw. Bundesstelle liegen, wobei die im Folgenden dargestellten, bundeseinheitlichen Kriterien für die Auswahl einzelner Expertinnen und Experten und für die Zusammensetzung der Fachkommission als Rahmenbedingungen vorgegeben werden sollten. Zur Steigerung der Akzeptanz des Verfahrens gegenüber den Expertinnen und Experten der Fachkommission wird eine bundesweit einheitliche Regelung zur Aufwandsentschädigung, die gleichermaßen für alle Expertinnen und Experten gilt, empfohlen. Auch im IQTIG-LQS-Treffen am 18. Juni 2019 wurde eine einheitliche Aufwandsentschädigung seitens einiger Teilnehmerinnen und Teilnehmer befürwortet (siehe Anhang, Kapitel 2). In Abschnitt 6.7.1 werden zunächst die Kriterien beschrieben, die von den einzelnen Expertinnen und Experten persönlich erfüllt sein sollten, während in Abschnitt 6.7.2 erläutert wird, wie die Fachkommission als Gruppe zusammengesetzt sein sollte.

### **6.7.1 Kriterien für die Auswahl der Expertinnen und Experten**

Bei den Kriterien für die Auswahl der einzelnen Expertinnen und Experten kann zwischen fachlich-inhaltlichen Anforderungskriterien, die sich auf die fachliche Eignung (Qualifikation und berufliche Erfahrung) beziehen, und formalen Anforderungskriterien (z. B. Zusicherung der Vertraulichkeit, Angaben zu Interessenkonflikten) unterschieden werden. Die folgenden Anforderungskriterien sollen sowohl die fachliche Kompetenz als auch den Einbezug möglichst unabhängiger Mitglieder der Fachkommission gewährleisten. Es wird empfohlen, die Kriterien erst ab der nächsten Neubesetzung der Fachkommissionen anzulegen.

#### **6.7.1.1 Fachlich-inhaltliche Anforderungskriterien**

Um eine möglichst nachvollziehbare Beurteilung der besonderen Versorgungskonstellationen vornehmen zu können, sollten die Fachexpertinnen und -experten nachweislich über die für das jeweilige QS-Verfahren relevante klinisch-praktische Erfahrung verfügen. Die fachliche Expertise und die Kenntnis der Versorgungspraxis ist dabei eine notwendige Voraussetzung für die Beurteilung der in den Stellungnahmen vorgebrachten Gründe und die Bewertung dahingehend, ob es sich dabei um eine besondere Versorgungskonstellation handelt. Angesichts des sich schnell entwickelnden medizinischen Wissens und der Vielzahl an existierenden, fachlich unterschiedlichen Versorgungsbereichen ist die LAG bzw. Bundesstelle darauf angewiesen, sich von Expertinnen und Experten beraten zu lassen, deren Wissen und Erfahrungen möglichst aktuell sind. Die

LAG bzw. Bundesstelle sollte die jeweiligen Fachgesellschaften über eine ausstehende Neubenennung der Mitglieder einer Fachkommission informieren, damit eine Bewerbung von Mitgliedern der entsprechenden Fachgesellschaften möglich ist. Die zu berufenden ärztlichen Fachexpertinnen und -experten sollten nach der Facharztausbildung während der letzten drei Jahre vor ihrem Mitwirken in der Fachkommission im entsprechenden Versorgungsbereich praktisch tätig gewesen sein. Auch nicht medizinische und an der Leistungserbringung beteiligte Fachexpertinnen und -experten (z. B. Pflegekräfte, Logopädinnen/Logopäden, Ergotherapeutinnen/Ergotherapeuten, Physiotherapeutinnen/Physiotherapeuten) sollten nach ihrer qualifizierenden Ausbildung während der letzten drei Jahre im entsprechenden Versorgungsbereich tätig gewesen sein. Expertinnen und Experten, die während ihrer Mitgliedschaft in der Fachkommission in den Ruhestand wechseln oder eine Tätigkeit außerhalb der Gesundheitsversorgung ausüben, können bis zu vier Jahre nach ihrem Ausscheiden aus der Tätigkeit in der Gesundheitsversorgung mitwirken. Expertinnen und Experten, die im Ausland tätig sind oder waren, sollten mit den nationalen fachlichen Standards vertraut sein und mindestens fünf Jahre als Klinikerinnen/Kliniker in Deutschland tätig gewesen sein. Bei längeren Auslandsaufenthalten ist die Eignung der Expertinnen und Experten im Einzelfall abzuwägen. Um Empfehlungen für qualitätsverbessernde Maßnahmen abgeben zu können, sind Kenntnisse der Qualitätssicherung und des internen Qualitätsmanagements hilfreich.

#### **6.7.1.2 Formale Anforderungskriterien**

Eine Bewerbung sollte unter Übermittlung folgender Unterlagen (Bewerbungsunterlagen) an die LAG bzw. Bundesstelle erfolgen:

- Aussagekräftiger Lebenslauf (mit notwendigen Fachkenntnisse und -erfahrungen, ggf. auch in den Bereichen Qualitätssicherung und Qualitätsmanagement)
- Offenlegung von Interessenkonflikten (inkl. Angaben zu Gremienarbeit)
- Einwilligung, dass Interessenkonflikte ggf. bei Berufung öffentlich genannt werden
- Verpflichtungserklärung zur Wahrung der Vertraulichkeit
- Zusicherung einer regelmäßigen Beteiligung an den Sitzungen der Fachkommission
- Anerkennung der Geschäftsordnung des jeweiligen Gremiums
- Bei Angestelltenverhältnis: Zustimmungserklärung des Arbeitgebers

#### **Lebenslauf**

Anhand des Lebenslaufs sind die notwendigen Fachkenntnisse und Erfahrungen zu prüfen. Nur bei Erfüllung der beschriebenen fachlich-inhaltlichen Kriterien darf eine Expertin oder ein Experte ausgewählt werden. Werden Expertinnen und Experten trotz unzureichender Erfüllung der fachlich-inhaltlicher Anforderungen von der LAG bzw. Bundesstelle benannt, sind die Gründe für die Benennung schriftlich niederzulegen.

#### **Interessenkonflikte**

Bei Akteuren im Gesundheitswesen können primäre und sekundäre Interessen das Handeln beeinflussen und im Konflikt zueinanderstehen. Das primäre Interesse im Rahmen des Stellung-

nahmeverfahrens sollte die sachgemäße Bewertung der Versorgungsqualität anhand der Indikatoren sein. Damit eine unabhängige Bewertung der Versorgungsqualität gewährleistet werden kann, müssen Konflikte mit sekundären Interessenslagen weitestgehend vermieden werden. Solche Interessenkonflikte können beispielsweise auffällige Indikatorergebnisse der eigenen Einrichtung sein oder eine Konkurrenzsituation geografisch benachbarter Einrichtungen.

Ein wesentlicher Aspekt des Umgangs mit Interessenkonflikten ist deren Transparenz. Daher müssen Fachexpertinnen und -experten bereit sein, dass ihre Interessenkonflikte im Vorfeld einer Berufung zumindest dem Lenkungsgremium des Verfahrens in der LAG bzw. Bundesstelle offengelegt werden. Können Interessenkonflikte nicht ausreichend vermieden werden, sind sie ggf. zu veröffentlichen. Die Prüfung, Einordnung und Veröffentlichung der Interessenkonflikte liegt in der Verantwortung der LAG bzw. Bundesstelle.

### **Vertraulichkeit**

Die sich bewerbenden Expertinnen und Experten müssen sich zur strikten Wahrung der Vertraulichkeit von Unterlagen und Diskussionsinhalten verpflichten. Ein Verstoß dagegen soll zum Ausschluss mit sofortiger Wirkung führen. Eine Fachkommission kann bei Missbrauch der Vertraulichkeit von der LAG bzw. Bundesstelle aufgelöst werden.

### **Zeitliche Verfügbarkeit**

Die sich bewerbenden Expertinnen und Experten müssen im Vorfeld schriftlich bestätigen, dass sie zeitlich in der Lage sind, die Teilnahme an den Sitzungen sowie deren Vor- und Nachbereitung zu gewährleisten. Bei abhängig beschäftigten Personen ist eine Bestätigung des Arbeitgebers beizufügen, in der erklärt wird, dass der Arbeitgeber die Wahrnehmung dieses Amtes unterstützt und der Mitarbeiterin / dem Mitarbeiter die Teilnahme an den Sitzungen der Fachkommission ermöglichen wird. Dieses Kriterium ist zur Aufrechterhaltung einer konstanten Sitzungsbeteiligung insbesondere unter Einbeziehung von beruflich tätigen Mitgliedern notwendig.

### **Weitere Gründe, Mitteilung bei Änderung der Gegebenheiten**

Weiterhin hat eine bewerbende Person zu versichern, dass zusätzlich keine anderen Gründe vorliegen, die der Fachlichkeit oder dem Ansehen des Verfahrens entgegenstehen könnten. Ändert sich nach Berufung die Situation eines Mitglieds der Fachkommission so, dass ein neuer oder verstärkter Interessenkonflikt angenommen werden kann oder dass ein anderer relevanter Grund eintritt, so muss diese Änderung der LAG bzw. Bundesstelle gegenüber unverzüglich, spätestens aber zwei Wochen vor der nächsten Beratungsleistung, angezeigt werden. Die Nicht-Anzeige eines solchen Umstands kann von der LAG bzw. Bundesstelle als alleiniger Grund zum Anlass genommen werden, um eine Expertin bzw. einen Experten von der weiteren Mitwirkung fristlos auszuschließen.

#### **6.7.2 Kriterien für die Zusammensetzung der Fachkommission**

Im Vordergrund für die Zusammensetzung der Fachkommission steht der Erfolgsfaktor der *Interdisziplinarität*. Interdisziplinarität bedeutet im Kontext des Stellungnahmeverfahrens, dass

für eine sachgerechte Bewertung der Versorgungsqualität anhand der Indikatoren und für Empfehlungen zur Qualitätsförderung alle an den Versorgungsergebnissen beteiligten Berufsgruppen einbezogen werden sollen, die auf die Indikatorergebnisse wesentlichen Einfluss haben. Um die Interdisziplinarität zu wahren, sollte vor dem Hintergrund der einzelnen, sehr unterschiedlichen Versorgungsbereiche daher die Zusammensetzung der jeweiligen Fachkommission spezifisch ausgewählt werden. Dabei ist für jeden Versorgungsbereich zu identifizieren, welche Berufsgruppen durch ihre Versorgungsleistung einen wesentlichen Einfluss auf die Ergebnisse der Qualitätsindikatoren ausüben. Beispielsweise hat im QS-Verfahren *Geburtshilfe* die Berufsgruppe der Hebammen und Geburtshelfer einen Einfluss auf die Ergebnisse der Indikatoren zum Dambris. Durch die Einbindung relevanter Berufsgruppen soll einerseits vermieden werden, dass die Bewertung der Versorgungsqualität durch die Interessen einzelner Berufsgruppen verzerrt wird. Andererseits können durch eine interdisziplinäre Fachkommission berufsgruppenübergreifende Probleme identifiziert und entsprechende Fördermaßnahmen empfohlen werden. In Abschnitt 3.7 wurde zudem dargelegt, dass dabei auch Patientinnen und Patienten einbezogen werden müssen.

Für eine interdisziplinäre Fachkommission wird empfohlen, folgende **ständige Expertinnen und Experten** für jeweils alle QS-Verfahren zu benennen:

#### **Fachärztinnen und Fachärzte**

Fachärztinnen und Fachärzte tragen in der Regel die Hauptverantwortung für das Behandlungsergebnis. Daher sind mindestens zwei Fachärztinnen bzw. Fachärzte einzubeziehen. Für manche Versorgungsbereiche kann es sinnvoll sein, unterschiedliche Fachdisziplinen einzubinden.

#### **Pflegefachkräfte**

Pflegekräfte spielen in der gesundheitlichen Versorgung eine zentrale Rolle und beeinflussen damit die Ergebnisse der Qualitätsindikatoren. Daher ist in der Regel mindestens eine Pflegefachkraft in jeder Fachkommission einzubeziehen.

#### **Klinische Qualitätsmanagerin bzw. klinischer Qualitätsmanager**

Ausgehend von der Qualitätsbewertung sollten Empfehlungen für geeignete Maßnahmen der Qualitätsverbesserung vorgeschlagen werden. Diese Maßnahmen sollten idealerweise unter Berücksichtigung gängiger Qualitätsmanagementstrategien erfolgen, daher sollte mindestens eine Expertin bzw. ein Experte des klinischen Qualitätsmanagements hinzugezogen werden.

#### **Patientenvertreterinnen bzw. -vertreter**

Als Empfänger der Gesundheitsleistungen, deren Qualität gesichert werden soll, ist die Perspektive der Patientinnen und Patienten im Rahmen von Qualitätssicherungsverfahren unverzichtbar. Bei allen QS-Verfahren ist daher mindestens eine Patientenvertreterin bzw. ein Patientenvertreter mit Kenntnissen im Bereich des jeweiligen Versorgungsbereichs durch Vorschlag der Patientenvertretung nach § 140f SGB V einzubeziehen.

In Abhängigkeit vom jeweiligen QS-Verfahren sollten folgende **weitere Expertinnen und Experten** benannt werden:

#### **Weitere Angehörige der Gesundheitsfachberufe**

Zu den weiteren Gesundheitsfachberufen zählen z. B. Logopädinnen und Logopäden, Ergotherapeutinnen und Ergotherapeuten, Hebammen und Entbindungshelfer, Physiotherapeutinnen und Physiotherapeuten. Die Ergebnisse der Qualitätsindikatoren werden durch die Behandlungsergebnisse der Mitglieder dieser Berufsgruppen beeinflusst (z. B. im Versorgungsbereich Orthopädie und Unfallchirurgie<sup>73</sup>). Daher sollte in Abhängigkeit vom jeweiligen Versorgungsbereich mindestens eine Expertin bzw. ein Experte aus einem für das QS-Verfahren relevanten Gesundheitsfachberuf einbezogen werden.

#### **Krankenhaushygienikerin bzw. -hygieniker oder Hygienefachkraft**

Bei allen Versorgungsbereichen, bei welchen die Ergebnisse der Qualitätsindikatoren durch die Krankenhaushygiene beeinflusst werden, sollte mindestens ein Krankenhaushygieniker bzw. eine Hygienefachkraft eingebunden werden.

Hinsichtlich der Einbindung bestimmter Expertinnen und Experten nach der Art der Berufsgruppe ist auch die Anzahl der jeweiligen Expertinnen und Experten abhängig vom jeweiligen Versorgungsbereich und sollte jeweils bundeseinheitlich geregelt sein.

#### **Assoziierte Mitglieder**

Für spezifische Fragestellungen können zusätzlich auch weitere Personen für die jeweilige Sitzung und nicht über ein Verfahrensjahr hinausgehend von der LAG bzw. Bundesstelle beratend hinzugezogen werden. Dies können z. B. Assistenzärztinnen und -ärzte, medizinische Dokumentarinnen und Dokumentare und andere Vertreterinnen und Vertreter von Berufsgruppen der Gesundheitsversorgung sein, aber auch Expertinnen und Experten, die sich im Ruhestand noch aktiv an der Qualitätssicherung ihres Fachgebiets beteiligen. Die in Abschnitt 6.7.1.2 genannten formalen Anforderungskriterien sind für die assoziierten Mitglieder einer Fachkommission ebenfalls zu beachten. Die Berufung der assoziierten Mitglieder erfolgt durch die LAG bzw. Bundesstelle. Die Gründe für die Berufung der Expertinnen und Experten macht die LAG bzw. Bundesstelle in geeigneter Form transparent.

Die LAG bzw. Bundesstelle trägt die Verantwortung für die Prüfung und Einhaltung sowohl der Kriterien für die Auswahl der Expertinnen und Experten als auch der Kriterien zur Zusammensetzung der Fachkommission und veröffentlicht diese für die jeweilige Fachkommission auf der Webseite der Geschäftsstelle der LAG bzw. des IQTIG. Wenn von den obigen Kriterien abgewichen werden musste (z. B. da zu wenig Fachexpertinnen bzw. Fachexperten mit entsprechender fachlicher Qualifikation zur Verfügung standen), sind die Abweichung und deren Gründe zu dokumentieren und zu veröffentlichen.

---

<sup>73</sup> Zum Beispiel Qualitätsindikatoren zur Knie- und Hüftbeweglichkeit durch Maßnahmen der Physiotherapeutinnen und Physiotherapeuten.

## 6.8 Fazit

Das in den vorherigen Abschnitten beschriebene Vorgehen für die fachliche Bewertung stellt eine konsequente Umsetzung des Ziels dar, möglichst explizite Verfahrensregeln zu formulieren und auf heuristische Beurteilungsprozesse nur dort zurückzugreifen, wo sie unvermeidlich erscheinen. Es wurde dargestellt, wie ein Maximum an Standardisierung und damit Objektivität der fachlichen Bewertung theoretisch erreicht werden könnte. Dabei wurde deutlich, dass die Steigerung der Objektivität des Verfahrens durch Standardisierung mittels expliziter Regeln an Grenzen stößt. Erstens lassen sich aufgrund der qualitativen Informationsgrundlage in den Stellungnahmen nicht alle Schritte der fachlichen Bewertung standardisieren. Zweitens geht eine höhere Standardisierung auch mit teilweise komplexeren Prozessen einher (vgl. Abschnitt 3.6). So müssen dafür beispielsweise die in der Stellungnahme vorgebrachten Gründe systematisch klassifiziert werden, eine Nachberechnung vorgenommen und ein Abgleich mit einem (modifizierten) Referenzwert durchgeführt werden. Der im Anhang in Kapitel 8 dargestellte Gesamtalgorithmus bei einem möglichst expliziten Vorgehen verdeutlicht diese Komplexität. Allerdings müssten die meisten der hier vorgeschlagenen, explizit formulierten Schritte auch bei einem rein heuristischen Vorgehen so oder so ähnlich angewendet werden (siehe Abschnitt 6.6). Wird beispielsweise bei der Prüfung, ob der Ausschluss bestimmter Fälle dazu führt, dass der Indikatorwert den Referenzbereich erreicht, nicht nach explizit formulierten Regeln vorgegangen, ist das Vorgehen nur schwer überprüfbar und fehleranfällig.

Wie in Abschnitt 3.6 angedeutet, sollte für die Umsetzung der Empfehlungen für die fachliche Bewertung eine Kosten-Nutzen-Abwägung durchgeführt werden. Der Nutzen eines möglichst expliziten Vorgehens in Form größtmöglicher Objektivität und Transparenz muss gegenüber den Kosten, die die Umsetzung des Vorgehens bedeuten würde, abgewogen werden. Hinweise, wo mit vermehrtem Aufwand im Vergleich zum bisherigen Vorgehen zu rechnen ist, wurden in Abschnitt 6.6 ausgeführt. Der Nutzen von Entscheidungsregeln, die explizit gemacht werden, besteht u. a. darin, dass sie auch für Kritik zugänglich werden (Dawes et al. 1989). Implizite Entscheidungsregeln dagegen können nicht überprüft und demnach auch nicht kritisiert werden. Wie in Kapitel 2 ausgeführt, werden die Ergebnisse der indikatorgestützten externen Qualitätssicherung sowohl für die Qualitätsförderung als auch für Zwecke der *accountability* eingesetzt. Dabei stellt der zweitgenannte Verwendungsbereich besonders hohe Anforderungen an die Güte der Qualitätsmessungen (Chassin et al. 2010, Solberg et al. 1997). Die Güte der fachlichen Bewertung muss sich daher an diesen höheren Anforderungen an die Objektivität und Transparenz im Verwendungszweck *accountability* orientieren. Es wird daher empfohlen, in jedem Fall den expliziten Bewertungsalgorithmus gemäß Abbildung 33 für die Einstufung der Ergebnisse in „Qualitätsdefizit“ und „Hinweis auf Qualitätsdefizit entkräftet“ umzusetzen. Für die Klassifikation in die Ziffern des Bewertungsschemas (Unterkategorien) kann dagegen statt der empfohlenen expliziten Bewertungsalgorithmen (Abschnitte 6.4.4 und 6.4.5) auch ein heuristisches Vorgehen umgesetzt werden, wenn der Gewinn an Objektivität und Transparenz gegenüber dem Aufwand für die Umsetzung weniger stark gewichtet wird. Allerdings führt der Verzicht auf eine Differenzierung hier nur zu einer geringen Aufwandsreduktion.

## 7 Modul Qualitätsförderung

Maßnahmen zur Verbesserung der Versorgungsqualität stellen neben der Messung und Darstellung der Versorgungsqualität die zweite wichtige Säule der gesetzlich verpflichtenden Qualitätssicherung dar. Die empfohlene Standardisierung des Mess- und Bewertungsprozesses sowie die Trennung zwischen den Modulen „Qualitätsbewertung“ und „Qualitätsförderung“ sollen dazu beitragen, dass den LAG bzw. der Bundesstelle und den Leistungserbringern mehr Ressourcen für qualitätsverbessernde Maßnahmen zu Verfügung stehen (siehe hierzu auch Kapitel 8). Das Modul „Qualitätsförderung“ beinhaltet diejenigen Maßnahmen, die nach Abschluss des Moduls „Qualitätsbewertung“ zur Verbesserung der Versorgungsqualität eingesetzt werden können. Für die Verbesserung der Versorgungsqualität werden insbesondere die dialogischen Anteile des bisherigen Strukturierten Dialoges benötigt. Hierzu zählen z. B. Besprechungen und Begehungen. Während ein qualitatives, dialogisches Vorgehen für objektive und standardisierte Qualitätsmessungen als weniger geeignet angesehen wird, soll es berechtigter Weise im Rahmen der Qualitätsförderung in besonderem Maß zum Einsatz kommen. Für eine effektive Qualitätsförderung müssen Sachverhalte erklärt und Hypothesen über geeignete Maßnahmen zur Qualitätsverbesserung entwickelt werden, wozu sich in erster Linie qualitative Methoden eignen (IQTIG 2019a: Abschnitt 5.1). Folglich wird eine Verschiebung dieser Anteile vom Modul der Qualitätsbewertung zum Modul der Qualitätsförderung empfohlen.

Die auf die Qualitätsbewertung aufbauende Verbesserung der Versorgungsqualität durch den Leistungserbringer entspricht teilweise der Verwendung von Qualitätsmessungen im Rahmen der *improvement* Strategie nach Berwick et al. (Berwick et al. 2003). Dabei wird davon ausgegangen, dass die Ergebnisse von Qualitätsmessungen Leistungserbringer in die Lage versetzen, Defizite in ihrer Versorgung zu erkennen und durch geeignete Maßnahmen selbst beheben. Diese Strategie spiegelt sich auch in den Zielen der datengestützten Qualitätssicherung wider, die auf Grundlage der Messung der Versorgungsqualität und durch Unterstützung des einrichtungsinternen Qualitätsmanagements ein höheres Qualitätsniveau der Versorgung erreichen soll (vgl. § 1 DeQS-RL). Dabei ist eine wesentliche Voraussetzung für das Einleiten geeigneter Maßnahmen die indikatorgestützte Ursachenanalyse zu den Versorgungsdefiziten, um eine Qualitätsverbesserung zu erzielen. Ein deutlicher Unterschied in dem Vorgehen nach DeQS-RL und der *improvement* Strategie nach Berwick et al. besteht jedoch in der Freiwilligkeit der qualitätsverbessernden Maßnahmen (Berwick et al. 2003). Während die *improvement* Strategie eher von der freiwilligen Veränderung der Versorgungsgestaltung durch die Leistungserbringer auf Basis von Qualitätsergebnissen ausgeht (die sowohl intrinsisch als auch extern motiviert sein kann), kann die LAG bzw. Bundesstelle Maßnahmen für einen Leistungserbringer beschließen (vgl. § 17, Abs. 5 DeQS-RL). Für eine selbstkritische Ursachenanalyse eines Leistungserbringers sollten möglichst günstige Bedingungen geschaffen werden. Diese wurden als sogenannte Erfolgsfaktoren für qualitätsverbessernde Maßnahmen beschrieben, und deren Anwendbarkeit im Rahmen der externen Qualitätssicherung wurde in Abschnitt 3.7 erörtert. Dort wurde u. a. hervorgehoben, dass eine klare Trennung von Qualitätsbewertung und Qualitätsförderung als notwendig erachtet wird, um den Erfolg qualitätsverbessernder Maßnahmen zu steigern. Durch

diese Trennung befinden sich auch die Bewertenden in einer klaren Position, die dadurch gekennzeichnet ist, dass sie in einem kollegialen Dialog mit dem Leistungserbringer geeignete Maßnahmen entwickeln können ohne gleichzeitig eine strenge Prüfung der Versorgungsqualität vornehmen zu müssen.

Da der Schwerpunkt der Beauftragung auf der Standardisierung der Qualitätsbewertung liegt, werden im Folgenden hauptsächlich konkrete Empfehlungen für die zwingende Einleitung von qualitätsverbessernden Maßnahmen ausgesprochen. Eine Ausnahme stellen die sich anschließenden Empfehlungen zum Abschluss von Zielvereinbarungen (siehe Abschnitt 7.2) dar. Das IQTIG hat dem G-BA im Jahr 2018 ein Zielvereinbarungskonzept für die bundesbezogenen Verfahren vorgelegt, das insbesondere Empfehlungen zur Rückkopplung bezüglich des Standes der Zielerreichung zwischen IQTIG und dem Unterausschuss des G-BA beinhaltet. Um auch hier eine bundesweite Vereinheitlichung der Vorgehensweise zu erreichen, wird dieses Zielvereinbarungskonzept (siehe Abschnitt 7.2) auch für die Verwendung in den landesbezogenen Verfahren empfohlen.

### **7.1 Empfehlungen zur Einleitung qualitätsverbessernder Maßnahmen**

Maßnahmen zur Qualitätsverbesserung sollten auf eine abgeschlossene und aussagekräftige Qualitätsbewertung aufbauen. Es wird grundsätzlich davon ausgegangen, dass das interne Qualitätsmanagement der Leistungserbringer ein Qualitätsdefizit zum Anlass nimmt, Prozesse und Strukturen zu prüfen und bei Bedarf anzupassen, um das Qualitätsdefizit zu beheben. Nach Abschluss der Qualitätsbewertung wird dem internen Qualitätsmanagement bei erstmaliger Feststellung eines Qualitätsproblems zunächst die Möglichkeit eingeräumt, intern und damit eigenverantwortlich geeignete Maßnahmen einzuleiten ohne dass die LAG bzw. Bundesstelle tätig wird. Nur bei wiederholten Qualitätsmängeln soll die LAG bzw. die Bundesstelle aktiv werden, da wiederholte Qualitätsdefizite bei einem oder inhaltlich ähnlichen Indikatoren<sup>74</sup> darauf hindeuten, dass die leistungserbringerinternen Anstrengungen, das Qualitätsproblem zu beheben, nicht ausreichend wirksam waren. Handelt es sich daher um ein wiederholtes Qualitätsdefizit, wird empfohlen, zwingend eine oder mehrere Maßnahmen der Maßnahmenstufe 1 und/oder der Maßnahmenstufe 2 gemäß § 17 Abs. 3 der DeQS-RL im Sinne einer Qualitätsförderung durchzuführen (siehe Abbildung 38).

---

<sup>74</sup> Auf Anregung der Landesebene (siehe Anhang, Kapitel 1) sollten die LAG auch bei wiederholten Qualitätsdefiziten bei inhaltlich ähnlichen Indikatoren die Möglichkeit haben, Maßnahmen einzuleiten. Die LAG bzw. Bundesstelle entscheidet unter Einbezug der Fachkommission, ob es sich um inhaltlich ähnliche Indikatoren handelt.

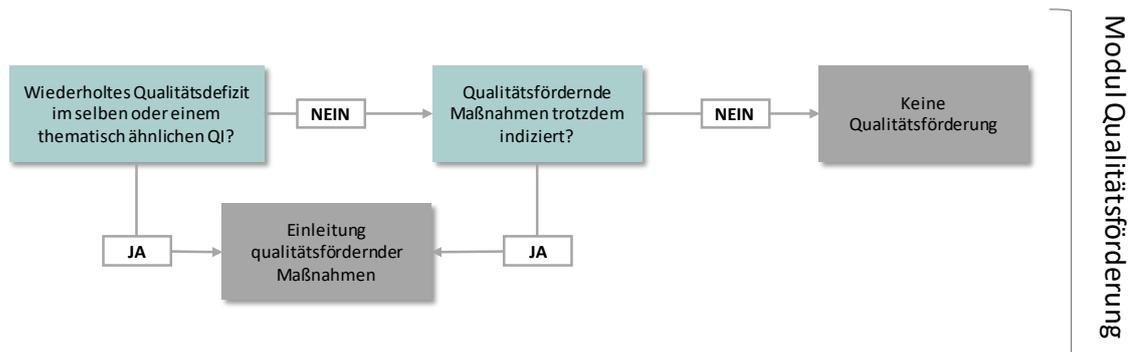


Abbildung 38: Empfehlungen zur Einleitung qualitätsverbessernder Maßnahmen (Ausschnitt aus Abbildung 9 in Kapitel 4)

Es wird daher unter Berücksichtigung der Angaben der LAG bzw. Bundesstelle zur Notwendigkeit entsprechender Maßnahmen empfohlen, dass ab dem zweiten Qualitätsdefizit bei einem Indikatorergebnis oder bei inhaltlich ähnlichen Indikatoren innerhalb von drei Jahren die LAG bzw. Bundesstelle immer qualitätsverbessernde Maßnahmen einleitet (siehe Anhang 1). Die LAG bzw. Bundesstelle kann bereits bei einem erstmalig auftretenden Qualitätsdefizit unter Angabe von Gründen jederzeit in einem Qualitätsindikator qualitätsverbessernde Maßnahmen bei einem Leistungserbringer einleiten (siehe Abbildung 38).

In § 17 Abs. 3 DeQS-RL werden qualitätsfördernde Maßnahmen in zwei Stufen unterteilt. Zu den Maßnahmen Stufe 1 zählen Vereinbarungen zwischen der LAG bzw. Bundesstelle und dem Leistungserbringer bezüglich:

- der Teilnahme an Fortbildungen, Fachgesprächen und Kolloquien,
- der Teilnahme an Qualitätszirkeln,
- der Implementierung von Behandlungspfaden,
- der Durchführung von Audits,
- der Durchführung von Peer Reviews und
- der Implementierung von Handlungsempfehlungen anhand von Leitlinien.

Nach den hier vorliegenden Empfehlungen sollen Gespräche und Begehungen zukünftig nicht mehr im Rahmen des Stellungnahmeverfahrens vorgenommen werden. Daher wird empfohlen, diese qualitätsverbessernde Maßnahmen der Maßnahmenstufe 1 zuzuordnen.

Am 18. April 2019 hat der G-BA die Richtlinie zur Förderung der Qualität und zu Folgen der Nichteinhaltung sowie zur Durchsetzung von Qualitätsanforderungen (Qualitätsförderungs- und Durchsetzungs-Richtlinie; QFD-RL) beschlossen. Diese gilt auch für die QS-Verfahren der DeQS-RL. Eine Konkretisierung anhand themenspezifischer Empfehlungen für die einzelnen QS-Verfahren soll gemäß der QFD-RL vorgenommen werden. Hinsichtlich der Maßnahmen zur Qualitätsförderung sieht die QFD-RL im Vergleich zu der Maßnahmenstufe 1 der DeQS-RL folgende weitere Beratungs- und Unterstützungsleistungen vor:

- Schriftliche Empfehlung
- Zielvereinbarungen
- Implementierung von Vorgaben für das interne Qualitätsmanagement

- Implementierung von Standard Operating Procedures (SOPs)
- Prüfung unterjähriger Auswertungsergebnisse

Beide Richtlinien sehen die Möglichkeit vor, im Anschluss an die Maßnahmen der Stufe 1 unter bestimmten Umständen weitere Maßnahmen der Stufe 2 bzw. Durchsetzungsmaßnahmen (§ 5 QFD-RL) zu veranlassen. Dazu gehören u. a. die Information zur Bewertung des Leistungserbringers und sich daraus ergebende Empfehlungen für Vergütungsabschläge des Leistungserbringers (§ 17 Abs. 4 DeQS-RL, § 5 QFD-RL)

Über die Vorgaben der beiden Richtlinien hinaus wird empfohlen, die Möglichkeit einzuräumen, qualitätsverbessernde Maßnahmen einzuleiten, wenn diese vom Leistungserbringer oder von der LAG bzw. Bundesstelle für notwendig und sinnvoll erachtet werden, um eine weitere Verbesserung der Qualität zu erreichen.

Der LAG bzw. Bundesstelle und der Fachkommission kommt damit hinsichtlich der Qualitätsförderung eine entscheidende Rolle zu. Qualitätsverbessernde Maßnahmen sollten möglichst individuell auf den einzelnen Leistungserbringer angepasst sein, um das spezifische Qualitätsproblem optimal zu adressieren. Das bedeutet, dass die LAG bzw. Bundesstelle unter Beteiligung der Fachkommission nach Erstellung einer Ursachenanalyse durch den Leistungserbringer mit diesem in Kontakt tritt, um gemeinsam geeignete Maßnahmen zu erörtern. Der Kontakt kann z. B. im Rahmen einer Besprechung oder Begehung stattfinden. Zudem trägt die LAG bzw. Bundesstelle die Verantwortung für die Umsetzung der externen Qualitätssicherung. Daher hat sie, unterstützt von der Fachkommission, neben der Bewertung der Versorgungsqualität darüber zu entscheiden, ob, wann und falls ja, welche qualitätsverbessernde Maßnahmen eingeleitet werden müssen und wie diese umzusetzen sind. Zudem wird empfohlen, den Einsatz und Umsetzungsgrad der qualitätsverbessernde Maßnahmen zu dokumentieren (siehe Kapitel 9). Die Geschäftsstelle der LAG bzw. das IQTIG hat dem Lenkungsgremium regelmäßig über den Stand der Maßnahmen zu berichten.

In Abschnitt 3.7 wurde die fachliche Kompetenz der Peers als Erfolgsfaktor für qualitätsverbessernde Maßnahmen identifiziert (siehe auch die Empfehlungen in Abschnitt 6.6). Da die fachliche Kompetenz der Peers nicht nur im Rahmen der Qualitätsbewertung für die Akzeptanz des Gesamtverfahrens von Bedeutung ist, sondern auch im Setting der Qualitätsförderung, erscheint es zukünftig von Bedeutung ein „Lernen von den Besseren“ für diesen Bereich zu etablieren. Ein erster Schritt und zugleich die Voraussetzung, um ein Lernen von den Besseren zu ermöglichen, ist die in Abschnitt 6.1 ausgesprochene Empfehlung, die Methodik dahingehend weiterzuentwickeln, dass zureichende Qualität bezogen auf das Qualitätsziel eines Indikators festgestellt werden kann. Somit bestünde grundsätzlich die Möglichkeit, diejenigen Leistungserbringer, bei denen mit hoher Sicherheit davon ausgegangen werden kann, dass sie die Qualitätsanforderungen erfüllt haben, zu identifizieren und als Peers für qualitätsverbessernde Maßnahmen bei anderen Leistungserbringern zu gewinnen. So lange das Verfahren zur Qualitätsmessung jedoch nur das Nicht-Erreichen von Qualitätsanforderungen feststellen kann, ist ein Lernen von den Besseren nur schwierig zu realisieren.

## 7.2 Zielvereinbarungen

Im Rahmen von Zielvereinbarungen einigen sich ein Leistungserbringer und die LAG bzw. Bundesstelle auf Empfehlungen zur Behebung von Qualitätsdefiziten (fachliches Ziel) in einem festgelegten Zeitraum (Zeitziel) sowie ggf. auf begleitende Maßnahmen, die dieser Zielerreichung dienlich sind (flankierende Umsetzungsziele).

Abgeschlossene Zielvereinbarungen weisen dem Leistungserbringer also konkrete Handlungsempfehlungen zu, die für eine Verbesserung der Versorgungsqualität von den Beteiligten als notwendig erachtet werden. Bis zur Zielerreichung wird der weitere Verlauf vom internen Qualitätsmanagement der Einrichtung beobachtet und der Erfolg der Maßnahmen dem Lenkungsgremium der LAG nach § 5 DeQS-RL berichtet.

### 7.2.1 Formulierung von Zielvereinbarungen

Als erster Schritt zum Abschluss von Zielvereinbarungen erfragt die LAG bzw. Bundesstelle, welche Maßnahmen die Einrichtung vorschlägt, um bestehende Qualitätsdefizite zu beheben. Auch die prüfende Fachkommission kann z. B. in einer ihrer Sitzungen Rückmeldungen geben, wo sie Möglichkeiten für Verbesserungen sieht. Diese Überlegungen finden in kollegialer Weise statt und die Vorschläge werden gesammelt und in Zielvereinbarungen zusammengetragen.

Zielvereinbarungen sollten die folgenden Hauptinhalte enthalten:

- kurze Zusammenfassung der Situation
- Aufzählung der mit einem Qualitätsdefizit bewerteten Indikatoren, einschließlich des Zeitraums, innerhalb dessen erste Besserungen eintreten sollen
- Aufzählung der von der Einrichtung weiter durchzuführenden Ursachenanalysen der mit einem Qualitätsdefizit bewerteten Indikatoren einschließlich der Berichtszeitpunkte
- Aufzählung der vereinbarten Maßnahmen zur Unterstützung der Behebung der Qualitätsdefizite einschließlich der Berichtszeitpunkte über die Durchführung der Maßnahmen und der Erfolgskontrollen
- weitere spezifische Vereinbarungen, sofern erforderlich

Folgende ergänzende Inhalte können Zielvereinbarungen darüber hinaus enthalten:

- Vereinbarung zusätzlicher Maßnahmen zur weiteren Klärung der Ursachen des Qualitätsdefizits, z. B. Bereitstellung von weiteren Materialien und/oder Begehung vor Ort
- Vereinbarung flankierender unterstützender Maßnahmen (z. B. Peer Review, Einbindung von Expertinnen und Experten)
- Vereinbarung zur Lieferung weiterer Informationen durch den Leistungserbringer, die den Fortgang der Verbesserungsmaßnahmen dokumentieren (interne Statistiken, revidierte SOPs, Bericht zu strukturellen Maßnahmen etc.)
- Vereinbarung von Maßnahmen zur Abschluss- oder Nachhaltigkeitskontrolle

### 7.2.2 Übersicht über den Stand der Zielvereinbarungen

Die Geschäftsstelle der LAG bzw. das IQTIG gibt dem Lenkungsgremium nach routinemäßig oder bei besonderem Anlass eine Übersicht zum Sachstand aller jeweils aktiven Zielvereinbarungen

(siehe Kapitel 9). Darüber hinaus wird zu Zielvereinbarungen, in deren Verlauf ein Qualitätsdefizit bei Indikatoren gemessen wird, gesondert berichtet.

Im Anhang findet sich ein Formular zur Standardisierung eines solchen Einzelberichts mit Hinweisen und Erläuterungen zur Nutzung (siehe Anhang, Kapitel 8).

Zu jeder seiner Sitzungen wird dem Lenkungsgremium über den aktuellen Stand von Zielvereinbarungen berichtet. Dies erfolgt in tabellarischer Form und unter Wahrung der Anonymität der Leistungserbringer.

Aufgeführt wird:

- die Anzahl der zum jeweils aktuellen Zeitpunkt aktiven Zielvereinbarungen
- die Anzahl der aktiven Zielvereinbarungen, die zum aktuellen Zeitpunkt eingehalten werden
- die Anzahl der aktiven Zielvereinbarungen, die zum aktuellen Zeitpunkt nicht eingehalten werden
  - bei Nicht-Einhaltung von mind. einer Zielvereinbarung wird dies in der Spalte „Erläuterung“ näher beschrieben
- die Anzahl der von der LAG vorgesehenen, jedoch nicht abgeschlossenen Zielvereinbarungen
  - Der Hintergrund, weshalb Zielvereinbarungen nicht abgeschlossen wurden, wird in der Spalte „Erläuterung“ näher beschrieben (z. B. Verweigerung der Unterzeichnung durch den Leistungserbringer)

*Tabelle 15: Vorlage für den Bericht an das Lenkungsgremium zum aktuellen Stand aktiver Zielvereinbarungen*

	Anzahl	Erläuterung
<b>aktive Zielvereinbarungen</b>		...
davon: Zielvereinbarung zum aktuellen Zeitpunkt eingehalten		...
davon: Zielvereinbarung zum aktuellen Zeitpunkt nicht eingehalten		...
<b>geplante, nicht abgeschlossene Zielvereinbarungen</b>		...
<b>seit dem letzten Bericht abgeschlossene Zielvereinbarungen</b>		...

## 8 Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens

Die Beauftragung sieht vor, dass das IQTIG Optionen zur Verkürzung des Verfahrens<sup>75</sup> prüft. Gemäß der hier vorliegenden Ausführungen und Empfehlungen (siehe Kapitel 2, 4 und 7) soll das Verfahren zukünftig in die verschiedenen Module Qualitätsbewertung, Bewertung der Dokumentationsqualität und Qualitätsförderung aufgeteilt werden. Da aussagekräftige Qualitätsbewertungen die Ausgangsbasis für alle folgenden Verwendungszwecke darstellen, wurden Optionen zur Verkürzung des Verfahrens vor allem mit Blick auf das Modul Qualitätsbewertung geprüft. Erstens ist aus der Perspektive der Patientinnen und Patienten eine zeitnahe Verfügbarkeit der abschließenden Qualitätsbewertung wünschenswert, beispielsweise für Auswahlentscheidungen. Eine zeitnahe Qualitätsbewertung würde zweitens auch eine frühere Einleitung von qualitätsverbessernden Maßnahmen und mehr Zeit für entsprechende Maßnahmen ermöglichen, deren Erfolg unter Umständen dann bereits im Jahr nach der Datenerfassung (Erfassungsjahr+1) (EJ+1) quantifizierbar wäre.

### 8.1 Das Verfahren gemäß DeQS-RL

Da neben dem Stellungnahmeverfahren selbst auch die vor- und nachgelagerten Phasen einen Einfluss auf die Dauer des Stellungnahmeverfahrens haben, werden in Abbildung 39 die Phasen des Verfahrens sowie die Fristen der DeQS-RL schematisch dargestellt. Für die Qualitätsindikatoren der Verfahren nach DeQS-RL werden zukünftig unterschiedliche Datenquellen herangezogen werden (einrichtungs- und fallbezogene QS-Dokumentation, Sozialdaten und Patientenbefragungen). Da bisher jedoch nur einrichtungs- und fallbezogenen QS-Dokumentation eingesetzt werden, beziehen sich die nachfolgenden Kapitel nur auf diese Datenquellen.

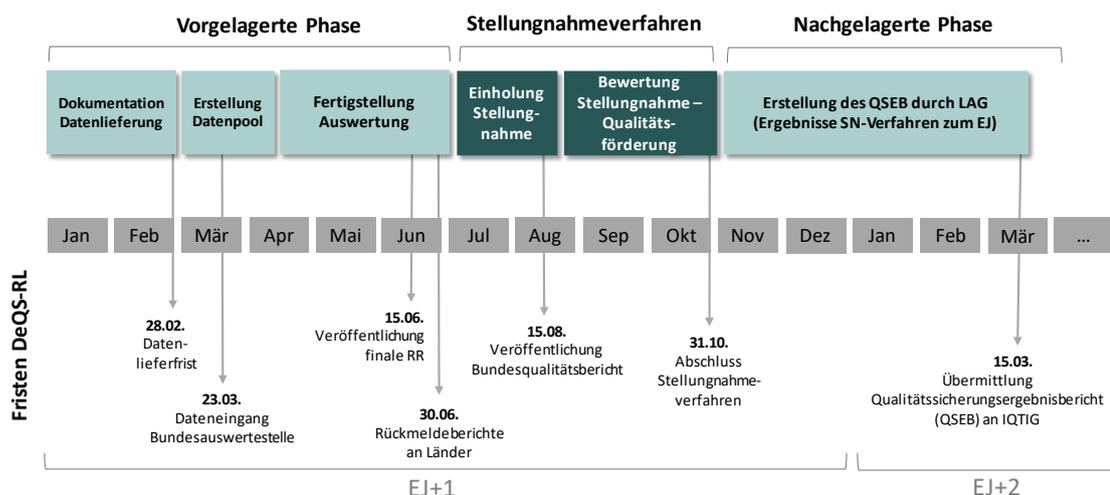


Abbildung 39: Phasen und Fristen des Verfahrens gemäß DeQS-RL.

<sup>75</sup> Unter Verfahren wird das Gesamtverfahren von der Dokumentation und Datenlieferung bis zur Berichterstattung verstanden.

In der vorgelagerten Phase können drei Prozesse unterschieden werden. Hierzu zählen die Dokumentation und Datenlieferung der Leistungserbringer bzw. die Datenübermittlung an die Bundesauswertungsstelle. Gemäß den themenspezifischen Bestimmungen der DeQS-RL (für QS WI und QS PCI) ist die Frist für die Datenlieferung des gesamten Erfassungsjahres (EJ) der 28. Februar des darauffolgenden Jahres (EJ + 1). Nach Ablauf der Korrekturfrist am 15. März werden die Daten von der Vertrauensstelle pseudonymisiert, sodass die Daten erst am 23. März bei der Bundesauswertungsstelle eingehen. Anschließend folgt die Erstellung des Bundesdatenpools durch das IQTIG. Der letzte Schritt in der vorgelagerten Phase ist die Fertigstellung der Qualitätsindikatoren datenbank (QIDB) und die Auswertungserstellung gemäß den finalen Rechenregeln (einschließlich der Entwicklung von Risikoadjustierungsmodellen anhand der Daten aus dem Erfassungsjahr) bis zum 15. Juni des EJ + 1. Der Versand der Rückmeldeberichte an die LAG erfolgt im Anschluss bis 30. Juni des EJ + 1. Parallel zu den genannten Prozessen erfolgt die Erstellung des Bundesqualitätsberichts, der am 15. August des EJ + 1 veröffentlicht wird und der neben den reinen Indikatorergebnissen des EJ auch die Ergebnisse des Stellungnahmeverfahrens zum Bezugsjahr EJ – 1 enthält (siehe Abbildung 39).

Im Stellungnahmeverfahren, welches gemäß DeQS-RL für die Krankenhäuser bis zum 31. Oktober des EJ + 1 abzuschließen ist, können zwei Prozesse unterschieden werden. Der erste Prozess bezieht sich auf das Einholen der Stellungnahmen, in welchem die Stellungnahmen angefordert, erstellt und abgegeben werden. Im sich anschließenden Prozess findet die Bewertung der Stellungnahmen sowie die Entscheidung darüber, welche qualitätsverbessernde Maßnahmen eingeleitet und umgesetzt werden sollen sowie die Umsetzung (durch die LAG bzw. das IQTIG z. B. in Form kollegialer Gespräche, Begehungen, Zielvereinbarungen etc.) selbst statt. Da analog zur QSKH-RL auch gemäß DeQS-RL die Qualitätsbewertung nicht zwingend vor den Maßnahmenstufen abgeschlossen sein muss, beinhaltet der letzte Prozess eben auch die Umsetzung der Maßnahmen der Qualitätsverbesserung.

In der nachgelagerten Phase erfolgt die Veröffentlichung der qualitativen Bewertung nach dem Stellungnahmeverfahren und der durchgeführten oder zur Durchführung empfohlenen Qualitätssicherungsmaßnahmen im Bundesqualitätsbericht bis zum 15. August. Um den Bundesqualitätsbericht erstellen zu können werden die entsprechenden Ergebnisse im Qualitätssicherungsergebnisbericht (QSEB) zusammengefasst und bis zum 15. März des EJ + 2 an das IQTIG übermittelt (siehe Abbildung 39).

## **8.2 Empfehlungen zur Verkürzung des Stellungnahmeverfahrens**

Eine Verkürzung des Stellungnahmeverfahrens kann grundsätzlich auf drei unterschiedliche Weisen erreicht werden. Eine Möglichkeit ist, den Aufwand für die LAG bzw. Bundesstelle zu reduzieren, indem z. B. die Anzahl zu prüfender Qualitätsindikatorergebnisse verringert wird. Dies kann über die Anzahl an ausgewerteten Qualitätsindikatoren sowie über die Methodik zur Berechnung der quantitativen Auffälligkeitseinstufung für diese Qualitätsindikatoren gesteuert werden. Dadurch stünden mehr Ressourcen für die Bearbeitung der verbleibenden Indikatorergebnisse zur Verfügung und die Bearbeitung könnte unter Umständen beschleunigt werden. Eine zweite Möglichkeit besteht darin, die Prozesse des Stellungnahmeverfahrens effizienter zu

gestalten. Drittens besteht die Möglichkeit, mehr Ressourcen zur Verfügung zu stellen, z. B. indem mehr Personal bei den LAG eingesetzt wird, um die Bearbeitung zu beschleunigen. Im Folgenden werden Optionen zur Verkürzung des Verfahrens über eine Aufwandsreduktion und Effizienzsteigerung der Prozesse ausgehend von den hier vorgelegten Empfehlungen und den Rückmeldungen der Vertreterinnen und Vertreter der LAG und LQS im Rahmen des Workshops zur Weiterentwicklung des Strukturierten Dialogs erörtert.

### **8.2.1 Aufwandsreduktion durch Volumenreduktion**

Eine Reduktion des Aufwands für die LAG bzw. Bundesstelle kann einerseits erreicht werden, indem z. B. die Anzahl an Qualitätsindikatorergebnissen, für die ein Stellungnahmeverfahren einzuleiten ist, verringert wird. Diesbezüglich werden in Kapitel 5 Empfehlungen ausgesprochen, die die Anzahl an quantitativ auffälligen Ergebnissen im Vergleich zur bisher verwendeten Methodik der rechnerischen Auffälligkeit zielgerichtet reduzieren könnte. Dadurch würde sich auch der Arbeitsaufwand bei den LAG, der Bundesstelle und Leistungserbringern im Rahmen des Stellungnahmeverfahrens reduzieren. Allerdings soll zukünftig die Möglichkeit, Hinweise zu versenden, entfallen und bei allen Qualitätsindikatorergebnissen, die einen hinreichenden Hinweis für ein Qualitätsdefizit aufweisen, soll obligatorisch eine Stellungnahme eingeholt werden. Inwieweit eine Aufwandsreduktion bei spezifischerer Auslösung eines Stellungnahmeverfahrens einerseits und gleichzeitig obligatorischer Einholung einer Stellungnahme andererseits erreicht werden kann, wird beispielhaft am QS-Verfahren HEP in Abschnitt 10.4 vorgestellt.

Eine Aufwandsreduktion kann außerdem auch durch eine Verringerung der Anzahl an Qualitätsindikatoren erreicht werden. Das IQTIG sieht vor, sukzessive die bestehenden Qualitätsindikatoren anhand der Eignungskriterien (siehe Methodische Grundlagen V1.1) zu prüfen und gegebenenfalls Empfehlungen zur Aussetzung von Indikatoren auszusprechen (siehe auch Abschnitt 11.1). Im Workshop zur Einbindung der Vertreterinnen und Vertreter der LAG und LQS äußerten diese, dass viele Indikatoren aus fachlicher Sicht nicht mehr geeignet seien und daher eine Reduzierung der Anzahl dieser Indikatoren als notwendig erachtet würde (siehe Anhang, Kapitel 1).

### **8.2.2 Effizienzsteigerung durch Standardisierung im Vorgehen**

Das vorliegende Konzept sieht im Sinne einer Standardisierung und Effizienzsteigerung ein einheitliches schriftliches Stellungnahmeverfahren zur Qualitätsbewertung vor. Eine Standardisierung des Verfahrens erfordert für die Einholung der Stellungnahmen konkrete formale und inhaltliche Anforderungen an die Stellungnahmen der Leistungserbringer (siehe Abschnitt 6.5). Es wird außerdem empfohlen, dass die LAG bzw. die Bundesstelle bei Nicht-Einhaltung der Anforderungen direkt eine abschließende Bewertung ohne weitere Einholung von Präzisierungen vornehmen soll (siehe Kapitel 6). Diese Empfehlung soll die LAG bzw. Bundesstelle einerseits dadurch entlasten, dass diese die Anonymisierung und Pseudonymisierung nicht nachträglich selbst vornehmen müssen, falls dies seitens der Leistungserbringer nicht erfolgt ist. Andererseits würde dieses Vorgehen den Aufwand der LAG bzw. Bundesstelle auch dadurch reduzieren, dass eine weitere inhaltliche Prüfung aufgrund der Einholung von Präzisierungen entfallen würde. Dabei ist zu berücksichtigen, dass auch die formale Prüfung einen gewissen Aufwand darstellt, der jedoch geringer ausfallen dürfte als eine inhaltliche Prüfung von Präzisierungen.

Neben den Empfehlungen zu Fristen im Rahmen der formalen Anforderungen an das Stellungnahmeverfahren wurden in den vorherigen Kapiteln weitere Empfehlungen zur Standardisierung des Bewertungsprozesses gemacht. Eine Standardisierung von Prozessen steigert z. B. dadurch die Effizienz, dass Entscheidungen, die zuvor von den LAG bzw. der Bundesstelle (Versand von Hinweisen) oder den Fachkommissionen getroffen wurden, zukünftig einheitlich geregelt wären. So entfielen beispielsweise durch die Einholung einer Stellungnahme bei Vorliegen eines hinreichenden Hinweises auf ein Qualitätsdefizit der Bedarf für eine Fachkommissionssitzung, da die Entscheidung über die Einholung einer Stellungnahme nach statistischen Kriterien einheitlich getroffen würde. Auf der anderen Seite geht die empfohlene Umstellung der fachlichen Bewertung weg von einem bisher eher heuristischen Vorgehen hin zu einem stärker expliziten Vorgehen einschließlich Nachberechnung des Indikatorwerts und Vergleich mit dem Referenzwert vermutlich in der Anfangsphase mit einem höheren Aufwand als bisher einher (vgl. Abschnitt 6.6).

### **8.2.3 Auswirkungen der Empfehlungen auf den Beginn der Bewertung von Stellungnahmen**

Die hier vorgelegten Empfehlungen zu bundesweit einheitlichen Fristen für die Abgabe einer Stellungnahme von fünf Wochen würden vor dem Hintergrund des Beginns des Stellungnahmeverfahrens am 30. Juni des EJ + 1 einen klar definierten Stellungnahmezeitraum ohne Zeitverzug ermöglichen. Nach Eingang der Rückmeldeberichte bei den LAG benötigen diese Zeit, die Aufforderungsschreiben zu erstellen und zu versenden, sodass die Aufforderung zur Stellungnahme eine Woche später (am 07. Juli) beim Leistungserbringer eingehen könnten. Dieses Zeitfenster erscheint auch vor dem Hintergrund der vorliegenden Empfehlungen, die bei bestehendem hinreichendem Hinweis auf ein Qualitätsdefizit keinen Ermessensspielraum bei der Einholung einer Stellungnahme erlauben, als ausreichend. Anders als in den themenspezifischen Bestimmungen gemäß § 12 DeQS-RL festgelegt, würde damit die Einholung der Stellungnahmen nicht erst von der Fachkommission beraten werden, sondern unmittelbar durch die LAG auf Basis des statistischen Kriterien erfolgen, woraus sich eine deutliche Zeitersparnis ergeben dürfte (siehe Abschnitt 8.3). Zudem ist eine Aufwandsreduktion dadurch möglich, dass die Aufforderungen zur Stellungnahme nicht mehr individuell angepasst werden müssen, sondern einheitlich für alle QI wären (siehe Abschnitt 6.5 und Anhang, Kapitel 7). Somit müsste von den LAG ausschließlich die Auswertungen für einen Leistungserbringer zusammen mit einem einheitlichen Begleitschreiben weitergeleitet werden. Unter Berücksichtigung des Stellungnahmezeitraums von fünf Wochen und einem einwöchigen Zeitraum bis zum Eingang der Aufforderung zur Stellungnahme bei den Leistungserbringern könnte somit die Abgabefrist für Stellungnahmen bzw. die Bewertung der Stellungnahmen sechs Wochen nach dem 30.06. des EJ+1 also auf den 11. August beginnen.

### **8.2.4 Auswirkungen der Empfehlungen auf den Abschluss des Stellungnahmeverfahrens**

Eine frühere Verfügbarkeit der Qualitätsergebnisse kann auch über die in Kapitel 4 beschriebene konzeptuelle und chronologische Trennung von Qualitätsbewertung und -förderung erreicht werden. Diese Trennung würde neben einer Effektivitätssteigerung der Qualitätsförderung

(siehe Kapitel 4) auch den Aufwand im bisherigen Stellungnahmeverfahrens reduzieren, da der Prozess der Qualitätsförderung mit den Leistungserbringern in der nachgelagerte Phase zu verorten wäre.

Derzeit beträgt der Zeitraum für das gesamte Stellungnahmeverfahren ab Versand der Rückmeldeberichte an die LAG bis Abschluss des Stellungnahmeverfahrens gemäß DeQS-RL vom 30. Juni bis 31. Oktober, also vier Monate (siehe Abbildung 39). Unbekannt ist, innerhalb welches Zeitrahmens die LAG derzeit die Bewertung der Stellungnahmen einschließlich der Aufbereitung der Stellungnahmen vornehmen. Es ist jedoch bekannt, dass neben dem Zeitraum für die Bewertung der Stellungnahmen innerhalb des 4-monatigen Zeitraums außerdem folgende Aufgaben anfallen:

1. die Entscheidung, von welchen Leistungserbringern Stellungnahmen eingeholt werden sollen
2. die Erstellung und der Versand der Aufforderungsschreiben für die Stellungnahmen
3. der Stellungnahmezeitraum
4. mögliche Präzisierungen

Für 1. kann unter der Berücksichtigung, dass die Fachkommission gemäß DeQS-RL einzubeziehen ist (siehe Abbildung 40), vermutlich eine Woche geschätzt werden. Für 2. muss zum Eingang der Aufforderung zur Stellungnahme beim Leistungserbringer vermutlich mindestens eine weitere Woche geschätzt werden. Dieses Zeitintervall ist sicherlich stark von dem Grad der Automatisierung dieses Prozesses bei den LAG abhängig. Hinsichtlich 3. wurden im Workshop Zeiträume von vier oder fünf Wochen, als umsetzbar diskutiert (siehe Anhang, Kapitel 1). Für das Einholen und das Verfassen von Präzisierungen zu Stellungnahmen können weitere ein bis zwei Wochen geschätzt werden. Werden die geschätzten Zeitangaben von dem 4-monatigem Zeitraum subtrahiert, bedeutet das, dass für die Bewertung der Stellungnahmen und Präzisierungen (inkl. Aufbereitung der Stellungnahmen durch die LAG bzw. Bundesstelle) ca. acht Wochen zur Verfügung stehen.

Ausgehend von einem Start der Bewertung der Stellungnahmen am 11. August, der Verlagerung der Qualitätsförderung in die nachgelagerte Phase sowie des reduzierten Bearbeitungsbedarfs durch den Wegfall der Einholung und inhaltlichen Prüfung von Präzisierungen (siehe Abschnitt 8.2.2) erscheint eine Verkürzung der Dauer für die Bewertung der Stellungnahmen um einen Monat zunächst möglich. Folglich könnte das Stellungnahmeverfahren grundsätzlich am 30. September des EJ + 1 abgeschlossen werden (siehe Abbildung 42). Wie oben beschrieben steht diese Schätzung jedoch unter Vorbehalt, da keine Erfahrungen mit der Umsetzung der empfohlenen Methodik vorliegen. Da gerade in der Anfangsphase der Konzepteinführung jedoch mit einem höheren Aufwand zu rechnen ist, sollte der Zeitraum für die Bewertung der Stellungnahmen zunächst nicht verkürzt werden. Eine Verkürzung des Stellungnahmeverfahrens von vier auf drei Monaten sollte nur unter der Voraussetzung in Erwägung gezogen werden, dass Erfahrungen mit dem neuen Verfahren vorliegen, dieses mit einer entsprechenden IT-Unterstützung umgesetzt ist und das Stellungnahmeverfahrens vor der Sommerpause startet (siehe hierzu auch Abschnitt 8.4).

Das Modul Qualitätsförderung könnte sich direkt daran anschließen und parallel zur Berichterstattung für die Qualitätsbewertung erfolgen. Durch die Aufteilung der Prozesse in zwei Module stünde den LAG bzw. der Bundesstelle im Jahr nach der Erfassung somit mehr Zeit für die Einleitung und Umsetzung von Maßnahmen zur Qualitätsverbesserung zur Verfügung. Für den Gesamtprozess bestehend aus den Modulen der Qualitätsbewertung und der Qualitätsförderung stünden somit insgesamt sechs Monate zur Verfügung, also zwei Monate mehr als aktuell durch die DeQS-RL vorgegeben (vgl. Abbildung 39 und Abbildung 42). Zudem könnte der Erfolg mancher Maßnahmen aus dem Vorjahr ggf. bereits im EJ + 1 gemessen werden.

### 8.3 Vorschläge der Vertreterinnen und Vertreter der LAG und der LQS

Im Rahmen des Workshops zur Einbeziehung der Vertreterinnen und Vertreter der LAG und der LQS wurden zwei weitere Aspekte genannt, die eine Verkürzung des Verfahrens erreichen können.

Der erste Aspekt bezieht sich auf die aktuellen Entscheidungsprozesse und teilweise ungeklärten Zuständigkeiten innerhalb der LAG. Abbildung 40 zeigt den Ablauf der Bewertung von Auffälligkeiten gemäß DeQS-RL.

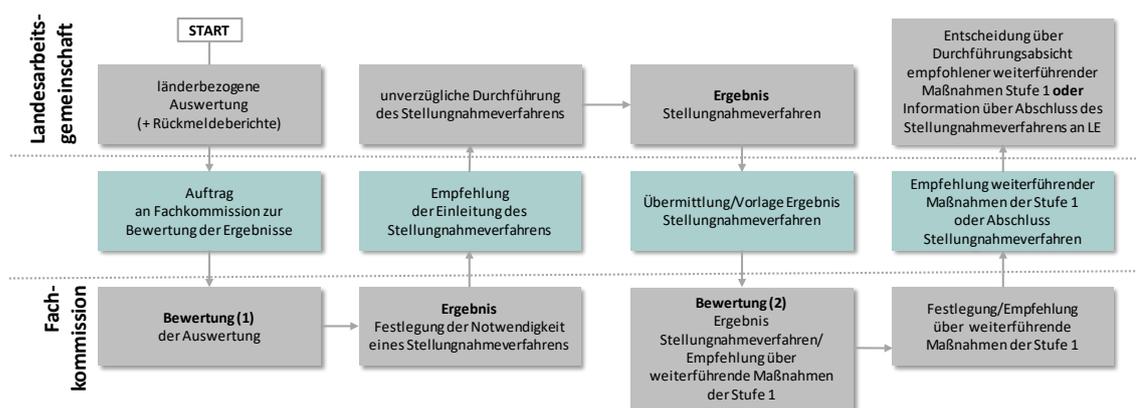


Abbildung 40: Schematischer Ablauf der Bewertung von Auffälligkeiten gemäß DeQS-RL.

Die DeQS-RL sieht aktuell vor, dass die Landesarbeitsgemeinschaften in einem Prozess wechselnder Zuständigkeiten mit der Fachkommission die Bewertung der auffälligen Indikatorergebnisse vornehmen und Empfehlungen für qualitätsverbessernde Maßnahmen abgeben (Abbildung 40): Die LAG beauftragen die Fachkommissionen damit, die Ergebnisse der Qualitätsindikatoren zu bewerten. Nach dieser Bewertung durch die Fachkommission soll diese die Notwendigkeit eines Stellungnahmeverfahrens festlegen und Empfehlungen zur Einleitung des Stellungnahmeverfahrens sowie zur Art und Weise (schriftliche Stellungnahme, Gespräch oder Begehung) einschließlich des Zeitrahmens aussprechen. Die LAG sollen das Stellungnahmeverfahren durchführen und der Fachkommission das Ergebnis vorlegen, sodass diese die Ergebnisse bewerten kann. Im Weiteren empfiehlt die Fachkommission der LAG im Anschluss weiterführende Maßnahmen der Maßnahmenstufe 1 oder den Abschluss des Stellungnahmeverfahrens. Abschließend entscheidet die LAG darüber, weiterführende Maßnahmen der Maßnahmenstufe 1 durchzuführen oder gibt dem Leistungserbringer die Information, dass das Stellungnahmeverfahren abgeschlossen

ist. Die Umsetzung dieses Prozesses in der Praxis wurde im Workshop als sehr zeitaufwendig beschrieben (siehe Anhang, Kapitel 1).

Im Einklang mit dieser Rückmeldung der Vertreterinnen und Vertreter der LAG und LQS empfiehlt das IQTIG eine Vereinfachung dieses Prozesses. Zum einen entfällt der erste Entscheidungsschritt, da bei einem hinreichenden Hinweis auf ein Qualitätsdefizit immer eine Stellungnahme einzuholen ist und dies mittels der statistischen Methodik ohne Ermessensspielraum eindeutig operationalisiert ist (siehe Abschnitt 6.1). Zum anderen übernimmt die LAG bzw. Bundesstelle die vorgeschaltete formale Prüfung. Damit erscheint eine vergleichsweise zeitnahe und zügige fachliche Bewertung der Stellungnahmen mit anschließender Empfehlung für die abschließende Einstufung in das Bewertungsschema durch die jeweilige Fachkommission möglich (siehe Abschnitte 6.4 und 8.2.3). Dieser auf Basis der vorliegenden Empfehlungen vereinfachte Ablauf wird in Abbildung 41 dargestellt. Hier wird auch deutlich, dass nach der formalen Prüfung und der Sitzung der Fachkommission direkt qualitätsverbessernde Maßnahmen durch die LAG bzw. Bundesstelle eingeleitet werden könnten. Welche Zuständigkeiten bei der Durchführung der Qualitätsbewertung und der Durchführung der qualitätsverbessernde Maßnahmen empfohlen werden, wird in den Kapitel 6 und 7 erläutert.

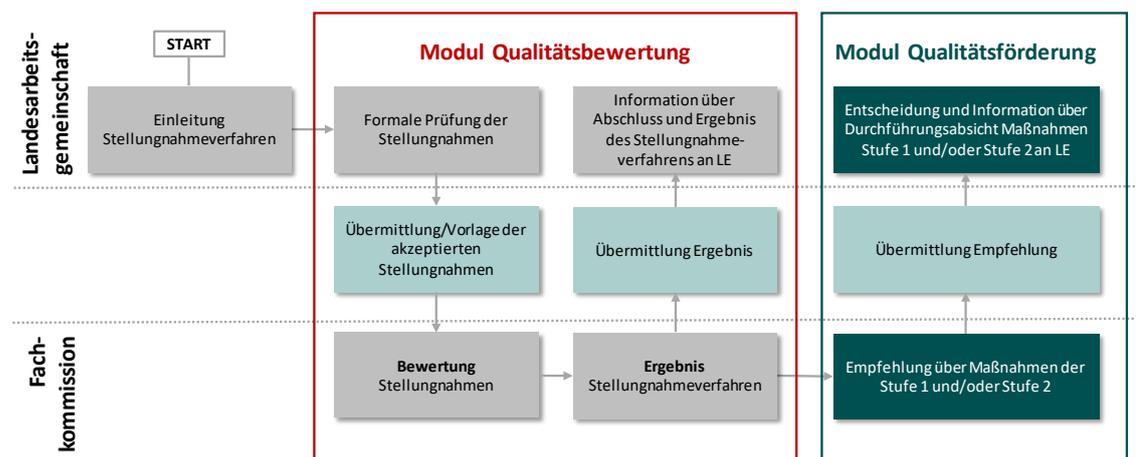


Abbildung 41: Schematischer Ablauf der Bewertung von Auffälligkeiten gemäß Empfehlungen.

Der zweite Aspekt, den die LAG und LQS hinsichtlich der Verkürzung des Verfahrens nannten, ist ein früherer Start des Stellungnahmeverfahrens (siehe Anhang, Kapitel 1). Hintergrund ist, dass aufgrund der aktuellen Fristen der DeQS-RL das Stellungnahmeverfahren in die Sommermonate fällt. Sowohl die LAG als auch die Leistungserbringer wären durch eine Vorverlagerung des Stellungnahmeverfahrens mit mehr personellen Ressourcen ausgestattet, wenn dieser Prozess nicht während der Sommerurlaubszeit mit reduziertem Personal erfolgen müsste (siehe Anhang, Kapitel 1).

## 8.4 Verkürzung des Gesamtverfahrens durch früheren Start des Stellungnahmeverfahrens

Eine Möglichkeit für einen früheren Start des Stellungnahmeverfahrens wäre, die Korrekturfrist (vgl. Abschnitt 8.1) für die Daten zu streichen. Angesichts der quartalsweisen Auswertungen und Rückmeldeberichte und der damit verbundenen Möglichkeit, nötige Korrekturen rechtzeitig und bereits während des EJ durchzuführen, erscheint eine zusätzliche Korrekturfrist nicht notwendig. Darüber hinaus bestünde eine weitere Möglichkeit das Stellungnahmeverfahren vorzulegen darin, dass man die Datenlieferfrist vorzieht. Gemäß der DeQS-RL berücksichtigt der Datensatz des EJ nur Daten der Patientinnen und Patienten, die noch im gleichen EJ entlassen werden, „Überlieger“ werden dem EJ nicht mehr zugerechnet. Dadurch ist die Dokumentation der letzten Fälle bereits am 31. Dezember des EJ möglich, sodass prinzipiell eine frühere Datenannahmefrist und damit auch ein früherer Eingang der Daten bei der Bundesauswertestelle beispielsweise bis zum 15. Februar des EJ + 1 festgelegt werden könnte. In der Folge könnte zukünftig auch der Versand der Rückmeldeberichte an die Länder früher stattfinden.

Ein früherer Start des Stellungnahmeverfahrens ist auch angesichts der Überführung der Verfahren der QSKH-RL in die DeQS-RL theoretisch möglich. Die DeQS-RL legt in § 10 Abs. 2 fest, dass das IQTIG die Auswertungen und Überprüfung der Daten auch für alle länderbezogenen Verfahren vornehmen soll. Neben einer Aufwandsreduktion für die LAG durch den Wegfall der Indikatorberechnungen (innerhalb des Prozesses „Fertigstellung Auswertung“) ergäbe sich auch eine Zeitersparnis insgesamt, da die einzelnen Prozessschritte gleichzeitig stattfinden könnten. Die Aufforderungsschreiben für die auffälligen Indikatorergebnisse könnten quasi zeitgleich mit der Übermittlung der Auswertungen von der LAG an die Leistungserbringer versendet werden (siehe Anhang, Kapitel 7), falls entsprechende IT-seitige Strukturen aufgebaut werden würden. Das bedeutet, dass mit Erstellung und Versand der Auswertung an die LAG diese direkt mit dem Stellungnahmeverfahren beginnen könnten. Demnach könnte das Stellungnahmeverfahren, welches mit der Aufforderung einer Stellungnahme bei einem hinreichenden Hinweis auf ein Qualitätsdefizit beginnt, voraussichtlich bereits mit Übermittlung der Rückmeldeberichte erfolgen. Folglich könnte auch der Start der Bewertung der Stellungnahmen eine Woche früher, am 04. August anstatt am 11. August beginnen (vgl. Abschnitt 8.2.3).

In Abbildung 42 sind die möglichen Änderungen für das Modul Qualitätsbewertung und damit die Verfügbarkeit der abschließenden Qualitätsergebnisse dargestellt.

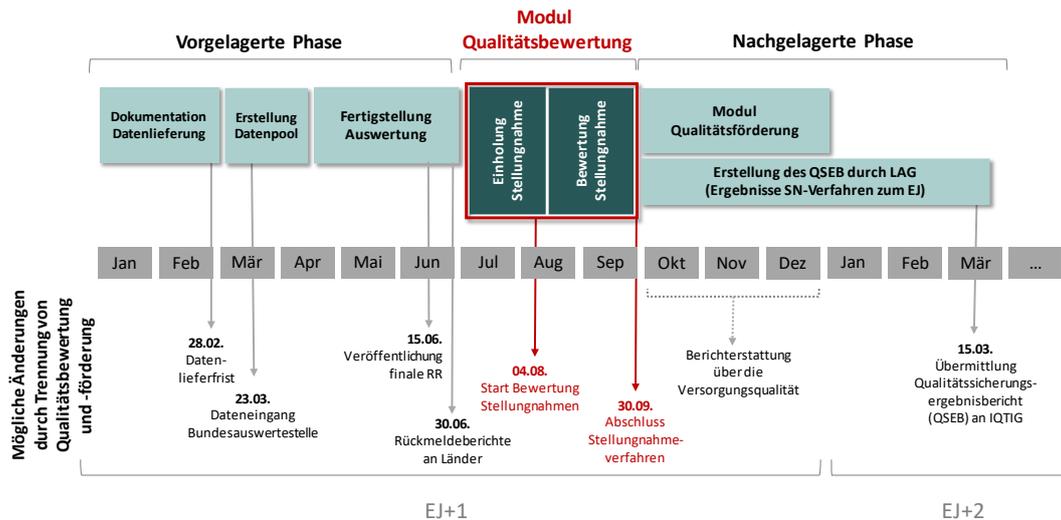


Abbildung 42: Mögliche Phasen und Änderungen des Verfahrens nach Erfahrung mit der Umsetzung der Empfehlungen.

In Kombination mit einem früheren Dateneingang bei der Bundesauswertungsstelle und damit einer möglichen früheren Übermittlung der Rückmeldeberichte bestünde grundsätzlich die Möglichkeit das Stellungnahmeverfahren noch vor der Sommerpause beginnen zu lassen. Ein früherer Beginn des Moduls Qualitätsbewertung würde sich, wie oben beschrieben, einerseits günstig auf das Stellungnahmeverfahren auswirken und andererseits eine frühzeitigere Berichterstattung zur Versorgungsqualität ermöglichen. Eine notwendige Voraussetzung für die Festlegung entsprechender Regelungen ist jedoch eine intensive Abstimmung unter Beteiligung der unterschiedlichen Akteure (z. B. LAG, Leistungserbringer, IQTIG, Softwareanbieter, usw.). Die Abstimmung müsste beispielsweise die unterschiedlichen Abläufe der verschiedenen Verfahren (z. B. landes- und bundesbezogene Verfahren sowie plan. QI-RL-Verfahren) aber auch für die unterschiedlichen Erfassungsinstrumente (fall- und einrichtungsbezogene QS-Dokumentation, Sozialdaten und Patientenbefragung) berücksichtigen. Da ein Start des Moduls Qualitätsbewertung vor 30. Juni des EJ + 1 weitreichende Änderungen für alle Beteiligten bedeuten würde, sollten entsprechende Änderungen nicht ohne weitere umfangreiche Prüfungen der Umsetzbarkeit und der Konsequenzen vorgenommen werden (siehe hierzu auch Abschnitt 11.5.2).

In der Zusammenschau ergeben sich eine Reihe von Möglichkeiten, um einerseits das Gesamtverfahren und andererseits vor allem das Modul Qualitätsbewertung früher abzuschließen. Mit Blick auf das gesamte Verfahren entstehen im bisherigen Vorgehen Verzögerungen vor allem durch viele und teilweise längere Korrekturfristen (für Daten und für Stellungnahmen). Mit Blick auf das Modul Qualitätsbewertung sind die vorliegenden Empfehlungen zur Standardisierung der Abläufe geeignet, einige Prozesse zu automatisieren (z. B. die Entscheidung, für welchen Leistungserbringer eine Stellungnahme eingeholt werden soll) und komplexe Abstimmungsprozess zwischen Fachkommission und der LAG bzw. Bundesstelle zu vereinfachen.

## 9 Berichterstattung

Neben der Vereinheitlichung der Vorgehensweise im Strukturierten Dialog, der Optimierung der Effizienz des Verfahrens und der Entwicklung eines Rahmenkonzepts beinhaltet die Beauftragung auch, Empfehlungen für eine optimierte Berichterstattung zu den Ergebnissen des Strukturierten Dialogs zu erarbeiten. Dies beinhaltet Vorschläge, wie eine einheitliche Datengrundlage für den Bericht zum Strukturierten Dialog, den das IQTIG auf Basis der Berichte der Stellen auf Landesebene erstellt, gewährleistet werden kann (Punkt 2.a der Beauftragung) sowie Mindestanforderungen, die an diese Berichte der Landesstellen (Punkt 2.c der Beauftragung) gestellt sind. Während im Bericht zu Stufe 1 der Beauftragung noch Empfehlungen für die Berichte im Rahmen der QSKH-RL gemacht wurden, wird wegen des absehbaren Außerkrafttretens der QSKH-RL davon abgesehen, diese Empfehlungen weiterzuentwickeln. Daher werden an dieser Stelle Empfehlungen mit Blick auf die DeQS-RL gegeben. Somit ist ein Vorschlag für eine Vereinheitlichung der bisher bundesweit nicht einheitlich standardisierten Berichte über die Ergebnisse des Strukturierten Dialogs sowie des Datenvalidierungsverfahrens (Erstellung von Mustertabellen, Templates, wie unter Punkt 2.c der Beauftragung genannt) nicht mehr erforderlich, da sich die Inhalte dieser Berichte mit Überführung aller Verfahren in die DeQS-RL dann in den Empfehlungen zu den QSEB wiederfinden. Vor dem Hintergrund der Regelungen in der DeQS-RL, welche vorsehen, dass das IQTIG die Auswertungen für die Landesebene erstellt (§ 10 DeQS-RL), erübrigt sich auch die beauftragte Gewährleistung einer einheitlichen Datengrundlage für den Qualitätsreport und den Bericht zum Strukturierten Dialog.

Die DeQS-RL sieht vor, dass in den sog. Qualitätssicherungsergebnisberichten (QSEB) die LAG die Ergebnisse der Stellungnahmeverfahren der QS-Verfahren der externen stationären Qualitätssicherung an das Institut nach § 137a SGB V in maschinenlesbarer und -verwertbarer Form übermitteln (§19 DeQS-RL). Diese Berichte sind bis 15. März des EJ + 2 abzugeben und lösen die Berichte über die Ergebnisse des Strukturierten Dialogs sowie des Datenvalidierungsverfahrens der QSKH-RL ab. Der Bundesqualitätsbericht (BQB) nach § 20 DeQS-RL umfasst die Ergebnisse der Bundesauswertung, die Ergebnisse der Datenvalidierung, die Ergebnisse aus den QSEB sowie die Ergebnisse einer wissenschaftlichen Begleitung der Verfahren. Die bisherigen sog. „Mai-Berichte“, in denen das IQTIG die Ergebnisse der Datenvalidierung und der Strukturierten Dialoge zusammengefasst hat, werden somit zukünftig im BQB berichtet. Der Bundesqualitätsbericht ist standardisiert und wird dem G-BA in maschinenlesbarer und -verwertbarer Form übermittelt (aktuell am 15. August zu den Verfahren *QS PCI* und *QS WI*) (siehe Kapitel 8, Abb. 1)

In Kapitel 8 dieses Berichts werden Empfehlungen zur Verfahrensverkürzung gegeben. Werden diese Empfehlungen berücksichtigt, ergeben sich daraus einerseits neue Möglichkeiten für die Berichterstattung und andererseits Implikationen für die zugehörigen Datenflüsse.

Bisher liegen die Informationen aus dem Stellungnahmeverfahren, zu qualitätsverbessernden Maßnahmen und zur Datenvalidierung erst im EJ + 2 vor. Die in Kapitel 1 und Kapitel 8 vorgelegten Empfehlungen, die eine Trennung der Qualitätsbewertung und der Qualitätsförderung vor-

sehen, ermöglichen eine frühzeitigere Berichterstattung zu den Ergebnissen der Qualitätsbewertung im EJ + 1 mit einer nachgelagerten Berichterstattung zur Qualitätsförderung im EJ + 2. Es wird daher empfohlen, die Ergebnisse der fachlichen Bewertung inkl. der Begründung der Einstufung getrennt von den Empfehlungen zur Qualitätsförderung (einzuleitende Maßnahmen und deren Ergebnisse) zu berichten. Diese Trennung ermöglicht, dass die Qualitätsergebnisse früher zur Verfügung gestellt werden können, wohingegen für die Ursachenanalyse und die Durchführung qualitätsverbessernder Maßnahmen weiterhin mehr Zeit zur Verfügung stehen soll.

Es wird daher empfohlen, dass die Berichterstattung zukünftig aus den folgenden beiden „Berichtsarten“ besteht, die beide zu unterschiedlichen Zeiten in den BQB einfließen:

- Berichterstattung zur Versorgungsqualität: Qualitätsergebnisse am Ende des Moduls „Qualitätsbewertung“
- Berichterstattung über die Maßnahmen der Qualitätsverbesserung: Empfehlungen für die einzuleitenden qualitätsverbessernden Maßnahmen und die Ergebnisse aus der Umsetzung dieser Maßnahmen

Für beide Berichtsarten wird eine Spezifikation dessen, was genau berichtet werden soll, benötigt. Für die Verfahren gemäß DeQS-RL liegt bereits eine Spezifikation für die Qualitätssicherungsergebnisberichte (QSEB) vor. Die Spezifikation für die QSEB kann als Grundlage für die zukünftig einheitliche Berichterstattung für alle Verfahren herangezogen werden und entspricht der im Auftrag genannten Festlegung von Mindestanforderungen für die Länderberichte zum Strukturierten Dialog inkl. der Mustertabellen (Punkt 2.c der Beauftragung). Die bisherige Spezifikation müsste jedoch an die im hier vorliegenden Konzept zur Qualitätsbewertung und -förderung vorgeschlagene Vorgehensweise angepasst werden. Darüber hinaus müssen bisher nicht abgefragte Informationen zu Maßnahmen der Qualitätsverbesserung ergänzt werden. Die konkrete Ausarbeitung kann erst nach dem Beschluss über die Umsetzung des hier vorgeschlagenen Konzepts erfolgen, da die Erstellung der Spezifikation ein aufwändiger Prozess ist, der zudem erst nach Beschluss über die Inhalte sinnvoll ist.

Bisher werden die Ergebnisse zur Versorgungsqualität und zu den qualitätsverbessernden Maßnahmen im EJ + 2 berichtet. Wünschenswert ist jedoch, dass die Ergebnisse der Qualitätsbewertung aus dem Stellungnahmeverfahren (Ende Modul Qualitätsbewertung) schon früher vorliegen. Je nach angestrebtem Erscheinungstermin für den Bundesqualitätsbericht wäre es denkbar, nach Abschluss des Stellungnahmeverfahrens die Ergebnisse der Qualitätsbewertung im Bundesqualitätsbericht zu berichten. Hierfür müsste zeitnah eine eigene Datenübermittlung der Ergebnisse des Stellungnahmeverfahrens an das IQTIG erfolgen. Für die Prüfung, Aufbereitung, Auswertung und Verschriftlichung der Ergebnisse der Qualitätsbewertung für den BQB müssen jedoch mindestens 6 bis 8 Wochen Zeit nach Abschluss des Stellungnahmeverfahrens eingeplant werden. Die Maßnahmen zur Qualitätsverbesserung sowie die Ergebnisse der Datenvalidierung würden wie bisher im EJ + 2 berichtet werden.

Sollte das hier vorgelegte Konzept zur Qualitätsbewertung und -förderung beschlossen werden, wird diesbezüglich eine Anpassung der DeQS-RL erforderlich sein (z. B. hinsichtlich der Etablierung eines Datenflusses von den LAG zum IQTIG für die Ergebnisse am Ende des Moduls Qualitätsbewertung mit der Festlegung von Fristen).

## 9.1 Berichterstattung zur Versorgungsqualität

Wie bereits im vorangegangenen Kapitel dargelegt, lassen sich aus den Empfehlungen zum Vorgehen bei der Qualitätsbewertung im Stellungnahmeverfahren (siehe Kapitel 6) die Vorgaben für die Spezifikation ableiten. Es ist davon auszugehen, dass Änderungen der bisherigen Spezifikation zu den in den Verfahren gemäß DeQS-RL zu übermittelnden Informationen notwendig sind. Es wird eine Leistungserbringer- und QI-bezogene Übermittlung der Ergebnisse des Stellungnahmeverfahrens empfohlen, d. h. es müssen Datensätze übermittelt werden, die für jeden Leistungserbringer und jeden QI eine Angabe beinhalten. Insbesondere müssen hier auch die Begründungen für die fachlichen Bewertungen mitgeteilt werden. Eine Aggregation der Daten (z. B. übergeordnet für das jeweilige Bundesland oder als Ländervergleich etc.) sowie eine Aus- und Bewertung erfolgt dann durch das IQTIG für den Bundesqualitätsbericht.

## 9.2 Berichterstattung über die Maßnahmen der Qualitätsförderung

Neben der abschließenden Bewertung der Indikatorergebnisse sind lt. DeQS-RL in den QSEB die folgenden Inhalte zu Maßnahmen der Qualitätsverbesserung zu berichten:

- die Art, Häufigkeit und Ergebnisse aller durchgeführten oder zur Durchführung empfohlenen Qualitätssicherungsmaßnahmen
- die längsschnittliche verlaufsbezogene Darstellung des Erfolgs von Qualitätssicherungsmaßnahmen
- Erfahrungsberichte aus den Fachkommissionen
- Angaben über wiederholte oder besonders schwerwiegende Auffälligkeiten

Hinsichtlich der Maßnahmen, die zur Qualitätsförderung zu ergreifen sind, finden sich Regelungen sowohl in der DeQS-RL selbst als auch übergreifend in der Richtlinie zur Förderung der Qualität und zu Folgen der Nichteinhaltung sowie zur Durchsetzung von Qualitätsanforderungen des Gemeinsamen Bundesausschusses gemäß § 137 Abs. 1 SGB V (Qualitätsförderungs- und Durchsetzungs-Richtlinie; QFD-RL). In der DeQS-RL werden zwei Maßnahmenstufen unterschieden. Maßnahmenstufe 1 enthält beispielhaft Maßnahmen, die in § 4 QFD-RL als „Maßnahmen zur Beratung und Unterstützung“ festgelegt sind; Maßnahmenstufe 2 nach DeQS-RL entspricht den „Durchsetzungsmaßnahmen“ nach § 5 QFD-RL. Wegen der übergeordneten Bedeutung der QFD-RL wird empfohlen, die dort festgelegten Maßnahmen als Grundlage für die Berichterstattung vorzusehen (zu den Maßnahmen der Qualitätsförderung siehe auch Kapitel 7).

Die Spezifikation für die Übermittlung der relevanten Informationen zu qualitätsfördernden Maßnahmen ist teilweise bereits in der Spezifikation für den QSEB umgesetzt. Dort werden Maßnahmen der Stufe 1 bzw. nach QFD-RL § 4 (Beratung und Unterstützung) abgefragt. Abfragen zu § 5 (Durchsetzung), die den Maßnahmen der Stufe 2 nach DeQS-RL entsprechen, fehlen noch.

Neben den eingeleiteten Maßnahmen selbst soll auch deren Umsetzung und das Umsetzungsergebnis berichtet werden. Dazu gehört auch ein Monitoring der Zielvereinbarungen (siehe Abschnitt 7.2). Es wird empfohlen, nach entsprechender Beschlussfassung die Informationen wie in Abschnitt 7.2 dargelegt zu berichten.

Im G-BA-Auftrag sind unter Punkt 2 „Optimierung der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung“ auch ergänzende Punkte für die Berichterstattung genannt, die die Bereitstellung von Informationen für die Öffentlichkeit und für teilnehmende Krankenhäuser umfassen (Punkt 2.d). Im Rahmen des aktuellen Konzepts kann dies die Berichterstattung zu den Ergebnissen sowohl des Moduls Qualitätsbewertung als auch zum Modul Qualitätsförderung betreffen. Hinsichtlich der Aufarbeitung von Informationen für die Öffentlichkeit wird auf den Auftrag zum G-BA-Qualitätsportal verwiesen, in dem die Grundlage für eine allgemeinverständliche und für Patientinnen und Patienten nutzbringende Darstellung der Qualitätsergebnisse (Ergebnisse des Moduls Qualitätsbewertung) erarbeitet wird. Im Bundesqualitätsbericht wird die Öffentlichkeit über die Ergebnisse aus beiden Modulen informiert.

Darüber hinaus sollen zukünftig den Leistungserbringern ein gegenseitiges Lernen aus den Analysen auffälliger Ergebnisse, identifizierter Verbesserungspotenziale und beispielhaft positiver Versorgungsmodelle ermöglicht werden (Punkt 2.d und 2.e der Beauftragung, siehe auch Kapitel 7, „Lernen von den Besten“).

Der Schwerpunkt des vorliegenden Rahmenkonzepts liegt jedoch auf dem Modul „Qualitätsbewertung“. Wie oben dargelegt muss das Modul „Qualitätsförderung“ für die Berichterstattung noch im Detail spezifiziert werden. Bis die Rückmeldungen der Einrichtungen zu den mehrstufigen Maßnahmen und Ergebnissen der Qualitätsförderung tatsächlich vorliegen, werden mehrere Jahre vergehen. Daher erscheint es nicht sinnvoll, bereits in diesem Bericht und ohne einheitliche Datengrundlage Vorschläge im Detail zu entwickeln, wie diese Analyseergebnisse und Verbesserungsmaßnahmen auf einer übergeordneten Ebene allen beteiligten Einrichtungen zur Verfügung gestellt und für Fortbildungszwecke genutzt werden können. Es sollten erst die Rückmeldungen der LAG zu den getroffenen Maßnahmen und deren Erfolg etabliert sein, bevor ein Konzept für weiterführende Auswertungen dieser Daten erstellt wird.

An dieser Stelle können daher nur Eckpunkte für die zukünftige Berichterstattung zur Versorgungsqualität und zu Maßnahmen der Qualitätsverbesserung skizziert werden. Ein umfassendes Konzept mit Spezifikation für die zu übermittelnden Daten/Inhalte und zu den Zeitpunkten der Veröffentlichung muss erarbeitet werden, nachdem das vorliegende Rahmenkonzept zur Qualitätsbewertung und -förderung beschlossen wurde. Die Umsetzung dieses Rahmenkonzepts zur Qualitätsbewertung und -förderung und der dazugehörigen Berichterstattung erfordert im Vorfeld insbesondere ein schlüssiges, zeitlich machbares Ineinandergreifen aller Verfahrensabschnitte der DeQS-Verfahren und ggf. auch eine Änderung der DeQS-RL zu den dann einzuhaltenen Fristen. Dabei müssen auch Aufwand und Machbarkeit bedacht werden für die Zeit nach der Überführung aller Verfahren in die DeQS-RL. Die AG DeQS hat im Hinblick auf eine Neugestaltung des Fristengerüsts das IQTIG gebeten, die die Verfahrensabläufe der DeQS-Verfahren zu prüfen und zu überarbeiten. Die Ergebnisse dazu müssen auch im Hinblick auf die Fristen und möglichen Inhalte der Berichterstattung einbezogen und berücksichtigt werden. Die Vorschläge

zur Neugestaltung des Fristengerüsts werden aktuell parallel zum vorliegenden Bericht erarbeitet und können derzeit noch nicht in den hier vorgelegten Abschlussbericht aufgenommen werden.

## 10 Empfehlungen zur Evaluation des neuen Konzepts

Laut Beauftragung sollen auch Empfehlungen zur Evaluation des neuen Konzepts gegeben werden. Im folgenden Abschnitt werden Empfehlungen hinsichtlich möglicher Schwerpunkte und der Vorgehensweise einer Evaluierung des vorliegenden Konzepts dargelegt. Als Schwerpunkte der Evaluation bieten sich hierbei einerseits die Erfassung des Umsetzungsgrads und andererseits die Untersuchung der Zielerreichung an. Als Ziele der Weiterentwicklung des Strukturierten Dialogs wurden definiert:

- die Einheitlichkeit der Vorgehensweise
- die Transparenz und Nachvollziehbarkeit der Entscheidungsfindung
- die Effizienz im Sinne einer Verschlankung des Verfahrens

Eine genaue Definition der Evaluationsschwerpunkte wäre erforderlich und sollte im Hinblick auf eine mögliche Handlungsrelevanz erfolgen. Daher müsste im nächsten Schritt die Erstellung ein Evaluationsplan beauftragt werden. Dieser soll unter anderem die genauen Evaluationsfragestellungen, die Untersuchungsdesigns, die vorgesehenen Datenquellen sowie ein Wirkmodell des Weiterentwicklungskonzepts beinhalten. Darstellung und Differenzierung der verschiedenen Mechanismen des Weiterentwicklungskonzepts helfen dabei, dessen Wirkungen und dessen Komponenten mit den zugrunde liegenden theoretischen Zusammenhängen zu verknüpfen, um ein möglichst umfassendes Bild von der Wirkweise des Weiterentwicklungskonzepts zu erhalten. Das Wirkmodell dient zudem als Unterstützung dabei, die relevanten Fragestellungen der Evaluation zu entwickeln (W. K. Kellogg Foundation 2004). Die Entwicklung eines detaillierten Wirkmodells kann somit erst mit der Erstellung eines Evaluationsplans erfolgen.

### 10.1 Evaluationsdesign

Je nach Schwerpunkt einer Evaluation liegt das Hauptaugenmerk auf einer formativen oder summarischen Evaluation. Für das vorliegende Konzept sind beide Arten von Nutzen. Die genaue Vorgehensweise sowie eine umfassende Festlegung der Schwerpunkte der Evaluation sollten vom Auftraggeber mit dem Augenmerk auf handlungsrelevante Ergebnisse festgelegt werden. Im Folgenden werden beide empfohlenen Evaluationstypen vorgestellt.

Eine prozessbegleitende, formative Evaluation des Weiterentwicklungskonzepts ist zu empfehlen als Maßnahme, um bereits vorgeschlagene Änderungen des Strukturierten Dialogs während dessen Implementierung zu begleiten und gegebenenfalls zu optimieren.

Mögliche Ziele einer formativen Evaluation sind

- Umsetzungsgrad und -hindernisse der einzelnen Maßnahmen zu untersuchen sowie
- die Nützlichkeit (Verständlichkeit, Praktikabilität etc.) der vorgeschlagenen Instrumente aus Sicht der beteiligten Stellen (Landesarbeitsgemeinschaften und Leistungserbringer) zu erfassen.

Bei der Definition von Schwerpunkten der formativen Evaluation ist die Handlungsrelevanz von entscheidender Bedeutung. So kann ein durch eine formative Evaluation festgestelltes Optimierungspotenzial zu einer zusätzlichen Verbesserung der weiterentwickelten Vorgehensweise führen. Beispielsweise könnte eine prozessbegleitende Befragung der Leistungserbringer die Verständlichkeit und Nützlichkeit der entwickelten Empfehlungen zu Form und Inhalt der Einholung von Stellungnahmen erfragen und gegebenenfalls zu Verbesserungsvorschlägen führen.

Eine summative Evaluation nach Umsetzung des vorliegenden Konzepts eröffnet zusätzlich zur Untersuchung von Veränderungen der Prozessqualität (Veit et al. 2013) auch die Untersuchung der Zielerreichung des vorliegenden Konzepts. Einerseits kann in einer summativen Evaluation der Umsetzungsgrad nach Abschluss der Weiterentwicklung gemessen werden. Andererseits kann die Zielerreichung der Weiterentwicklung gemessen werden. Nur bei gleichzeitiger Erreichung eines hohen Umsetzungsgrads und der Erreichung der Ziele der Weiterentwicklung des Strukturierten Dialogs kann von einer erfolgreichen Weiterentwicklung ausgegangen werden. Methodisch wäre sowohl eine retrospektive Beobachtungsstudie als auch eine randomisiert kontrollierte Vergleichsstudie denkbar. Letztere ist trotz des Vorteils der Unabhängigkeit von zeitabhängigen Störgrößen nicht zu empfehlen, da mit der Randomisierung ein deutlich erhöhter Aufwand bei den LAG entstehen würde. Aus diesem Grund wird eine retrospektive Beobachtungsstudie empfohlen. Der Effekt der Weiterentwicklung wird dabei in einem Vorher-Nachher-Vergleich ermittelt. Dabei wird der Zustand nach der Weiterentwicklung sowie der Zustand vor jeglicher Veränderung erfasst und die Differenz berechnet. Zeitlich unabhängige individuelle Störgrößen werden so eliminiert. Somit kann sich die Evaluation auf die zeitliche Veränderung bei Einführung der Weiterentwicklung konzentrieren. Damit kann, unter der Annahme der Abwesenheit von zeitabhängigen Störgrößen, von einem kausalen Zusammenhang ausgegangen werden.

Im Folgenden werden mögliche Schwerpunkte einer summativen Evaluation des ersten Teils des Weiterentwicklungskonzepts dargelegt. Hierfür bieten sich in erster Linie die in Abschnitt 2.5 beschriebenen und im dazugehörigen Anhang 4 vertieften Fragestellungen zur Heterogenität in der Einstufung von Leistungserbringern durch die für die Durchführung des Strukturierten Dialogs beauftragten Stellen an. Daraus ergeben sich folgende mögliche Evaluationsfragestellungen:

- **Fragestellung 1:** Wie groß ist die bewertungsstellenabhängige Heterogenität bezüglich der Aufnahme eines Stellungnahmeverfahrens und hat sich diese durch die Weiterentwicklung reduziert?
- **Fragestellung 2:** Wie groß ist die bewertungsstellenabhängige Heterogenität bezüglich der Bewertung als qualitativ auffällig nach Aufnahme eines Stellungnahmeverfahrens und hat sich diese durch die Weiterentwicklung reduziert?

Konkret empfiehlt sich eine Auswertung basierend auf dem Regressionsmodell in Formel 1 des zweiten Abschnitts in Kapitel 4 des Anhangs. Von primärem Interesse für die Evaluation ist dabei die Entwicklung der jeweiligen Heterogenität im Vergleich zum Status vor der Weiterentwicklung. Dabei ist zu empfehlen, das Regressionsmodell auf alle Qualitätsindikatoren eines Leistungsbereichs zu erweitern, da damit die statistische Teststärke (Power) steigt. Die im vierten

Kapitel des Anhangs beschriebenen Methodik lässt sich in dieser Hinsicht erweitern. Eine Messung aller Leistungsbereiche sollte das Problem multipler Messungen berücksichtigen, ist aber zu empfehlen, um die Unsicherheit der Messung des zeitlichen Trends der Heterogenität weiter zu reduzieren. Die Operationalisierung dieser Fragestellungen mithilfe von Evaluationskennziffern würde auf den Ausführungen des vierten Kapitels des Anhangs basieren. Dies sollte in einem zu erstellenden Evaluationsplan dargestellt werden.

Darüber hinaus sollten weitere handlungsrelevante Schwerpunkte wie etwa Fragestellungen zum Ziel der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung definiert und anhand eines Wirkmodells im Rahmen des Evaluationsplans ausgearbeitet werden.

## **10.2 Datenquellen und Zeitschiene**

Naheliegende routinemäßig verfügbare Datenquellen sind die Indikatorergebnisse der Qualitätssicherung und die Bewertungsergebnisse im Modul „Qualitätsbewertung“. Darüber hinaus ist die Befragung der LAG sowie der Leistungserbringer, inklusive eines Zeitpunktes vor der Weiterentwicklung, zu empfehlen.

Eine formative Evaluation sollte prozessbegleitend mit der Implementation der Empfehlungen stattfinden, eine summative Evaluation nach Abschluss der Implementation. Einerseits gilt es, den Zustand nach dem Einpendeln der Implementation, d. h., bei hinreichendem Umsetzungsgrad zu messen. Andererseits sollte der zeitliche Abstand einer Vorher-Nachher-Messung bei einer retrospektiven Beobachtungsstudie und das damit einhergehende Risiko einer Verzerrung durch zeitveränderliche Störgrößen minimiert werden. Unter Berücksichtigung des zu messenden Umsetzungsgrades und des Umstands, dass dadurch die Effekte der Weiterentwicklung möglicherweise noch nicht in vollem Umfang zum Tragen gekommen sind, wird die Messung im ersten Kalenderjahr nach Einführung der Weiterentwicklung empfohlen.

## **10.3 Zusammenfassung der Empfehlungen zur Evaluation des vorliegenden Konzepts**

Es wird empfohlen, sowohl eine formative als auch eine summative Evaluation des vorliegenden Konzepts durchzuführen. Die formative Evaluation soll die Weiterentwicklung unterstützen, indem prozessbegleitend der Umsetzungsgrad der Empfehlungen und die Nützlichkeit aus Sicht der Beteiligten gemessen werden sollen. Die summative Evaluation im Rahmen einer retrospektiven Kohortenstudie ist geeignet, den Umsetzungsgrad und die Zielerreichung im Vergleich zu einer Nullpunktmessung zu erfassen. Empfohlene Evaluationsschwerpunkte sind unter anderem die Untersuchung der Veränderung der Heterogenität in der Einstufung von Leistungserbringern durch die LAG. Es wird empfohlen, im Rahmen der Erstellung eines Evaluationsplans, ein detailliertes Wirkmodell zu erarbeiten, um alle Aspekte der Wirkweise des vorliegenden Weiterentwicklungskonzepts auch in dessen Umfeld in vollständiger Weise untersuchen zu können.

## 10.4 Evaluation der quantitativen Qualitätsbewertung

In Abschnitt 5.3.1 wurden drei alternative Operationalisierungen des hinreichenden Hinweises für ein Qualitätsdefizit vorgestellt: die rechnerische Auffälligkeitseinstufungsmethode, die statistisch signifikante Einstufungsmethode und die statistisch relevante Einstufungsmethode. In Abschnitt 5.6 wurde empfohlen, die Einstufung nach der statistisch signifikanten Einstufungsmethode vorzunehmen. Des Weiteren wurden in Abschnitt 5.4 Erweiterungen dieser Klassifikationsmethode im Rahmen einer 2-Jahres-Einstufung vorgeschlagen. Im Folgenden werden die empfohlene 1-Jahres-Klassifikation nach der statistisch signifikanten Auffälligkeit sowie die 2-Jahres-Version dieser Einstufungsmethode exemplarisch anhand der Indikatorergebnisse der QS-Verfahren *Hüftendoprothesenversorgung* (HEP) und *Nierentransplantation* (NTX) auf Basis der Ergebnisse des Strukturierten Dialogs für das Erfassungsjahr 2017 verglichen. Ziel dieses Kapitels ist es dabei, die Auswirkungen der gewählten Einstufungsmethode analog zu Abschnitt 2.6.1 abzuschätzen, um zu illustrieren, welche Aufwandsreduktion (im Sinne einer Reduktion der quantitativen Auffälligkeiten) erreicht werden würde. Darüber hinaus werden aufwandsorientierte Vorschläge für den in der Klassifikationsmethode zu wählenden „Tuning-Parameter“ in Form des Signifikanzniveaus erarbeitet. Für die Illustration wurden dabei ein QS-Verfahren mit einer großen Anzahl an Leistungserbringern und behandelten Fällen (HEP) sowie ein Verfahren mit wenigen Leistungserbringern und behandelten Fällen (NTX) ausgewählt.

Der Vergleich dient zum einen dazu, eine Abschätzung vorzunehmen, wie die Anzahl an quantitativen Auffälligkeiten von dem Tuning-Parameter der Einstufungsmethode abhängt. Weiterhin wird untersucht, wie viele der rechnerischen Auffälligkeiten aus dem Erfassungsjahr 2017, die anschließend im Strukturierten Dialog als „qualitativ auffällig“ bewertet wurden, auch anhand der hier untersuchten Methode als statistisch signifikant auffällig eingestuft worden wären. Die Klassifikationsmethode sowie die 2-Jahres-Einstufung werden im Folgenden hinsichtlich der folgenden vier Aspekte betrachtet:

- Die über den vom Referenzwert hinaus tolerierte Anzahl an Fällen mit unerwünschten Ereignissen, die nicht im Strukturierten Dialog analysiert wurden bzw. worden wären, da das Leistungserbringerergebnis nicht als quantitativ auffällig eingestuft wird
- Die Anzahl der quantitativen Auffälligkeiten
- Die Anzahl an qualitativen Auffälligkeiten, die durch die quantitative Einstufungsmethodik vorhergesagt wird (unter der Annahme, dass der darauffolgende Bewertungsprozess dem bisherigen SD-Prozess entspricht)
- Positive-Predictive-Value (PPV) als Maß für die Treffsicherheit des Einstufungsalgorithmus gegeben der bisherigen SD-Ergebnissen

Die Auswertungen werden dabei jeweils differenziert danach, ob es sich um Indikatoren mit festem oder verteilungsabhängigem Referenzbereich handelt. Für verteilungsabhängige Referenzbereiche wird für die statistisch signifikante Auffälligkeitseinstufung der Referenzwert so gewählt, dass immer  $q \cdot 100\%$  aller Standorte quantitativ auffällig werden, wobei  $q$  dem gewählten Anteil an quantitativen Auffälligkeiten für den Indikator entspricht – vgl. Abschnitt

5.3.4. Da die Anzahl an quantitativen Auffälligkeiten unabhängig vom Tuning-Parameter konstant ist, ist eine Aufwandsreduktion im Sinne einer Reduktion der quantitativen Auffälligkeiten hier – anders als bei Indikatoren mit festem Referenzwert – nicht möglich.

#### 10.4.1 Limitationen der Evaluation

Bei der folgenden Evaluation ist grundsätzlich zu beachten, dass die qualitative Auffälligkeitseinstufung im Rahmen des Strukturierten Dialogs nicht als Goldstandard für die Bewertung der Qualität eines Leistungserbringers betrachtet werden sollte, da das Ergebnis der qualitativen Bewertung nicht nur von der Qualität der Leistungserbringung sondern auch von anderen Faktoren abhängt, sodass die Bewertung auch eine deutliche bewertungsstellenabhängige Heterogenität ausweist (vgl. auch Abschnitt 2.5). Schon allein diese Heterogenität im Vorgehen zwischen den den Strukturierten Dialog durchführenden Stellen zeigt, dass es unterschiedliche Vorgehensweisen und Maßstäbe für die qualitative Bewertung zu geben scheint, statt eines einheitlichen Goldstandards.

Außerdem erscheint es plausibel, dass im bisherigen Vorgehen teilweise eine Qualitätsbewertung vorgenommen werden wird, die über das Qualitätsmerkmal des jeweiligen Indikators hinausgeht. So ist es beispielsweise denkbar, dass die durchführenden Stellen auch Ergebnisse des Leistungserbringers in anderen Qualitätsindikatoren für die Bewertung heranziehen, sowie Ergebnisse aus den Vorjahren. Sind beispielsweise im Rahmen eines Strukturierten Dialogs Zielvereinbarungen oder qualitätsverbessernde Maßnahmen mit einem Leistungserbringer vereinbart worden, ist es auch denkbar, dass bei der qualitativen Bewertung im anschließenden Jahr berücksichtigt wird, dass diese Maßnahmen in der Regel eine zeitversetzte Wirkung haben und deshalb auf ein Stellungnahmeverfahren im direkt anschließenden Erfassungsjahr verzichtet wird. Sollte also im bisherigen Vorgehen teilweise eine Qualitätsbewertung vorgenommen worden sein, die über das Qualitätsmerkmal des jeweiligen Indikators hinausgeht, führt dies in Einzelfällen zu Einstufungsergebnissen, die größtenteils unabhängig von dem numerischen Indikatorergebnis sind. Bei einem grundsätzlich validen Indikator und einem Stellungnahmeverfahren, das sich auf das dem jeweiligen Indikator zugrunde liegende Qualitätsmerkmal bezieht, sollte es über alle Stellungnahmeverfahren hinweg einen statistischen Zusammenhang zwischen numerischem Indikatorergebnis und der Wahrscheinlichkeit für eine qualitativ auffällige Bewertung geben. Zeigt sich ein solcher Zusammenhang nicht, ist das Ergebnis der Qualitätsbewertung unabhängig vom Indikatorergebnis. Die damit einhergehenden Probleme sind in Abschnitt 3.3 beschrieben. Der Vergleich der verschiedenen statistischen Operationalisierungen des hinreichenden Hinweises für ein Qualitätsdefizit mit den bisherigen Ergebnissen des Strukturierten Dialogs ist daher auch nur bedingt aussagekräftig

Darüber hinaus ist zu beachten, dass die Sensitivität und Spezifität der qualitativen Bewertung vor dem Hintergrund begrenzter Ressourcen bei den durchführenden Stellen und den Fachkommissionen vermutlich stark von dem Gesamtaufwand aller Dialoge abhängt. Deshalb könnte beispielsweise eine Reduktion der quantitativen Auffälligkeiten selbst schon zu einer Erhöhung der Sensitivität und Spezifität der qualitativen Einstufung führen.

Aus diesen Gründen sollte die vorliegende Evaluation der vorgeschlagenen Klassifikationsmethode nicht als Vergleich gegen einen Goldstandard angesehen werden und die Wahl der Methode und des Tuning-Parameters nicht auf Grundlage dieses Vergleichs, sondern vor allem auf Basis theoretischer Überlegungen getroffen werden. Die retrospektive Evaluation dient daher vor allem dazu, einen Anhaltspunkt für einen Vergleich zum derzeitigen Vorgehen zu ermöglichen und aufzuzeigen, wie groß die Reduktion der quantitativen Auffälligkeiten bei der Wahl verschiedener Werte für den Tuning-Parameter in etwa gewesen wäre, wenn die statistisch signifikante Auffälligkeitseinstufungs-Methodik im Erfassungsjahr 2017 zur Anwendung gekommen wäre, und die anschließende fachliche Bewertung stattdessen der Vorgehensweise des Strukturierten Dialogs für das Erfassungsjahr 2017 entsprochen hätte. Grundsätzlich bleibt offen, ob und inwieweit sich die Evaluationsergebnisse auf die in diesem Bericht vorgeschlagene Neugestaltung des der quantitativen Einstufung nachgelagerten Stellungnahmeverfahrens übertragen lassen.

Außerdem ist zu beachten, dass die Evaluation anhand zweier exemplarisch ausgewählter QS-Verfahren vorgenommen wird und die Ergebnisse nicht zwingend auf andere QS-Verfahren übertragbar sind.

#### 10.4.2 1-Jahres-Einstufung

Betrachtet wird als erstes die Anzahl an Fällen mit unerwünschten Ereignissen (über den Referenzbereich hinaus) in Abhängigkeit vom Tuning-Parameter der statistisch signifikanten Auffälligkeitseinstufung. Diese wird definiert als

$$M_{\text{stat.sig}}(\alpha) = \sum_{i=1}^I E[l_i(\cdot)] \cdot \left(1 - I\left(s_{\text{stat.sig}}(o_i, J_i, R, \alpha)\right)\right),$$

wobei  $I(\cdot)$  die Indikatorfunktion bezeichnet, die den Wert 1 annimmt, wenn ein Leistungserbringerergebnis statistisch signifikant auffällig wird.  $l(\cdot)$  ist wiederum die über die vom Referenzbereich hinaus tolerierte Anzahl an Fällen mit unerwünschten Ereignissen hinausgeht:  $(\theta - R)_+ \cdot J$ .

$M_{\text{stat.sig}}$  ist damit die über alle Leistungserbringer erwartete Anzahl an Fällen, die über die vom Referenzbereich hinaus tolerierte Anzahl an Fällen mit unerwünschten Ereignissen hinausgeht und für die kein Stellungnahmeverfahren durchgeführt wird, da das Leistungserbringerergebnis nicht als quantitativ auffällig eingestuft wird. Die Größe  $M$  ist somit ein Maß für die Anzahl an Fällen mit potentiellen Mängeln in der Behandlungsqualität, die jedoch nicht im Stellungnahmeverfahren nachverfolgt werden. Im Fall der rechnerischen Auffälligkeitseinstufung werden alle Leistungserbringer mit mehr unerwünschten Ereignissen als vom Referenzbereich toleriert als quantitativ auffällig eingestuft. Da jedoch im aktuellen Strukturierten Dialog nicht immer eine Stellungnahme angefordert wird, wird auch die Anzahl an Fällen mit unerwünschtem Ereignis, die durch den Verzicht auf das Anfordern einer Stellungnahme trotz rechnerischer Auffälligkeit in Kauf genommen wird, betrachtet. Sie ist definiert als

$$M_{\text{Hinweis}} = \sum_{i=1}^I E[l_i(\cdot)] \cdot I(S_{\text{rech}}(o_i, J_i, R) \wedge \text{keine Stellungnahme}).$$

Abbildung 43 stellt  $M_{\text{stat.sig}}$  und  $M_{\text{Hinweis}}$  in Abhängigkeit des Tuning-Parameters  $\alpha$  dar, der das Signifikanzniveau für die statistisch signifikante Auffälligkeit darstellt. Die oberen beiden Grafiken zeigen den Verlauf für Qualitätsindikatoren mit festem Referenzwert, die unteren beiden Grafiken den für Indikatoren mit perzentilbasiertem Referenzwert.  $M_{\text{stat.sig}}$  und  $M_{\text{Hinweis}}$  wurden dabei gepoolt über alle 11 ratenbasierten Qualitätsindikatoren des QS-Verfahrens *Hüftendoprothesenversorgung* bzw. alle 7 ratenbasierten Qualitätsindikatoren des QS-Verfahrens *Nierentransplantation* berechnet.

In den oberen beiden Grafiken für Indikatoren mit festem Referenzwert wird dabei deutlich, dass die Kurve für  $M_{\text{stat.sig}}$  für  $\alpha = 0,47$  bei 0 beginnt, da in diesem Fall alle rechnerisch auffälligen Ergebnisse auch mit der statistisch signifikanten Einstufungsmethode als quantitativ auffällig eingestuft werden und dadurch alle Standorte mit mehr unerwünschten Ereignissen als durch den Referenzbereich zugelassen in den Strukturierten Dialog kommen würden (unter der Annahme, es würden immer Stellungnahmen angefordert und keine Hinweise versendet werden). Als Vergleichswert wird die erwartete Anzahl  $M_{\text{Hinweis}}$  dargestellt, die im Strukturierten Dialog zum Erfassungsjahr 2017 durch das Versenden von Hinweisen bzw. den Verzicht auf das Anfordern einer Stellungnahme entsteht. Da diese Anzahl unabhängig vom Tuning-Parameter ist, entsteht in der Abbildung 43 für diese Methode eine Gerade.

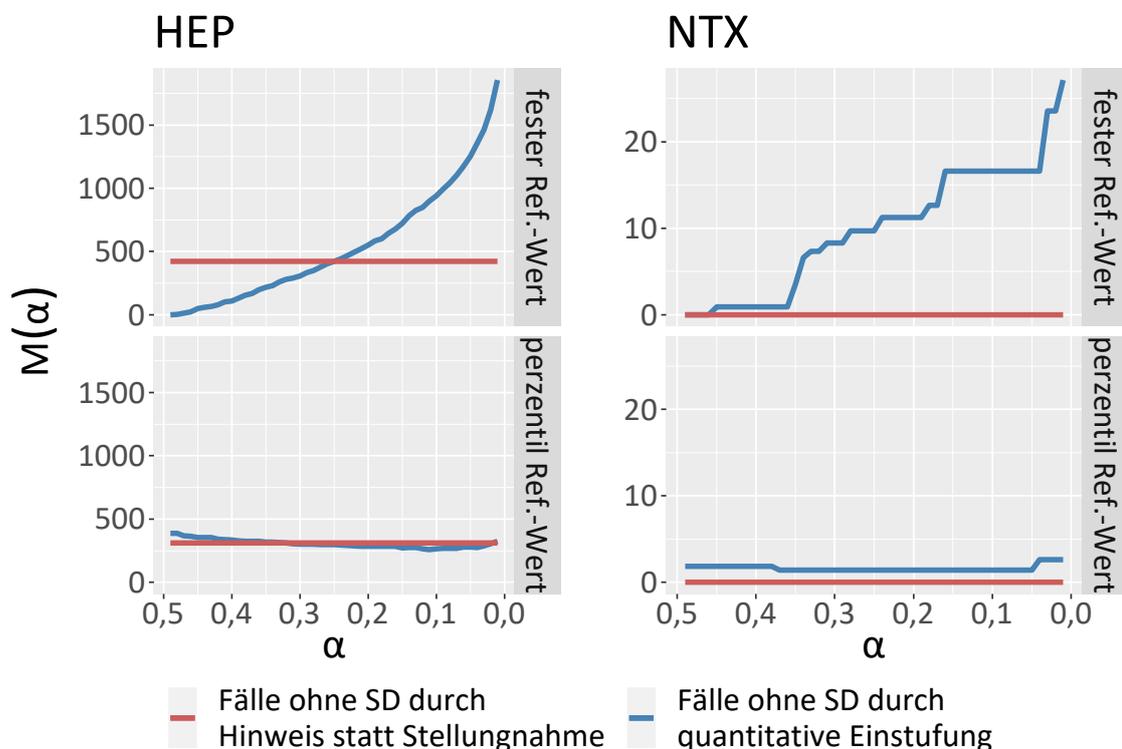


Abbildung 43: Über die vom Referenzwert hinaus tolerierte Anzahl an Fällen mit unerwünschtem Ereignis in Abhängigkeit des Tuning-Parameters ( $M(\alpha)$ )

Bei festen Referenzwerten zeigt sich: Je mehr statistische Evidenz für eine quantitativ auffällige Einstufung verlangt wird (d. h. je kleiner  $\alpha$ ), desto größer ist  $M$ . Bei HEP entspricht  $M_{\text{stat.sig}}$  bei etwa  $\alpha = 0,25$  dem Wert von  $M_{\text{Hinweis}}$ , d. h. der Anzahl an Fällen, die in der jetzigen Durchführung des Strukturierten Dialogs durch das Versenden eines Hinweises statt dem Anfordern einer Stellungnahme nach einer rechnerischen Auffälligkeit toleriert wird. Da bei NTX bei jeder rechnerischen Auffälligkeit auch eine Stellungnahme angefordert wird, entspricht  $M_{\text{stat.sig}}$  bei etwa  $\alpha = 0,47$  dem Wert von  $M_{\text{Hinweis}}$ .

Bei perzentilbasierten Referenzwerten, bei denen beispielsweise 5 % der Leistungserbringer quantitativ auffällig werden, hat  $\alpha$  nur in Ausnahmefällen einen Einfluss auf  $M$ . Da in der rechnerischen Auffälligkeitseinstufung und Berechnung von perzentilbasierten Referenzwerten in der Regel mehr als die nominell erwarteten 5 % der Leistungserbringer auffällig wurden, ist  $M$  auch bei großem  $\alpha$  zum Teil größer, als die Anzahl, die durch das Versenden von Hinweisen entsteht. Es besteht also ein nicht monotoner Zusammenhang zwischen dem jeweiligen Tuning-Parameter und der Maßzahl  $M$ . Dies trifft grundsätzlich auf Qualitätsindikatoren mit perzentilbasiertem Referenzwert zu, da bei diesen die Anzahl an quantitativen Auffälligkeiten festgelegt ist und die Klassifikationsmethoden sich nur darin unterscheiden können, welche (aber nicht wie viele) Leistungserbringerergebnisse quantitativ auffällig werden. Siehe auch Abschnitt 6.4.3 für eine Diskussion der methodischen Limitationen von perzentilbasierten Referenzbereichen.

#### 10.4.2.1 Anzahl an quantitativen Auffälligkeiten

Im Folgenden wird die Anzahl an quantitativen Auffälligkeiten in Abhängigkeit von  $\alpha$  dargestellt. In Abbildung 44 wird dabei wieder zwischen ratenbasierten Indikatoren mit festem (in den oberen beiden Grafiken) und perzentilbasiertem Referenzwert (in den unteren beiden Grafiken) unterschieden. Auf der x-Achse ist  $\alpha$  dargestellt, auf der y-Achse die absolute Anzahl an quantitativen Auffälligkeiten über alle ratenbasierten Indikatoren. Schwarz eingezeichnet ist die Anzahl an rechnerischen Auffälligkeiten und in Rot die Anzahl an Stellungnahmen im Strukturierten Dialog.

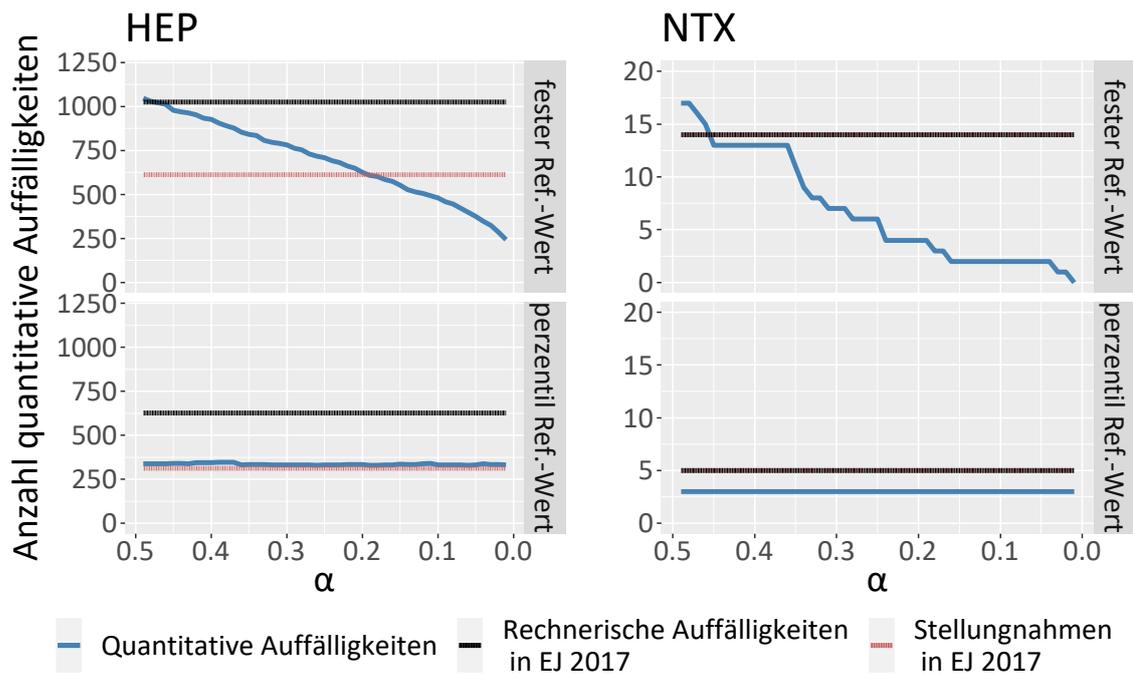


Abbildung 44: Anzahl an quantitativen Auffälligkeiten in Abhängigkeit des Tuning-Parameters

Dabei fällt auf, dass bei perzentilbasierten Referenzwerten unabhängig vom Tuning-Parameter die Anzahl an zu bearbeitenden Stellungnahmen bei HEP nahezu konstant auf dem Wert liegt, der aktuell von den Bewertungsstellen bearbeitet wird, bei NTX wiederum konstant knapp unterhalb dieses Wertes. Dies ist deshalb nicht verwunderlich, da durch die Perzentilreferenzwerte die letztliche Anzahl an quantitativen Auffälligkeiten bereits festgelegt ist.

Bei festen Referenzwerten ist der Einfluss des Tuning-Parameters  $\alpha$  auf die Aufwandseinsparung dagegen deutlich erkennbar.

#### 10.4.2.2 Anzahl an qualitativen Auffälligkeiten

Im Folgenden wird verglichen, welcher Anteil an Leistungserbringerergebnissen, die im Anschluss an eine rechnerische Auffälligkeit im Strukturierten Dialog für das Erfassungsjahr 2017 als qualitativ auffällig eingestuft wurden, mit der statistisch signifikanten Klassifikationsmethode als quantitativ auffällig bewertet worden wäre. Abbildung 45 zeigt den Anteil in Abhängigkeit von  $\alpha$ , jeweils getrennt nach ratenbasierten Qualitätsindikatoren mit festem Referenzwert (obere Grafiken) und Qualitätsindikatoren mit perzentilbasiertem Referenzwert (untere Grafiken).

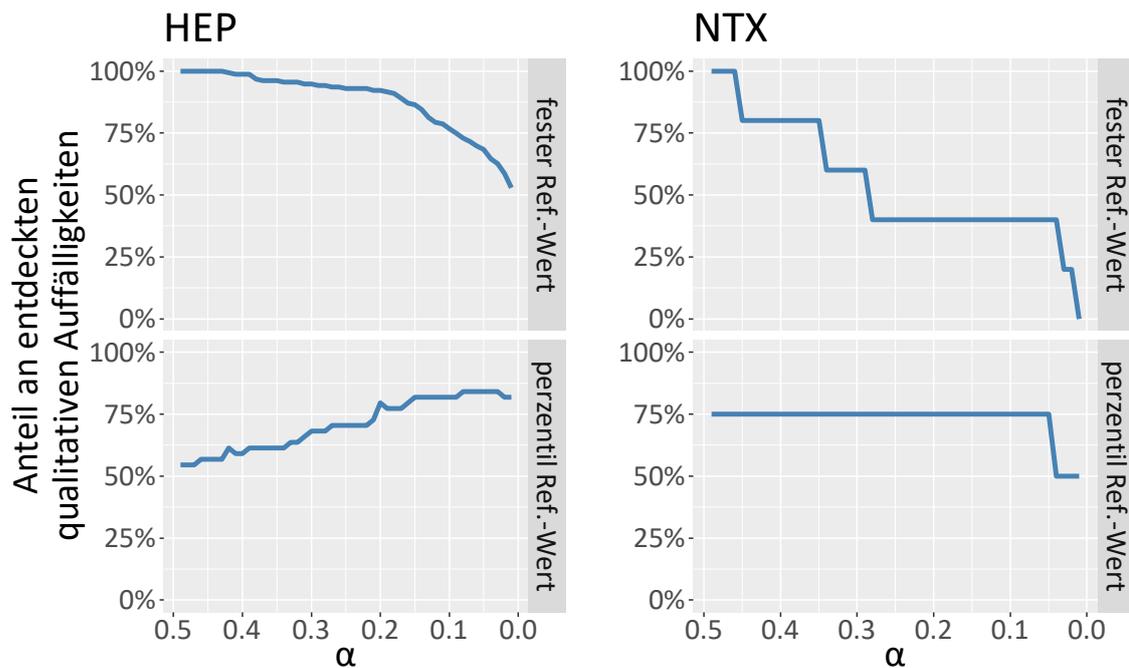


Abbildung 45: Entdeckte qualitative Auffälligkeiten in Abhängigkeit des Tuning-Parameters

Bei HEP würden bei Werten von  $\alpha = 0,05$  im hypothetischen Vergleich ca. 75 % der qualitativen Auffälligkeiten aus dem SD 2017 auch als quantitativ auffällig eingestuft werden. Bei NTX fällt der Anteil an entdeckten Auffälligkeiten mit kleiner werdendem  $\alpha$  bei Indikatoren mit festem Referenzwert deutlich schneller ab und liegt schon bei einem Wert von  $\alpha = 0,25$  bei unter 50 %. In Anbetracht der in Abschnitt 10.4.1 genannten Limitationen dieser Analyse sind diese Ergebnisse jedoch mit großer Zurückhaltung zu interpretieren. Insbesondere muss bemerkt werden, dass eine strengere Klassifikation, d.h. ein kleineres  $\alpha$ , automatisch immer mit einer Reduktion der quantitativen Auffälligkeiten und damit der entdeckten qualitativen Auffälligkeiten gemäß SD 2017 einhergeht.

#### 10.4.2.3 Positive-Predictive-Value

Im Folgenden wird der Positive-Predictive-Value (PPV) für die empfohlene Klassifikationsmethode als Anteil an qualitativen Auffälligkeiten unter allen als quantitativ auffällig eingestuften Leistungserbringern betrachtet.

Für perzentilbasierte Referenzbereiche ist für die Evaluation mit den Ergebnissen des Strukturierten Dialogs aus dem Erfassungsjahr 2017 dabei zu beachten, dass in seltenen Ausnahmefällen Leistungserbringerergebnisse mit der neu vorgeschlagenen Klassifikationsmethode quantitativ auffällig werden, die damals nicht rechnerisch auffällig waren. In diesen Fällen liegt keine Qualitätsbewertung aus dem Strukturierten Dialog zum Erfassungsjahr 2017 vor. Der PPV könnte dementsprechend höher oder niedriger sein; dies wird im Folgenden dadurch illustriert, dass der Bereich zwischen Best- und Worst-Case-Szenario (d. h. alle bzw. keiner der quantitativ auffällig aber nicht rechnerisch auffälligen Standorte wäre auch qualitativ auffällig) farblich dargestellt wird. Es handelt sich dabei nicht um ein Konfidenzband.

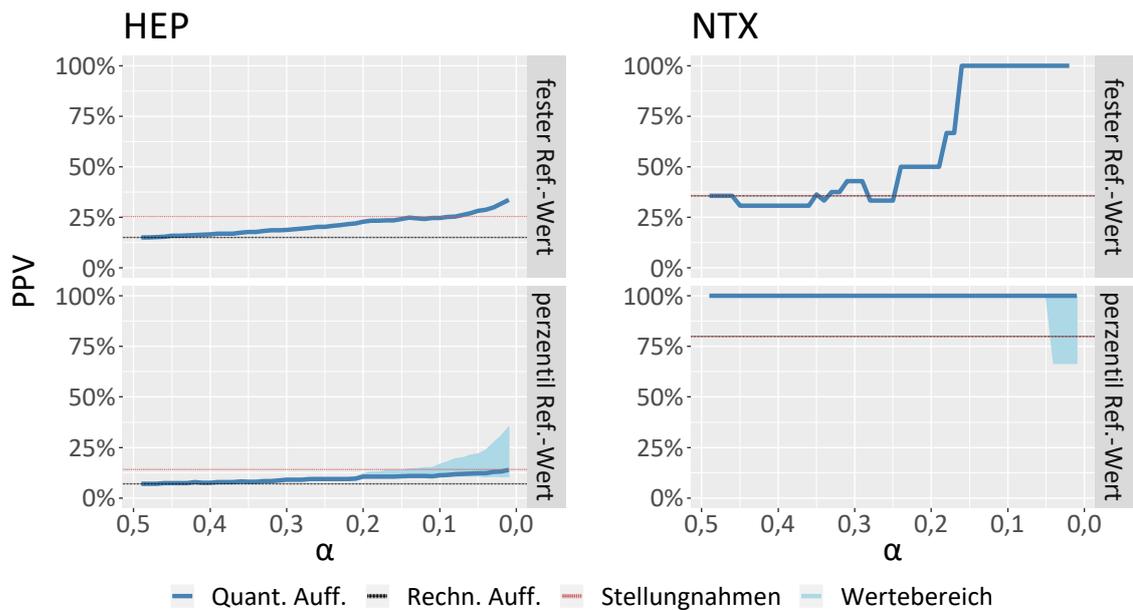


Abbildung 46: PPV in Abhängigkeit des Tuning-Parameters

Bei festen Referenzwerten steigt der PPV mit sinkendem  $\alpha$  und ist für HEP niemals geringer als der PPV der rechnerischen Auffälligkeit. Bei NTX ist die Schwankung des PPV über die verschiedenen Werte von  $\alpha$  größer und liegt teilweise unter der rechnerischen Auffälligkeit, was unter anderem auch durch die kleine Anzahl an betrachteten Standorten erklärt werden kann. Ab einem  $\alpha < 0,15$  steigt der PPV im Beispiel NTX jedoch auf 100 %, was bedeutet, dass durch die statistisch signifikante Einstufung ausschließlich Standorte quantitativ auffällig werden, die letztlich auch als qualitativ auffällig bewertet wurden.

Bei HEP entsteht ein Unterschied im PPV abhängig davon, ob alle rechnerischen Auffälligkeiten als Grundlage herangezogen werden, oder lediglich jene, bei denen nachaktuelles Vorgehen gemäß QSKH-RL Stellungnahmen angefordert wurden (vgl. Linien „Rechn. Auff.“ vs. „Stellungnahmen“ in Abbildung 46). Betrachtet man den PPV des aktuellen Vorgehens gemäß QSKH-RL („Stellungnahmen“), so wird deutlich, dass eine ähnliche Steigerung des PPV durch das neue Einstufungsverfahren im Vergleich zur rechnerischen Auffälligkeit erreicht werden kann. Bei der Wahl von  $\alpha = 0,05$  ist der PPV bei Indikatoren aus HEP mit festem Referenzwert bei über 25 % und damit fast doppelt so groß, wie der PPV der rechnerischen Auffälligkeit, jedoch nur knapp über dem PPV der Stellungnahmen.

#### 10.4.2.4 Abwägung von Aufwandsreduktion und Sensitivität

Der oben grafisch dargestellte Vergleich der Methoden zur aktuellen rechnerischen Auffälligkeit mit den durchgeführten Stellungnahmeverfahren im Erfassungsjahr (EJ) 2017 ist in Tabelle 16 für die statistisch signifikante Auffälligkeitseinstufungsmethode mit ausgewählten Werten von  $\alpha$  noch einmal tabellarisch für ratenbasierte HEP-Indikatoren mit festen Referenzbereichen illustriert. Da diese Tabelle nur eine vereinfachende Zusammenfassung der oben dargestellten Grafiken ist, wird hier auf die Darstellung des Leistungsbereichs NTX verzichtet.

Tabelle 16: Stellungnahmeverfahren (STNV), quantitative Auffälligkeiten, entdeckte qualitative Auffälligkeiten und PPV für ausgewählte Werte von  $\alpha$

$\alpha$	Durchgeführte STNV für das EJ2017	quantitative Auffälligkeiten (stat.sig. Auffälligkeit)	Verhältnis quant. Auffälligkeiten zu STNV (in %)	Anteil an vorhergesagten qualitativen Auffälligkeiten des EJ2017	PPV der quantitativen Auffälligkeit
0,20	613	629	103 %	92 %	23 %
0,10	613	484	79 %	77 %	25 %
0,05	613	378	62 %	68 %	28 %

Wäre beispielsweise die statistisch signifikante Auffälligkeit mit  $\alpha = 0,2$  als Klassifikationsmethodik verwendet worden, wären im Erfassungsjahr 2017 genau 629 Leistungserbringerergebnisse quantitativ auffällig geworden, was in etwa der Anzahl an durchgeführten Stellungnahmeverfahren entspricht. Bei diesem Signifikanzniveau wären außerdem 92 % der später als qualitativ auffällig bewerteten Ergebnisse auch durch die quantitative Auffälligkeit vorhergesagt worden. Bei  $\alpha = 0,05$  verringert sich die Anzahl an quantitativen Auffälligkeiten auf etwa 62 % der durchgeführten Stellungnahmeverfahren und es wären noch 68 % der als qualitativ auffällig bewerteten Ergebnisse durch die quantitative Auffälligkeit vorhergesagt worden. Dabei ist, wie oben beschrieben, grundsätzlich zu beachten, dass die qualitative Auffälligkeit keinen Goldstandard der tatsächlichen Qualitätsbewertung darstellt. Außerdem ist zu berücksichtigen, dass diese Werte in Abhängigkeit des betrachteten QS-Verfahrens stark variieren, wie der oben dargestellte Vergleich der Verfahren HEP und NTX illustriert. Daher kann es unter Umständen sinnvoll sein, den Tuning-Parameter  $\alpha$  verfahrens- oder QI-spezifisch zu wählen.

### 10.4.3 Fazit

Die Evaluation der empfohlenen Einstufungsmethodik anhand der QS-Verfahren HEP und NTX gibt Anhaltspunkte für die zu erwartende Anzahl an quantitativen Auffälligkeiten und vorhergesagter qualitativer Auffälligkeiten in Abhängigkeit des Tuning-Parameters  $\alpha$ . Im bisherigen Strukturierten Dialog wurde eine Abwägung zwischen der Anzahl an quantitativen Auffälligkeiten und dem Aufdecken qualitativer Auffälligkeiten zumindest schon teilweise durch das Versenden von Hinweisen in manchen QS-Verfahren, wie z. B. in HEP, getroffen. Obige Analysen zeigen, dass durch eine entsprechende Wahl des Tuning-Parameters die quantitativen Auffälligkeiten auf etwa jene Anzahl an Standorten reduziert werden könnte, von der auch im bisherigen Vorgehen eine Stellungnahme angefordert wurde. Dies würde gegenüber der bisherigen Praxis zu einer Vereinheitlichung über die Bundesländer und Qualitätsindikatoren zu führen. So wäre eine Vereinheitlichung im Vorgehen bei Beibehaltung des bisherigen Volumens an Stellungnahmen erreicht. Aufgrund der Heterogenität im bisherigen Vorgehen würde dies jedoch zu Volumenerhöhung für diejenigen Stellen führen, die bisher mehr Hinweise versendet haben als der Durchschnitt und es käme zu einer Volumenreduktion für diejenigen Stellen, die bisher weniger Hinweise versendet haben. Eine noch größere Volumenreduktion der quantitativen Auffälligkeiten ist darüber hinaus möglich, wenn noch strengere Werte als  $\alpha = 0,2$  gewählt werden.

Die qualitative Auffälligkeitseinstufung im Rahmen des Strukturierten Dialogs stellt bei dieser Evaluation dabei keinen Goldstandard für die Bewertung der zugrunde liegenden Qualität eines Leistungserbringers dar. Dies ist allein schon an der Heterogenität des Vorgehens der verschiedenen, den Strukturierten Dialog durchführenden Institutionen auf Landesebene erkennbar, welche nahe legt, dass die Maßstäbe für eine qualitative Auffälligkeit stark variieren (vgl. Abschnitt Heterogenitätsanalyse Abschnitt 2.5). Die vorgeschlagene Methodik für die quantitative Auffälligkeitseinstufung kann daher anhand dieser qualitativen Bewertungen nicht gegen einen Goldstandard evaluiert werden und die Wahl des Tuning-Parameters  $\alpha$  sollte nicht nur auf dem oben dargestellten Vergleich zur gegenwärtigen Praxis getroffen, sondern auch von theoretischen Überlegungen gestützt werden. Der obige Vergleich gibt jedoch Anhaltspunkte für die Anzahl an quantitativen Auffälligkeiten und vorhergesagter qualitativer Auffälligkeiten der empfohlenen Methode in Abhängigkeit vom verwendeten Tuning-Parameter.

In QS-Verfahren, in denen jede rechnerische Auffälligkeit zu einem Stellungnahmeverfahren geführt hat, führt die empfohlene Einstufungsmethodik immer zu einer Aufwandsreduktion, sofern ein  $\alpha$  von kleiner 0,47 gewählt wird, da dieser Wert in etwa der rechnerischen Auffälligkeitseinstufungsmethode entspricht. Die Abwägung zwischen einer Aufwandsreduktion und dem Verlust an entdeckten qualitativen Auffälligkeiten sollte dabei QS-verfahrensspezifisch (und möglicherweise QI-spezifisch) getroffen werden, d. h. der Tuning-Parameter  $\alpha$  sollte Verfahrens- und möglicherweise QI-spezifisch wählbar sein.

#### 10.4.4 Mehrjahreseinstufung

Analog zur Klassifikationseinstufung basierend auf einem Erfassungsjahr wird im Folgenden die für die sequentielle Klassifikationsmethodik zweier Erfassungsjahre vorgeschlagene Methodik anhand der Indikatorergebnisse und der Ergebnisse des Strukturierten Dialogs der QS-Verfahren *Hüftendoprothesenversorgung* und *Nierentransplantation* für die Erfassungsjahre 2016 und 2017 illustriert. Wie in Abschnitt 5.4.2 dargelegt, wird ein Leistungserbringerergebnis dabei dann als quantitativ auffällig eingestuft, wenn es entweder im aktuellen Erfassungsjahr statistisch signifikant auffällig ist, oder es im aktuellen Erfassungsjahr rechnerisch auffällig ist und die gepoolten Daten aus dem aktuellen Erfassungsjahr und dem Vorjahr statistisch signifikant auffällig sind<sup>76</sup>.

Wie bei der Evaluation der 1-Jahres-Einstufung werden auch hier wieder ratenbasierte Qualitätsindikatoren mit festem Referenzbereich von jenen mit verteilungsabhängigem Referenzbereich getrennt betrachtet und dargestellt. Verteilungsabhängige Referenzbereiche mit Perzentilwert  $q \cdot 100 \%$  werden dabei, wie in Abschnitt 5.4.3 dargelegt, so bestimmt, dass im aktuellen Erfassungsjahr 2017  $q \cdot 100 \%$  der Leistungserbringerergebnisse statistisch signifikant auffällig werden.

---

<sup>76</sup> Wobei mit statistisch signifikanter Auffälligkeit die Bayesianische Version  $S_{\text{stat.sig.bayes}}$  aus Abschnitt 5.3.1 gemeint ist

#### 10.4.4.1 Anzahl an unerwünschten Ereignissen ohne Stellungnahmeverfahren

Abbildung 47 zeigt die Anzahl an Fällen mit unerwünschten Ereignissen (über den Referenzbereich hinaus), die nicht im Stellungnahmeverfahren diskutiert werden, in Abhängigkeit vom Tuning-Parameter  $\alpha$  für die sequentielle Entscheidungsstrategie.

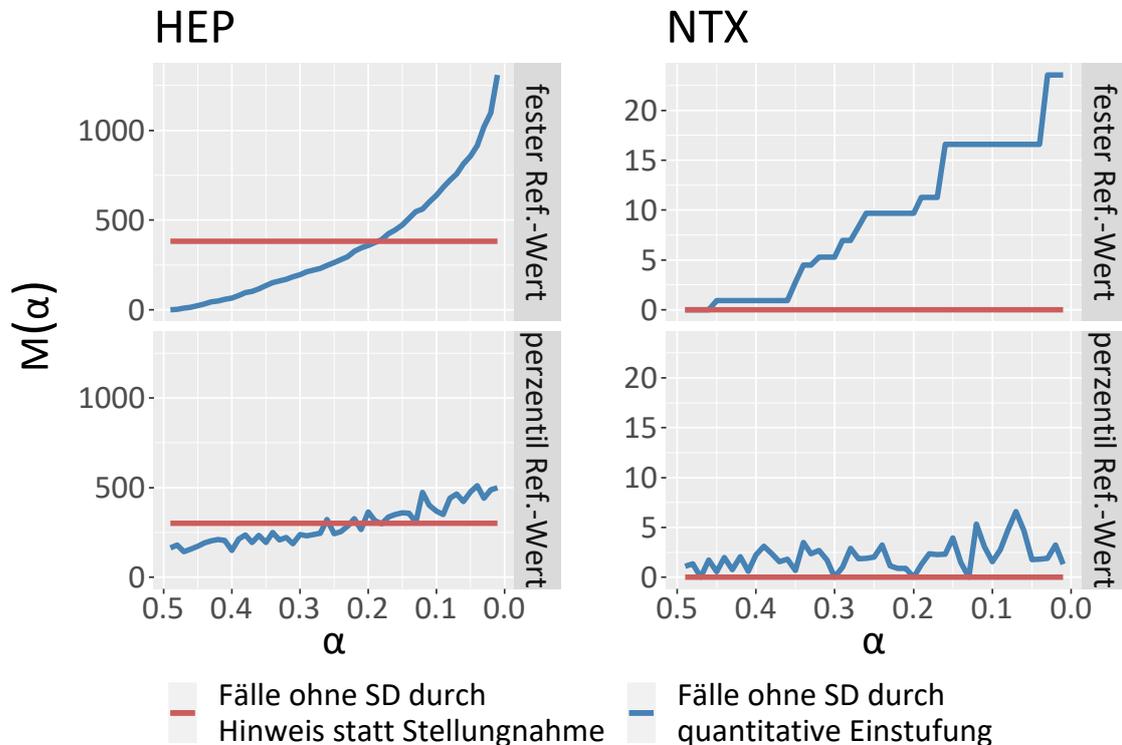


Abbildung 47: Über die vom Referenzwert hinaus tolerierte Anzahl an Fällen mit unerwünschtem Ereignis in Abhängigkeit vom Tuning-Parameter

Betrachtet werden hier nur Standorte, für die auch ein Vorjahresergebnis vorliegt, weshalb die Anzahl an Leistungserbringern ohne Strukturierten Dialog aufgrund der Versendung von Hinweisen statt Stellungnahmeanforderungen etwas niedriger ist als bei der einjährigen Einstufung. Insgesamt ist der Zusammenhang zwischen dem Tuning-Parameter  $\alpha$  und der Maßzahl  $M$  sehr ähnlich wie bei der einjährigen Einstufung über die statistisch signifikante Auffälligkeit. Für feste Referenzbereiche steigt die Anzahl an überschüssigen Fällen mit unerwünschtem Ereignis, für die kein Stellungnahmeverfahren eingeleitet wird, mit kleiner werdendem  $\alpha$ . Bei HEP entspricht bei etwa  $\alpha = 0,2$   $M$  der Anzahl, die durch das Versenden von Hinweisen derzeit schon in Kauf genommen wird. Dieser Schnittpunkt liegt bei der einjährigen Einstufung bei  $\alpha = 0,25$ , d. h. bei gleichem  $\alpha$  führt die sequentielle Einstufung zu einem kleineren  $M$ . Dies ist dadurch zu erklären, dass bei der 2-Jahres-Einstufung per Design alle Standortergebnisse auffällig werden, die auch in der einjährigen Einstufung quantitativ auffällig sind, sowie zusätzlich jene Standortergebnisse, deren aktuelles Ergebnis rechnerisch auffällig und deren gepoolte Daten aus den EJ 2017 und 2016 statistisch signifikant auffällig sind. Das heißt bei gleichem  $\alpha$  werden bei der sequentiellen Einstufung in der Regel mehr Standortergebnisse auffällig, was zu einem kleineren  $M$  führt. Auch bei NTX führt die Klassifikation über die 2-Jahreseinstufung bei gleichem  $\alpha$  zu etwas kleinerem  $M$  gegenüber der einjährigen Klassifikation.

Bei perzentilbasierten Referenzwerten liegt kein monotoner Zusammenhang zwischen  $\alpha$  und M vor und für HEP liegt M wieder durchgängig bei etwa dem Wert, der aktuell durch das Versenden von Hinweisen toleriert wird.

#### 10.4.4.2 Anzahl an quantitativen Auffälligkeiten

Wie oben dargelegt führt die betrachtete 2-Jahres-Klassifikation zu mindestens so vielen quantitativen Auffälligkeiten wie die einjährige Einstufung, was in Abbildung 48 bei Qualitätsindikatoren mit festem Referenzwert erkennbar ist. In Einzelfällen kann es dabei jedoch dadurch zu einer geringeren Anzahl an quantitativen Auffälligkeiten gegenüber der einjährigen Einstufung kommen, dass im Folgenden nur Standorte betrachtet werden, für die für beide betrachteten Erfassungsjahre Daten vorliegen.

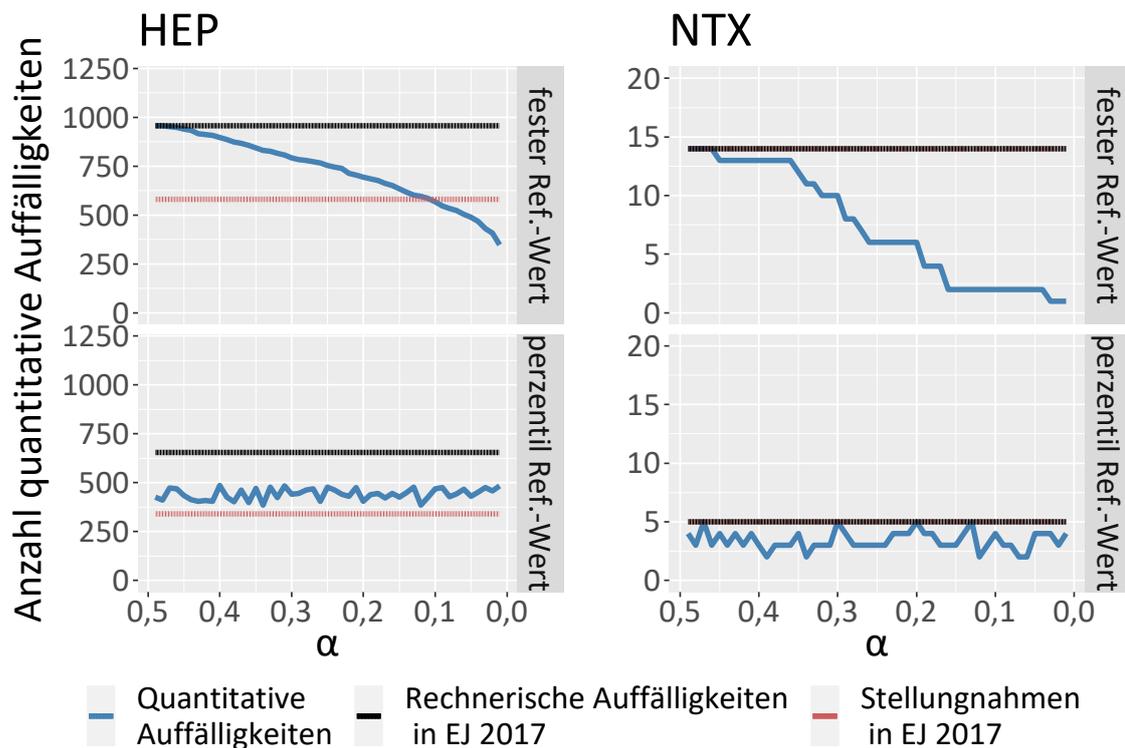


Abbildung 48: Anzahl an quantitativen Auffälligkeiten in Abhängigkeit vom Tuning-Parameter

Für Indikatoren mit festen Referenzwerten nimmt die Anzahl an Stellungnahmeverfahren mit abnehmendem  $\alpha$  ab und bei HEP wird bei etwa  $\alpha = 0,1$  der Wert erreicht, der durch den Versand von Hinweisen erreicht wird. Dies bedeutet, dass die sequentielle Einstufung bei gleichem  $\alpha$  zu mehr Stellungnahmeverfahren führt, als die einjährige Klassifikation, aus den oben erläuterten Gründen. Dies gilt auch für NTX, wobei hier die Anzahl an durchgeführten Stellungnahmen wiederum identisch ist zur Anzahl rechnerischer Auffälligkeiten.

Bei Indikatoren mit perzentilbasiertem Referenzwert liegt wieder kein monotoner Zusammenhang zwischen der Anzahl an Stellungnahmeverfahren und dem Parameter  $\alpha$  vor. Die Anzahl an quantitativen Auffälligkeiten liegt aber wieder durchgängig über jener der einjährigen Auffälligkeit, da die sequentielle Einstufung per Design zu gleich vielen oder mehr quantitativen Auffälligkeiten führt.

### 10.4.4.3 Anzahl qualitativer Auffälligkeiten

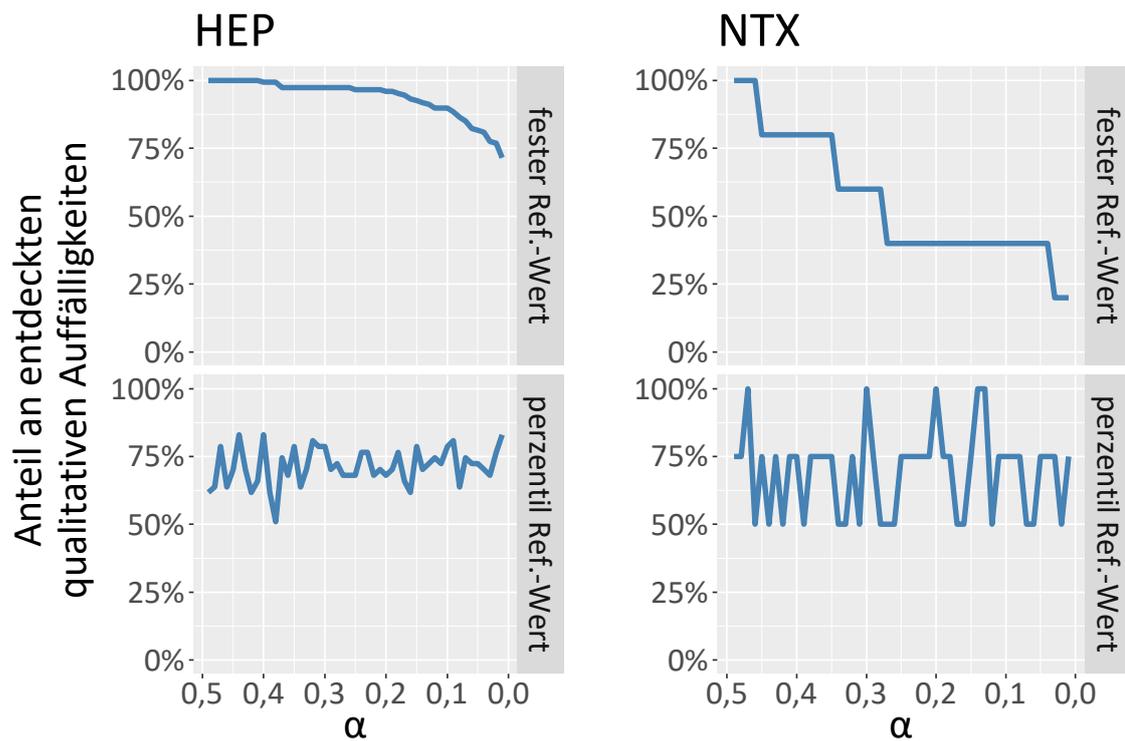


Abbildung 49: Entdeckte qualitative Auffälligkeiten in Abhängigkeit vom Tuning-Parameter

Auch der Zusammenhang zwischen  $\alpha$  und dem Anteil an entdeckten qualitativen Auffälligkeiten ist ähnlich wie bei der einjährigen Einstufung über die statistisch signifikante Auffälligkeit. Der Anteil ist bei der sequentiellen Einstufung jedoch durchgängig etwas größer, was wiederum dadurch zu erklären ist, dass die Anzahl an quantitativen Auffälligkeiten bei gleichem  $\alpha$  bei der sequentiellen Einstufung größer ist.

### 10.4.4.4 Positive-Predictive-Value

In Abbildung 50 dargestellt ist der PPV für die sequentielle Einstufung:

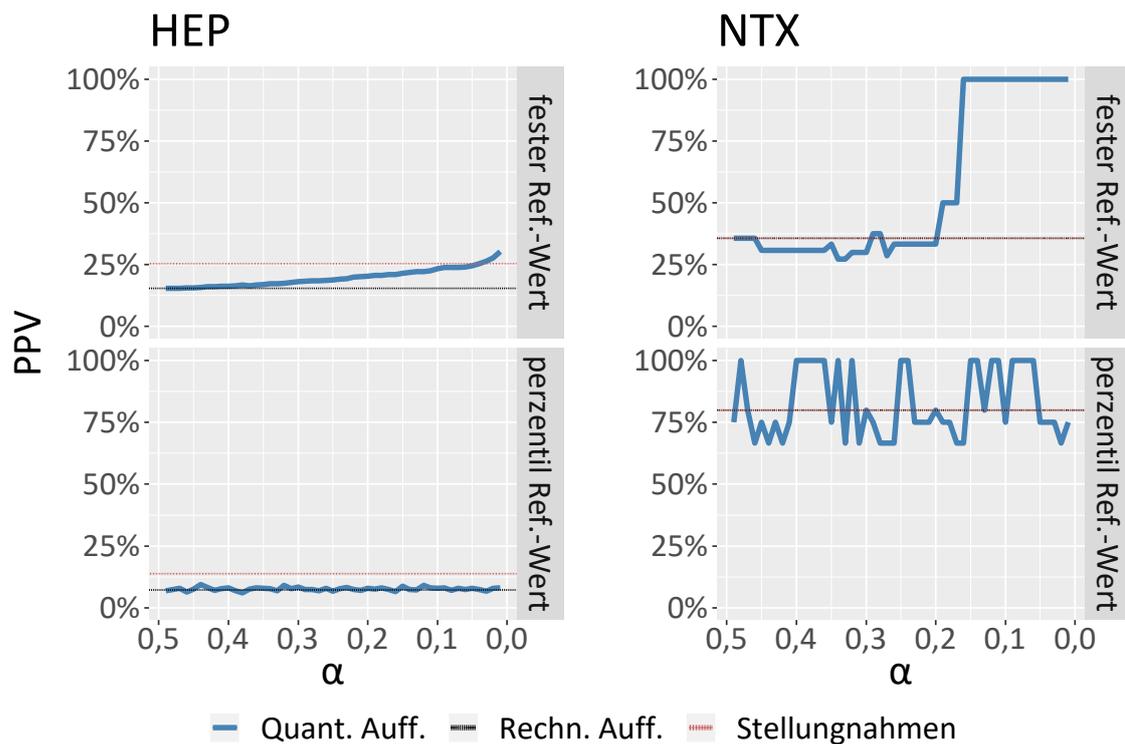


Abbildung 50: PPV in Abhängigkeit vom Tuning-Parameter

Abbildung 50 zeigt, dass der PPV bei HEP durchgängig etwas niedriger ist, als bei der einjährigen Einstufung. Dies bedeutet, dass die zusätzlichen quantitativen Auffälligkeiten durch die gepoolten Daten aus den Erfassungsjahren 2017 und 2016 einen geringeren PPV haben, als die quantitativen Auffälligkeiten aus dem Erfassungsjahr 2017 allein, wenn PPV in der Metrik der SD-Ergebnisse gemessen wird. Dies erscheint insofern plausibel, als dass schwerwiegende Qualitätsmängel im Erfassungsjahr 2017, welches im Strukturierten Dialog als qualitativ auffällig bewertet wurde, schon über die einjährige Einstufung des Erfassungsjahres 2017 quantitativ auffällig werden sollten. Alle *zusätzlichen* quantitativen Auffälligkeiten aus der Zweijahresbetrachtung sind daher insbesondere nicht statistisch signifikant auffällig alleinig auf Basis des Erfassungsjahres 2017 und wurden daher im SD zum Erfassungsjahr 2017 auch seltener als qualitativ auffällig bewertet. Insbesondere wurden gerade diese zusätzlichen quantitativen Auffälligkeiten eventuell häufig nur mit Hinweisen bearbeitet, da einige den Strukturierten Dialog durchführenden Stellen bereits ein Verfahren auf Basis eines statistischen Signifikanztests auf Basis eines Erfassungsjahres für die Durchführung von Stellungnahmen gegenüber dem Versenden von Hinweisen verwenden. Dies trägt dazu bei, dass die Einstufungen aus der Einjahres-Betrachtung einen höheren PPV aufweisen, als die Einstufungen basierend auf zwei Erfassungsjahren. Bei NTX ist dagegen kein deutlicher Unterschied im PPV zwischen der einjährigen und zweijährigen Einstufung zu erkennen.

#### 10.4.4.5 Abwägung von Aufwandsreduktion und Sensitivität

Wie für die 1-Jahres-Einstufung sind in Tabelle 17 die Anzahl an quantitativen Auffälligkeiten und der Anteil an vorhergesagten qualitativen Auffälligkeiten für den Leistungsbereich HEP für ausgewählte Werte von  $\alpha$  tabellarisch dargestellt:

$\alpha$	Durchgeführte Stellungnahme- verfahren	Quantitative Auf- fälligkeiten (2- Jahres stat. sig. Auffälligkeit)	Verhältnis quant. Auf- fälligkeiten zu STNV	Anteil an vorherge- sagten quali- tativen Auf- fälligkeiten	PPV der quantitati- ven Auffäl- ligkeit
0,2	580	695	120 %	96 %	20 %
0,1	580	567	98 %	90 %	23 %
0,05	580	488	84 %	82 %	25 %

Tabelle 17: Quantitative Auffälligkeiten, entdeckte qualitative Auffälligkeiten und PPV für ausgewählte Werte von  $\alpha$

Im Vergleich zur 1-Jahres-Klassifikation auf Basis der statistisch signifikanten Auffälligkeit wären bei der 2-Jahres-Einstufung und  $\alpha = 0,2$  genau 695 Leistungserbringerergebnisse quantitativ auffällig, im Vergleich zu 629 bei der 1-Jahres-Einstufung. Die Zahl der Stellungnahmeverfahren wird dabei bei etwa  $\alpha = 0,1$  erreicht, wobei 90 % der qualitativ auffälligen Ergebnisse abgedeckt worden wären.

#### 10.4.5 Fazit

Bei der 1-Jahres-Einstufung wurde deutlich, dass eine Aufwandsreduktion im Sinne einer Reduktion der quantitativen Auffälligkeiten gegenüber der aktuellen Anzahl an Stellungnahmeverfahren beim QS-Verfahren HEP erst ab etwa  $\alpha = 0,2$  erreicht wird. Dies ist dadurch begründet, dass auch bislang nicht für alle rechnerisch auffälligen Standorten eine Stellungnahmen im SD angefordert wurde, sondern stattdessen auch vermehrt lediglich Hinweise verschickt wurden. Ein Reduktion der quantitativen Auffälligkeiten im Vergleich zur Anzahl rechnerisch auffälliger Standorte wird somit dagegen bereits ab einem  $\alpha = 0,5$  erreicht.

Die Anzahl an quantitativen Auffälligkeiten bei der statistisch signifikanten Einstufungsmethodik über 2 Jahre führt, bei gleicher Wahl des Tuning-Parameters  $\alpha$ , zu einer etwas höheren Anzahl an quantitativen Auffälligkeiten als die statistisch signifikante Einstufungsmethodik basierend auf einem Erfassungsjahr. Bei etwa  $\alpha = 0,1$  wird im Beispiel von HEP die aktuelle Anzahl an durchgeführten Stellungnahmeverfahren erreicht. Der PPV der quantitativen Auffälligkeit ist dabei etwas geringer als jener der 1-Jahres-Einstufung; dies könnte darin begründet sein, dass erhebliche Qualitätsdefizite —die letztlich zu einer qualitativen Auffälligkeit geführt haben— auch bereits zu einer quantitativen Auffälligkeit bei der Betrachtung nur eines Erfassungsjahres führen, da die den Strukturierten Dialog durchführenden Institutionen bereits eine Art der statistischen Auffälligkeit nutzen, um eine Entscheidung zwischen dem Versenden von Hinweisen und dem Führen von Stellungnahmen zu treffen.

Bei NTX führt dagegen jede Wahl von  $\alpha$  kleiner als 0,47 zu einer Aufwandsreduktion gegenüber der aktuellen Anzahl an durchgeführten Stellungnahmeverfahren, da dort jede rechnerische Auffälligkeit auch im SD behandelt wird. Auch bei NTX führt die zweijährige Einstufung dabei zu einer etwas höheren Anzahl an quantitativen Auffälligkeiten, als die einjährige Einstufung, bei etwa identischem PPV. Generell sind die hier dargestellten Analyseergebnisse am Beispiel des QS-Verfahrens NTX deutlich wackliger, da bei insgesamt lediglich 14 rechnerischen Auffälligkeiten jede eingesparte Stellungnahme anteilmäßig einen enormen Effekt hat. Gleichmaßen wiegt auch jede qualitative Auffälligkeit, die somit potentiell nicht entdeckt würde, relativ schwer. Jedoch gilt wie für HEP auch für das verhältnismäßig kleine QS-Verfahren NTX, dass ein kleiner Tuning-Parameter  $\alpha$  eine größere Reduktion der ausgelösten Stellungnahmeverfahren und einen tendenziellen Anstieg des PPV in Bezug auf vorliegende Qualitätsmängel bedeutet.

Wie schon bei der Evaluation der 1-Jahres-Klassifikation sollte auch bei der Evaluation der Auswertungsmethodik basierend auf zwei Erfassungsjahren bedacht werden, dass die qualitative Bewertung im Rahmen des Strukturierten Dialogs kein Goldstandard für das tatsächliche Vorliegen von Qualitätsmängeln darstellt und der oben dargestellte Vergleich höchstens eine grobe Abschätzung liefern kann, mit welcher Reduktion an quantitativen Auffälligkeiten gegenüber der gegenwärtigen Praxis eine neue Klassifikationsmethode einhergeht.

## 11 Zusammenfassungen der Empfehlungen und Konsequenzen

Die vorliegenden Empfehlungen für ein richtlinienübergreifendes Vorgehen bei der Aus- und Bewertung von Indikatorergebnissen sind an den Zielen der Beauftragung ausgerichtet:

- Optimierung der Einheitlichkeit der Vorgehensweise (Objektivität des Verfahrens)
- Optimierung der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung (Objektivität des Verfahrens)
- Optimierung der Effizienz (Standardisierung)

Umgesetzt wurden diese Ziele in Form einer Methodik für die indikatorbasierte Qualitätsmessung und -bewertung, die die weiterführenden Handlungsanschlüsse, wie etwa öffentliche Berichterstattung (*accountability*) und qualitätssteigernde Maßnahmen (*improvement*) berücksichtigt. Das Konzept wurde aus den in Abschnitt 1.2 dargelegten Gründen mit Blick auf die Implementation in der DeQS-RL entwickelt. Durch die modulare Betrachtungsweise lassen sich die Empfehlungen für das Modul „Qualitätsbewertung“ jedoch prinzipiell auch in andere Richtlinien integrieren.

### 11.1 Anforderungen an und Weiterentwicklung von Qualitätsindikatoren

Die Beauftragung sieht vor dem Hintergrund der gestiegenen Anforderungen an die Prozesse und Ergebnisse der Qualitätssicherung auch eine methodische Weiterentwicklung der bestehenden Qualitätsindikatoren vor. Diese soll laut Beauftragung die Erfahrungen der LAG mittels einer Umfrage berücksichtigen. Eine Überarbeitung der mehr als 200 Qualitätsindikatoren war im Rahmen der Weiterentwicklung des Strukturierten Dialogs jedoch nicht möglich. Eine methodische Überarbeitung ist allerdings auch aus Sicht des IQTIG vor dem Hintergrund der gestiegenen methodischen Anforderungen an die Qualitätsindikatoren erforderlich. Dafür wird nachfolgend ein Vorgehen vorgeschlagen.

Um Qualitätsindikatoren weiterzuentwickeln muss das Ziel der Überarbeitung klar definiert sein. Aus Sicht des IQTIG ist die Maximierung der methodischen Güte (Validität) jedes Qualitätsindikators das Ziel einer solchen Weiterentwicklung. Gemäß den Methodischen Grundlagen V1.1 (IQTIG 2019a) definieren die Eignungskriterien die methodische Güte von Qualitätsindikatoren. Ein Qualitätsindikator ist dann methodisch optimal, wenn er alle Eignungskriterien der Methodischen Grundlagen erfüllt. Je höher die methodische Güte eines Qualitätsindikators ist, desto weniger häufig sollten tendenziell auch besondere Versorgungskonstellationen auftreten, die den Rückschluss vom Indikatorergebnis auf die Versorgungsqualität in Zweifel ziehen. Je seltener solche Konstellationen auftreten, desto geringer fällt auch der Aufwand für das Stellungnahmeverfahren aus, das die Funktion hat, solche Konstellationen zu identifizieren.

Eine methodische Weiterentwicklung der Indikatoren beinhaltet auch die Überprüfung der Referenzbereiche. Da verteilungsbezogene Referenzbereiche keine Aussagen über die Erfüllung definierter Standards erlauben, sondern lediglich vergleichende Aussagen relativ zu den

Ergebnissen aller Leistungserbringer, sind sie für die Qualitätsbewertung nur eingeschränkt nutzbar (vgl. Abschnitt 6.4.3) (IQTIG 2019a, S. 161 f.). Daher sollte im Rahmen der Weiterentwicklung der Indikatoren für alle Indikatoren mit verteilungsbezogenen (perzentilbasierten) Referenzbereichen geprüft werden, ob diese in feste Referenzbereiche überführt werden können.

Für die Weiterentwicklung der Qualitätsindikatoren wird daher ein zweischrittiges Vorgehen empfohlen. Im ersten Schritt werden die Qualitätsindikatoren hinsichtlich der Eignungskriterien geprüft. Für diesen Schritt hat das IQTIG bereits eine Methodik entwickelt, die zur Anwendung im Regelbetrieb kommen soll. Diese Methodik sieht vor, dass das IQTIG unter Einbezug der Bundesfachgruppen eine Einschätzung für jeden Qualitätsindikator hinsichtlich jedes Eignungskriterium vornimmt. Als Ergebnis der Eignungsprüfung jedes Qualitätsindikators steht eine individuelle Einschätzung des Indikators je Eignungskriterium sowie eine zusammenfassende Bewertung mit einer Empfehlung, entweder zur Abschaltung oder zur Überarbeitung.

Dieser Schritt sollte als erster und damit vor der Befragung der LAG erfolgen, da vermutlich schon in diesem Schritt manche Qualitätsindikatoren als nicht mehr geeignet beurteilt werden. So ist beispielsweise für einige Qualitätsindikatoren bekannt, dass das Kriterium „Potenzial zur Verbesserung“ (IQTIG 2019a, S. 140) nicht mehr erfüllt ist. Für diese Indikatoren muss dann keine Rückmeldung von den LAG eingeholt werden. Durch diesen ersten Schritt kann der Aufwand für eine Befragung der 16 LAG zu über 200 Indikatoren reduziert werden. Im zweiten Schritt sollen dann die Erfahrungen der LAG mit den nicht schon im ersten Schritt für die Abschaltung empfohlenen Qualitätsindikatoren eingeholt werden. Diese Erfahrungen sollen dann zusammen mit den Erkenntnissen aus der Eignungsprüfung für die Weiterentwicklung der Indikatoren verwendet werden.

## **11.2 Gleichsetzung quantitativer Auffälligkeit mit qualitativer Auffälligkeit**

Im Rahmen der Beauftragung soll das IQTIG auch Qualitätsindikatoren identifizieren, für die eine rechnerische Auffälligkeit mit einer qualitativen Auffälligkeit gleichgesetzt werden kann (Punkt 3.d der Beauftragung). Dieser Beauftragungspunkt wird so verstanden, dass das IQTIG prüfen soll, unter welchen Umständen ein Stellungnahmeverfahren verzichtbar ist und ob diese Umstände auf bestehende Qualitätsindikatoren zutreffen.

Aus den methodischen Vorüberlegungen und den Gütekriterien für Bewertungsprozesse (vgl. Kapitel 3) ergibt sich die klar umschriebene Funktion des Stellungnahmeverfahrens als nachgeschalteter Prüfschritt zur Sicherstellung der Aussagekraft eines Indikatorergebnisses (Validität je Messung), wenn dieses hinreichende Evidenz für ein Qualitätsdefizit nahelegt (vgl. Kapitel 6). Damit könnte eine Gleichsetzung des quantitativen Indikatorergebnisses mit einer abschließenden Qualitätsbewertung einerseits erfolgen, wenn der Bedarf für diesen Prüfschritt entfielen – also wenn die Aussagekraft eines quantitativen Indikatorergebnisses (im Sinne des Soll-Ist-Vergleichs mit dem Referenzwert ggf. mittels eines statistischen Verfahrens) als ausreichend hoch eingeschätzt würde. Vor dem Hintergrund des Weiterentwicklungsbedarfs der bestehenden Qualitätsindikatoren besteht aus Sicht des IQTIG allerdings weiterhin Bedarf für diesen Prüfschritt (siehe auch Rückmeldungen im Rahmen des Workshops zur Einbindung der Vertreterinnen und Vertreter der LAG und der LQS, Anhang, Kapitel 1).

Andererseits kann eine Entscheidung gegen diesen Prüfschritt auch aus Aufwand-Nutzen-Überlegungen heraus erfolgen, wenn etwa der Aufwand für diesen Prüfschritt nicht durch dessen Nutzen gerechtfertigt erscheint (vgl. Kapitel 5 zu Verlustfunktionen). Um zu einer solchen Einschätzung zu gelangen, müssen Aufwand und Nutzen eines Vorgehens mit diesem Prüfschritt und denen eines hypothetischen Vorgehens ohne diesen Prüfschritt verglichen werden. Der Nutzen dieses Prüfschritts besteht in einer Steigerung der Spezifität des Vorgehens bezogen auf die Wahrscheinlichkeit, „tatsächliche“ Qualitätsdefizite zu detektieren im Vergleich zu einem Vorgehen ohne diesen Prüfschritt (vgl. Abschnitt 2.1). Dieser Nutzen kann aus verschiedenen Perspektiven unterschiedlich bewertet werden. Aus Sicht der Leistungserbringer könnte eine höhere Spezifität wünschenswert sein, da die Wahrscheinlichkeit sinkt, irrtümlicherweise unzureichende Qualität attestiert zu bekommen. Aus Sicht von Patientinnen und Patienten könnte die Sensitivität des Verfahrens höher gewichtet werden als die Spezifität, da eher Leistungserbringer mit „tatsächlich“ unzureichender Qualität identifiziert würden. Im nächsten Schritt müsste die Frage beantwortet werden, wie groß der Gewinn an Spezifität durch diesen Prüfschritt ausfällt. Dies ist nicht ohne weitere Annahmen möglich. Eine solche weitergehende Analyse geht jedoch über den Rahmen dieser Beauftragung hinaus.

Der Aufwand des Prüfschritts entsteht für verschiedene Beteiligte (Leistungserbringer, LAG bzw. Bundesstelle) durch alle nicht automatisierbaren Schritte des Moduls „Qualitätsbewertung“, angefangen von der Einholung der Stellungnahmen (LAG bzw. Bundesstelle), des Verfassens der Stellungnahmen (Leistungserbringer), über deren formale Prüfungen (LAG bzw. Bundesstelle), die inhaltliche Bewertung (Fachexpertinnen und -experten) bis hin zur abschließenden Einstufung des Indikatorergebnisses (siehe auch Kapitel 6). Nicht berücksichtigt ist hier der Aufwand für die Einleitung und Durchführung qualitätsverbessernder Maßnahmen bei den Leistungserbringern, da diese nicht zu dem Modul Qualitätsbewertung zählen. Eine Motivation für die Weiterentwicklung stellt laut Beauftragung auch der hohe Aufwand des bisherigen Vorgehens im zweiten Prüfschritt (dem Stellungnahmeverfahren) sowohl für die Leistungserbringer als auch für die LAG bzw. Bundesstelle dar. Eine teilweise Reduktion dieses Aufwands kann durch methodisch weiterentwickelte Qualitätsindikatoren erreicht werden, da sich dadurch deren Spezifität erhöht und damit der Aufwand für den folgenden Prüfschritt reduziert (siehe auch Abschnitt 11.2). Unter der Annahme, dass die bestehenden Qualitätsindikatoren teilweise eher als Aufgreifkriterien mit eher hoher Sensitivität aber niedriger Spezifität konstruiert wurden, erscheint der Bedarf für einen zweiten, die Spezifität des Vorgehens zur Qualitätsmessung steigenden Prüfschritt jedoch weiterhin gegeben.

Grundsätzlich ergibt sich hier jedoch ein Zielkonflikt zwischen einer maximal fairen Qualitätsmessung einerseits und einem möglichst geringen Aufwand andererseits, vor allem für die Leistungserbringer. Wie in Abschnitt 2.1 dargelegt, erhöht ein Stellungnahmeverfahren, das die Aussagekraft eines auffälligen Indikatorergebnisses überprüft, die Fairness der Qualitätsmessung zugunsten der Leistungserbringer, da es falsch-positive Ergebnisse (siehe Tabelle 1) reduziert und einheitliche a priori definierte Kriterien anlegt. Demgegenüber steht der Aufwand für die Leistungserbringer, für jedes Indikatorergebnis mit dem Hinweis auf ein Qualitätsdefizit eine Stellungnahme zu verfassen. Eine Steigerung der Spezifität und der damit einhergehenden Stei-

gerung der Fairness der Qualitätsmessung (siehe Abschnitt 2.1), wird erkaufte durch einen höheren Aufwand im Vergleich zu einem Vorgehen ohne Stellungnahmeverfahren. Diese Aufwandssteigerung durch ein Stellungnahmeverfahren im Vergleich zu einem Vorgehen ohne Stellungnahmeverfahren ist auch im Hinblick auf die vertragsärztlichen und vertragspsychotherapeutischen Leistungserbringer mit dem Gewinn an Fairness durch das Stellungnahmeverfahren abzuwägen. Im Gegensatz zu Leistungserbringern des stationären Sektors verfügen Leistungserbringer des ambulanten Sektors nicht über eigene QM-Abteilungen, die die Erstellung einer Stellungnahme unterstützen können. Eine Steigerung der Spezifität des Vorgehens der Qualitätsbewertung ohne weitere Aufwände bei den Leistungserbringern ist jedoch nicht möglich. Eine Quantifizierung dieser Aufwände ist im Rahmen dieses Projekts allerdings nicht möglich gewesen. Da weder konkrete Nutzen- noch Aufwandsabschätzungen für den Prüfschritt vorgenommen werden konnten, kann hier auch kein Aufwand-Nutzen-Vergleich eines Vorgehens mit und ohne zweiten qualitativen Prüfschritt vorgenommen werden. Anhand dieser Überlegungen wird jedoch deutlich, dass die Entscheidung für oder gegen den zweiten Prüfschritt entscheidend von der Bewertung des Gewinns an Spezifität im Verhältnis zu dem Aufwand abhängig ist. Da sich außerdem sowohl Aufwand als auch Nutzen zwischen den verschiedenen Beteiligten unterscheiden, existieren mindestens zwei verschiedene Perspektiven auf dieses Aufwand-Nutzen-Verhältnis, in Anbetracht der geringeren Personalressourcen pro Leistungserbringer im ambulanten Sektor können vermutlich drei Perspektiven unterschieden werden: die der Patientinnen und Patienten, die der stationären Leistungserbringer und die der ambulanten Leistungserbringer. Vor dem Hintergrund des Weiterentwicklungsbedarfs einiger bestehender Qualitätsindikatoren und dem damit einhergehenden Bedarf an einer (unbekannten) Steigerung der Spezifität des Vorgehens durch den zweiten, qualitativen Prüfschritt wird empfohlen, das Stellungnahmeverfahren beizubehalten.

### 11.3 Indizes als Aufgreifkriterien

Die Beauftragung sieht auch vor, dass das IQTIG prüfen soll, ob Indizes als Aufgreifkriterien für die Bewertung ganzer Leistungsbereiche verwendbar sind (Punkt 3.e der Beauftragung). Indizes in diesem Sinn werden verstanden als eine (gewichtete) Aggregation der Ergebnisse mehrerer Indikatorergebnisse pro Leistungserbringer und Leistungsbereich. Unter der Annahme, dass bei der Verwendung von Indizes als Aufgreifkriterien für die Bewertung der Leistungserbringung in ganzen Leistungsbereichen keine Betrachtung der einzelnen Indikatorergebnisse mehr nötig ist, könnte dieses Vorgehen zu einer Effizienzsteigerung beitragen, da weniger Ergebnisse bewertet werden müssten. Diese Annahme wird vom IQTIG jedoch nicht geteilt (siehe Abschnitt 4.2).

Wie in Abschnitt 2.2 dargelegt, ist aus Sicht des IQTIG vor dem Hintergrund der gestiegenen Anforderungen an die Qualitätsmessung ein Wandel geboten im Verständnis von Qualitätsindikatoren weg von Aufgreifkriterien hin zu quantitativen Größen für Qualität. Daher sieht das hier vorgeschlagene Konzept vor, dass das Ergebnis jedes einzelnen Qualitätsindikators abschließend bewertet werden soll. Schon heute werden in der gesetzlich verpflichtenden Qualitätssicherung Indikatoren eingesetzt, die selbst ein Index sind (z. B. QI 51901 Qualitätsindex der Frühgebo-

renenversorgung). Bei diesen Indikatoren findet die Qualitätsbewertung jedoch erst auf Indexebene statt – nicht auf Ebene der konstituierenden Kennzahlen oder Variablen. Durch die Entscheidung einen Referenzbereich zu setzen, auf einem Index oder auf einer Kennzahl, fällt per Definition die Entscheidung, auf der jeweiligen Aggregationsebene eine Qualitätsbewertung vorzunehmen. Da in dem hier vorliegenden Konzept die Funktion eines Stellungnahmeverfahrens darin gesehen wird, falsch-positive Qualitätsbewertungen (siehe Abschnitt 2.1) zu reduzieren und damit die Spezifität der Qualitätsbewertung zu erhöhen, folgt aus der Entscheidung einen Referenzbereich zu setzen logisch die Notwendigkeit für ein Stellungnahmeverfahren auf der Aggregationsebene, auf der der Referenzbereich gesetzt wurde.

Dies ist unter anderem auch deshalb notwendig, da die Ergebnisse der einzelnen Qualitätsindikatoren für andere Zwecke weiterverwendet werden. Dieses Vorgehen ist damit auch im Einklang mit dem Vorgehen gemäß plan. QI-RL. Die diesem Punkt der Beauftragung scheinbar zugrunde liegende Annahme, dass eine Bewertung einzelner Indikatorergebnisse verzichtbar sei, wird daher nicht geteilt. Aus diesen Gründen wird eine Verwendung von aggregierten Qualitätsindikatorergebnissen als Aufgreifkriterien nicht empfohlen. Gemäß der Methodischen Grundlagen V1.1 können Indizes jedoch prinzipiell geeignet sein, um die Versorgungsqualität eines Leistungserbringers in einem Leistungsbereich zu einer Messgröße zu verdichten (Profit et al. 2010). Auch für weiterführende Verwendungszwecke, wie etwa eine über einen Qualitätsindikator hinausgehende Qualitätsbewertung, beispielsweise einer Fachabteilung oder wie bei der Entwicklung einer Internetplattform für eine Vergleiche ermöglichende Qualitätsdarstellung, können Indizes geeignet sein.

#### **11.4 Leistungsbereichübergreifende Qualitätsindikatorensets**

Des Weiteren beinhaltet die Beauftragung auch die Prüfung von Optionen für leistungsbereichsübergreifende Indikatorensets (Punkt 3.c der Beauftragung). Dieser Punkt der Beauftragung wird so verstanden, dass das IQTIG prüfen soll, ob diagnose- und/oder prozedurunabhängige Qualitätsindikatoren sinnvoll im Rahmen der gesetzlich verpflichtenden QS eingesetzt werden können. Es wird nicht davon ausgegangen, dass solche leistungsbereichübergreifenden Indikatorensets als „Aufgreifkriterien“ verstanden werden.

Das Verständnis von leistungsbereichübergreifenden Indikatoren soll am Beispiel zweier Indikatoren erläutert werden, die jeweils den Anteil von Patientinnen und Patienten mit Komplikationen abbilden, einmal nach Hüftendoprothesenversorgung und einmal nach Knieendoprothesenversorgung. Mittels beider Indikatoren könnten mehr oder weniger für diesen Eingriff spezifische Komplikationen abgebildet werden. Ein leistungsbereichübergreifender Indikator in diesem Beispiel würde den Anteil von Patientinnen und Patienten mit Komplikationen unabhängig von dem Eingriff zusammenfassen. Als Weiterführung dieses Beispiels wäre ein Indikator vorstellbar, der den Anteil von Patientinnen und Patienten mit Komplikationen nach jeglichem Eingriff abbildet. Hier wird deutlich, dass die zugrunde liegende Frage nach diesem Verständnis von leistungsbereichübergreifenden Indikatoren die des Auflösungslevels des Indikators ist. Am einen Ende des Kontinuums steht ein für eine bestimmte Prozedur/Diagnose hochspezifi-

scher Indikator (z. B. „Lockerung der Hüftendoprothese“), am anderen Ende steht ein sehr allgemeiner Komplikationsindikator. Eine Effizienzsteigerung im Sinne der Beauftragung würde in diesem Verständnis dadurch entstehen, dass ein Leistungserbringer nur noch in einem statt in zwei Indikatoren auffällig werden könnte und dementsprechend auch nur ein Stellungnahmeverfahren eingeleitet und durchgeführt werden müsste.

Ob ein solches Vorgehen sinnvoll ist, hängt von zwei Faktoren ab: dem Verwendungszweck der Qualitätsmessung und der Annahme, dass das dem Indikator zugrunde liegende Qualitätsmerkmal über die beteiligten Leistungsbereiche hinweg homogen ist. Für die Ableitung von Maßnahmen zur Qualitätsverbesserung sind spezifischere Indikatoren geeigneter, da sie den Suchraum für die Problemursache eingrenzen. Für Verwendungszwecke im Rahmen von *accountability* können hingegen auch weniger spezifische Indikatoren geeignet sein, beispielsweise um Vergütungszuschläge an allgemein niedrige Komplikationsraten zu knüpfen. Für Patienteninformation zum Zweck der Auswahlentscheidung ist dagegen das Informationsbedürfnis der Patientinnen und Patienten zu berücksichtigen. Denn typischerweise werden Auswahlentscheidungen mit Blick auf einen bestimmten elektiven Eingriff getroffen – der spezifischere Indikator im obigen Beispiel wäre in diesem Fall aussagekräftiger.

Das zweite Kriterium betrifft die Frage, ob angenommen wird, dass das dem Indikator zugrunde liegende Qualitätsmerkmal über die Leistungsbereiche hinweg homogen ist (vgl. auch reflektive vs. formative Konstrukte; Shwartz et al. 2015). Unterscheiden sich in dem obigen Beispiel die Komplikationsraten zwischen den beiden Eingriffen substantiell, können solche Unterschiede in einem allgemeinen Indikator verdeckt werden. Ob eine solche gegenseitige Kompensation eine erwünschte oder eine unerwünschte Eigenschaft eines Indikators ist, hängt unter Umständen von seinem Verwendungszweck ab. Im Rahmen einer Konzeptentwicklung können die Bedingungen für einen sinnvollen Einsatz von leistungsbereichübergreifenden Indikatoren beschrieben werden.

## 11.5 Zusammenfassung der zentralen Empfehlungen

Die Anforderungen an die Prozesse und Ergebnisse der gesetzlich verpflichtenden Qualitätssicherung sind in den letzten Jahren deutlich gestiegen (vgl. Abschnitt 2.2). Die Ergebnisse von Qualitätsmessungen sollen beispielsweise für die Planung von Versorgungsstrukturen (plan. QI-RL) oder für die leistungserbringerbezogene Veröffentlichung von Qualitätsergebnissen verwendet werden. Das hier vorgeschlagene Konzept greift die gestiegenen Anforderungen an Qualitätsindikatoren auf und zeigt Wege für eine vor diesem Hintergrund notwendige Weiterentwicklung der Qualitätsbewertung auf.

Mit Hinblick auf das Ziel frühzeitigere und aussagekräftigere Qualitätsergebnisse zu erhalten, wird empfohlen, die Instrumente der Qualitätsbewertung und der Qualitätsförderung konzeptuell und zeitlich voneinander zu trennen. Hierzu wird eine modulare Betrachtungsweise vorgeschlagen, die die Prozesse der Qualitätsbewertung, der Qualitätsförderung und der Bewertung der Dokumentationsqualität voneinander trennt. Durch stärkere Standardisierung soll der bisher teilweise aufwendige und unterschiedlich ausgestaltete Prozess der Qualitätsbewertung ob-

jektiver gestaltet und verschlankt werden. Es wird empfohlen, den Begriff des Stellungnahmeverfahrens nur für den Teil des Vorgehens nach § 17 DeQS-RL zu verwenden, der die Einholung und abschließende Bewertung der Stellungnahmen umfasst. Die darauf aufbauenden Maßnahmen der Qualitätsförderung sollen im Sinne der Trennung von Qualitätsbewertung und Qualitätsförderung nicht als Teil des Stellungnahmeverfahrens verstanden werden. Diese Trennung kann auch dazu beitragen, dass die Qualitätsergebnisse früher verfügbar sind als bisher.

### **Kriterium für die Einholung von Stellungnahmen**

Zur Steigerung der Spezifität und der Effizienz des Verfahrens wird eine statistische Operationalisierung des hinreichenden Hinweises auf ein Qualitätsdefizit auf Basis des quantitativen Indikatorergebnisses empfohlen. Dafür wird in Kapitel 5 ein Rahmenkonzept für die statistische Auswertungsmethodik von Qualitätsindikatoren dargestellt, welches die Festlegung von Herangehensweise, Stichprobenart, Berechnungsart und Bewertungsart erfordert. Bei einer analytischen Herangehensweise sollen stochastische Einflüsse und die damit einhergehende statistische Unsicherheit auf die Ergebnisse berücksichtigt werden. Diese Herangehensweise trifft aus Sicht des IQTIGs für die meisten fallbasierten Qualitätsindikatoren zu. Aufbauend auf dem Konzept wird ein formaler Rahmen beschrieben, innerhalb dessen sich, unter bestimmten Annahmen, die Optimalität der Methode der statistischen Auffälligkeit zeigen lässt. Durch den Einsatz dieser Methode können Verzerrungen bei der Feststellung eines hinreichenden Hinweises auf ein Qualitätsdefizit, wie sie bei Einsatz der einfachen rechnerischen Auffälligkeit bei kleinen Fallzahlen auftreten, vermieden werden. Die bayesianische Version der statistischen Auffälligkeitseinstufung bietet darüber hinaus einen flexibleren Rahmen als derjenige der frequentistischen Auffälligkeitseinstufung, sodass auch komplexere Indikatorarten in einem einheitlichen Rahmen ausgewertet werden können. Auch die Interpretation der Ergebnisse ist deutlich einfacher im bayesianischen Rahmen, z. B. kann von der Wahrscheinlichkeit der Nullhypothese gesprochen werden, d. h. wie wahrscheinlich es ist, nach Beobachtung der Daten, dass der Kompetenzparameter des Leistungserbringers im Referenzbereich liegt. Eine solche bayesianische Auswertungsmethodik wurde z. B. bereits für die Auswertung der Daten von Patientenbefragungen in den Abschlussberichten für die Patientenbefragungen im Verfahren *QS PCI* und im Verfahren *Versorgung von volljährigen Patienten und Patientinnen mit Schizophrenie, schizotypen und wahnhaften Störungen* empfohlen. Für Ratenindikatoren und risikoadjustierte Indikatoren per Poisson-Approximation, wie sie z. Zt. in der plan. QI-RL zum Einsatz kommen, unterscheiden sich die beiden Ansätze nur minimal. Daher empfiehlt das IQTIG auch für die Auswertung der übrigen Indikatoren im Rahmen der DeQS-RL die bayesianische Version der statistischen Auffälligkeit als Kriterium für die Entscheidung, ob ein Leistungserbringer das durch den Referenzbereich definierte Soll in einem Qualitätsindikator erreicht hat.

Durch die Berücksichtigung von Unsicherheit bei der statistischen Entscheidung, ob eine Stellungnahme eingeholt werden soll, wird durch das Signifikanzniveau eine zusätzliche Stellschraube eingeführt, ab wann von hinreichender statistischen Evidenz für ein mögliches Qualitätsdefizit ausgegangen werden kann. Die andere wichtige Stellschraube ist der Referenzwert für den Qualitätsindikator. Bei der in Kapitel 5 vorgeschlagenen statistischen Auffälligkeitseinstufungsmethode wird im bayesianischen Rahmen die Wahrscheinlichkeit berechnet, dass der

latente Kompetenzparameter des Leistungserbringers im Referenzbereich liegt. Das Signifikanzniveau der Auffälligkeitseinstufungsmethode legt dann fest, wie klein diese Wahrscheinlichkeit sein darf, bevor eine Entscheidung für die Alternativhypothese, d. h. der Kompetenzparameter liegt außerhalb des Referenzbereiches, getroffen wird. Die Frage ist, welcher Grad an Evidenz bzw. welches Signifikanzniveau in der Praxis für diese Entscheidung gewählt werden soll. Eine Diskussion dazu findet sich in Abschnitt 5.6 und ist eng mit der Methodik zur Festlegung der Referenzwerte bei den Qualitätsindikatoren verknüpft. Eine sehr pragmatische Vorgehensweise wäre z. B. übergangsweise das Signifikanzniveau so zu wählen, dass in etwa der gleiche Aufwand durch Stellungnahmen resultiert wie im bisherigen Rahmen des Strukturierten Dialogs.

Im Rahmen der DeQS-RL werden alle Auswertungen durch die Bundesauswertungsstelle vorgenommen. Der Aufwand für die Umsetzung der Empfehlungen zur statistischen Operationalisierung ergibt sich somit hauptsächlich für das IQTIG. Da durch die statistische Operationalisierung eindeutig vorgegeben sein soll, für welche Indikatorergebnisse eine Stellungnahme einzuholen ist, bestünde damit kein Ermessensspielraum mehr, wodurch sich ein Gewinn an Objektivität ergibt. Außerdem ergibt sich ein Effizienzgewinn im Vorgehen, da eine zu häufige Einholung von Stellungnahmen bei kleinen Fallzahlen vermieden wird und der Entscheidungsprozess, für welches Indikatorergebnis eine Stellungnahme einzuholen ist, praktisch automatisiert wird.

Durch die statistische Operationalisierung des hinreichenden Hinweises auf ein Qualitätsdefizit ist auch das Versenden von Hinweisen nicht mehr nötig. Bisher wurden Hinweise vermutlich versendet, um bei eher schwachen Hinweisen für Qualitätsdefizite (z. B. 1-Fall-Regel gemäß § 10 QSKH-RL) aufgrund der Auslösung durch die einfache rechnerische Auffälligkeit keine Stellungnahmen anzufordern. Da die Leistungserbringer schon durch die Rückmeldeberichte auf ihre Indikatorergebnisse aufmerksam gemacht werden und erwartet wird, dass das interne Qualitätsmanagement diese Ergebnisse von sich aus heranzieht und gegebenenfalls tätig wird, ist das Versenden von zusätzlichen Hinweisen verzichtbar.

Aus methodischer Sicht ist die Verwendung einer Mehrjahresdatenbasis bei der statistischen Auffälligkeitseinstufung entscheidend, um die statistische Unsicherheit bei der Auslösung des Stellungnahmeverfahrens zu reduzieren. In Abschnitt 5.4 werden dazu mehrere, aus der statistischen Prozesskontrolle inspirierte Verfahren vorgeschlagen. Es wird gezeigt, dass diese Verfahren in einem theoretischen Setting eine bessere Trefferquote haben, als die 1-Jahres-Verfahren. Die entscheidende Frage ist, ob die Erweiterung der Datenbasis auch praktikable Lösungen für die Diskussionen im Stellungnahmeverfahren bietet. Klar ist, dass mit mehr Fällen besser Strukturen und Muster erkannt werden können. Wichtig wäre es somit, erste Erfahrungen in der Praxis mit einer solchen Vorgehensweise zu machen.

### **Fachliche Bewertung von Stellungnahmen**

Ergibt sich aus einem Indikatorergebnis ein hinreichender Hinweis für ein Qualitätsdefizit gemäß der empfohlenen statistischen Methodik, soll dem Leistungserbringer die Gelegenheit eingeräumt werden, diesen Hinweis zu entkräften. Damit hat das Stellungnahmeverfahren die Funktion, das Vorliegen von Einflussfaktoren, die den Rückschluss vom Indikatorergebnis auf die Versorgungsqualität in Frage stellen können, aber noch nicht (genügend) in der

Berechnungsvorschrift des Indikators berücksichtigt sind, zu klären (Prüfung der Validität je Messung, vgl. Abschnitt 3.3). Diese Funktion von Stellungnahmen impliziert einen Bedeutungswandel des Stellungnahmeverfahrens weg von der umfassenden Bewertung der Versorgungsqualität hin zur Bewertung der Validität des Indikatorergebnisses im speziellen Fall. Durch die explizite Vorgabe von Entscheidungsregeln für diesen Prozess soll auch hier eine Steigerung der Objektivität des Verfahrens erreicht werden. Werden Qualitätsindikatoren als Aufgreifkriterien für einen nachfolgenden Peer-Review-Prozess verstanden, ist eine solche über das Qualitätsziel des Indikators hinausgehende Qualitätsbewertung sinnvoll. Im Verwendungskontext von *accountability* sind jedoch transparente und einheitliche Bewertungsregeln, wie sie hier vorgeschlagen werden, zentral (vgl. Abschnitt 3.2). Für die fachliche Bewertung der in den Stellungnahmen vorgebrachten Einflussfaktoren wird daher ein explizites Vorgehen einschließlich Nachberechnung des Indikatorergebnisses und der Vergleich mit einem (korrigierten) Referenzwert empfohlen. Damit sollen die Entscheidungsregeln, die zu einer Entkräftung der hinreichenden Evidenz für ein Qualitätsdefizit führen können, so transparent wie möglich gemacht werden. Auch aus wissenschaftlicher Sicht hat eine transparente Darstellung expliziter Entscheidungsregeln einen hohen Wert, da diese damit nachvollziehbar und kritisierbar werden. Eine solche Transparenz kann dadurch auch zu einer Steigerung der Akzeptanz des Verfahrens führen. Darüber hinaus soll die Akzeptanz des Verfahrens auch durch die fachliche Kompetenz, die Unabhängigkeit sowie die Interdisziplinarität der Fachkommissionen gesteigert werden. Kollegiale Gespräche und Begehungen (sowie die weiteren Förderinstrumente nach § 17 DeQS-RL) sollen im Sinne der Trennung von Qualitätsbewertung und -förderung zukünftig nicht mehr für die Ermittlung der Qualität der Versorgung eines Leistungserbringers eingesetzt werden. Dies steht im Einklang mit der Empfehlung, zukünftig nur noch beim Vorliegen überzeugender Gründe Indikatorergebnisse, die hinreichende Evidenz für ein Qualitätsdefizit nahelegen, in der fachlichen Bewertung mit „Kein Hinweis auf ein Qualitätsdefizit“ zu bewerten. Die überzeugende schriftliche Darlegung solcher Einflussfaktoren auf das Indikatorergebnis sollte Leistungserbringern, bei denen kein Qualitätsdefizit vorlag, auch ohne zusätzliche Begehung und kollegiales Gespräch möglich sein. Dies soll vor allem auch den Arbeitsaufwand bei der Bewertung der Stellungnahmen reduzieren.

### **Qualitätsförderung**

Die Empfehlungen für qualitätsverbessernde Maßnahmen zielen darauf ab, bei den begrenzten Ressourcen der LAG, die Maßnahmen auf diejenigen Leistungserbringer zu fokussieren, die die größten Qualitätsdefizite und vermutlich auch das größte Verbesserungspotenzial aufweisen. Dabei soll jedoch den LAG bzw. Bundesstelle möglichst freie Hand bei der Wahl der Fördermaßnahme gelassen werden, da diese auf die individuelle Situation vor Ort angepasst werden muss. Denn im Gegensatz zu einem möglichst standardisierten und damit objektiven Vorgehen bei der Qualitätsbewertung ist für qualitätsverbessernde Maßnahmen ein möglichst auf den entsprechenden Leistungserbringer zugeschnittenes Vorgehen unter Berücksichtigung der in Abschnitt 3.7 beschriebenen Erfolgsfaktoren sinnvoll.

### 11.5.1 Überblicksliste der Empfehlungen

Im Folgenden werden die Empfehlungen stichpunktartig zusammengefasst:

#### Ermittlung einer statistischen Auffälligkeit und ihre Einordnung

- Die fallabhängigen Qualitätsindikatoren der DeQS-RL werden nach einer analytischen Herangehensweise berechnet und ausgewertet, d. h. durch stochastische Einflüsse entstehende Unsicherheit ist bei der Berechnung der Indikatorergebnisse und beim Abgleich mit dem Referenzbereich zu berücksichtigen.
- Der hinreichende Hinweis auf ein Qualitätsdefizit wird als vorgegebener Schwellenwert (entsprechend dem Signifikanzniveau) für die Posteriori-Verteilung der Null-Hypothese im Rahmen eines bayesianischen Hypothesentests statistisch operationalisiert.
- Der Auswertungs- und Bewertungsprozess der QI-Ergebnisse erfolgt jährlich.
- Ergibt sich anhand der vorliegenden Daten eines Leistungserbringers kein hinreichender Hinweis auf ein Qualitätsdefizit, wird das Indikatorergebnis mit „kein hinreichender Hinweis auf ein Qualitätsdefizit“ (U0) bewertet.
- Ergibt sich anhand der vorliegenden Daten eines Leistungserbringers ein hinreichender Hinweis auf ein Qualitätsdefizit, dann erhält der Leistungserbringer die Möglichkeit zur Stellungnahme.
- Das Signifikanzniveau  $\alpha$  kann zunächst pro QS-Verfahren so gewählt werden, dass die Anzahl an anzufordernden Stellungnahmen konstant gegenüber der Anzahl der im vorangegangenen Jahr angeforderten Stellungnahmen ist.
- Als Teil der Begleitung der Umsetzung trifft das IQTIG unter Aufwand-Nutzen-Überlegungen eine inhaltlich-methodisch begründete Festlegung des Signifikanzniveaus  $\alpha$ . Berücksichtigt werden dabei die Güte der Indikatoren, der Aufwand und vorhandene Ressourcen für die Durchführung des Stellungnahmeverfahrens, Konsequenzen für Leistungserbringer und Patienten und Patientinnen bei Fehlklassifikationen, die Möglichkeit unterschiedlicher Signifikanzniveaus für unterschiedliche QI und/oder QS-Verfahren.
- Um eine höhere Treffsicherheit bei der statistischen Operationalisierung des hinreichenden Hinweises auf ein Qualitätsdefizit zu gewährleisten, soll perspektivisch eine Mehrjahresbetrachtung bei der Klassifikation der Leistungserbringerergebnisse stattfinden.

#### Formaler Ablauf des Stellungnahmeverfahrens

- Das Anforderungsschreiben für eine Stellungnahme soll am 30. Juni an die Leistungserbringer versendet werden.
- Der Leistungserbringer erhält ab Eingang des Anforderungsschreibens fünf Wochen Zeit, um eine qualifizierte Stellungnahme einzureichen.
- Erklärt der Leistungserbringer den Verzicht auf Abgabe einer Stellungnahme, wird das Indikatorergebnis mit „Qualitätsdefizit – Hinweis auf Qualitätsdefizit im Stellungnahmeverfahren bestätigt oder Verzicht auf Einreichung einer Stellungnahme“ (A3) bewertet.
- Geht die Stellungnahme nicht innerhalb der 5-wöchigen Frist ein, wird das Indikatorergebnis mit „Qualitätsdefizit – Stellungnahme nicht fristgerecht eingereicht oder Stellungnahme entsprach nicht den formalen Anforderungen“ (A0) bewertet.

- Stellungnahmen, die nicht den formalen Kriterien entsprechen, werden nicht berücksichtigt und das Indikatorergebnis wird mit „Qualitätsdefizit – Stellungnahme nicht fristgerecht eingereicht oder Stellungnahme entsprach nicht den formalen Anforderungen“ (A0) bewertet.
- Bis zum Ablauf der fünfwöchigen Frist kann der Leistungserbringer Korrekturen seiner Stellungnahme einreichen.

### **Formale Kriterien für Stellungnahmen**

- Für jedes Indikatorergebnis, welches nach dem obigen statistischen Verfahren einen hinreichenden Hinweis auf ein Qualitätsdefizit liefert, ist eine Stellungnahme abzugeben.
- Die Stellungnahme einschließlich eventuell notwendiger Anhänge ist in Schriftform zu erstellen und einzureichen.
- Belege, die die Argumentation des Leistungserbringers unterstützen, können in angemessenem Umfang ergänzend zur Stellungnahme eingereicht werden, ersetzen diese jedoch nicht. Belege, die ohne erläuternden Text übermittelt werden, sollen nicht anerkannt werden.
- Die fachlich verantwortliche Instanz des Leistungserbringers (z. B. die Chefärztin/der Chefarzt, Praxisinhaberin/Praxisinhaber) ist über den Hinweis auf ein Qualitätsdefizit zu informieren und sie soll die angefertigte Stellungnahme autorisieren.
- Die Anonymität sowohl der Patientinnen und Patienten (personenidentifizierende Daten) als auch ggf. weiterer Beteiligter (z. B. Zuweiser) ist zu wahren. Bei Nichteinhaltung der Vorgaben zur Wahrung der Anonymität soll die Stellungnahme aus Datenschutzgründen vernichtet werden und der Leistungserbringer über diesen Vorgang benachrichtigt werden.
- Der Leistungserbringer soll eine Erklärung darüber abgeben, ob die übermittelten QS-Daten, auf denen das Indikatorergebnis basiert, aus seiner Sicht korrekt sind.
- Die Erfüllung der formalen Kriterien für Stellungnahmen prüft und dokumentiert die LAG bzw. die Bundesstelle bevor sie die Stellungnahme an die Fachkommission weiterleitet.

### **Inhaltliche Anforderungen an Stellungnahmen**

- Der Leistungserbringer soll für jeden Grund, der aus seiner Sicht zur Abweichung vom Referenzbereich geführt hat, angeben, auf welche seiner Behandlungsfälle dieser Grund zutrifft bzw. ob der Grund für alle Behandlungsfälle zutrifft.
- Für die Zuordnung der Gründe zu den Behandlungsfällen sind die Vorgangsnummern der entsprechenden Behandlungsfälle mit der Stellungnahme zu übermitteln. Gleiches gilt auch für diejenigen Behandlungsfälle, bei denen kein interessierendes Ereignis (Behandlungsfälle, die nur im Nenner des Indikators vorkommen) aufgetreten ist.

### **Fachliche Bewertung**

- Die Prüfung, ob Gründe für die statistische Auffälligkeit des Leistungserbringers vorliegen, die nicht vom Leistungserbringer zu vertreten sind und nicht schon in der Rechenregel des Indikators oder in dem Risikoadjustierungsmodell des Indikators berücksichtigt sind, heißt „fachliche Bewertung“ und wird von einer Fachkommission auf Basis der schriftlichen Stellungnahme des Leistungserbringers durchgeführt.

- Begehungen und Besprechungen bzw. kollegiale Gespräche werden ausschließlich im Rahmen der Qualitätsförderung eingesetzt.
- In der Stellungnahme angeführte Gründe, die aus Sicht des Leistungserbringers nicht von ihm zu verantworten waren, sollen (mittels schriftlichem Beleg) nachgewiesen werden.
- Die Fachkommission klassifiziert die angeführten Gründe dahingehend, ob sie vom Leistungserbringer zu verantworten sind oder nicht.
- Gründe, bei denen aus der Stellungnahme nicht hervorgeht, in welchem Maße sie auch bei Behandlungsfällen ohne interessierendes Ereignis vorgelegen haben, sollen in der fachlichen Bewertung nicht berücksichtigt werden. Die Fachkommission führt auf Basis der Klassifikation der angeführten Gründe mittels technischer Unterstützung eine Nachberechnung des Indikatorergebnisses für den Leistungserbringer durch. Dabei werden die Fälle, bei denen die Fachkommission zu der Einschätzung gelangt, dass die angeführten Gründe nicht vom Leistungserbringer zu verantworten sind, aus der Grundgesamtheit des Indikators ausgeschlossen.
- Liegt nach der Nachberechnung kein hinreichender statistischer Hinweis auf ein Qualitätsdefizit vor, dann erfolgt eine Bewertung des Indikatorergebnisses mit „Kein Hinweis auf Qualitätsdefizit“ (U).
- Liegt nach der Nachberechnung ein hinreichender statistischer Hinweis auf ein Qualitätsdefizit vor, prüft die Fachkommission, ob weitere, nicht eindeutig klassifizierbare Gründe vom Leistungserbringer geltend gemacht wurden.
- Wurden vom Leistungserbringer weitere, nicht eindeutig klassifizierbare Gründe geltend gemacht, beurteilt die Fachkommission auf Basis ihrer fachlichen Expertise, ob diese Gründe hinreichend sind, um die statistische Auffälligkeit nach der Nachberechnung zu erklären.
- Wenn die Fachkommission zu der Einschätzung gelangt, dass die angeführten Gründe hinreichend sind, um die statistische Auffälligkeit zu erklären, empfiehlt sie das Indikatorergebnis als „Kein Hinweis auf Qualitätsdefizit“ (U) zu bewerten.
- Wenn die Fachkommission zu der Einschätzung gelangt, dass die angeführten Gründe nicht hinreichend sind, um die statistische Auffälligkeit nach der Nachberechnung zu erklären, empfiehlt sie das Indikatorergebnis als „Qualitätsdefizit“ (A) zu bewerten.
- Die Fachkommission empfiehlt eine abschließende Bewertung des Indikatorergebnisses gemäß dem Bewertungsschema einschließlich der Einstufung in die Ziffern.
- Die LAG bzw. die Bundesstelle trifft die abschließende Entscheidung über die Einordnung des Indikatorergebnisses in das Bewertungsschema.
- Weicht die LAG bzw. die Bundesstelle von der Einstufungsempfehlung der Fachkommission ab, begründet und dokumentiert sie diese Abweichung.
- Mit der abschließenden Entscheidung in Bezug auf die Einordnung des Indikatorergebnisses in das Bewertungsschema ist die Qualitätsbewertung abgeschlossen.
- In Ausnahmefällen kann die LAG bzw. die Bundesstelle das Indikatorergebnis eines Leistungserbringers mit „Sonstiges – ohne Bewertung“ (S) bewerten. Der Sachverhalt, der eine Qualitätsbewertung unmöglich machte, ist durch die LAG bzw. die Bundesstelle anzugeben.

### Zusammensetzung der Fachkommissionen

- Die LAG bzw. die Bundesstelle trägt die Verantwortung für die Auswahl der Expertinnen und Experten sowie die Zusammensetzung der Fachkommission.
- Die Fachkommissionen sollen interdisziplinär zusammengesetzt werden.
- In Abhängigkeit vom jeweiligen QS-Verfahren sind folgende Berufsgruppen einzubinden: Fachärztinnen und Fachärzte, Pflegefachkräfte, Klinische Qualitätsmanagerinnen und Qualitätsmanager, Patientenvertreterinnen und -vertreter, weitere Angehörige der Gesundheitsberufe, Krankenhaushygienikerinnen und -hygieniker oder Hygienefachkräfte

### Qualitätsverbessernde Maßnahmen

- Qualitätsverbessernde Maßnahmen der LAG bzw. der Bundesstelle schließen sich nur an eine abgeschlossene Qualitätsbewertung an.
- Die Fachkommission empfiehlt unter Berücksichtigung des Indikatorergebnisses und seiner abschließenden Bewertung den Einsatz von Maßnahmen zur Qualitätsverbesserung.
- Die LAG bzw. die Bundesstelle leitet ab dem zweiten mit „Qualitätsdefizit“ (A) bewerteten Indikatorergebnis bei dem gleichen oder einem inhaltlich ähnlichen Indikator innerhalb von drei Jahren immer qualitätsverbessernde Maßnahmen ein.
- Es liegt im Ermessen der Fachkommission, unter Angabe von Gründen bereits bei nur einem mit „Qualitätsdefizit“ (A) bewerteten Indikatorergebnis qualitätsverbessernde Maßnahmen der LAG bzw. der Bundesstelle bei einem Leistungserbringer zu empfehlen.
- Die LAG bzw. Bundesstelle entscheidet über die Art und Einleitung von notwendigen Maßnahmen gemäß DeQS-RL und QFD-RL.
- Begehungen und Besprechungen bzw. kollegiale Gespräche werden der Maßnahmenstufe 1 zugeordnet.
- Der Einsatz und Umsetzungsgrad der qualitätsverbessernden Maßnahmen ist zu dokumentieren.

#### 11.5.2 Fazit

Zusammenfassend würde eine Umsetzung der vorliegenden Empfehlungen deutliche Änderungen am bisherigen Vorgehen und an der DeQS-RL mit einer Annäherung an das Vorgehen nach plan. QI-RL bedeuten. Gleichzeitig wäre eine frühzeitigere Berichterstattung zur Versorgungsqualität möglich. Diese Änderungen müssten von verschiedenen Akteuren an verschiedenen Stellen umgesetzt werden. Während Änderungen in der statistischen Methodik für die Auswertung der Indikatorergebnisse im Rahmen der DeQS-RL vom IQTIG umgesetzt werden müssten, müssten die Änderungen, die sich für die Erstellung der Stellungnahmen ergeben, den Leistungserbringern erst umfassend kommuniziert werden. Die Nachberechnung im Rahmen der fachlichen Bewertungen durch die LAG und Fachkommission in Echtzeit setzt eine umfassende technische Unterstützung voraus, sodass die notwendige IT-seitige Entwicklungsarbeit bei einer möglichen Beauftragung der Umsetzung des Konzeptes berücksichtigt werden muss. Für die LAG, die Bundesstelle sowie die Fachkommissionen ergeben sich vermutlich die meisten Änderungen. Viele der empfohlenen Veränderungen zielen dabei auf eine Standardisierung des Vor-

gehens und auf eine Verschiebung der aufwendigen dialogischen Elemente von der Qualitätsbewertung hin zur Qualitätsförderung ab. Inwieweit sich die theoretisch zu erwartenden Effizienzsteigerungen in der praktischen Umsetzung realisieren lassen, lässt sich jedoch ohne erste praktische Erfahrungen mit den Empfehlungen nicht ohne Weiteres quantifizieren. Eine enge Abstimmung mit den LAG ist daher Voraussetzung für eine erfolgreiche Umsetzung. Die in Kapitel 10 gegebenen Empfehlungen zu einer formativen, also die Umsetzung begleitenden Evaluation spielen hier eine große Rolle. Beispielsweise könnte eine prozessbegleitende Befragung der LAG Hürden bei der Umsetzung der Empfehlungen zur Bewertung von Stellungnahmen erfragen und zu Verbesserungsvorschlägen führen.

Aufgrund der Vielzahl der Änderungen, die die Umsetzung des empfohlenen Konzepts mit sich bringen würde und aufgrund der Vielzahl der beteiligten Akteure, deren Handeln aufeinander abgestimmt werden müsste, empfiehlt das IQTIG eine strukturierte Begleitung der Umsetzung durch das IQTIG. Im Rahmen einer solchen Umsetzungsbegleitung sollte unter anderem ein Informationskonzept für die Leistungserbringer zu den geänderten Modalitäten des Stellungnahmeverfahrens umgesetzt, weitere Workshops mit den LAG-Vertreterinnen und -Vertreter sowie die formative Evaluation der Empfehlungen durchgeführt werden. Außerdem sollte in Abhängigkeit der umzusetzenden Empfehlungen eine schrittweise Umsetzung erfolgen und gegebenenfalls Übergangsregelungen getroffen werden.

## Literatur

- Adams, JL (2009): The Reliability of Provider Profiling. A Tutorial. Technical Report. [Stand:] 17.09.2019. Santa Monica, US-CA: RAND Corporation. Document Number: TR-653-NCQA. DOI: 10.7249/TR653.
- Agresti, A (2013): Categorical Data Analysis. Third Edition. (Wiley Series in Probability and Statistics). Hoboken, US-NJ: Wiley. ISBN: 978-0-470-46363-5.
- Altman, DG; Bland, JM (1995): Statistics Notes. Absence of evidence is not evidence of absence. *BMJ* 311:485. DOI: 10.1136/bmj.311.7003.485.
- AQUA [Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen] (2013): Bericht zum Strukturierten Dialog 2012. Erfassungsjahr 2011. Stand: 03.06.2013. Göttingen: AQUA. Signatur: 13-SQG-012. URL: [https://sqg.de/upload/CONTENT/Themen/Strukturierter\\_Dialog/Bericht\\_Strukturierter\\_Dialog\\_2012.pdf](https://sqg.de/upload/CONTENT/Themen/Strukturierter_Dialog/Bericht_Strukturierter_Dialog_2012.pdf) (abgerufen am: 27.02.2019).
- Ash, AS; Fienberg, SE; Louis, TA; Normand, SLT; Stukel, TA; Utts, J (2012): Statistical Issues in Assessing Hospital Performance. Commissioned by the Committee of Presidents of Statistical Societies [*White paper*]. Revised: 27.01.2012. Baltimore, US-MD: Centers for Medicare and Medicaid Services; Committee of Presidents of Statistical Societies. URL: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf> (abgerufen am: 09.10.2019).
- Åström, KJ (1965): Optimal Control of Markov Processes with Incomplete State Information. *Journal of Mathematical Analysis and Applications* 10(1): 174-205. DOI: 10.1016/0022-247x(65)90154-x.
- BÄK [Bundesärztekammer] (2014): Leitfaden. Ärztliches Peer Review. (Texte und Materialien der Bundesärztekammer zur Fortbildung und Weiterbildung, Band 31). Berlin: BÄK. URL: [https://www.bundesaerztekammer.de/fileadmin/user\\_upload/downloads/Leitfaden\\_Aerztliches-Peer-Review\\_2014.pdf](https://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/Leitfaden_Aerztliches-Peer-Review_2014.pdf) (abgerufen am: 27.02.2019).
- Berger, JO (2010): Statistical Decision Theory and Bayesian Analysis. Second Edition. (Springer Series in Statistics). New York, US-NY [u. a.]: Springer. ISBN: 978-1-4419-3074-3.
- Berwick, DM (1991): Controlling Variation in Health Care: A Consultation from Walter Shewhart. *Medical Care* 29(12): 1212-1225.
- Berwick, DM; James, B; Coye, MJ (2003): Connections Between Quality Measurement and Improvement. *Medical Care* 41(1 Suppl.): I-30-I-38. DOI: 10.1097/00005650-200301001-00004.
- Brown, LD; Cai, TT; DasGupta, A (2001): Interval Estimation for a Binomial Proportion. *Statistical Science* 16(2): 101-133. DOI: 10.1214/ss/1009213286.

- Carlin, BP; Louis, TA (2008): *Bayesian Methods for Data Analysis Third Edition*. (Texts in Statistical Science). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC. ISBN: 978-1-58488-697-6.
- Carter, N (1989): Performance Indicators: 'backseat driving' or 'hands off' control? *Policy and Politics* 17(2): 131-138. DOI: 10.1332/030557389782454857.
- Chang, W; Cheng, J; Allaire, JJ; Xie, Y; McPherson, J (2019): shiny: Web Application Framework for R [*Open Source Software*]. Version 1.3.2. Published: 22.04.2019. Vienna, AT: R Foundation for Statistical Computing. URL: <https://CRAN.R-project.org/package=shiny> [Downloads: shiny\_1.3.2.tar.gz] (abgerufen am: 09.10.2019).
- Chassin, MR; Loeb, JM; Schmaltz, SP; Wachter, RM (2010): Accountability Measures – Using Measurement to Promote Quality Improvement. *The New England Journal of Medicine* 363(7): 683-688. DOI: 10.1056/NEJMs1002320.
- Chop, I; Eberlein-Gonska, M (2012): Übersichtsartikel zum Peer Review Verfahren und seine Einordnung in der Medizin. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 106(8): 547-552. DOI: 10.1016/j.zefq.2012.08.017.
- Christiansen, CL; Morris, CN (1996): Fitting and Checking a Two-Level Poisson Model: Modeling Patient Mortality Rates in Heart Transplant Patients. Kapitel 18. In: Berry, DA; Stangl, D; Hrsg.: *Bayesian Biostatistics*. (Statistics: textbooks and monographs, volume 151). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC, 467-501. ISBN: 0-8247-9334-X.
- Dawes, RM; Faust, D; Meehl, PE (1989): Clinical Versus Actuarial Judgment. *Science* 243(4899): 1668-1674. DOI: 10.1126/science.2648573.
- DeGEval [Gesellschaft für Evaluation] (2008): Standards für Evaluation. In: DeGEval; Hrsg.: *Standards für Evaluation*. 4. unveränderte Auflage. [Stand:] Juli 2008. Mainz: DeGEval, 10-13. ISBN: 3-00-009022-3. URL: [https://www.degeval.org/fileadmin/user\\_upload/Sonstiges/STANDARDS\\_2008-12.pdf](https://www.degeval.org/fileadmin/user_upload/Sonstiges/STANDARDS_2008-12.pdf) (abgerufen am: 09.10.2019).
- Deming, WE (1953): On the Distinction between Enumerative and Analytic Surveys. *Journal of the American Statistical Association* 48(262): 244-255. DOI: 10.1080/01621459.1953.10483470.
- Dimick, JB; Staiger, DO; Birkmeyer, JD (2010): Ranking Hospitals on Surgical Mortality: The Importance of Reliability Adjustment. *Health Services Research* 45: 1614-1629. DOI: 10.1111/j.1475-6773.2010.01158.x.
- Djulbegovic, B; Beckstead, JW; Elqayam, S; Reljic, T; Hozo, I; Kumar, A; et al. (2014): Evaluation of Physicians' Cognitive Styles. *Medical Decision Making* 34(5): 627-637. DOI: 10.1177/0272989x14525855.
- Donabedian, A (2003): *An Introduction to Quality Assurance in Health Care*. Oxford, GB [u. a.]: Oxford University Press. ISBN: 978-0-19-515809-0.
- Eberlein-Gonska, M; Rink, O; Winklmaier, C (2017): Wie wir Qualität verbessern. Kapitel 5. In: Eberlein-Gonska, M; Martin, J; Zacher, J; Hrsg.: *Handbuch IQM. Konsequenz transparent –*

*Qualität mit Routinedaten*. 2. aktualisierte und erweiterte Auflage. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft, 53-65. ISBN: 978-3-95466-115-2. URL: [https://www.initiative-qualitaetsmedizin.de/mediapool/1736/media\\_file/sendfile/](https://www.initiative-qualitaetsmedizin.de/mediapool/1736/media_file/sendfile/) (abgerufen am: 09.10.2019).

- Freeman, T (2002): Using performance indicators to improve health care quality in the public sector: a review of the literature. *Health Services Management Research* 15(2): 126-137. DOI: 10.1258/0951484021912897.
- G-BA [Gemeinsamer Bundesausschuss] (2018a): Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des IQTIG: Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern. [Stand:] 18.01.2018. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/39-261-3196/2018-01-18\\_IQTIG-Beauftragung\\_Weiterentwicklung-strukturierter-Dialog.pdf](https://www.g-ba.de/downloads/39-261-3196/2018-01-18_IQTIG-Beauftragung_Weiterentwicklung-strukturierter-Dialog.pdf) (abgerufen am: 27.02.2019).
- G-BA [Gemeinsamer Bundesausschuss] (2018b): Beschluss des Gemeinsamen Bundesausschusses über eine Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung. [Stand:] 19.07.2018. Berlin: G-BA. BAnz AT 18.12.2018 B3. URL: [https://www.g-ba.de/downloads/39-261-3419/2018-07-19\\_DeQS-RL\\_Erstfassung.pdf](https://www.g-ba.de/downloads/39-261-3419/2018-07-19_DeQS-RL_Erstfassung.pdf) (abgerufen am: 28.11.2018).
- G-BA [Gemeinsamer Bundesausschuss] (2019): Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des IQTIG mit einer Ursachenanalyse der Auffälligkeiten sowie der methodischen Weiterentwicklung der Qualitätsindikatoren zur präoperativen Verweildauer bei der Versorgung der hüftgelenknahen Femurfraktur. [Stand:] 17.01.2019. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/39-261-3666/2019-01-17\\_IQTIG-Beauftragung\\_QS-Indikatoren-h%C3%BCftgelenknahe-Femurfraktur.pdf](https://www.g-ba.de/downloads/39-261-3666/2019-01-17_IQTIG-Beauftragung_QS-Indikatoren-h%C3%BCftgelenknahe-Femurfraktur.pdf) (abgerufen am: 10.10.2019).
- Gardner, K; Olney, S; Dickinson, H (2018): Getting smarter with data: understanding tensions in the use of data in assurance and improvement-oriented performance management systems to improve their implementation. *Health Research Policy and Systems* 16:125. DOI: 10.1186/s12961-018-0401-2.
- George, EI; Ročková, V; Rosenbaum, PR; Satopää, VA; Silber, JH (2017): Mortality Rate Estimation and Standardization for Public Reporting: Medicare's Hospital Compare. *Journal of the American Statistical Association* 112(519): 933-947. DOI: 10.1080/01621459.2016.1276021.
- Gerber-Grote, A; Sandmann, FG; Zhou, M; ten Thoren, C; Schwalm, A; Weigel, C; et al. (2014): Decision making in Germany: Is health economic evaluation as a supporting tool a sleeping beauty? *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 108(7): 390-396. DOI: 10.1016/j.zefq.2014.06.018.
- Gerlach, F (2001): Konsequenzen: Empfehlungen für die eigene Tätigkeit in Praxis und Klinik. Kapitel 8. In: Gerlach, F: *Qualitätsförderung in Praxis und Klinik. Eine Chance für die Medizin*. Stuttgart [u. a.]: Georg Thieme, 267-285. ISBN: 978-3-13-125891-5.

- Griem, C; Kleudgen, S; Diel, F (2013): Instrumente der kollegialen Qualitätsförderung. *Deutsches Ärzteblatt* 110(26): A1310-A1313, A5. URL: <https://www.aerzteblatt.de/pdf.asp?id=141971> (abgerufen am: 09.10.2019).
- Hengelbrock, J; Höhle, M (2019): Evaluating quality of hospital care using time-to-event endpoints based on patient follow-up data. *Health Services and Outcomes Research Methodology*, Epub 08.07.2019. DOI: 10.1007/s10742-019-00202-7.
- Hensen, P (2016): Qualitätsmanagement im Gesundheitswesen. Grundlagen für Studium und Praxis. Wiesbaden: Springer Gabler. ISBN: 978-3-658-07744-0.
- Höhle, M; Paul, M (2008): Count data regression charts for the monitoring of surveillance time series. *Computational Statistics & Data Analysis* 52(9): 4357-4368. DOI: 10.1016/j.csda.2008.02.015.
- Huang, W; Reynolds, MR; Wang, S (2012): A Binomial GLR Control Chart for Monitoring a Proportion. *Journal of Quality Technology* 44(3): 192-208. DOI: 10.1080/00224065.2012.11917895.
- Hudson, DW; Holzmueller, CG; Pronovost, PJ; Gianci, SJ; Pate, ZT; Wahr, J; et al. (2012): Toward Improving Patient Safety Through Voluntary Peer-to-Peer Assessment. *American Journal of Medical Quality* 27(3): 201-209. DOI: 10.1177/1062860611421981.
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2016): Planungsrelevante Qualitätsindikatoren. Abschlussbericht zur Auswahl und Umsetzung. Stand: 31.08.2016. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2016/IQTIG\\_Planungsrelevante-Qualitaetsindikatoren\\_Abschlussbericht.pdf](https://iqtig.org/downloads/berichte/2016/IQTIG_Planungsrelevante-Qualitaetsindikatoren_Abschlussbericht.pdf) (abgerufen am: 09.10.2019).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2017a): Bericht zum Strukturierten Dialog 2016. Erfassungsjahr 2015. Stand: 15.05.2017. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2015/IQTIG\\_Bericht-zum-Strukturierten-Dialog-2016.pdf](https://iqtig.org/downloads/berichte/2015/IQTIG_Bericht-zum-Strukturierten-Dialog-2016.pdf) (abgerufen am: 27.02.2019).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2017b): Bericht zum Strukturierten Dialog 2016. Erfassungsjahr 2015. Anhang. Stand: 15.05.2017. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2015/IQTIG\\_Bericht-zum-Strukturierten-Dialog-2016\\_Anhang.pdf](https://iqtig.org/downloads/berichte/2015/IQTIG_Bericht-zum-Strukturierten-Dialog-2016_Anhang.pdf) (abgerufen am: 27.02.2019).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2017c): Ereigniszeitanalyse-Methodik für die Follow-up-Indikatoren nach QSKH-RL. Stand: 06.04.2017. Berlin: IQTIG. URL: [https://iqtig.org/dateien/berichte/2017/IQTIG\\_Ereigniszeitanalyse-Methodik-fuer-Follow-up-Indikatoren-nach-QSKH-RL\\_2017-04-06.pdf](https://iqtig.org/dateien/berichte/2017/IQTIG_Ereigniszeitanalyse-Methodik-fuer-Follow-up-Indikatoren-nach-QSKH-RL_2017-04-06.pdf) (abgerufen am: 23.09.2019).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2017d): Methodische Grundlagen V1.0. Stand: 15.09.2017. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2017/IQTIG\\_Methodische-Grundlagen-V1.0.pdf](https://iqtig.org/downloads/berichte/2017/IQTIG_Methodische-Grundlagen-V1.0.pdf) (abgerufen am: 15.05.2018).

- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018a): Bericht zum Strukturierten Dialog 2017. Erfassungsjahr 2016. Stand: 24.08.2018. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2016/IQTIG\\_Bericht-zum-Strukturierten-Dialog-2017\\_2018-08-24\\_barrierefrei.pdf](https://iqtig.org/downloads/berichte/2016/IQTIG_Bericht-zum-Strukturierten-Dialog-2017_2018-08-24_barrierefrei.pdf) (abgerufen am: 09.10.2019).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018b): Biometrische Methodik für die Follow-up-Indikatoren nach QSKH-RL für die Bundesauswertung des Auswertungsjahres 2017. Stand: 23.10.2018. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2017/IQTIG\\_Biometrische-Methodik-FU-Indikatoren\\_BuAW-AJ-2017\\_2018-10-23.pdf](https://iqtig.org/downloads/berichte/2017/IQTIG_Biometrische-Methodik-FU-Indikatoren_BuAW-AJ-2017_2018-10-23.pdf) (abgerufen am: 23.09.2019).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018c): Entwicklung einer Befragung von Patienten und Patientinnen mit Schizophrenie, schizotypen und wahnhaften Störungen. Entwicklung einer Patientenbefragung im Rahmen der Aktualisierung und Erweiterung des QS-Verfahrens *Versorgung von volljährigen Patienten und Patientinnen mit Schizophrenie, schizotypen und wahnhaften Störungen*. Abschlussbericht. Stand: 15.12.2018. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2018/IQTIG\\_Patientenbefragung\\_QS-Verfahren-Schizophrenie\\_Abschlussbericht-mit-AT\\_2018-12-15.pdf](https://iqtig.org/downloads/berichte/2018/IQTIG_Patientenbefragung_QS-Verfahren-Schizophrenie_Abschlussbericht-mit-AT_2018-12-15.pdf) (abgerufen am: 29.01.2020).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018d): Entwicklung von Patientenbefragungen im Rahmen des Qualitätssicherungsverfahrens *Perkutane Koronarintervention und Koronarangiographie*. Abschlussbericht. Stand: 15.12.2018. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2018/IQTIG\\_Patientenbefragung\\_QS-PCI\\_Abschlussbericht-mit-AT\\_2018-12-15.pdf](https://iqtig.org/downloads/berichte/2018/IQTIG_Patientenbefragung_QS-PCI_Abschlussbericht-mit-AT_2018-12-15.pdf) (abgerufen am: 20.01.2020).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018e): Hüftendoprothesenversorgung. Beschreibung der Qualitätsindikatoren für das Erfassungsjahr 2017. Indikatoren 2017. Stand: 06.04.2018. Berlin: IQTIG. URL: [https://iqtig.org/downloads/auswertung/2017/hep/QSKH\\_HEP\\_2017\\_QIDB\\_V01\\_2018-04-06.pdf](https://iqtig.org/downloads/auswertung/2017/hep/QSKH_HEP_2017_QIDB_V01_2018-04-06.pdf) (abgerufen am: 26.06.2019).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018f): Kriterien für den gezielten Datenabgleich in der Datenvalidierung nach QSKH-RL. Abschlussbericht. Stand: 19.12.2018. Berlin: IQTIG. [unveröffentlicht].
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2019a): Methodische Grundlagen V1.1. Stand: 15.04.2019. Berlin: IQTIG. URL: [https://iqtig.org/dateien/dasiqtig/grundlagen/IQTIG\\_Methodische-Grundlagen-V1.1\\_barrierefrei\\_2019-04-15.pdf](https://iqtig.org/dateien/dasiqtig/grundlagen/IQTIG_Methodische-Grundlagen-V1.1_barrierefrei_2019-04-15.pdf) (abgerufen am: 08.05.2019).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2019b): Methodische Grundlagen V1.1s. Würdigung der Stellungnahmen. Stand: 15.04.2019. Berlin: IQTIG. URL: [https://iqtig.org/dateien/dasiqtig/grundlagen/IQTIG\\_Methodische-Grundlagen-V1-1s\\_Wuerdigung-der-Stellungnahmen\\_2019-04-15.pdf](https://iqtig.org/dateien/dasiqtig/grundlagen/IQTIG_Methodische-Grundlagen-V1-1s_Wuerdigung-der-Stellungnahmen_2019-04-15.pdf) (abgerufen am: 12.09.2019).

- Jensen, FV; Nielsen, TD (2007): Bayesian Networks and Decision Graphs. Second Edition. (Information Science & Statistics). Berlin [u. a.]: Springer. ISBN: 978-0-387-68281-5.
- Kaelbling, LP; Littman, ML; Cassandra, AR (1998): Planning and acting in partially observable stochastic domains. *Artificial Intelligence* 101(1-2): 99-134. DOI: 10.1016/s0004-3702(98)00023-x.
- Kauermann, G; Küchenhoff, H (2011): Stichproben: Methoden und praktische Umsetzung mit R. Berlin [u. a.]: Springer. ISBN: 978-3-642-12317-7.
- Keiding, N; Clayton, D (2014): Standardization and Control for Confounding in Observational Studies: A Historical Perspective. *Statistical Science* 29(4): 529-558. DOI: 10.1214/13-STS453.
- Krell, RW; Hozain, A; Kao, LS; Dimick, JB (2014a): Reliability of Risk-Adjusted Outcomes for Profiling Hospital Surgical Quality. *JAMA Surgery* 149(5): 467-474. DOI: 10.1001/jama-surg.2013.4249.
- Krell, RW; Staiger, DO; Dimick, JB (2014b): Reliability of Surgical Outcomes for Predicting Future Hospital Performance. *Medical Care* 52(6): 565-571. DOI: 10.1097/mlr.0000000000000138.
- Krohne, HW; Hock, M (2015): Psychologische Diagnostik. Grundlagen und Anwendungsfelder. 2., überarbeitete und aktualisierte Auflage. Stuttgart: Kohlhammer. ISBN: 978-3-17-025255-4.
- Kruglanski, AW; Gigerenzer, G (2011): Intuitive and Deliberate Judgments Are Based on Common Principles. *Psychological Review* 118(1): 97-109. DOI: 10.1037/a0020762.
- Kuncel, NR; Klieger, DM; Connelly, BS; Ones, DS (2013): Mechanical Versus Clinical Data Combination in Selection and Admissions Decisions: A Meta-Analysis. *Journal of Applied Psychology* 98(6): 1060-1072. DOI: 10.1037/a0034156.
- Lauritzen, SL; Nilsson, D (2001): Representing and Solving Decision Problems with Limited Information. *Management Science* 47(9): 1235-1251. DOI: 10.1287/mnsc.47.9.1235.9779.
- Lieb, K; Klemperer, D; Koch, K; Baethge, C; Ollenschläger, G; Ludwig, WD (2011): Interessenkonflikte in der Medizin. Mit Transparenz Vertrauen stärken. *Deutsches Ärzteblatt* 108(6): A256-A260. URL: <http://www.aerzteblatt.de/int/article.asp?id=80790> [PDF-Version > Download] (abgerufen am: 30.09.2019).
- Liu, J; Louis, TA; Pan, W; Ma, JZ; Collins, AJ (2003): Methods for Estimating and Interpreting Provider-Specific Standardized Mortality Ratios. *Health Services and Outcomes Research Methodology* 4(3): 135-149. DOI: 10.1023/B:HSOR.0000031400.77979.b6.
- Montgomery, DC (2013): Statistical Quality Control. A Modern Introduction. Seventh Edition. Hoboken, US-NJ: Wiley. ISBN: 978-1-118-32257-4.
- Morris, CN (1983): Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association* 78(381): 47-55. DOI: 10.2307/2287098.

- Nimptsch, U; Peschke, D; Mansky, T (2016): Der Einfluss von Qualitätsmessung, Transparenz und Peer Reviews auf die Krankenhaussterblichkeit – Retrospektive Vorher-Nachher-Studie mit 63 Kliniken. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 115-116: 10-23. DOI: 10.1016/j.zefq.2016.05.007.
- Paddock, SM (2014): Statistical Benchmarks for Health Care Provider Performance Assessment: A Comparison of Standard Approaches to a Hierarchical Bayesian Histogram-Based Method. *Health Services Research* 49(3): 1056-1073. DOI: 10.1111/1475-6773.12149.
- Pearl, J (2009): Causality: Models, Reasoning, and Inference. Second Edition. Cambridge, GB [u. a.]: Cambridge University Press. ISBN: 978-0-521-89560-6.
- Profit, J; Typpo, KV; Hysong, SJ; Woodard, LD; Kallen, MA; Petersen, LA (2010): Improving benchmarking by using an explicit framework for the development of composite indicators: an example using pediatric quality of care. *Implementation Science* 5:13. DOI: 10.1186/1748-5908-5-13.
- Raats, VM; Moors, JJA (2003): Double-checking auditors: a Bayesian approach. *Journal of the Royal Statistical Society. Series D (The Statistician)* 52(3): 351-365. DOI: 10.1111/1467-9884.00364.
- Rink, O (2013): Das IQM-Peer-Review-Verfahren. Verbesserung der Ergebnisqualität, Strategie und Ergebnisse. *Prävention und Gesundheitsförderung* 8(1): 22-28. DOI: 10.1007/s11553-012-0374-x.
- Schmidt-Atzert, L; Amelang, M (2012): Psychologische Diagnostik. 5. Auflage. Berlin [u. a.]: Springer. ISBN: 978-3-642-17000-3.
- Shenoy, PP (1992): Valuation-Based Systems for Bayesian Decision Analysis. *Operations Research* 40(3): 463-484. DOI: 10.1287/opre.40.3.463.
- Shwartz, M; Cohen, AB; Restuccia, JD; Ren, ZJ; Labonte, A; Theokary, C; et al. (2011): How Well Can We Identify the High-Performing Hospital? *Medical Care Research and Review* 68(3): 290-310. DOI: 10.1177/1077558710386115.
- Shwartz, M; Restuccia, JD; Rosen, AK (2015): Composite Measures of Health Care Provider Performance: A Description of Approaches. *The Milbank Quarterly* 93(4): 788-825. DOI: 10.1111/1468-0009.12165.
- Solberg, LI; Mosser, G; McDonald, S (1997): The Three Faces of Performance Measurement: Improvement, Accountability, and Research. *The Joint Commission Journal on Quality Improvement* 23(3): 135-147. DOI: 10.1016/S1070-3241(16)30305-4.
- Spiegelhalter, D; Sherlaw-Johnson, C; Bardsley, M; Blunt, I; Wood, C; Grigg, O (2012): Statistical methods for healthcare regulation: rating, screening and surveillance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 175(1): 1-47. DOI: 10.1111/j.1467-985X.2011.01010.x.
- Spiegelhalter, DJ (2005): Handling over-dispersion of performance indicators. *Quality and Safety in Health Care* 14(5): 347-351. DOI: 10.1136/qshc.2005.013755.

- Stroup, WW (2013): Generalized Linear Mixed Models. Modern Concepts, Methods and Applications. (Texts in Statistical Science). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC. ISBN: 978-1-4398-1512-0.
- Tartakovsky, A; Nikiforov, I; Basseville, M (2015): Sequential Analysis. Hypothesis Testing and Change-point Detection. (Monographs on Statistics and Applied Probability 136). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC. ISBN: 978-1-4398-3820-4.
- Tenenbein, A (1970): A Double Sampling Scheme for Estimating from Binomial Data with Misclassifications. *Journal of the American Statistical Association* 65(331): 1350-1361. DOI: 10.2307/2284301.
- Veit, C; Lüken, F; Bungard, S; Trümner, A; Tewes, C; Hertle, D (2013): Rahmenkonzept Evaluation bezogen auf Evaluationen nach § 137b SGB V. Version 1.1. Entwurf vom 17.07.2013. Düsseldorf: BQS [Institut für Qualität & Patientensicherheit]. [unveröffentlicht].
- W. K. Kellogg Foundation (2004): Logic Model Development Guide. Using Logic Models to Bring Together Planning, Evaluation, and Action. Updated January 2004. Battle Creek, US-MI: W. K. Kellogg Foundation. URL: <https://www.wkcf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide> [Download PDF - Angabe von Name und E-Mail erforderlich] (abgerufen am: 09.10.2019).
- Westen, D; Weinberger, J (2004): When Clinical Description Becomes Statistical Prediction. *American Psychologist* 59(7): 595-613. DOI: 10.1037/0003-066X.59.7.595.
- Winkler-Komp, G; Misselwitz, B; Kupfernagel, F; van Emmerich, C; Döbler, K (2014): Maßnahmen zur Qualitätssicherung in Krankenhäusern: Strukturiertes Dialog – Strukturen und Prozesse. Ergebnis einer Umfrage des Gemeinsamen Bundesausschusses bei den auf Landesebene beauftragten Stellen und der Institution nach § 137 a SGB V. *Das Krankenhaus* 2014(3): 198-205.
- Wood, SN (2006): Generalized Additive Models. An Introduction with R. (Texts in Statistical Science). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC. ISBN: 978-1-58488-474-3.



Institut für Qualitätssicherung und  
Transparenz im Gesundheitswesen

# Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern

Abschlussbericht zu Stufe 1 und Stufe 2  
Anhang

Erstellt im Auftrag des  
Gemeinsamen Bundesausschusses

Stand: 31. Januar 2020

---

# Impressum

**Thema:**

Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern. Abschlussbericht zu Stufe 1 und Stufe 2. Anhang

**Ansprechpartnerin:**

Andrea Wolf

**Auftraggeber:**

Gemeinsamer Bundesausschuss

**Datum des Auftrags:**

18. Januar 2018

**Datum der Abgabe:**

31. Januar 2020

**Herausgeber:**

IQTIG – Institut für Qualitätssicherung  
und Transparenz im Gesundheitswesen

Katharina-Heinroth-Ufer 1  
10787 Berlin

Telefon: (030) 58 58 26-0  
Telefax: (030) 58 58 26-999

[info@iqtig.org](mailto:info@iqtig.org)

<https://www.iqtig.org>

# Inhaltsverzeichnis

Tabellenverzeichnis.....	5
Abbildungsverzeichnis.....	6
1 Protokoll zum Workshop zur Weiterentwicklung des Strukturierten Dialogs am 27. Mai 2019.....	8
2 Auszug des Protokolls zum LQS-IQTIG-Treffen am 18. Juni 2019 .....	17
3 Ergänzende Informationen zu Kapitel 2 – Analyse der Ausgangssituation.....	19
4 Biometrische Analyse: Heterogenität der Vorgehensweise und Bewertung im Strukturierten Dialog – Exemplarische Analyse anhand von Daten aus dem Erfassungsjahr 2016 .....	22
4.1 Methodik.....	22
4.2 Ergebnisse .....	26
4.3 Zusammenfassung und Diskussion .....	32
4.4 Details zur statistischen Modellierung.....	33
4.4.1 Fixed- vs. Random-Effects-Modelle .....	33
4.4.2 Operationalisierung des Begriffes vergleichbarer Leistungserbringer .....	34
4.4.3 Modelle ohne Berücksichtigung der Leistungserbringer-Ergebnisse.....	37
4.4.4 Implementation.....	39
4.5 Deskriptive Auswertungen zu den Ergebnissen des Strukturierten Dialogs zum Erfassungsjahr 2016 für ausgewählte Indikatoren .....	40
4.5.1 HEP: QI 54003 Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur.....	41
4.5.2 09/1: QI 54139 Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen .....	44
4.5.3 DEK: QI 52010 Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4 .....	47
5 Beispielfälle für die Qualitätsbewertung im Stellungnahmeverfahren.....	50
5.1 Fachlich-inhaltliche Einflussfaktoren .....	50
5.2 Datenfehler .....	54
6 Textvorschläge Einholung Stellungnahmen .....	56
7 Beispiel zur Übermittlung von Vorgangsnummern in Stellungnahmen.....	60
8 Beispiel zur Berücksichtigung mehrerer Einflussfaktoren bei der partiellen Nachberechnung .....	62

9	Berechnungsvorschrift für die partielle Nachberechnung .....	64
9.1	Berechnung des informativen Indikatorwerts .....	64
9.2	Berechnung des bereinigten Referenzbereiches .....	65
9.3	Retrospektive Schätzung der in Referenzbereichen enthaltenen Toleranz für Qualitätsdefizite .....	66
10	Gesamtalgorithmus für die fachliche Bewertung .....	69
11	Zielvereinbarungen .....	70
11.1	Formular zum Sachstandsbericht einzelner Zielvereinbarungen .....	70
11.2	Erläuterung zum Formular .....	72
	Literatur.....	75

## Tabellenverzeichnis

Tabelle 1: Anzahl von Fachgruppen und Mitgliederzahl pro Fachgruppe in der externen stationären Qualitätssicherung (Winkler-Komp et al. 2014).....	19
Tabelle 2: Einstufungsschema für rechnerische Auffälligkeiten bei Qualitätsindikatoren nach Abschluss des Strukturierten Dialogs (bisheriges Vorgehen) .....	19
Tabelle 3: Anteil von Hinweisen an allen rechnerisch auffälligen Ergebnissen nach Bundesland zum Erfassungsjahr 2015 (modifizierte Tabelle aus dem Bericht zum Strukturierten Dialog 2016; IQTIG 2017: 52) .....	21
Tabelle 4: Entscheidungsverfahren in den Bundesländern, erfasst anhand der Frage: „Wie wird festgelegt, ob ein Ergebnis als qualitativ auffällig oder unauffällig beurteilt wird?“ (Winkler-Komp et al. 2014).....	21
Tabelle 5: Parameterschätzungen der drei Modelle zu Fragestellung 1: Heterogenität bezüglich des Anforderns einer Stellungnahme .....	26
Tabelle 6: Parameterschätzungen der drei Modelle zu Fragestellung 2: Heterogenität bezüglich der Bewertung als qualitativ auffällig.....	27
Tabelle 7: (Konditionaler) AIC der Modellvarianten $M_1$ - $M_4$ zu Frage 1 .....	38
Tabelle 8: (Konditionaler) AIC der Modellvarianten $M_1$ - $M_4$ zu Frage 2 .....	38
Tabelle 9: Parameterschätzungen Modellvariante $M_1$ der drei Modelle zu Fragestellung 1: Heterogenität bezüglich des Anforderns einer Stellungnahme .....	39
Tabelle 10: Parameterschätzungen Modellvariante $M_1$ der drei Modelle zu Fragestellung 2: Heterogenität bezüglich der Bewertung als qualitativ auffällig .....	39
Tabelle 11: Fiktives Beispiel für die Angabe von Vorgangsnummern zur Stellungnahme zu einem Indikator.....	60
Tabelle 12: Fiktives Beispiel für die fallbezogene Berücksichtigung von Einflussfaktoren bei der partiellen Neuberechnung.....	62

## Abbildungsverzeichnis

Abbildung 1: Ergebnisse des Strukturierten Dialoges zum Qualitätsindikator HEP: QI 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ für das Erfassungsjahr 2016 aus zwei exemplarisch ausgewählten Bewertungsstellen. ....	23
Abbildung 2: HEP QI 54003 (ratenbasierter Indikator). Streuung der geschätzten Bewertungswahrscheinlichkeit in Quartilen der Leistungserbringerergebnisse.....	30
Abbildung 3: DEK QI 52010 (Sentinel-Event-Indikator). Streuung der geschätzten Bewertungswahrscheinlichkeit in Quartilen der Leistungserbringerergebnisse.....	31
Abbildung 4: 09/1 QI 54139 (ratenbasierter Indikator). Streuung der geschätzten Bewertungswahrscheinlichkeit in Quartilen der Leistungserbringerergebnisse.....	31
Abbildung 5: Illustration zu alternativen Definitionen von <i>rlb</i> anhand einer nach Fallzahl differenzierten Darstellung der rechnerischen Ergebnisse. Segmente gleicher Farbe markieren Gruppen von Leistungserbringern deren Ergebnisse als vergleichbar eingestuft würden. Links die p-Wert-basierten Definition; rechts die ratenbasierte Definition. ....	35
Abbildung 6: Qualitätsindikator HEP: 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“. Anfordern einer Stellungnahme nach rechnerischer Auffälligkeit.....	41
Abbildung 7: Qualitätsindikator HEP: 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“. Bewertungen nach Einleitung des Stellungnahmeverfahrens.....	42
Abbildung 8: Ergebnisse des strukturierten Dialoges zum Qualitätsindikator HEP: 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ für das Erfassungsjahr 2016 differenziert nach Fallzahl und Bewertungsstelle. Diese Daten sind die Grundlage für die modellbasierten Auswertungen in Abschnitt 2.2.....	43
Abbildung 9: Qualitätsindikator 09/1: 54134 „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“. Anfordern einer Stellungnahme nach rechnerischer Auffälligkeit.....	44
Abbildung 10: Qualitätsindikator 09/1: 54134 „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“. Bewertungen nach Einleitung des Stellungnahmeverfahrens.....	45
Abbildung 11: Ergebnisse des Strukturierten Dialogs zum Qualitätsindikator 09/1: 54134 „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“ für das Erfassungsjahr 2016 differenziert nach Fallzahl und Bewertungsstelle. Diese Daten sind die Grundlage für die modellbasierten Auswertungen in Abschnitt 2.2. ....	46
Abbildung 12: Qualitätsindikator DEK: 52010 „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“. Anfordern einer Stellungnahme nach rechnerischer Auffälligkeit.....	47

Abbildung 13: Qualitätsindikator DEK: 52010 „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“. Bewertungen nach Einleitung des Stellungnahmeverfahrens .....	48
Abbildung 14: Ergebnisse des Strukturierten Dialogs zum Qualitätsindikator DEK: 52010 „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“ für das Erfassungsjahr 2016 differenziert nach Fallzahl und Bewertungsstelle. Diese Daten sind die Grundlage für die modellbasierten Auswertungen in Abschnitt 2.2.....	49

# 1 Protokoll zum Workshop zur Weiterentwicklung des Strukturierten Dialogs am 27. Mai 2019

TOP	Thema
<b>TOP 1</b>	<p><b>Begrüßung</b></p> <p>Das IQTIG begrüßt die Teilnehmerinnen und Teilnehmer des Workshops und stellt die Beauftragung durch den G-BA vor. Die Ziele des Workshops sind die Vorstellung des methodischen Hintergrunds und des Rahmenkonzepts sowie die Einholung der Expertise seitens der mit der Durchführung des Strukturierten Dialogs Betrauten, um anschließend mögliche Herausforderungen zu identifizieren und ggf. Lösungsansätze zu entwickeln.</p> <p>Zunächst werden die Erwartungen/Wünsche der Teilnehmerinnen und Teilnehmer an den Workshop eingeholt und besprochen. Die Teilnehmerinnen und Teilnehmer wünschen, dass ihr Input aufgenommen wird und in die Weiterentwicklung einfließt. Ebenso wird ein transparenterer Umgang hinsichtlich der Entscheidung darüber, welche Aspekte aufgenommen/nicht aufgenommen werden, gewünscht. Hinsichtlich des Konzepts wird sich Praxistauglichkeit und Entbürokratisierung gewünscht. Das IQTIG versichert, dass die Anmerkungen der Landesvertreterinnen und -vertreter eingehend geprüft werden und nach Möglichkeit für die Stufe 2 der „Weiterentwicklung des Strukturierten Dialogs“ berücksichtigt werden.</p>
<b>TOP 2</b>	<p><b>Präsentation methodischer Hintergrund und Rahmenkonzept</b></p> <p>Das IQTIG erläutert den methodischen Hintergrund und das Rahmenkonzept (siehe Präsentationsfolien 6–17). Vor dem Hintergrund des richtlinienunabhängigen Konzepts wird erläutert, dass die Begriffe Strukturiertes Dialog und Stellungnahmeverfahren synonym verwendet werden. Da die QSKH-RL bald ausläuft, sieht das IQTIG vor, den Schwerpunkt auf Empfehlungen für die DeQS-Richtlinie zu legen.</p> <p>Im methodischen Hintergrund werden die Funktionen des bisherigen Strukturierten Dialogs vorgestellt. Das IQTIG hat die Methodik des bisherigen Verfahrens analysiert. Hierzu wurde u. a. eine biometrische Analyse zur Heterogenität (bewertungsstellenabhängig bezüglich der Einholung der Stellungnahme und bezüglich der Bewertung als qualitativ auffällig) im Strukturierten Dialog durchgeführt.</p> <p>Ziele der Weiterentwicklung sind einerseits eine aussagekräftige Qualitätsbewertung für alle Indikatorergebnisse zu erlangen und andererseits durch einheitliche Verfahrensregeln das Stellungnahmeverfahren mit abschließender Qualitätsbewertung effizienter zu gestalten, um eine Ressourcenverschiebung von Qualitätsbewertung hin zur Qualitätsförderung zu erreichen.</p> <p>Im methodischen Hintergrund wurden Gütekriterien für die Bewertungsprozesse zur Minimierung der Heterogenität eingeführt. Vor diesem Hintergrund empfiehlt das IQTIG zukünftig eine Trennung von Qualitätsbewertung und -förderung.</p> <p>Einige Teilnehmerinnen und Teilnehmer weisen darauf hin, dass die Richtlinien (plan. QI-RL und DeQS-RL) untereinander abgestimmt werden müssen.</p> <p>Das IQTIG erläutert, dass diese Abstimmung laut Beauftragung vorgesehen ist.</p>

TOP	Thema
TOP 3	<p data-bbox="389 255 1374 322"><b>Analyse Herausforderungen und Diskussion Lösungen – „Vor der Bewertung der Stellungnahme“</b></p> <p data-bbox="389 338 1374 405"><b>I. Operationalisierung des „Hinreichenden Verdachts auf ein Qualitätsdefizit“</b></p> <p data-bbox="389 421 1385 629">Das IQTIG präsentiert Möglichkeiten zur Effizienzsteigerung hinsichtlich der Auslösung des Stellungnahmeverfahrens anhand der Qualitätsindikatoren (siehe Präsentationsfolien 19–26). Am Beispiel der HEP-Indikatoren wird der große Arbeitsaufwand auf Landesebene im Sinne der Anzahl an rechnerisch auffälligen Ergebnissen dargestellt, sowie die Diskrepanz zwischen der Anzahl an rechnerisch auffälligen Ergebnissen und der Anzahl an qualitativ auffälligen Ergebnissen verdeutlicht.</p> <p data-bbox="389 645 1385 786">Die statistische Operationalisierung des Kriteriums „Hinreichender Verdacht auf ein Qualitätsdefizit“ erfolgt nach einer Taxonomie bestehend aus Studienart, Stichprobenart und Berechnungsart und führt immer zu einem Indikatorwert für den Standort, ggf. einem Unsicherheitsintervall, und einer binären Klassifikation.</p> <p data-bbox="389 801 879 831">Herausforderungen und Lösungsansätze:</p> <ol data-bbox="389 846 1385 1946" style="list-style-type: none"> <li data-bbox="389 846 1385 1249">1. Bei einigen Teilnehmerinnen und Teilnehmern besteht die Sorge, dass der Begriff „hinreichender Verdacht auf ein Qualitätsdefizit“ bei den Leistungserbringern als Generalverdacht, welcher entkräftet werden müsse, interpretiert werden könnte. <ul data-bbox="421 1003 1385 1249" style="list-style-type: none"> <li data-bbox="421 1003 1385 1249">▫ Das IQTIG erläutert, dass es sich dabei eben nicht um einen Generalverdacht handelt, sondern im Gegenteil zunächst von guter Versorgungsqualität ausgegangen wird. Das Ziel ist, eine spezifischere Auslösung anhand der Qualitätsindikatoren zu erreichen, um eine Verschlinkung des Stellungnahmeverfahrens zu erzielen. Das bedeutet, dass nur bei denjenigen Leistungserbringern ein Stellungnahmeverfahren eingeleitet werden soll, bei denen ein Qualitätsdefizit anhand der Daten begründet zu vermuten ist.</li> </ul> <p data-bbox="421 1265 1385 1361">Einige Teilnehmerinnen und Teilnehmer weisen darauf hin, dass unter dem Aspekt der Akzeptanzförderung Begriffe, die scheinbar aus dem Strafrecht entlehnt sind, vermieden werden sollten.</p> <ul data-bbox="421 1377 1385 1444" style="list-style-type: none"> <li data-bbox="421 1377 1385 1444">▫ Das IQTIG wird den Begriff „Hinreichender Verdacht auf ein Qualitätsdefizit“ beraten.</li> </ul> </li> <li data-bbox="389 1467 1385 1646">2. Die aktuelle Anzahl an Qualitätsindikatoren wird von einigen Teilnehmerinnen und Teilnehmern als zu groß eingeschätzt. <ul data-bbox="421 1545 1385 1646" style="list-style-type: none"> <li data-bbox="421 1545 1385 1646">▫ Das IQTIG erläutert, dass im Rahmen der beauftragten methodischen Überarbeitung der Qualitätsindikatoren ggf. auch Indikatoren, die die Eignungskriterien nicht im ausreichenden Maß erfüllen, zur Streichung empfohlen werden.</li> </ul> </li> <li data-bbox="389 1668 1385 1946">3. Einige Teilnehmerinnen und Teilnehmer merken an, dass mit den bestehenden Qualitätsindikatoren Qualität nicht abgebildet werden könne. Zudem wird von einigen Teilnehmerinnen und Teilnehmern angemerkt, dass eine Vielzahl der aktuellen Qualitätsindikatoren nicht geeignet sei, um zwischen guter und schlechter Qualität zu unterscheiden und daher eine Anwendung des Konzepts anhand der aktuellen Indikatoren kritisch gesehen wird. <ul data-bbox="421 1881 1385 1946" style="list-style-type: none"> <li data-bbox="421 1881 1385 1946">▫ Das IQTIG erläutert, dass Qualitätsindikatoren gemäß der Methodischen Grundlagen des Instituts geeignet sein müssen, um Qualität abzubilden. Vor</li> </ul> </li> </ol>

TOP	Thema
	<p>dem Hintergrund, der unterschiedlichen Verwendungszwecke von Qualitätsmessungen in der gesetzlichen Qualitätssicherung ist es erforderlich, zu jedem Qualitätsindikatorergebnis eine abschließende Qualitätsaussage zu erhalten. Das IQTIG bestätigt, dass viele der aktuellen Qualitätsindikatoren gemäß den Kriterien der Methodischen Grundlagen zu überarbeiten sind und hierzu bereits ein Konzept erarbeitet wird. Es wird zudem erläutert, dass es notwendig ist, ein Konzept zur Weiterentwicklung des Strukturierten Dialogs auf der Basis geeigneter Qualitätsindikatoren (siehe Eignungskriterien von Qualitätsindikatoren, Methodische Grundlagen des IQTIG 2019) zu erstellen und nicht auf der Basis ungeeigneter Indikatoren.</p> <p>4. Die Risikoadjustierung erfolgt anhand der Ergebnisse der Vorjahre. Im Stimmnahmeverfahren wird nach patientenbezogenen Besonderheiten gefragt. Es wird die Frage gestellt, inwiefern das einbezogen wird.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG erläutert, dass bei der Indikatorergebnisberechnung nicht-berücksichtigte Risikofaktoren können im Stimmnahmeverfahren als Ausnahmetatbestände geltend gemacht werden.</li> </ul> <p>5. Eine Teilnehmerin fragt, ob die, in den Qualitätssicherungsergebnisberichten (DeQS) ausgewiesenen Dokumentationsfehler in der vorgestellten statistischen Auswertungsmethodik berücksichtigt werden. Zudem wird berichtet, dass es bei dem PCI-Verfahren bspw. einen hohen Anteil an Dokumentationsproblemen gegeben habe. Aus Sicht der Teilnehmerin wäre es hilfreich, darauf zurückzugreifen, um abzuleiten, bei welchen Indikatoren Nachbesserungsbedarf besteht.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG erläutert, dass die Problematik zu vieler Dokumentationsfehler nicht mittels statistischer Verfahren im Rahmen der QI-Berechnung gelöst werden könne.</li> </ul> <p>6. Hinsichtlich der gesundheitsökonomischen Betrachtungsweise und der Frage: „Ist ein entdecktes Qualitätsdefizit bei einem kleinen Leistungserbringer gleichermaßen ‚teuer‘ wie ein Qualitätsdefizit bei einem großen Leistungserbringer (da mehr Patientinnen und Patienten gefährdet werden)?“ (siehe Präsentationsfolie 24) wurde angemerkt, dass die Erfahrungen gezeigt haben, dass Häuser mit geringen Fallzahlen eher Unterstützung benötigten und eher in die Qualitätsförderung einbezogen werden sollten.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG wird diesen Hinweis beraten.</li> </ul> <p>7. Eine Teilnehmerin merkt an, dass 1/3 der Abweichungen vom Referenzwert auf 1 oder 2 Fällen beruhten. Der Effekt verstärkte sich bei perzentilbasierten Referenzbereichen. Dies sei ein Problem, dass sich mathematisch nicht lösen lasse. Es wird weiterhin berichtet, dass bei einem Fall im Zähler die Kommunikation mit den Leistungserbringern schwierig sei.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG weist darauf hin, dass dies bei der zu entwickelnden statistische Auswertungsmethodik berücksichtigt wird, speziell bei der Methodik zur Bestimmung von Perzentilreferenzbereichen. Bei den HEP-Indikatoren sind ca. 13 % der rechnerischen auffälligen Ergebnisse bei Standorten mit 1–4 Fällen festgestellt worden.</li> </ul> <p>Ein Teilnehmer berichtet, dass bspw. die Akzeptanz für Qualitätsindikatorergebnisse zur Cholezystektomie (Landesverfahren in Hessen), die mittels statistischer Auswertungsmethodik (Signifikanztest) berechnet werden größer sei.</p>

TOP	Thema
	<p data-bbox="389 248 995 277"><b>II. Formale Kriterien für die Stellungnahmen</b></p> <p data-bbox="389 293 1386 360">Das IQTIG präsentiert die Empfehlungen zu den formalen Anforderungskriterien an Stellungnahmen (siehe Präsentationsfolie 28).</p> <p data-bbox="389 376 879 405">Herausforderungen und Lösungsansätze:</p> <ol data-bbox="389 421 1386 1986" style="list-style-type: none"> <li data-bbox="389 421 1386 568">1. Ein Teilnehmer fordert, dass die Angaben der Leistungserbringer (insbesondere bei plan. QI), die in den Stellungnahmen gemacht wurden, grundsätzlich als korrekt erachtet werden sollten. <ul data-bbox="421 539 879 568" style="list-style-type: none"> <li data-bbox="421 539 879 568">▫ Das IQTIG wird den Hinweis beraten.</li> </ul> </li> <li data-bbox="389 584 1386 1570">2. Einige Teilnehmerinnen und Teilnehmer stellen die Frage, ob die Nicht-Einhaltung formaler Vorgaben (z. B. des Datenschutzes, digitale Stellungnahme) dazu führen dürfe, dass das Ergebnis des Leistungserbringers als qualitativ auffällig bewertet wird. Im Nachhinein könne nicht mehr differenziert werden, ob ein Leistungserbringer aufgrund nicht-ordnungsgemäßer Stellungnahmen oder aufgrund unzureichender Versorgungsqualität als qualitativ auffällig bewertet wurde. Es wird die Sorge geäußert, dass dies zu einer „Beschädigung“ des Verfahrens führen könnte. Zudem wird in Frage gestellt, inwieweit die vorgestellten formalen Kriterien (z. B. digitale Stellungnahme, Unterschrift durch juristisch verantwortliche Instanz) für die Akzeptanz des Gesamtverfahrens förderlich sind. Die Teilnehmerinnen und Teilnehmer haben hierzu unterschiedliche Ansichten. Eine Teilnehmerin macht den Vorschlag die Kategorie A, Ziffer 4 (qualitativ auffällig aufgrund nicht-Einhaltung der formalen Kriterien) aus der Kategorie A herauszulösen und gesondert (nicht unter Kategorie A) auszugeben. Wenn an dieser Bewertung festgehalten werde, wird vermutet, dass die Akzeptanz der QS-Verfahren darunter leiden würde und das aufgebaute Vertrauen der Leistungserbringer Schaden nehme. Zudem wird vermutet, dass damit auch keine Transparenzerhöhung zu erreichen sei. <ul data-bbox="421 1240 1386 1570" style="list-style-type: none"> <li data-bbox="421 1240 1386 1375">▫ Das IQTIG wird die einzelnen Empfehlungen zu den formalen Kriterien prüfen. Die Wahrung der Anonymität der Patientinnen und Patienten ist jedoch auch im Rahmen des Stellungnahmeverfahrens (siehe Datenschutzgesetz) sicherzustellen.</li> <li data-bbox="421 1391 1386 1570">▫ Das IQTIG erläutert, dass die Konsequenzen bei Nicht-Einhaltung der formalen Kriterien, wie sie auch in anderen Kontexten umgesetzt werden, dazu führen, dass die Kriterien eingehalten werden. Zudem differenziert das Bewertungsschema bereits, ob eine qualitative auffällige Bewertung aufgrund der Nicht-Einhaltung der formalen Kriterien vergeben wurde.</li> </ul> </li> <li data-bbox="389 1585 1386 1800">3. Die Teilnehmerinnen und Teilnehmer diskutieren, ob im Rahmen der Qualitätsförderung die juristisch verantwortliche Instanz die Stellungnahme autorisieren muss. <ul data-bbox="421 1697 1386 1800" style="list-style-type: none"> <li data-bbox="421 1697 1386 1800">▫ Das IQTIG wird diesen Hinweis beraten. Im Rahmen der Qualitätsförderung ist ggf. auch eine Autorisierung durch die fachlich verantwortliche Person ausreichend.</li> </ul> </li> <li data-bbox="389 1816 1386 1986">4. Seitens der Teilnehmerinnen und Teilnehmer wird die im Konzept empfohlene 4-Wochen-Frist für die Abgabe einer Stellungnahme kontrovers diskutiert. Einige Teilnehmerinnen und Teilnehmer geben zu bedenken, dass der Strukturierte Dialog oder das Stellungnahmeverfahren genau über die Haupturlaubszeit abzuwickeln ist, in der die Krankenhäuser mit einer deutlich reduzierten Personaldecke</li> </ol>

TOP	Thema
	<p>die Patientenversorgung aufrechterhalten müssen und deshalb solche administrativen Aufgaben nicht erste Priorität erhalten könnten.</p> <ul style="list-style-type: none"> <li>▫ Der Vorschlag seitens einiger Teilnehmerinnen und Teilnehmer einer 3-Wochen-Frist plus 2-Wochen-Nachrichtfrist und weitere Alternativen werden durch das IQTIG geprüft werden.</li> </ul> <p>5. Inwieweit die Anforderung, die Stellungnahme digital einzureichen, ein sinnvolles formales Kriterium sei, wird unter den Teilnehmerinnen und Teilnehmern kontrovers diskutiert. Eine solche Anforderung könne für einzelne Leistungserbringer ein Problem sein.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG wird diesen Hinweis beraten.</li> </ul> <p>6. Einige Teilnehmerinnen und Teilnehmer berichten, dass es Fachgruppen gebe, die einen strukturierten Fragebogen (teilweise zwischen Geschäftsstellen abgestimmt) entwickelt und Anfragetexte zusammengestellt haben, welche dann scheinbar nicht mehr genutzt werden können, wenn die Texte zur Einholung von Stellungnahmen vorgeschrieben seien. Dies könne für die Akzeptanz des Verfahrens abträglich seien.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG wird diesen Hinweis beraten.</li> </ul> <p>7. Einige Teilnehmerinnen und Teilnehmer äußern die Sorge, dass durch eine Formalisierung der Druck auf die Leistungserbringer steigen könne. Zudem wird die Sorge geäußert, dass durch die Empfehlungen des IQTIG das „Formelle“ stärker als bislang in die Bewertung mit einbezogen werde.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG wird den Hinweis beraten und erläutert, dass die formalen Kriterien zu einer Effizienzsteigerung bei der Qualitätsbewertung beitragen können, um mehr Ressourcen für die Qualitätsförderung zur Verfügung zu haben.</li> </ul> <p>Das IQTIG erläutert abschließend, dass auch formale Kriterien für eine Vereinheitlichung des Vorgehens beauftragt wurden. Um eine Vereinheitlichung zu erreichen, seien einheitliche Verfahrensregeln dazu, wie sich Leistungserbringer zu auffälligen Indikatorergebnissen äußern können, unabdingbar.</p>
<b>TOP 4</b>	<p><b>Analyse Herausforderungen und Diskussion Lösungsansätze – „Bewertung der Stellungnahme“ und „Einleitung Maßnahmen zur Qualitätsförderung“</b></p> <p><b>I. Fachliche Bewertung</b></p> <p>Das IQTIG präsentiert das Ablaufschema zum Bewertungsprozess, die Definition des Ausnahmetatbestands und das Bewertungsschema (vgl. Präsentationsfolien 33–36).</p> <p>Herausforderungen und Lösungsansätze:</p> <p>1. Hinsichtlich der Definition des Ausnahmetatbestandes wird seitens einer Teilnehmerin angemerkt, dass die Qualitätsanforderung, die zu erfüllen ist, die Erfüllung des Indikatorziels sei. Da die Indikatoren teilweise veraltet und ungeeignet seien, könne die Erfüllung eines Indikatorziels nicht mit guter Qualität gleichgesetzt werden.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG weist darauf hin, dass aus methodischer Sicht bei Qualitätsindikatoren, die die Eignungskriterien erfüllen, die Erfüllung des Indikatorziels guter</li> </ul>

TOP	Thema
	<p>Qualität entspräche. Das IQTIG erläutert, dass aktuell ein Konzept zur Weiterentwicklung der Qualitätsindikatoren anhand der Eignungskriterien erarbeitet wird.</p> <ol style="list-style-type: none"> <li>2. Einige Teilnehmerinnen und Teilnehmer sprechen sich dafür aus, dass auch Aspekte der Pflege bei der Definition von Ausnahmetatbeständen berücksichtigt werden sollten. <ul style="list-style-type: none"> <li>▫ Das IQTIG wird über die Definition des Ausnahmetatbestandes beraten und prüfen, was unter medizinisch-inhaltlich (z. B. auch pflegerisch) zu verstehen ist.</li> </ul> </li> <li>3. Es wird der Hinweise gegeben, dass der Begriff „organisatorisch-strukturelle Gründe“ sehr viel Interpretationsspielraum lässt. <ul style="list-style-type: none"> <li>▫ Das IQTIG wird den Hinweis beraten.</li> </ul> </li> <li>4. Einige Teilnehmerinnen und Teilnehmer kritisieren den Begriff „verschuldet“, da er an das Strafrecht erinnere. Es wird der Vorschlag gemacht, diesen durch „verantwortet“ zu ersetzen. <ul style="list-style-type: none"> <li>▫ Das IQTIG wird den Vorschlag und weitere Alternativen zum Begriff „Schuld“ prüfen.</li> </ul> </li> <li>5. Ein Teilnehmer merkt an, dass bei unverschuldeten Fehlern in der Datengrundlage (U1) nicht bewertet werden könne, ob ein Leistungserbringer eine gute Versorgungsqualität habe. Daher könne keine unauffällige Versorgungsqualität bescheinigt werden. <ul style="list-style-type: none"> <li>▫ Das IQTIG erläutert, dass in Abhängigkeit davon, wie gravierend die Fehler sind ggf. noch eine Bewertung vorgenommen werden kann (siehe gestrichelte Linie im Prozessablauf der Bewertung Folie 33). Da grundsätzlich zunächst von guter Versorgungsqualität ausgegangen werde, gelte die Unschuldsvermutung, falls unverschuldete Datenfehler belegt werden können.</li> </ul> </li> <li>6. Es wird seitens der Teilnehmerinnen und Teilnehmer gefragt, wo die Grenze sei, zwischen Dokumentationsfehlern, die der Leistungserbringer verschuldet und solchen, die er nicht verschuldet habe. <ul style="list-style-type: none"> <li>▫ Das IQTIG wird den Hinweis beraten und ggf. Entscheidungshilfen zur Verfügung stellen. Das IQTIG erläutert in diesem Zusammenhang, dass die Leistungserbringer verpflichtet seien, korrekt zu dokumentieren. Die derzeit bestehende Lücke, bei Dokumentationsfehlern keine Qualitätsbewertung vornehmen zu können, soll mit diesem Konzept geschlossen werden.</li> </ul> </li> <li>7. Einige Teilnehmerinnen und Teilnehmer fordern eine deutlichere Differenzierung des Bewertungsschemas in Bezug auf Fehldokumentationen der Leistungserbringer (Kategorie „D“). Patientinnen und Patienten sollten ein Recht darauf haben, zu wissen, ob eine Abteilung falsch dokumentiert hat oder ob ein medizinisch-fachlicher Qualitätsmangel bestand. Hier ist die Kategorie D „Bewertung nicht möglich wegen fehlerhafter Dokumentation“ eine sinnvolle Alternative. <ul style="list-style-type: none"> <li>▫ Das IQTIG wird den Hinweis beraten.</li> </ul> </li> <li>8. Leistungserbringer mit einem unauffälligen Qualitätsindikatorergebnis auch als qualitativ unauffällig zu bewerten wird von einigen Teilnehmerinnen und Teilnehmern kritisch gesehen. Nach dem Konzept wird denjenigen, die nicht im Stel-</li> </ol>

TOP	Thema
	<p>lungnahmeverfahren geprüft werden, weil sie ein unauffälliges Indikatorergebnis aufweisen, eine unauffällige Qualität zugestanden. Dies sei nicht vertretbar, da nur nach einem Stellungnahmeverfahren ein qualitativ unauffälliges Ergebnis bescheinigt werden könne. Daher solle die Kategorie R (Ergebnis liegt innerhalb des Referenzbereichs) beibehalten werden.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG erläutert, dass per Definition von unauffälliger Qualität gesprochen werde, wenn ein Leistungserbringer den Referenzbereich erreiche. Gibt die Datengrundlage keinen Anlass (aufgrund ausreichender Evidenz) eine weitere Prüfung vorzunehmen, weil das „Erwartbare“ erfüllt wurde, könne daher von unauffälliger Qualität bezogen auf das jeweilige Qualitätsziel des Indikators gesprochen werden.</li> </ul> <p>9. Es wird die Frage gestellt, wie „Mischergebnisse“ zu bewerten seien (z. B. wenn bei 5 Fällen Dokumentationsfehler angegeben und 5 Fälle als Ausnahmetatbeständen bewertet wurden).</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG wird den Hinweis beraten.</li> </ul> <p>10. Eine Teilnehmerin merkt an, dass die Kategorie A0 und A4 zusammengefasst werden könne. Die übrigen Kategorien seien sinnvoll gewählt.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG wird den Hinweis beraten.</li> </ul> <p><b>II. Zusammensetzung der Expertengremien</b></p> <p>Das IQTIG präsentiert die Empfehlungen zur Qualifikationen Expertinnen und Experten und zur Zusammensetzung des Expertengremiums (vgl. Präsentationsfolien 38–39).</p> <p>Herausforderungen und Lösungsansätze:</p> <p>1. Eine Teilnehmerin stellt die Frage, ob diese Empfehlungen nur für neue Expertinnen und Experten gelten sollen oder ob alle bisher mitarbeitenden Expertinnen und Experten nochmal überprüft werden müssten. Die Anforderungen seien zu hoch, insbesondere in kleinen Ländern gebe es nicht genug Fachexpertinnen und Fachexperten, die diese Anforderungen erfüllen und bereit seien mitzuwirken. Zudem wird die Frage stellt, wer die fachliche Qualifikation bewerten und über die Mitwirkung der Expertinnen und Experten entscheiden solle.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG wird diesen Hinweis beraten.</li> </ul> <p>2. Aus Sicht einiger Teilnehmerinnen und Teilnehmer werden Expertinnen und Experten benötigt, die bereits QS-Bögen ausgefüllt und Erfahrungen mit der esQS haben bzw. Verantwortung dafür tragen, die „Kodierer“ im Haus zu schulen etc. Es sollten, neben der Fachexpertise, umfangreiche QS-Erfahrungen vorliegen.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG wird diesen Hinweis beraten.</li> </ul> <p>3. Einige Teilnehmerinnen und Teilnehmer erläutern, dass der Begriff „Zuständige Stelle“ und LAG vor dem Hintergrund der unterschiedlichen Richtlinien zu konkretisieren sei. Gegebenenfalls müssten Abläufe/Zeitpläne für die LAG geändert werden. Sinnvoll wäre eine Orientierung an der DeQS-RL und es müsse geklärt werden, welches Organ innerhalb der LAG welche Kompetenzen übertragen bekommen solle.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG wird den Hinweis beraten.</li> </ul>

TOP	Thema
	<p>4. Einige Teilnehmerinnen und Teilnehmer fordern, dass die Bewertung zwingend von der Fachgruppe durchgeführt werden solle. Die LQS/LAG sollte nicht die abschließende Entscheidung treffen und diese verantworten müssen.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG wird den Hinweis beraten, weist jedoch darauf hin, dass formal die LQS/LAG die Verantwortung für den Prozess trägt.</li> </ul> <p><b>III. Kriterien für das Einleiten von Maßnahmen zur Qualitätsförderung</b></p> <p>Das IQTIG präsentiert die Empfehlungen hinsichtlich der Kriterien für die Einleitung von Maßnahmen zur Qualitätsförderung (vgl. Präsentationsfolien 42–43).</p> <p>Herausforderungen und Lösungsansätze:</p> <ol style="list-style-type: none"> <li>1. Einige Teilnehmerinnen und Teilnehmer empfehlen, dass es sich bei der zwingenden Einleitung von Fördermaßnahmen bei einem wiederholt auffälligen Ergebnis um denselben Qualitätsindikator oder einen thematisch ähnlichen Indikator handeln sollte. <ul style="list-style-type: none"> <li>▫ Das IQTIG wird diesen Hinweis beraten.</li> </ul> </li> <li>2. Ein Teilnehmer weist darauf hin, dass die Chance, dass ein Leistungserbringer, der z. B. im EJ 2018 auffällig war und diese Information Mitte des Folgejahres (EJ 2019) erhält, bereits im EJ 2019 ein unauffälliges Ergebnis erzielen zu können, gering sei. <ul style="list-style-type: none"> <li>▫ Das IQTIG nimmt den Hinweis auf und erläutert, dass in der DeQS-RL quartalsweise Auswertungen vorgesehen sind, sodass das interne QM frühzeitig intervenieren könne. Es gibt weiterhin zu bedenken, dass es sich bei den qualitätsfördernden Maßnahmen auch um sehr niedrigschwellige Maßnahmen handeln könne.</li> </ul> </li> <li>3. Einige Teilnehmerinnen und Teilnehmer erachten einen 5-Jahres-Zeitraum bis zwingend qualitätsfördernde Maßnahmen eingeleitet werden müssen als zu lang. <ul style="list-style-type: none"> <li>▫ Der Vorschlag „innerhalb von 3 Jahren bei thematisch ähnlichen Indikatoren“ wird durch das IQTIG geprüft.</li> </ul> </li> <li>4. Einige Teilnehmerinnen und Teilnehmer wünschen Kriterien, die vorgeben, wann welche Fördermaßnahmen einzuleiten sind. <ul style="list-style-type: none"> <li>▫ Das IQTIG wird den Hinweis beraten.</li> </ul> </li> </ol>
TOP 5	<p><b>Analyse Herausforderungen und Diskussion Lösungsansätze – „Verkürzung Gesamtverfahrens (I.) und Strukturierter Dialog und Auffälligkeitskriterien (AK) (II.)“</b></p> <p><b>I. Verkürzung des Gesamtverfahrens</b></p> <p>Das IQTIG stellt die Frage, welche weiteren Maßnahmen zu einer „Verschlankung“ des Verfahrens im Sinne einer Verkürzung führen würden.</p> <p>Die Teilnehmerinnen und Teilnehmer äußern, dass die Auswertungen früher kommen müssten. Außerdem sei der in der DeQS-RL vorgesehene Prozess bis zur Anforderung einer Stellungnahme zu aufwendig.</p> <p><b>II. Der Strukturierte Dialog und Auffälligkeitskriterien</b></p> <ol style="list-style-type: none"> <li>1. Einige Teilnehmerinnen und Teilnehmer erklären, dass die Bewertung der Stellungnahmen zu den AK und Qualitätsindikatoren bisher in einem Prozess/Vorgang abgehandelt wurden und dies auch zukünftig beibehalten werden solle.</li> </ol>

TOP	Thema
	<p>Das Verfahren bei den AK solle sich nicht von dem für Qualitätsindikatoren unterscheiden.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG erklärt, dass das Konzept für den Strukturierten Dialog zu den AK in der Entwicklung ist. Die bisherige Bewertung der AK wurden in korrekte Dokumentation, fehlerhafte Dokumentation und „Sonstiges“ eingeteilt.</li> </ul> <p>2. Das IQIG bittet die Teilnehmerinnen und Teilnehmer, ihre Vorschläge zum Verfahren und zur Klassifikation/Bewertung der AK im Nachgang zum Workshop dem IQTIG schriftlich mitzuteilen.</p> <p>3. Bezogen auf die Frage des IQTIG, inwiefern qualitätsfördernde Maßnahmen bezüglich der AK eingeleitet wurden, nennen einzelne Teilnehmerinnen und Teilnehmer Maßnahmen zur Förderung der Dokumentationsqualität. Aus Sicht der Teilnehmerinnen und Teilnehmer wäre eine mögliche „dokumentationsqualitätsfördernde Maßnahme“ zu überlegen, inwiefern die Akten anzufordern seien, wenn es wiederholt nicht korrekte Dokumentationen gab. Auch der Aspekt zur Qualitätsförderung bzgl. der Dokumentationsqualität soll im Konzeptvorschlag bedacht werden.</p> <ul style="list-style-type: none"> <li>▫ Das IQTIG wird den Hinweis beraten.</li> </ul>
<b>TOP 6</b>	<p><b>Weiteres Vorgehen</b></p> <p>Das IQTIG teilt mit, dass das Protokoll den Workshop-Teilnehmerinnen und -Teilnehmern zeitnah zugesandt wird. Das Konzept wird am 28. Mai im G-BA und beim LQS-Treffen am 18. Juni 2019 vorgestellt.</p> <p>Das IQTIG teilt mit, dass im Nachgang eine E-Mail zur Erinnerungen zu den Detailfragen an die Teilnehmerinnen und Teilnehmer versendet wird.</p>
<b>TOP 7</b>	<p><b>Sonstiges</b></p> <p>Ein Teilnehmer äußert, dass das Konzept bislang keine Empfehlungen zum Umgang mit Best-Practice-Leistungserbringern beinhaltet. Das IQTIG erläutert, dass das Teil von Stufe 2 der Beauftragung sei.</p>

## 2 Auszug des Protokolls zum LQS-IQTIG-Treffen am 18. Juni 2019

TOP	Thema
TOP 3	Weiterentwicklung Strukturiertes Dialog
	<p><b>Präsentation der Ergebnisse des Workshops</b></p> <p>Das IQTIG präsentiert das Rahmenkonzept zur Weiterentwicklung des Strukturierten Dialogs (analog zur Präsentation des entsprechenden Workshops vom 27. Mai 2019) sowie die Workshop-Ergebnisse (Präsentationsfolien 5–30).</p> <p>Das IQTIG erläutert, dass das Protokoll des Workshops dazu dient, die Workshop-Ergebnisse zu sichern. Die seitens der Workshop-Teilnehmerinnen und -Teilnehmer eingegangenen Anmerkungen zum Protokoll werden geprüft und das Protokoll ggf. überarbeitet. Darüber hinaus wird das Workshop-Protokoll voraussichtlich Bestandteil des Abschlussberichts sein. Das IQTIG erläutert, dass das Protokoll und die eingehenden schriftlichen Hinweise der LAG- und LQS-Vertreterinnen und -Vertreter wichtige Dokumente für die weitere Entwicklung des Konzepts sind und das IQTIG nach eingehender Prüfung die Anmerkungen ggf. entsprechend berücksichtigen wird. Des Weiteren wird erläutert, dass es ein Stellungnahmeverfahren zum Abschlussbericht geben wird. Das IQTIG wird das finale Protokoll gemeinsam mit den Workshop-Folien zum Rahmenkonzept an alle LAG- und LQS-Vertreterinnen und -Vertreter senden.</p> <p>Einige Teilnehmerinnen und Teilnehmer merken an, dass ein rechnerisch unauffälliges Ergebnis nicht automatisch als qualitativ unauffälliges Ergebnis bewertet und dass die Nicht-Einhaltung formaler Kriterien nicht mit unzureichender medizinischer Versorgungsqualität gleichgesetzt werden dürfe. Auch im Hinblick auf das Qualitätsportal wird angemerkt, dass Patientinnen und Patienten anhand der Bewertung erkennen sollten, ob das qualitativ auffällige Ergebnis aufgrund formaler Fehler oder aufgrund unzureichender medizinischer Qualität zustande gekommen ist.</p> <p>Die Teilnehmerinnen und Teilnehmer äußern sich kontrovers über die Abgabefrist für Stellungnahmen.</p> <p>Es wird der Hinweis gegeben, dass sich das Rahmenkonzept an der DeQS-RL orientieren sollte und das Konzept die zeitlichen Vorgaben der Richtlinie berücksichtigen müsse. Das IQTIG bestätigt, dass sich die Empfehlungen im Abschlussbericht auf die DeQS-RL beziehen werden.</p> <p>Ein Teilnehmer macht folgende Anmerkungen:</p> <ol style="list-style-type: none"> <li>1. Ein Konzept zu „Lernen von den Besten“ ist bislang nicht im Konzept enthalten.</li> <li>2. Die Möglichkeit Hinweise versenden zu können wird vermisst.</li> <li>3. Auch die Sozialdaten sollten im Konzept berücksichtigt werden.</li> <li>4. Die QIDB sollte möglichst frühestmöglich geliefert werden, um den Strukturierten Dialog rechtzeitig zu starten. Ideal wäre z. B. Anfang Februar.</li> </ol>

---

5. Für die Anforderung der Stellungnahmen sollten konkrete und einheitliche Fragen formuliert werden.

Das IQTIG nimmt die Hinweise auf und wird diese prüfen.

Seitens einiger Teilnehmerinnen und Teilnehmer wird angemerkt, dass eine einheitliche Aufwandentschädigung aller Mitglieder der Expertengremien geleistet werden sollte. Des Weiteren wird die Begrifflichkeit „hinreichender Verdacht“ hinterfragt und gebeten diese zu prüfen.

Das Konzept müsse zudem berücksichtigen, dass Fördermaßnahmen, die eingeleitet wurden sich oftmals nicht direkt im Folgejahr in einem unauffälligem QI-Ergebnis niederschlagen.

Das IQTIG führt aus, dass die Operationalisierung des hinreichenden Verdachts auf ein Qualitätsdefizit im Abschlussbericht vorgenommen wird.

Einige Teilnehmerinnen und Teilnehmer merken an, dass ein Vorgehen im Umgang mit „Hybridstellungen“ (ATB, Dokumentationsfehlern, nicht akzeptierte Gründe) gefunden werden müsse.

Das IQTIG nimmt die Hinweise auf und wird diese prüfen.

Das IQTIG bittet alle LAG- und LQS-Vertreterinnen und -Vertreter ihre Vorschläge zum Rahmenkonzept und zu Optionen zur Verkürzung des Verfahrens bis zum **26. Juni 2019** einzureichen. Das IQTIG wird hierzu eine Erinnerungs-E-Mail versenden.

---

### 3 Ergänzende Informationen zu Kapitel 2 – Analyse der Ausgangssituation

Tabelle 1: Anzahl von Fachgruppen und Mitgliederzahl pro Fachgruppe in der externen stationären Qualitätssicherung (Winkler-Komp et al. 2014)

Bundesland	Anzahl der Fachgruppen	Mitgliederzahl / Fachgruppe	
		Durchschnitt	Spannweite
Baden-Württemberg	9	6,7	8–10
Bayern	5	9,4	6–8
Berlin	7	8	8
Brandenburg	8	4	3–6
Bremen	7	9,9	6–13
Hamburg	7	11	8–15
Hessen	7	7,1	4–13
Mecklenburg-Vorpommern	7	7,1	6–11
Niedersachsen	6	5,8	4–8
Nordrhein-Westfalen	7	9,4	5–18
Saarland	7	7,7	7–8
Sachsen	7	7,3	5–10
Sachsen-Anhalt	7	5,1	3–6
Schleswig-Holstein	9	5,8	4–8
Thüringen	8	7,5	6–8
Gesamt	114	7	3–18

Tabelle 2: Einstufungsschema für rechnerische Auffälligkeiten bei Qualitätsindikatoren nach Abschluss des Strukturierten Dialogs (bisheriges Vorgehen)

Kategorie	Einstufung	Ziffer	Begründung
N	Bewertung nicht vorgesehen	01	Qualitätsindikator ohne Ergebnis, da entsprechende Fälle nicht aufgetreten sind
		02	Referenzbereich ist für diesen Indikator nicht definiert
		99	Sonstiges (im Kommentar erläutert)

Kategorie	Einstufung	Ziffer	Begründung
R	Ergebnis liegt im Referenzbereich	10	Ergebnis rechnerisch unauffällig, daher kein Strukturierter Dialog erforderlich
H	Einrichtung auf rechnerisch auffälliges Ergebnis hingewiesen	20	Aufforderung an das einrichtungsinterne Qualitätsmanagement zur Analyse der rechnerischen Auffälligkeit
		99	Sonstiges (im Kommentar erläutert)
U	Bewertung nach Strukturiertem Dialog als qualitativ unauffällig	30	Korrekte Dokumentation wird bestätigt (Datenvalidierung)
		31	Besondere klinische Situation
		32	Das abweichende Ergebnis erklärt sich durch Einzelfälle
		33	Kein Hinweis auf Mängel der med. Qualität (vereinzelte Dokumentationsprobleme)
		99	Sonstiges (im Kommentar erläutert)
A	Bewertung nach Strukturiertem Dialog als qualitativ auffällig	40	Fehlerhafte Dokumentation wird bestätigt (Datenvalidierung)
		41	Hinweise auf Struktur- und Prozessmängel
		42	Keine (ausreichend erklärenden) Gründe für die rechnerische Auffälligkeit benannt
		99	Sonstiges (im Kommentar erläutert)
D	Bewertung nicht möglich wegen fehlerhafter Dokumentation	50	Unvollständige oder falsche Dokumentation
		51	Softwareprobleme haben eine falsche Dokumentation verursacht
		99	Sonstiges (im Kommentar erläutert)
S	Sonstiges	90	Verzicht auf Maßnahmen im Strukturierten Dialog
		91	Strukturierter Dialog noch nicht abgeschlossen
		99	Sonstiges (im Kommentar erläutert)

Tabelle 3: Anteil von Hinweisen an allen rechnerisch auffälligen Ergebnissen nach Bundesland zum Erfassungsjahr 2015 (modifizierte Tabelle aus dem Bericht zum Strukturierten Dialog 2016; IQTIG 2017: 52)

Bundesland	rechnerische Auffälligkeit  n	Hinweis  H20 + H99	
		n	%
Bayern	2759	1684	61,0
Baden-Württemberg	1556	901	57,9
Berlin	551	181	32,8
Brandenburg	525	71	13,5
Bremen	112	6	5,4
Hamburg	257	92	35,8
Hessen	977	461	47,2
Mecklenburg-Vorpommern	373	79	21,2
Niedersachsen	1410	165	11,7
Nordrhein-Westfalen	3357	1165	34,7
Rheinland-Pfalz	1152	491	42,6
Saarland	208	2	1,0
Sachsen	800	110	13,8
Sachsen-Anhalt	612	249	40,7
Schleswig-Holstein	439	105	23,9
Thüringen	397	177	44,6
IQTIG	371	0	0,0
<b>Gesamt</b>	<b>15856</b>	<b>5939</b>	<b>37,5</b>

Tabelle 4: Entscheidungsverfahren in den Bundesländern, erfasst anhand der Frage: „Wie wird festgelegt, ob ein Ergebnis als qualitativ auffällig oder unauffällig beurteilt wird?“ (Winkler-Komp et al. 2014)

Art der Entscheidung	Häufigkeit
im Konsens	9
teilweise durch Mehrheitsbeschluss	4
Konsens und teilweise durch Mehrheitsbeschluss	2
Konsens in LQS oder in Fachgruppe auf Landesebene	2

## 4 Biometrische Analyse: Heterogenität der Vorgehensweise und Bewertung im Strukturierten Dialog – Exemplarische Analyse anhand von Daten aus dem Erfassungsjahr 2016

Im Folgenden wird eine Methodik vorgestellt, welche ermöglicht die Heterogenität in den Bewertungen des Strukturierten Dialoges quantitativ zu analysieren. Auf Basis dieser Methodik wurden im Abschlussbericht in Abschnitt 2.5 Analyseergebnisse formuliert, die an dieser Stelle ausführlich hergeleitet und beschrieben werden. Angewandt wird die Methodik auf folgende Qualitätsindikatoren: „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ (QI-ID 54003), „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“ (QI-ID 54139) sowie „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“ (QI-ID 52010), vgl. Tabelle 2 im Abschlussbericht, Abschnitt 2.5.

Für die Analyse der Unterschiede zwischen den Bewertungsweisen der Bundesländer wurden folgende zwei Fragestellungen formuliert:

1. **Fragestellung 1:** Wie groß ist die landesstellenabhängige Heterogenität bezüglich der Aufnahme eines Stellungsnahmeverfahrens bei rechnerischer Auffälligkeit eines Leistungserbringers?
2. **Fragestellung 2:** Wie groß ist die landesstellenabhängige Heterogenität bezüglich der Bewertung als qualitativ auffällig nach Aufnahme eines Stellungsnahmeverfahrens?

Diese Fragestellungen sind jeweils unter Berücksichtigung der rechnerischen Ergebnisse der Leistungserbringer in ihrem Bundesland zu beantworten, um auszuschließen, dass sich die beobachteten Unterschiede durch Unterschiede zwischen den Leistungserbringern erklären. Es geht also darum zu charakterisieren, wie die jeweiligen Bewertungsstellen *vergleichbare rechnerische Ergebnisse* bewerten.<sup>1</sup>

### 4.1 Methodik

Zur Beantwortung von Fragestellung 1 und 2 ist zu untersuchen, wie verschiedene Bewertungsstellen vergleichbare rechnerische Ergebnisse von Leistungserbringern bewerten. Dazu ist eine Definition zu treffen, welche rechnerischen Ergebnisse als vergleichbar einzustufen sind. Die QSKH-Richtlinie sieht zwar keine Differenzierung der rechnerischen Ergebnisse etwa nach Größe der Leistungserbringer oder Ausmaß der Überschreitung des Referenzwertes vor.<sup>2</sup> Für eine differenzierte Analyse der Bewertungsergebnisse des Strukturierten Dialogs ist es allerdings unabdingbar die rechnerischen Ergebnisse genauer mit einzubeziehen. Ein geeigneter Ausgangspunkt

---

<sup>1</sup> Dabei dienen die rechnerischen Ergebnisse der Leistungserbringer für die Analyse der qualitativ auffälligen Bewertungen (Fragestellung 2) lediglich als Approximation, um tatsächliche Qualitätsunterschiede zwischen den Leistungserbringern zu modellieren (vgl. Abschnitt 2.3).

<sup>2</sup> Mit Ausnahme der sog. Ein-Fall-Regel

dafür ist eine nach Bundesland und Größe (Fallzahl) der Leistungserbringer differenzierte Auswertung der Ergebnisse des strukturierten Dialoges. Derartige Auswertungen werden exemplarisch in Abbildung 1 für einen Qualitätsindikator und zwei ausgewählte Bewertungsstellen dargestellt. In Abschnitt 4.5 finden sich vollständige Auswertungen für alle drei ausgewählten Qualitätsindikatoren. Dort finden sich auch weitere Darstellungen der Ergebnisse des strukturierten Dialoges zu diesen Indikatoren.

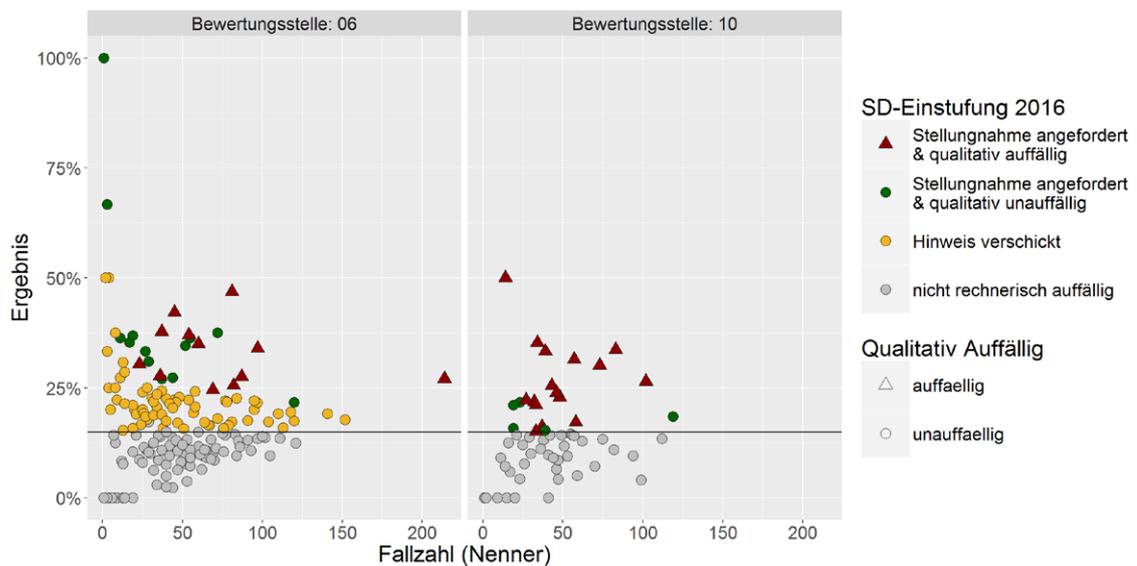


Abbildung 1: Ergebnisse des Strukturierten Dialoges zum Qualitätsindikator HEP: QI 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ für das Erfassungsjahr 2016 aus zwei exemplarisch ausgewählten Bewertungsstellen.

Es zeigt sich, dass in der Praxis mehrerer Bewertungsstellen inferenzstatistische Methoden (z. B. Konfidenzintervalle) zum Einsatz kommen, um die Entscheidung für die Aufnahme eines Stellungnahmeverfahrens zu treffen. Die Handhabung des Strukturierten Dialoges fällt somit i. d. R. bereits innerhalb einer Bewertungsstelle unterschiedlich aus, je nach Größe des Leistungserbringers und Ausmaß der Referenzbereichsüberschreitung. Die vorliegende Analyse untersucht, wie unterschiedlich die Bewertungen bei *ähnlichen* Fallzahlen und *ähnlichen* rechnerischen Ergebnissen der Leistungserbringer ausfallen.

Ein erster Ansatz die Heterogenität der Bewertungen des Strukturierten Dialogs zu quantifizieren könnte sein, Gruppen von Leistungserbringern mit vergleichbaren Ergebnissen zu bilden und diese anschließend nach Bundesland zu stratifizieren. Da es allerdings in einem einzelnen Qualitätsindikator selbst bei Indikatoren mit besonderem Handlungsbedarf aus den größten QS-Verfahren bundesweit nur wenige hundert rechnerisch auffällige Standorte gibt, kann man auf Ebene der Bundesländer keine statistisch aussagekräftigen Vergleiche durchführen, wenn man zusätzlich zu der Stratifizierung nach Bundesland weitere Untergruppen von Leistungserbringern mit ähnlichem Ergebnis betrachtet.

## Analysemodell

Ein alternativer Ansatz an Stelle einer Stratifizierung der Leistungserbringerergebnisse ist ein modellbasierter Ansatz. Um die Heterogenität im Strukturierten Dialog unter Berücksichtigung von Unterschieden zwischen den Leistungserbringern in ihren rechnerischen Ergebnissen und Unterschieden in ihren Fallzahlen zu analysieren, modellieren für einen Qualitätsindikator die Wahrscheinlichkeit, dass

1. von einem Leistungserbringer eine Stellungnahme angefordert wird (Fragestellung 1),
2. ein Leistungserbringer nach Anforderung einer Stellungnahme als qualitativ auffällig bewertet wird (Fragestellung 2),

in Abhängigkeit des rechnerischen Leistungserbringerergebnisses und der Bewertungsstelle. Sei  $Y_{lb}$  das interessierende Bewertungsergebnis im Strukturierten Dialog eines Leistungserbringers  $l$  durch Bewertungsstelle  $b$ . Je nach Fragestellung 1 oder 2 beschreibt  $Y_{lb}$  das Anfordern einer Stellungnahme (ja = 1, nein = 0) oder die Bewertung als qualitativ auffällig (ja = 1, nein = 0). Wir formulieren folgendes logistisches Regressionsmodell mit bewertungsstellenspezifischen Effekten:

$$\log \left[ \frac{P(Y_{lb} = 1 | r_{lb}, u_b)}{1 - P(Y_{lb} = 1 | r_{lb}, u_b)} \right] = \beta_0 + \beta_{Q_2} \cdot Q_2(r_{lb}) + \beta_{Q_3} \cdot Q_3(r_{lb}) + \beta_{Q_4} \cdot Q_4(r_{lb}) + u_b. \quad (1)$$

Hierbei sind  $\beta_0, \beta_{Q_2}, \beta_{Q_3}, \beta_{Q_4}$  und  $u_b$  zu bestimmende Modellparameter. Die globale Konstante  $\beta_0$  quantifiziert die Grundhäufigkeit von Stellungnahmen (Fragestellung 1) bzw. qualitativen Auffälligkeiten (Fragestellung 2) für den betrachteten Qualitätsindikator. Mit  $r_{lb}$  quantifizieren wir die statistische Evidenz, dass ein Leistungserbringer außerhalb des Referenzbereiches liegt. Je nach Bewertungsart des Indikators fließen dabei die rechnerischen Ergebnisse des Leistungserbringers unterschiedlich in  $r_{lb}$  ein. Für jeden Leistungserbringer ist dabei  $r_{lb}$  eine bekannte Zahl, welche von den beobachteten Ergebnissen des Leistungserbringers (Zähler und Nenner) abhängt. In der Art und Weise wie  $r_{lb}$  definiert wird, wird festgelegt welche Leistungserbringerergebnisse als vergleichbar eingestuft werden (siehe Anhang 4.4.2 für eine ausführliche Diskussion). Mit  $Q_2(r_{lb}), Q_3(r_{lb})$  und  $Q_4(r_{lb})$  bezeichnen wir das zweite bis vierte Quartil von  $r_{lb}$ . Die Koeffizienten  $\beta_{Q_2}, \beta_{Q_3}$  und  $\beta_{Q_4}$  quantifizieren somit, wie stark die Wahrscheinlichkeit  $P(Y_{lb} = 1 | r_{lb}, u_b)$  ansteigt, wenn ein Leistungserbringer im zweiten, dritten oder vierten Quartil der rechnerischen Ergebnisse  $r_{lb}$  liegt im Vergleich mit dem ersten Quartil.

### Wahl von $r_{lb}$

Für die beiden ratenbasierten Indikatoren wählen wir zur Definition von  $r_{lb}$  p-Werte eines einseitigen exakten Binomialtests [vgl. z. B. Spiegelhalter et al. (2012)], welcher sowohl das Ratenergebnis als auch die Fallzahl des Leistungserbringers berücksichtigt. Diese Vorgehensweise wird gewählt, da die Streuung der rechnerischen Leistungserbringerergebnisse von der Fallzahl abhängt und somit Fallzahl und Ratenergebnis keine unabhängigen Kovariablen darstellen.<sup>3</sup> Beispielsweise ist davon auszugehen, dass bei einem Referenzbereich  $\leq 15\%$  die Ergebnisse 2 von

<sup>3</sup> In einem ratenbasierten Qualitätsindikator streuen die Ergebnisse der Leistungserbringer mit kleiner Fallzahl i. d. R. breiter. Beispielsweise kann ein Leistungserbringer mit nur fünf behandelten Fällen nur die Ergebnisse

4 und 50 von 100 jeweils zu einer anderen Bewertung durch die Fachgruppen führen, obwohl beide einem Ergebnis von 50 % entsprechen. In der Regressionsmodellierung ist daher zwingend eine Interaktion zwischen Fallzahl und Ergebnis zu berücksichtigen. Der p-Wert ist eine einfache Möglichkeit diese Interaktion abzubilden, welche zudem auch zu einer klaren Interpretation der bewertungsstellenspezifischen Effekte  $u_b$  führt: Diese Stellen die Unterschiede in der Bewertung bei gleicher statistischer Evidenz der Leistungserbringerergebnisse dar. Die Parameter  $\beta_{Q_2}$ ,  $\beta_{Q_3}$  und  $\beta_{Q_4}$  beschreiben, wie sehr die stat. Evidenz  $r_{lb}$  in das Bewertungsergebnis  $Y_{lb}$  einfließen. Für den Sentinel-Event-Indikator ist  $r_{lb}$  definiert als die Anzahl beobachteter Sentinel-Events pro Leistungserbringer, ohne Berücksichtigung der Fallzahl, vgl. Abschnitt 4.4.2 für weitere Details.

### Random-Effects-Modell

Im oben beschriebenen Ansatz ist die statistische Schätzung der Modellparameter bei geringer Anzahl von Leistungserbringern eine Herausforderung, weshalb für die Modellierung des Einflusses der Bundesländer ein sogenanntes Random-Effects-Modell [vgl. z. B. Agresti (2013), Stroup (2013)] benutzt wird, welches die Schätzung der Effekte  $u_b$  vor allem bei kleiner Anzahl von Leistungserbringern pro Bundesland stabilisiert (vgl. Ash et al. (2012) für eine Diskussion zu Random-Effects-Modellen im Kontext der Qualitätsbewertung von Krankenhäusern). Dabei wird die Annahme getroffen, dass die bewertungsstellenspezifischen Effekte  $u_b$  einer übergeordneten Normalverteilung mit Varianz  $\tau^2$  unterliegen. Die Standardabweichung  $\tau$  ist dabei ein weiterer zu schätzender Modellparameter, welcher explizit die Heterogenität der Bewertungen zwischen den Bundesländern quantifiziert. Der Parameter  $\tau$  ist dabei ein geeignetes Maß für die Beantwortung von Fragestellung 1 und 2, da es quantifiziert, wie groß die Unterschiede in der Wahrscheinlichkeit  $P(Y_{lb} = 1 | r_{lb}, u_b)$  zwischen den Bundesländern sind, die nicht durch die rechnerischen Ergebnisse der Leistungserbringer erklärt werden können. Je größer der geschätzte Wert von  $\tau$  ausfällt, desto größer ist die auf die Bewertungsstellen zurückzuführenden Heterogenität. Bei der Interpretation der Standardabweichung  $\tau$  ist allerdings zu beachten, dass damit die Standardabweichung der  $u_b$  auf Ebene des linearen Prädiktors (Logit-Ebene) modelliert wird. Die Streuung der Wahrscheinlichkeiten  $P(Y_{lb} = 1 | r_{lb}, u_b)$  je nach Bewertungsstelle hängt neben dem Wert von  $\tau$  auch von den rechnerischen Leistungserbringerergebnissen  $r_{lb}$  ab. Beispielsweise ist zu erwarten, dass für extreme Ergebnisse  $r_{lb}$  eine geringere Streuung in den Wahrscheinlichkeiten  $P(Y_{lb} = 1 | r_{lb}, u_b)$  vorherrscht als für Ergebnisse, die den Referenzbereich nur knapp überschreiten.

Eine mögliche Interpretation der Heterogenitätsparameter  $\tau$  bieten sog. *Median Odds Ratios* [vgl. Larsen et al. (2000)]. Würde man zwei Leistungserbringer  $A$  und  $B$  aus dem gleichen Quartil der rechnerischen Ergebnisse von zwei zufällig ausgewählten Bewertungsstellen bewerten lassen, so wäre der Odds-Ratio zwischen Leistungserbringer  $A$  mit dem höheren Risiko und Leistungserbringer  $B$  im Median:

$$\text{MOR}(\tau) = \exp(\sqrt{2 \cdot \tau^2} \cdot \Phi^{-1}(3/4)),$$

---

0 %, 20 %, 40 %, 60 %, 80 %, 100 % erreichen. Ein einzelner Behandlungsfall hat somit einen größeren Einfluss auf das Leistungserbringerergebnis.

dabei ist  $\Phi^{-1}$  die inverse Verteilungsfunktion der Normalverteilung.

### Datengrundlage

Datengrundlage der Analyse zu Fragestellung 1 sind alle im Erfassungsjahr 2016 im jeweils betrachteten Qualitätsindikator rechnerisch auffällig gewordenen Standorte. Für die beiden ratenbasierten Qualitätsindikatoren (QI 54003 und 54139) wurden zusätzlich alle Standorte ausgeschlossen, welche nur auf Grund eines einzigen Zählerfalles rechnerisch auffällig wurden. In diesem Fall wird im strukturierten Dialog nach der sog. Ein-Fall-Regel fast immer ein Hinweis verschickt, und somit unterliegt das Anfordern einer Stellungnahme i. d. R. nicht dem Ermessen der Landesfachgruppen.

Datengrundlage der Analyse zu Fragestellung 2 sind alle Standorte, von welchen zum Erfassungsjahr 2016 eine Stellungnahme angefordert wurde und für welche keine D-Bewertung („Bewertung nicht möglich wegen fehlerhafter Dokumentation“) getroffen wurde.

## 4.2 Ergebnisse

Die Berechnung der Modelle zu Fragestellung 1 (Aufnahme eines Stellungnahmeverfahrens bei rechnerischer Auffälligkeit) ergab für die drei Qualitätsindikatoren die in Tabelle 5 dargestellten Parameterschätzungen.

Tabelle 5: Parameterschätzungen der drei Modelle zu Fragestellung 1: Heterogenität bezüglich des Anforderns einer Stellungnahme

Parameter	Schätzung	(Median-)Odds-Ratio	Konfidenzintervall (Median-)Odds-Ratio	
			2,5 %	97,5 %
<b>HEP: QI 54003 (ratenbasiert)</b>				
$\beta_0$	0,57	-	-	-
$\beta_{Q_2}$	1,28	3,6	1,1	14
$\beta_{Q_3}$	4,28	72	20	360
$\beta_{Q_4}$	7,21	$1,4 \cdot 10^3$	300	$8,4 \cdot 10^3$
$\tau$	4,67	86	15	$5,1 \cdot 10^3$
<b>DEK: QI 52010 (Sentinel Event)</b>				
$\beta_0$	13,65	-	-	-
$\beta_{Q_3}$	1,37	3,9	0,33	70
$\beta_{Q_4}$	6,27	530	68	$1,1 \cdot 10^4$
$\tau$	8,08	$2,2 \cdot 10^3$	19	$7,9 \cdot 10^{12}$
<b>09/1: QI 54139 (ratenbasiert)</b>				

Parameter	Schätzung	(Median-)Odds-Ratio	Konfidenzintervall (Median-)Odds-Ratio	
			2,5 %	97,5 %
$\beta_0$	0,43	-	-	-
$\beta_{Q_2}$	0,61	1,8	0,54	6,9
$\beta_{Q_3}$	4,04	57	15	270
$\beta_{Q_4}$	4,48	88	23	430
$\tau$	5,00	120	18	$9,2 \cdot 10^3$

Die Heterogenität zwischen den Bewertungsstellen lässt sich aus den Schätzungen für den Parameter  $\tau$  bzw. dem daraus resultierenden Median-Odds-Ratio ablesen. Für alle drei Indikatoren ergeben die modellbasierten Analysen einen Wert für  $\tau$ , dessen Konfidenzintervall 0 nicht einschließt. Dies bedeutet, dass bei allen drei Indikatoren signifikante Heterogenität darin vorliegt, ob bei vergleichbarem rechnerischen Ergebnis eine Stellungnahme eingeholt wird oder lediglich ein Hinweis versendet wird. Diese Heterogenität lässt sich dementsprechend nicht auf Unterschiede in der Fallzahl oder im Ausmaß der rechnerischen Auffälligkeit zwischen den Leistungserbringern der Bundesländer erklären. Der Modellparameter  $\beta_0$  quantifiziert die Grundwahrscheinlichkeit für den jeweiligen Indikator, dass eine Stellungnahme eingeholt wird. Die Modellparameter  $\beta_{Q_2}$  bis  $\beta_{Q_4}$  sind auch deutlich von 0 verschieden. Dies bedeutet, dass der in  $r_{lb}$  zusammengefasste Einfluss von Fallzahl und Ausmaß der rechnerischen Auffälligkeit einen messbaren Einfluss auf die Entscheidung hat, ob eine Stellungnahme eingeholt wird oder ein Hinweis versendet wird. Dieser in  $r_{lb}$  kombinierte Einfluss kann als die Evidenz für die Verfehlung des Referenzbereichs interpretiert werden. Von Quartil  $Q_1$  bis Quartil  $Q_4$  steigt demnach die Evidenz dafür, dass die Leistungserbringer im jeweiligen Quartil den Referenzbereich verfehlt haben. Die von  $Q_2$  bis  $Q_4$  größer werdenden  $\beta$ -Parameter zeigen, dass mit zunehmender Evidenz für die Verletzung des Referenzbereichs die Bundesländer mit zunehmender Wahrscheinlichkeit eine Stellungnahme einholen statt einen Hinweis zu versenden.

Die Ergebnisse zu Fragestellung 2 (Qualitative Auffälligkeit nach Aufnahme eines Stellungnahmeverfahrens) finden sich in Tabelle 6.

Tabelle 6: Parameterschätzungen der drei Modelle zu Fragestellung 2: Heterogenität bezüglich der Bewertung als qualitativ auffällig

Parameter	Schätzung	(Median-)Odds-Ratio	Konfidenzintervall (Median-)Odds-Ratio	
			2,5 %	97,5 %
<b>HEP: QI 54003 (ratenbasiert)</b>				
$\beta_0$	-2,71	-	-	-
$\beta_{Q_2}$	1,03	2,8	0,95	9,3

Parameter	Schätzung	(Median-)Odds-Ratio	Konfidenzintervall (Median-)Odds-Ratio	
			2,5 %	97,5 %
$\beta_{Q_3}$	1,91	6,8	2,5	21
$\beta_{Q_4}$	3,02	21	7,7	66
$\tau$	1,02	2,7	1,7	5,8
<b>DEK: QI 52010 (Sentinel Event)</b>				
$\beta_0$	-2,18	-	-	-
$\beta_{Q_3}$	1,18	3,2	1,5	7,3
$\beta_{Q_4}$	1,93	6,9	3,5	14
$\tau$	1,59	4,6	2,6	12
<b>09/1: QI 54139 (ratenbasiert)</b>				
$\beta_0$	-2,83	-	-	-
$\beta_{Q_2}$	0,93	2,5	0,64	11
$\beta_{Q_3}$	0,93	2,5	0,72	10
$\beta_{Q_4}$	2,054	7,8	2,2	33
$\tau$	1,289	3,4	1,5	14

Auch für Fragestellung 2 ergeben die modellbasierten Analysen für alle drei Indikatoren einen Wert für  $\tau$ , dessen Konfidenzintervall 0 nicht miteinschließt. Somit gibt es auch in der Bewertung der Leistungserbringer als qualitativ auffällig eine messbare Heterogenität zwischen den Bewertungsstellen, die sich nicht durch die rechnerischen Ergebnisse der Leistungserbringer erklären lässt. Mit dem Quartil  $Q_2$  bis  $Q_4$  größer werdende  $\beta$ -Parameter belegen, dass mit zunehmender Evidenz  $r_{lb}$  für die Verletzung des Referenzbereichs die Wahrscheinlichkeit steigt, dass ein Leistungserbringer mit rechnerischem Ergebnis im entsprechenden Quartil abschließend als qualitativ auffällig bewertet wird.

Zur besseren Interpretation der Ergebnisse stellen wir die Verteilung der durch das Modell geschätzten Wahrscheinlichkeiten für beide Fragestellungen und alle drei Qualitätsindikatoren in Abbildung 2 bis Abbildung 4 dar. Für jedes Quartil der Leistungserbringerergebnisse wird durch die Punkte dargestellt, was die Bewertungswahrscheinlichkeit aller 16 Bewertungsstellen wäre, wenn diese einen Leistungserbringer aus dem entsprechenden Quartil beurteilt hätten. Dies wird auf Grundlage der geschätzten Werte der Parameter  $\beta_0, \beta_{Q_2}, \beta_{Q_3}, \beta_{Q_4}$  und  $u_b$  für jede Bewertungsstelle  $b$  extrapoliert. In der grafischen Darstellung entspricht in jedem Quartil der Leistungserbringerergebnisse jeder Bewertungsstelle ein Punkt.

Vergleicht man die Ergebnisse der modellbasierten Analysen zwischen den Indikatoren und Fragestellungen, so stellt man fest, dass für die beiden ratenbasierten Indikatoren aus den QS-Verfahren 09/1 und HEP die Heterogenität beim Anfordern von Stellungnahmen (Fragestellung 1)

deutlich größer ist als in der Bewertung der Leistungserbringer als qualitativ auffällig (Fragestellung 2). In den Ergebnissen zu Fragestellung 1 (vgl. Abbildung 2 bis Abbildung 4 oben) lässt sich erkennen, dass sich die Bewertungsstellen tendenziell in zwei Gruppen einteilen: Bewertungsstellen, die fast immer Stellungnahmen einholen und Bewertungsstellen, welche seltener und nur bei deutlichen Überschreitungen des Referenzbereiches eine Stellungnahme anfordern. Diese bimodale Verteilung der Bewertungsstelleneffekte deutet auf eine Verletzung der Normalverteilungsannahme des Random-Effects-Modelles hin. Da es sich immer noch um eine symmetrische Verteilung handelt, kann der Parameter  $\tau$  allerdings immer noch als ein valides Heterogenitätsmaß angesehen werden.

Ein anderes Bild ergibt sich bei dem Sentinel-Event-Indikator (Dekubitus Grad 4): 15 von 16 Landesstellen nehmen bei rechnerischer Auffälligkeit eines Leistungserbringers fast immer ein Stellungnahmeverfahren auf, sodass sich hier (bis auf ein Bundesland) ein sehr homogenes Bild bzgl. Fragestellung 1 zeichnet. Die Standardabweichung  $\tau$  wird in diesem Fall allerdings sehr hoch geschätzt (mit breitem Konfidenzintervall), was sich vor allem durch das stark abweichende Verhalten einer Bewertungsstelle begründet. Auch hier ist davon auszugehen, dass die Normalverteilungsannahme des Random-Effects-Modells verletzt ist. Anders als bei den beiden ratenbasierten Indikatoren, stellt der Parameter  $\tau$  für diese Art von Verteilung mit einem starken Ausreißer kein verlässliches Heterogenitätsmaß dar. Während für diesen Sentinel-Event-Indikator das Anfordern von Stellungnahmen recht homogen im Vergleich zu den beiden ratenbasierten Indikatoren gehandhabt wird, ist die Heterogenität in der Bewertung der Leistungserbringer als qualitativ auffällig für den Sentinel-Event-Indikator am größten.

Zusätzlich zum Einfluss der Bewertungsstellen lässt sich übergreifend für alle drei Qualitätsindikatoren und beide Fragestellungen ein deutlicher Einfluss der rechnerischen Leistungserbringerergebnisse auf die modellierten Wahrscheinlichkeiten feststellen. Im Extremfall hat bei der Frage nach dem Anfordern einer Stellungnahme (Fragestellung 1) im HEP-Indikator „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ ein Leistungserbringer aus dem vierten Quartil der rechnerischen Ergebnisse  $r_{lb}$  gegenüber einem Leistungserbringer aus dem ersten Quartil einen Odds-Ratio von  $1,4 \cdot 10^3$  (95 % KI: [300;  $8,4 \cdot 10^3$ ]). Für den Sentinel-Event-Indikator fallen erstes und zweites Quartil der rechnerischen Leistungserbringerergebnisse (also die Anzahl der Sentinel-Event-Ereignisse) für beide Fragestellungen zusammen: Mehr als die Hälfte der Leistungserbringer sind durch ein einzelnes Sentinel-Event-Ereignis rechnerisch auffällig geworden. Gegenüber diesen Leistungserbringern haben Leistungserbringer mit 2 Sentinel-Event-Fällen (drittes Quartil  $Q_3$ ) und Leistungserbringer mit mehr als drei Sentinel-Event-Fällen (viertes Quartil  $Q_4$ ) bei Fragestellung 1 einen Odds-Ratio von 3,9 (95 % KI: [0,33; 70]) für das dritte Quartil bzw. 530 (95 % KI: [68;  $1,1 \cdot 10^4$ ]) für das vierte Quartil. Der Einfluss der rechnerischen Leistungserbringerergebnisse auf die Einstufung als qualitativ auffällig (Fragestellung 2) fällt insgesamt geringer aus als für Fragestellung 1. Auch hier stellt der Indikator „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ den Extremfall dar: Leistungserbringer im vierten Quartil der rechnerischen Leistungserbringerergebnisse haben gegenüber dem ersten Quartil einen Odds-Ratio von 21 (95 % KI: [7,7; 66]). Ein geringerer Effekt ist dagegen beim Indikator „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“ zu erkennen. Gegenüber dem ersten

Quartil der rechnerischen Leistungserbringerergebnisse  $r_{lb}$  liegt der Odds-Ratio für das vierte Quartil bei 7,8 (95 % KI: [2,2: 33]).

Bei allen Ergebnissen ist durch tendenziell große Konfidenzintervalle zu erkennen, dass die Parameterschätzungen einer deutlichen Unsicherheit unterliegen. Dies erklärt sich durch eine vergleichsweise kleine Datenbasis im Vergleich zur spezifizierten Modellkomplexität: In jedem Modell werden mindestens 20 Parameter geschätzt. Die Analysen zu beiden Fragestellungen basieren jedoch auf wenigen Hundert Datenpunkten (vgl. Abschnitt 2.5, Tabelle 2 im Abschlussbericht)

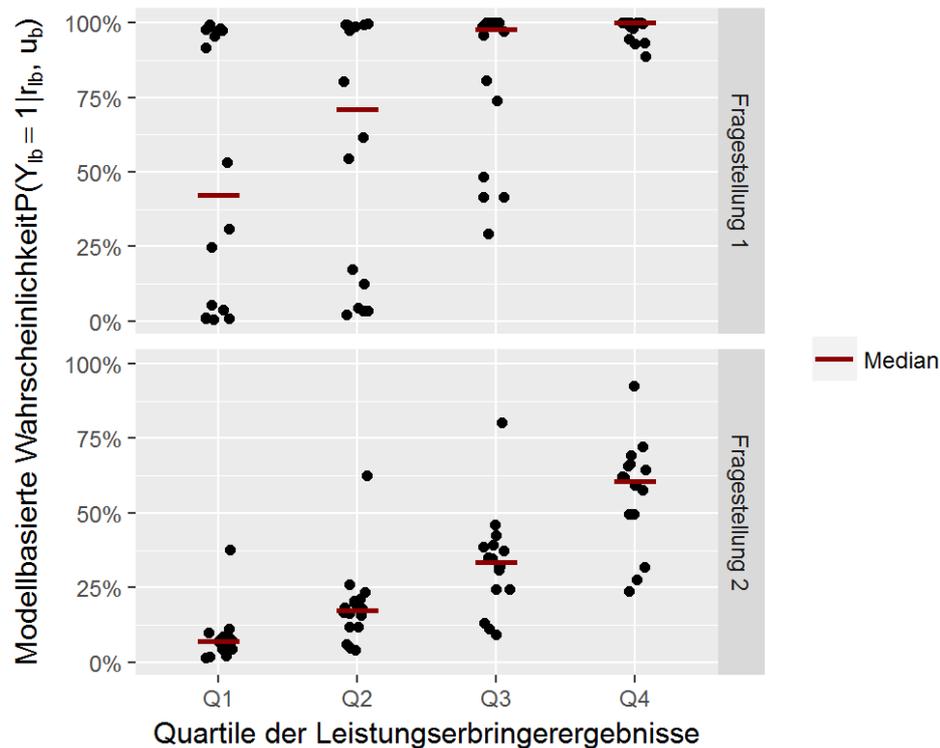


Abbildung 2: HEP QI 54003 (ratenbasierter Indikator). Streuung der geschätzten Bewertungswahrscheinlichkeit in Quartilen der Leistungserbringerergebnisse

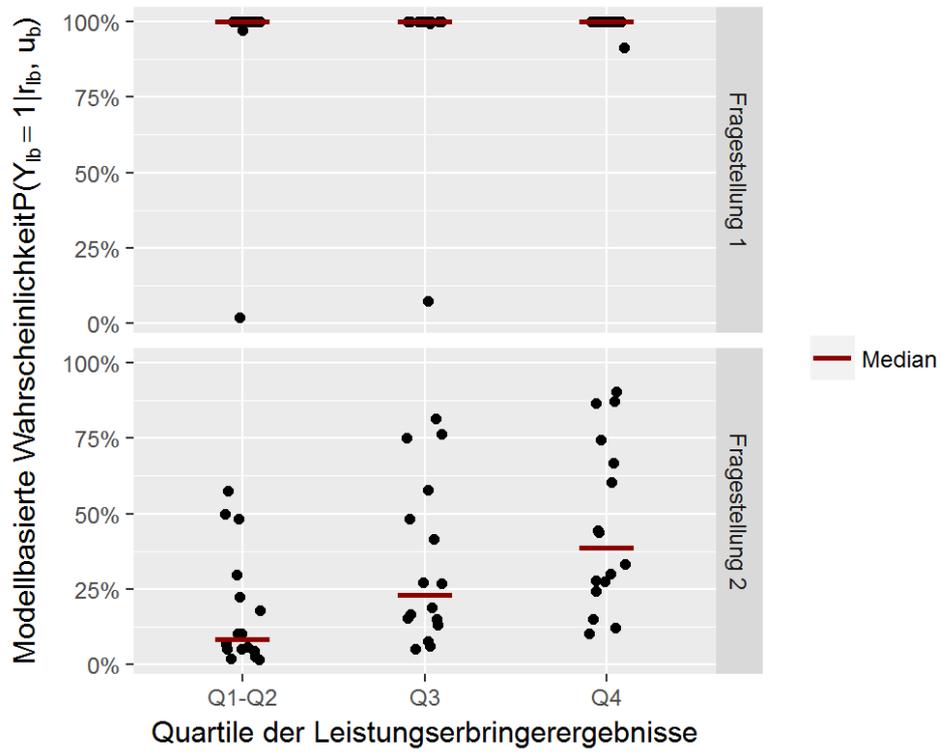


Abbildung 3: DEK QI 52010 (Sentinel-Event-Indikator). Streuung der geschätzten Bewertungswahrscheinlichkeit in Quartilen der Leistungserbringerergebnisse

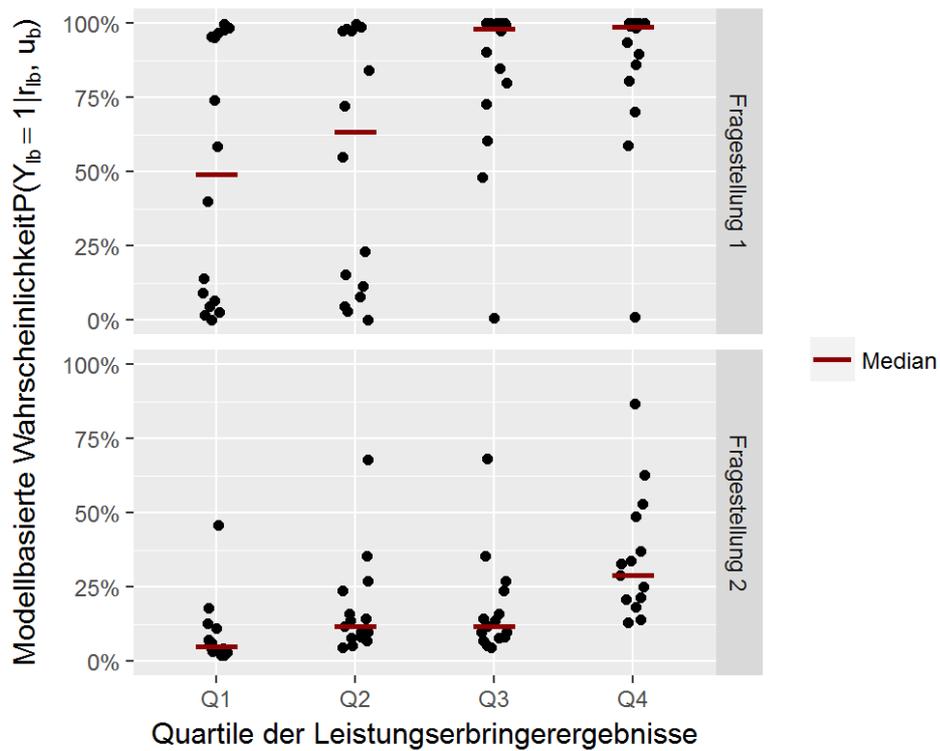


Abbildung 4: 09/1 QI 54139 (ratenbasierter Indikator). Streuung der geschätzten Bewertungswahrscheinlichkeit in Quartilen der Leistungserbringerergebnisse

### 4.3 Zusammenfassung und Diskussion

Es wurde eine Methodik vorgestellt, welche ermöglicht, die Heterogenität in der Aufnahme und Bewertungen des Strukturierten Dialoges quantitativ unter Berücksichtigung der rechnerischen Ergebnisse der Leistungserbringer zu untersuchen. Die Ergebnisse der Analyse zu beiden Fragestellungen belegen für die betrachteten Qualitätsindikatoren, dass auch unter Berücksichtigung der rechnerischen Leistungserbringerergebnisse eine starke Heterogenität zwischen den Bewertungsweisen der Landesstellen besteht. Dies gilt sowohl für die Frage nach der Aufnahme eines Stellungnahmeverfahrens als auch für die finale Bewertung der Leistungserbringer als qualitativ auffällig. Dabei wird die Einstufung als qualitativ auffällig weniger heterogen gehandhabt als die Aufnahme eines Stellungnahmeverfahrens.

Sensitivitätsanalysen mit Modellen, welche die rechnerischen Leistungserbringerergebnisse nicht berücksichtigen (vgl. Abschnitt 4.4.3), legen nahe, dass die Heterogenität in der Bewertung sogar *größer* wird, dadurch dass man die rechnerischen Ergebnisse der Leistungserbringer mit einbezieht. Die Hypothese, dass sich ein Teil der Heterogenität der Bewertungen des Strukturierten Dialogs aus tatsächlichen Unterschieden zwischen den Leistungserbringern erklärt, kann somit nicht bekräftigt werden.

Weiterhin zeigen Schätzungen der jeweiligen Parameter  $\beta_{Q_2}$ ,  $\beta_{Q_3}$  und  $\beta_{Q_4}$  zu den Leistungserbringerergebnissen, dass die rechnerischen Ergebnisse der Leistungserbringer sowie ihre Fallzahl (implizit oder explizit) in die Bewertung durch die Landesstellen einbezogen werden. Insbesondere belegen weitere Sensitivitätsanalysen, dass bei der Bewertung der Leistungserbringerergebnisse durch die Bewertungsstellen neben dem rechnerischen Ergebnis auch die Fallzahl des Leistungserbringers eine Rolle spielt (vgl. Abschnitt 4.4.2). Dabei wurde allerdings nicht untersucht, wie genau welche Bewertungsstelle die rechnerischen Ergebnisse in ihre Bewertungen einfließen lässt. Beispielsweise berücksichtigen manche Bewertungsstellen Konfidenzintervalle, während andere Bewertungsstellen keine inferenzstatistischen Methoden anwenden.

Grundsätzlich ist die Rolle der rechnerischen Leistungserbringerergebnisse in der Definition der Modelle (1) je nach Fragestellungen eine andere. Für die Frage nach der Heterogenität bezüglich des Anforderns einer Stellungnahme, stellen die rechnerischen Ergebnisse der Leistungserbringer die wichtigste Grundlage dar, anhand derer sich die Bewertungsstellen orientieren können. Es ist daher sinnvoll zu betrachten, wie vergleichbare rechnerische Ergebnisse jeweils in den Bewertungsstellen bewertet werden. Eine andere Rolle haben die rechnerischen Ergebnisse bei der Analyse der finalen Qualitätsbewertungen als qualitativ auffällig. Hier dienen die rechnerischen Ergebnisse dazu tatsächliche Qualitätsunterschiede zwischen den Leistungserbringern zu approximieren, da kein anderes Maß um die Landesergebnisse quantitativ vergleichbar zu machen zur Verfügung steht.

Eine Limitation der vorliegenden Analyse liegt in der Einschränkung auf drei exemplarisch ausgewählte Qualitätsindikatoren: Nur wenige Qualitätsindikatoren weisen so viele rechnerisch auffällige Standorte und Stellungnahmeverfahren auf, dass eine vergleichende statistische Analyse der Landesergebnisse auf Ebene einzelner Qualitätsindikatoren sinnvoll ist. Das Ergebnis

der Analyse ist daher nicht unbedingt repräsentativ für alle Qualitätsindikatoren. Es besteht allerdings die Möglichkeit zukünftig durch Verallgemeinerung der vorgestellten Modelle auch qualitätsindikatorenübergreifende Analysen durchzuführen. Dies kann etwa im Rahmen einer langfristigen Evaluation der getroffenen Maßnahmen zur Angleichung des strukturierten Dialoges sinnvoll sein. Die Varianz  $\tau^2$  der Bewertungsstelleneffekte aus dem Random-Effects-Modell wäre dabei auch für einen jahresübergreifenden Vergleich ein geeignetes Maß für die Heterogenität in der Bewertung des strukturierten Dialoges.

#### 4.4 Details zur statistischen Modellierung

Wir gehen im Folgenden auf weitere Details zum in Gleichung (1) beschriebenen Modell ein und diskutieren die Modellwahl.

Sei  $Y_{lb}$  das betrachtete Bewertungsergebnis (je nach Fragestellung 1 oder 2) eines Leistungserbringers  $l$  durch Bewertungsstelle  $b$ . In Abschnitt 4.1 wurde folgendes logistisches Regressionsmodell mit bewertungsstellenspezifischen Effekten formuliert:

$$\text{logit}[P(Y_{lb} = 1|r_{lb}, u_b)] = \beta_0 + \beta_{Q_2} \cdot Q_2(r_{lb}) + \beta_{Q_3} \cdot Q_3(r_{lb}) + \beta_{Q_4} \cdot Q_4(r_{lb}) + u_b.$$

Hierbei sind  $\beta_0, \beta_{Q_2}, \beta_{Q_3}, \beta_{Q_4}$  und  $u_b$  zu bestimmende Modellparameter,  $r_{lb}$  eine Transformation des rechnerischen Ergebnisses des Leistungserbringers. Mit  $Q_2(r_{lb}), Q_3(r_{lb})$  und  $Q_4(r_{lb})$  bezeichnen wir das zweite bis vierte Quartil von  $r_{lb}$ . Es handelt sich um ein hierarchisches Regressionsmodell bzw. Multilevelmodell [vgl. z. B. (Gelman und Hill 2007); (Stroup 2013)], da statistische Effekte auf zwei Ebenen (Bewertungsstelle, Leistungserbringer) berücksichtigt werden. Hierarchische Modelle werden insbesondere bei Analysen von Longitudinal- oder Cluster-Daten eingesetzt, bei denen Beobachtungen in bestimmten Gruppen auftreten. Die Beschreibung von Bewertungen des Strukturierten Dialogs, welche durch verschiedene Bewertungsstellen durchgeführt werden, ist somit ein Standardanwendungsfall hierarchischer Modelle.

##### 4.4.1 Fixed- vs. Random-Effects-Modelle

Bei hierarchischen Regressionsmodellen unterscheidet man sog. Fixed- und Random-Effects-Parameter [vgl. z. B. Gelman und Hill (2007), Stroup (2013)]. Der Unterschied zwischen diesen beiden Arten von Parametern liegt dabei in der Vorgehensweise bei der statistischen Schätzung. Während bei der Schätzung von Fixed-Effects a priori keine Beschränkung der Parameter vorgesehen ist (und diese somit beliebig hohe Werte annehmen können), wird der mögliche Wertebereich von Random-Effects durch eine sog. Penalisierung beschränkt. Man spricht an dieser Stelle auch von Shrinkage, da die Modellkoeffizienten in einer penalisierten Schätzung tendenziell eine geringere Streuung aufweisen als ohne Penalisierung. Der Grad der Penalisierung muss dabei durch einen weiteren Parameter festgelegt werden. In Random-Effects-Modellen werden die sog. Penalisierungsparameter jeweils für eine Gruppe an Parametern (z. B. die Gruppe der  $u_b$  in Gleichung (1) anhand der Daten bestimmt. Dabei macht man die Annahme, dass die gruppenspezifischen Effekte (im Beispiel  $u_b$ ) einer übergeordneten Normalverteilung  $\mathcal{N}(0, \tau^2)$  folgen. Der Penalisierungsparameter wird dann als Varianz der gruppenspezifischen Effekte festgelegt [vgl. Wood (2013)]. Je größer die Varianz zwischen den Gruppen ausfällt, desto weniger wird die

Schätzung penalisiert. Im Gegensatz zu Fixed-Effects-Parametern, welche meist unter Verwendung der Maximum-Likelihood-Methode geschätzt werden, verwendet man für Random-Effects-Parameter sog. *BLUP*-Schätzer<sup>4</sup> [vgl. z. B. Stroup (2013)]. In der Praxis führt die Verwendung von Random-Effects-Parametern mit Normalverteilungsannahme dazu, dass Ausreißereffekte im Vergleich zum Fixed-Effects-Modell reduziert und insbesondere die Parameter  $u_b$  von Gruppen mit geringer Fallzahl<sup>5</sup> zum gruppenübergreifenden Mittelwert „geschrumpft“ werden. Eine detaillierte Diskussion zur Verwendung von Random-Effects-Modellen im Kontext der Qualitätsbewertung von Krankenhäusern findet sich bei Ash et al. (2012).

In der vorliegenden Analyse wurde ein Random-Effects-Modell verwendet, da einerseits die Anzahl an Leistungserbringern in einigen Bundesländern so gering ist, dass eine Stabilisierung der Schätzung zur Vermeidung von numerischen Konvergenzproblemen nötig ist. Andererseits ist genau die Quantifizierung der Heterogenität in der Bewertung durch die Landesstellen ein Ziel der Analyse. Der Varianz-Parameter  $\tau^2$  stellt hierfür ein geeignetes Maß zur Verfügung.

#### 4.4.2 Operationalisierung des Begriffes vergleichbarer Leistungserbringer

Um zu analysieren wie unterschiedlich die Bewertung im Strukturierten Dialog für vergleichbare Leistungserbringer ausfällt, ist eine operative Definition des Begriffes „vergleichbare Leistungserbringer“ nötig. Dies ist unabhängig davon, ob man die Ergebnisse stratifiziert analysiert oder modellbasiert vorgeht. Auch wenn die Auslösung des strukturierten Dialoges nach QSKH-Richtlinie keine Differenzierung nach Fallzahl vorsieht, gehen wir im Folgenden davon aus, dass ein gleiches Ratenergebnis bei kleiner und großer Fallzahl (z. B. 2/4 und 50/100 bei einem Referenzbereich von  $\leq 15\%$ ) als unterschiedliche Evidenz zu bewerten ist. Wir berücksichtigen daher in der Modellierung Fallzahl und Ratenergebnis der Leistungserbringer. Eine Ausnahme davon machen wir für Sentinel-Event-Indikatoren. Hier zählt allein die absolute Anzahl an beobachteten Sentinel-Events, da in diesem Fall die Auslösung des strukturierten Dialoges allein anhand der „Zähler“-Fälle (also nicht ratenbasiert) erfolgt.

Es gibt mehrere Möglichkeiten, die rechnerischen Ergebnisse der Leistungserbringer in das Modell (1) einzubeziehen. Sei  $\hat{\pi}_{lb} = o_{lb}/n_{lb}$  das beobachtete Ratenergebnis (Zähler durch Nenner) von Leistungserbringer  $l$  in Bundesland bzw. Bewertungsstelle  $b$ . Wir verstehen  $\hat{\pi}_{lb}$  als Schätzung einer zu Grunde liegenden Rate  $\pi_{lb}$  des Leistungserbringers anhand der beobachteten Daten. Sei  $t$  der Referenzwert des betrachteten Qualitätsindikators. Ohne Beschränkung der Allgemeinheit wird angenommen, dass der Qualitätsindikator so definiert ist, dass hohe Werte der Rate schlechter Versorgungsqualität entsprechen. Analog zu Spiegelhalter et al. (2012) bezeichnen wir mit  $p(o_{lb}, n_{lb}, t)$  den p-Wert eines einseitigen exakten Binomialtests zur Null-Hypothese  $H_0: \pi_{lb} \leq t$  („Der Leistungserbringer liegt im Referenzbereich“).<sup>6</sup>  $p(o_{lb}, n_{lb}, t)$  quantifiziert somit die Evidenz, dass ein Leistungserbringer den Referenzbereich einhält bzw. verletzt.<sup>7</sup> Diesen

<sup>4</sup> Engl. Best Linear Unbiased Predictors.

<sup>5</sup> In diesem Anwendungsfall bedeutet „geringe Fallzahl“, dass nur wenige Leistungserbringer in einem Bundesland rechnerisch auffällig wurden.

<sup>6</sup> Wir verwenden dabei auch die von Spiegelhalter et al. (2012) vorgeschlagene mid-P-Erweiterung.

<sup>7</sup> So wie die Null-Hypothese definiert wurde, indiziert ein kleiner p-Wert eine Verletzung des Referenzbereiches.

p-Wert verwenden wir dabei in Modell (1) als Maß, wie ähnlich die Ergebnisse zweier Leistungserbringer sind. Genauer definieren wir die in Gleichung (1) einfließende Kovariable  $r_{lb}$  als

$$r_{lb} := p(o_{lb}, n_{lb}, t). \quad (2)$$

Eine anschauliche Interpretation dieser Definition von  $r_{lb}$  liefert ein nach Leistungserbringerfallzahl differenzierter Plot der Ergebnisse (vgl. Abbildung 5). Würde man nur die rechnerischen Ratenergebnisse ohne zusätzliche Berücksichtigung der Fallzahl verwenden um Leistungserbringer in vergleichbare Gruppen einzuteilen, würde man den Plot in (fallzahlunabhängige) horizontale Segmente einteilen (Abbildung 5 rechts). Dem entgegen entspricht die Verwendung der p-Werte einer Unterteilung des Plots in hyperbelförmige Segmente. Dies wird Abbildung 5 verdeutlicht. Die schwarzen Punkte zeigen die Ergebnisse hypothetischer Leistungserbringer. Gezeigt wird dabei nur der Ausschnitt rechnerisch auffälliger Leistungserbringer (oberhalb von 15 %). Leistungserbringerergebnisse in einem gleichen Farbsegment werden je nach Definition von  $r_{lb}$  als „ähnlich“ eingestuft.

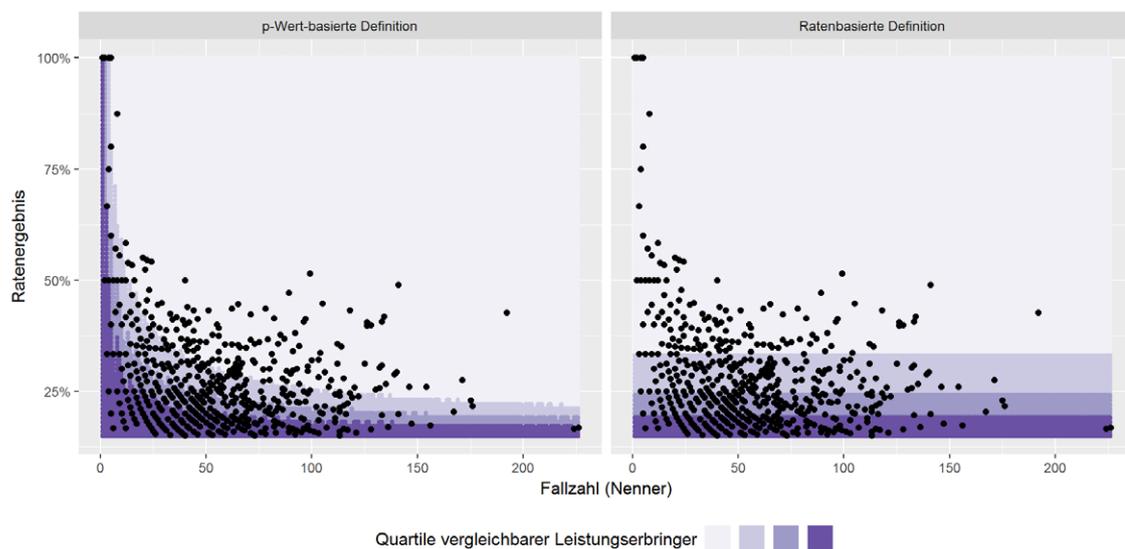


Abbildung 5: Illustration zu alternativen Definitionen von  $r_{lb}$  anhand einer nach Fallzahl differenzierten Darstellung der rechnerischen Ergebnisse. Segmente gleicher Farbe markieren Gruppen von Leistungserbringern deren Ergebnisse als vergleichbar eingestuft würden. Links die p-Wert-basierten Definition; rechts die ratenbasierte Definition.

Die obige Modellierung ist ohne Definition eines Referenzbereiches  $t > 0$  für Sentinel-Event-Indikatoren nicht sinnvoll. Die Auslösung des Strukturierten Dialoges passiert hier ausschließlich durch das Zähler-Ergebnis  $o_{lb} \geq 1$ . Daher fließt in diesem Fall in die Modellierung auch nur das Zählerergebnis ein und wir definieren in (1)  $r_{lb}$  fallzahlunabhängig als

$$r_{lb} := o_{lb}.$$

Die Interpretation der Koeffizienten  $\beta_{Q_2}$ ,  $\beta_{Q_3}$  und  $\beta_{Q_4}$  bleibt allerdings gleich: Sie quantifizieren, wie stark die Wahrscheinlichkeit  $P(Y_{lb} = 1 | o_{lb}, n_{lb}, u_b)$  ansteigt, wenn ein Leistungserbringer im zweiten, dritten oder vierten Quartil der Ergebnisse  $r_{lb}$  liegt.

Alternativ zur Definition von  $r_{lb}$  in Gleichung (2) wurden zur Sensitivitätsanalyse Modelle mit unterschiedlichen Definitionen von  $r_{lb}$  berechnet. Konkret wurden dabei folgende Alternativ-Modelle mit Random-Effects berechnet:

1. Ausgangsmodell:

$$\text{logit}[P(Y_{lb} = 1 | o_{lb}, n_{lb}, u_b)] = \beta_0 + \beta_{Q_2} \cdot Q_2(r_{lb}) + \beta_{Q_3} \cdot Q_3(r_{lb}) + \beta_{Q_4} \cdot Q_4(r_{lb}) + u_b.$$

2. Variante 1: Modell mit stetiger Berücksichtigung des p-Wertes

$$\text{logit}[P(Y_{lb} = 1 | o_{lb}, n_{lb}, u_b)] = \beta_0 + \beta_1 \cdot \text{logit}[p(o_{lb}, n_{lb}, t)] + u_b.$$

3. Variante 2: Modell ohne Berücksichtigung der Fallzahl, mit stetiger Berücksichtigung des Ratenergebnisses

$$\text{logit}[P(Y_{lb} = 1 | o_{lb}, n_{lb}, u_b)] = \beta_0 + \beta_1 \cdot \text{logit}\left(\frac{o_{lb}}{n_{lb}}\right) + u_b.$$

4. Variante 3: Modell mit separatem dichotomen Fallzahl-Effekt ( $n_{lb} \leq 20$  ja/ nein)<sup>8</sup>

$$\text{logit}[P(Y_{lb} = 1 | o_{lb}, n_{lb}, u_b)] = \beta_0 + \beta_1 \cdot \text{logit}\left(\frac{o_{lb}}{n_{lb}}\right) + \beta_2 \cdot I(n_{lb} \leq 20) + u_b.$$

5. Variante 4: Modell mit separatem stetigen Fallzahl-Effekt  $f(n_{lb})$  in einem generalisierten gemischten additiven Modell (Wood 2006)

$$\text{logit}[P(Y_{lb} = 1 | o_{lb}, n_{lb}, u_b)] = \beta_0 + \beta_1 \cdot \text{logit}(\hat{\pi}_{lb}) + f(n_{lb}) + u_b.$$

Die logit-Transformation in Variante 1 und Variante 2 ist dabei ein gängiges Mittel die zwischen 0 und 1 liegenden p-Werte und Ratenergebnisse für die Regressionsmodellierung zu linearisieren [vgl. z. B. Hosmer et al. (2013)]. Die Modelle wurden für die beiden ratenbasierten Qualitätsindikatoren zu beiden Fragestellungen anhand des konditionalen AIC (*Aikake Information Criterion*) verglichen, einer Verallgemeinerung des AIC für Random-Effects-Modelle welches zur Modellselektion verwendet werden kann [vgl. Säfken et al. (2018)]. Somit wurde unter den betrachteten Modellen genau jenes Modell gewählt, welches am besten (nach AIC) zu den beobachteten Bewertungen im strukturierten Dialog passt. In fast allen Konstellationen hatte das Ausgangsmodell den niedrigsten AIC und wird daher favorisiert. Die AIC-Differenzen zwischen den Modellvarianten sind allerdings nicht in allen Konstellationen gleich ausgeprägt, was darauf schließen lässt, dass die Definition von  $r_{lb}$  je nach Qualitätsindikator und Fragestellung eine unterschiedliche Bedeutung haben kann. Die Tatsache, dass meist jenes Modell favorisiert wurde, welches den p-Wert-basierten Ansatz zur Definition von  $r_{lb}$  verwendet, kann man so interpretieren, dass einerseits viele Bewertungsstellen bei der Auslösung eines Stellungnahmeverfahrens explizit inferenzstatistische Methoden (etwa Konfidenzintervalle) berücksichtigen. Andererseits kann vermutet werden, dass Entscheidungen in den Bewertungsstellen implizit (also ohne konkrete Rechenvorschrift) die Fallzahl der Leistungserbringer in die Bewertung mit einbeziehen und dass dieses Verhalten wiederum durch den p-Wert-basierten Ansatz gut erklärt wird.

<sup>8</sup> Wir bezeichnen mit  $I(A)$  die Indikatorfunktion, die den Wert 1 annimmt, wenn die Bedingung  $A$  zutrifft und sonst den Wert 0 annimmt.

Von allen fünf Varianten wurden für die beide ratenbasierten Qualitätsindikatoren zu beiden Fragestellungen auch die Schätzungen der Varianzparameter  $\tau^2$  verglichen. Ähnlich wie beim AIC zeigt sich auch hier, dass je nach Fragestellung und Qualitätsindikator die Schätzungen von  $\tau^2$  unterschiedlich stark von der Definition von  $r_{lb}$  abhängen. Unter Berücksichtigung der Konfidenzintervalle ergeben sich jedoch keine relevanten Unterschiede zwischen den Modellvarianten. Die Schätzung von  $\tau^2$  hängt somit nicht maßgeblich von der genauen Definition von  $r_{lb}$  ab.

Für den Sentinel-Event-Indikator wurden ebenfalls alternative Modelld Definitionen mit und ohne Berücksichtigung der Fallzahl geprüft. Hier stellte sich heraus, dass eine Berücksichtigung der Fallzahl in der Definition von  $r_{lb}$  keine Verringerung des AIC in den Modellen zu Fragestellung 1 (Anforderung einer Stellungnahme) ergab. Die Modellierung von Fragestellung 2 (Bewertung als qualitativ auffällig) mit einem generalisierten additiven Modell mit stetigem Fallzahl-Effekt  $f(n_{lb})$  analog zu Variante 3 ergab eine geringfügige Verringerung des AIC. In Fragestellung 2 ist also auch für den Sentinel-Event-Indikator nicht auszuschließen, dass die Fallzahl des Leistungserbringers in der abschließenden Qualitätsbewertung eine Rolle spielt. Um die Interpretation des Modelles zu erleichtern, wurde allerdings davon abgesehen diesen nichtlinearen Effekt durch die Fallzahl des Leistungserbringers zu modellieren.

#### 4.4.3 Modelle ohne Berücksichtigung der Leistungserbringer-Ergebnisse

Eine weitere wichtige Sensitivitätsanalyse besteht im Vergleich von Modellen mit und ohne Berücksichtigung der Leistungserbringer-Ergebnisse. Im Folgenden stellen wir pro Indikator und Fragestellung jeweils vier Modelle gegenüber:

- $M_1$ : Random-Effect Modell ohne Berücksichtigung der Leistungserbringer (LE)-Ergebnisse

$$\text{logit}[P(Y_{lb} = 1|u_b)] = \beta_0 + u_b$$

- $M_2$ : Random-Effect Modell mit Berücksichtigung der LE-Ergebnisse

$$\text{logit}[P(Y_{lb} = 1|r_{lb}, u_b)] = \beta_0 + \beta_1 \cdot r_{lb} + u_b$$

- $M_3$  Berücksichtigung der LE-Ergebnisse ohne Random-Effects

$$\text{logit}[P(Y_{lb} = 1|r_{lb})] = \beta_0 + \beta_1 \cdot r_{lb}$$

- $M_4$  Triviales Modell ohne Random-Effects und ohne LE-Ergebnisse

$$\text{logit}[P(Y_{lb} = 1)] = \beta_0$$

mit  $u_b \sim \mathcal{N}(0, \tau^2)$ . Zum Vergleich der Modelle betrachten wir erneut die AIC-Werte. In der Situation von Random-Effects-Modellen ist zu beachten, dass man den konditionalen AIC betrachtet [vgl. Säfken et al. (2018)]. Die Ergebnisse finden sich in Tabelle 7 und Tabelle 8: (Konditionaler) AIC der Modellvarianten  $M_1$  -  $M_4$  zu Frage 2. Es überrascht nicht, dass das flexibelste Modell  $M_2$  für alle drei Qualitätsindikatoren und beide Fragestellungen die Daten am besten erklärt.

Sehr überraschend sind allerdings die Ergebnisse der Parameter-Schätzungen in der Modellvariante  $M_1$  ohne Berücksichtigung der rechnerischen Leistungserbringerergebnisse. Die Parameter-Schätzungen zu Fragestellung 1 und 2 finden sich in Tabelle 9 und Tabelle 10. Vergleicht man

dies mit den Ergebnissen aus Tabelle 5 und Tabelle 6, so fällt auf, dass die Heterogenität zwischen den Bewertungsstellen *geringer* ausfällt, wenn man die Ergebnisse der Leistungserbringer nicht berücksichtigt.

Tabelle 7: (Konditionaler) AIC der Modellvarianten  $M_1$ -  $M_4$  zu Frage 1

	AIC mit Random-Effects	AIC ohne Random-Effects
<b>HEP: QI 54003</b>		
ohne LE-Ergebnisse ( $M_1$ und $M_4$ )	469	757
mit LE-Ergebnissen ( $M_2$ und $M_3$ )	238	620
<b>DEK: QI 52010</b>		
ohne LE-Ergebnisse ( $M_1$ und $M_4$ )	120	406
mit LE-Ergebnissen ( $M_2$ und $M_3$ )	80	382
<b>09/1: QI 54139</b>		
ohne LE-Ergebnisse ( $M_1$ und $M_4$ )	267	600
mit LE-Ergebnissen ( $M_2$ und $M_3$ )	179	581

Tabelle 8: (Konditionaler) AIC der Modellvarianten  $M_1$  -  $M_4$  zu Frage 2

	AIC mit Random-Effects	AIC ohne Random-Effects
<b>HEP: QI 54003</b>		
ohne LE-Ergebnisse ( $M_1$ und $M_4$ )	439	473
mit LE-Ergebnissen ( $M_2$ und $M_3$ )	389	421
<b>DEK: QI 52010</b>		
ohne LE-Ergebnisse ( $M_1$ und $M_4$ )	354	443
mit LE-Ergebnissen ( $M_2$ und $M_3$ )	324	414
<b>09/1: QI 54139</b>		
ohne LE-Ergebnisse ( $M_1$ und $M_4$ )	188	203
mit LE-Ergebnissen ( $M_2$ und $M_3$ )	185	202

Tabelle 9: Parameterschätzungen Modellvariante M\_1 der drei Modelle zu Fragestellung 1: Heterogenität bezüglich des Anforderns einer Stellungnahme

Parameter	Schätzung	(Median-)Odds-Ratio	Konfidenzintervall (Median-)Odds-Ratio	
			2,5 %	97,5 %
<b>HEP: QI 54003</b>				
$\beta_0$	2,2	-	-	-
$\tau$	2,7	13	4,7	140
<b>DEK: QI 52010</b>				
$\beta_0$	10,3	-	-	-
$\tau$	54	170	8,7	9,5 · 10 <sup>9</sup>
<b>09/1: QI 54139</b>				
$\beta_0$	1,9	-	-	-
$\tau$	3,4	26	7,1	510

Tabelle 10: Parameterschätzungen Modellvariante M\_1 der drei Modelle zu Fragestellung 2: Heterogenität bezüglich der Bewertung als qualitativ auffällig

Parameter	Schätzung	(Median-)Odds-Ratio	Konfidenzintervall (Median-)Odds-Ratio	
			2,5 %	97,5 %
<b>HEP: QI 54003</b>				
$\beta_0$	-0,58	-	-	-
$\tau$	0,92	2,4	1,6	4,7
<b>DEK: QI 52010</b>				
$\beta_0$	-1,23	-	-	-
$\tau$	1,48	4,1	2,5	10
<b>09/1: QI 54139</b>				
$\beta_0$	-1,51	-	-	-
$\tau$	0,95	2,5	1	8

#### 4.4.4 Implementation

Alle Berechnungen wurden in der Statistiksoftware R Version 3.5.0 durchgeführt. Zur Berechnung der logistischen Regressionsmodelle mit Random-Effects wurde das Paket lme4 (Bates et al. 2015) in der Version 1.1-17 verwendet. Die Integralapproximation der log-Likelihood erfolgte durch eine adaptive Gauss-Hermite-Approximation mit jeweils 25 Stützpunkten pro Achse. Zur Optimierung der log-Likelihood wurde der BOBYQA-Algorithmus aus dem minqa-Paket, Version

1.2.4, genutzt. Die Konfidenzintervalle der Parameter-Schätzungen wurden durch die Likelihood-Profile-Methode berechnet, implementiert durch das lme4-Paket. Zur Berechnung der p-Werte für die einseitigen Binomialtests in der Definition von  $r_{lb}$  in Gleichung (2) wurde die Implementation des IQTIG aus dem IQTIGpvci-Paket in der Version 1.0.0 verwendet (IQTIG 2018). Die Modellvergleiche anhand des konditionalen AIC wurden mithilfe des cAIC4-Paketes Version 0.4 berechnet Säfken et al. (2018).

#### **4.5 Deskriptive Auswertungen zu den Ergebnissen des Strukturierten Dialogs zum Erfassungsjahr 2016 für ausgewählte Indikatoren**

Im Folgenden werden als Hintergrund deskriptive Auswertungen zu den Ergebnissen des strukturierten Dialoges der drei ausgewählten Qualitätsindikatoren dargestellt, welche die modellbasierten Auswertungen aus Abschnitt 4.2 ergänzen. Ziel dieser Auswertungen ist es Transparenz über die Datengrundlage der vorgestellten modellbasierten Analysen zu schaffen. Dies soll eine Einordnung der Heterogenität in den Bewertungen des strukturierten Dialoges ermöglichen, die anders als die Analysen aus Abschnitt 4.2 keinerlei zusätzliche Modellannahmen erfordert. Insbesondere die nach Bewertungsstelle, Fallzahl und Indikatorergebnis stratifizierte Darstellungen zeigen, dass Bewertungen im strukturierten Dialog sowohl abhängig vom Indikatorergebnis als auch abhängig von der Fallzahl getroffen werden.

#### 4.5.1 HEP: QI 54003 Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur

##### Anteil an Stellungnahmen an allen rechnerisch auffälligen Standorten

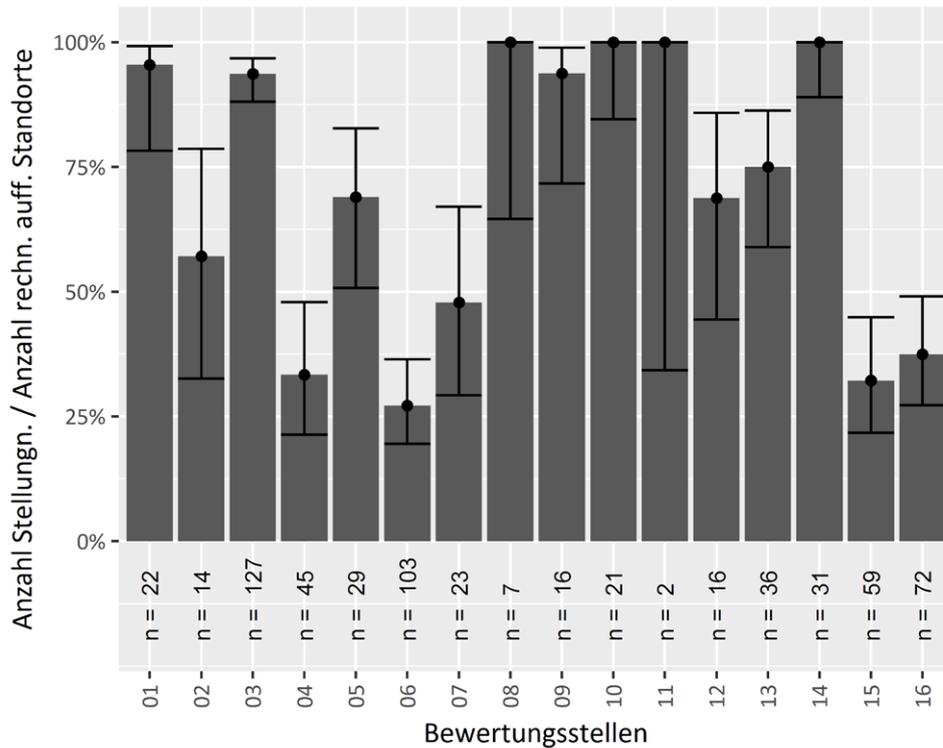


Abbildung 6: Qualitätsindikator HEP: 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“. Anfordern einer Stellungnahme nach rechnerischer Auffälligkeit

**Anteil an qualitativ auffälligen Bewertungen an allen Stellungnahmen**

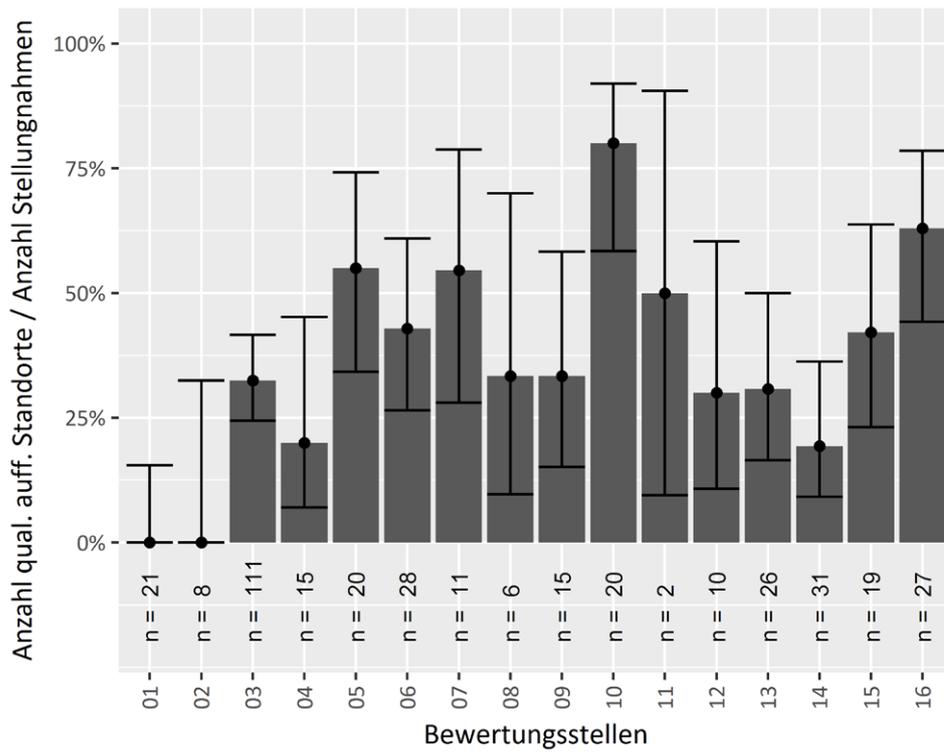


Abbildung 7: Qualitätsindikator HEP: 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“. Bewertungen nach Einleitung des Stellungnahmeverfahrens

**Darstellung der Standortergebnisse nach Bewertungsstelle und Fallzahl**

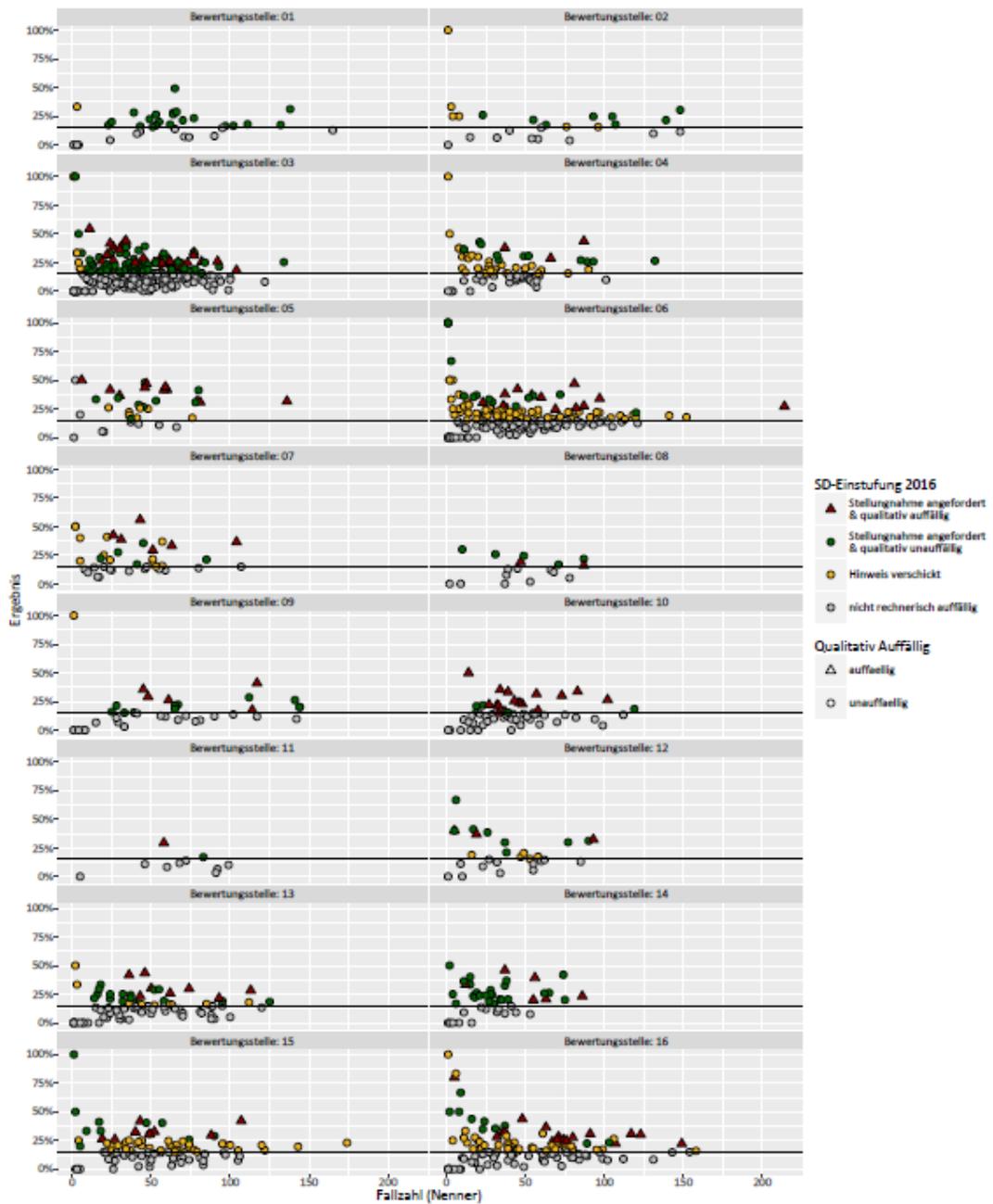


Abbildung 8: Ergebnisse des strukturierten Dialoges zum Qualitätsindikator HEP: 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ für das Erfassungsjahr 2016 differenziert nach Fallzahl und Bewertungsstelle. Diese Daten sind die Grundlage für die modellbasierten Auswertungen in Abschnitt 2.2.

#### 4.5.2 09/1: QI 54139 Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen

##### Anteil an Stellungnahmen an allen rechnerisch auffälligen Standorten

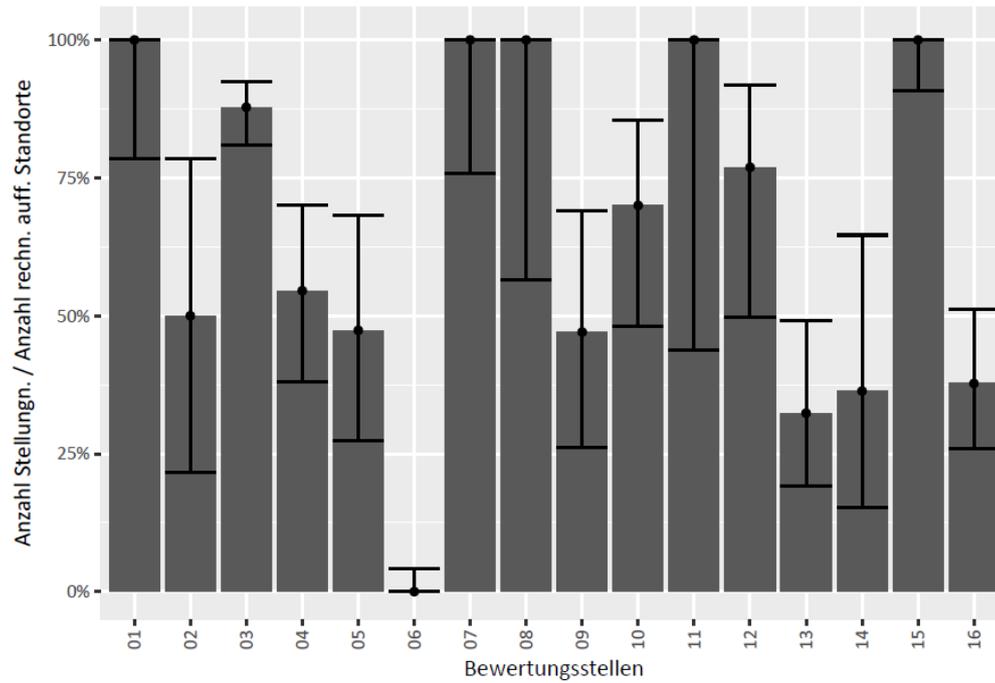


Abbildung 9: Qualitätsindikator 09/1: 54134 „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“. Anfordern einer Stellungnahme nach rechnerischer Auffälligkeit

**Anteil an qualitativ auffälligen Bewertungen an allen Stellungnahmen**

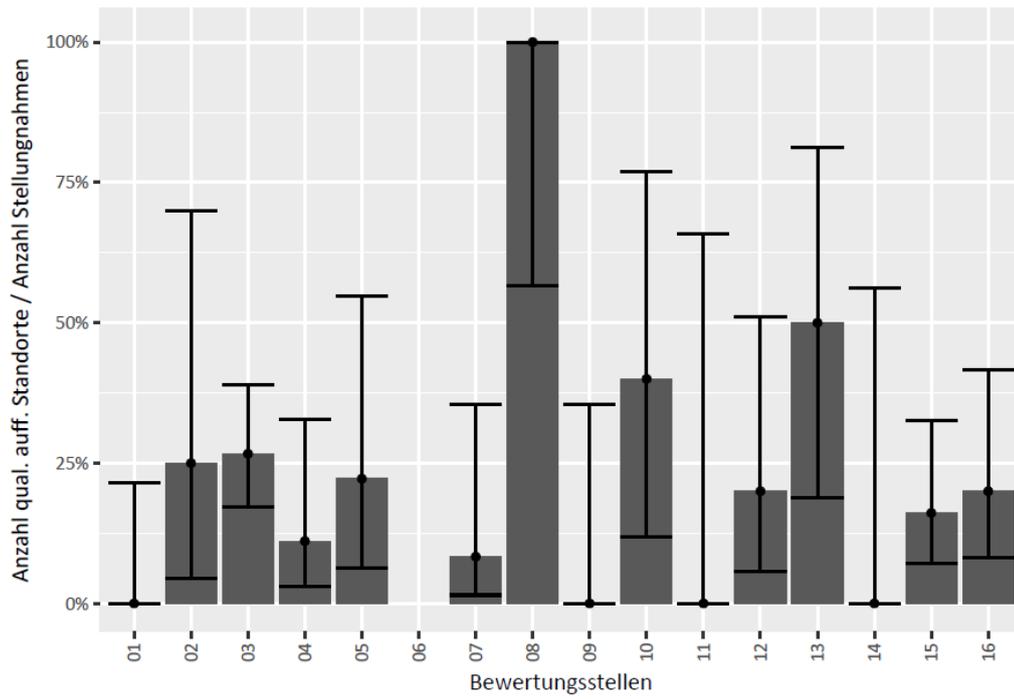


Abbildung 10: Qualitätsindikator 09/1: 54134 „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“. Bewertungen nach Einleitung des Stellungnahmeverfahrens

**Darstellung der Standortergebnisse nach Bewertungsstelle und Fallzahl**

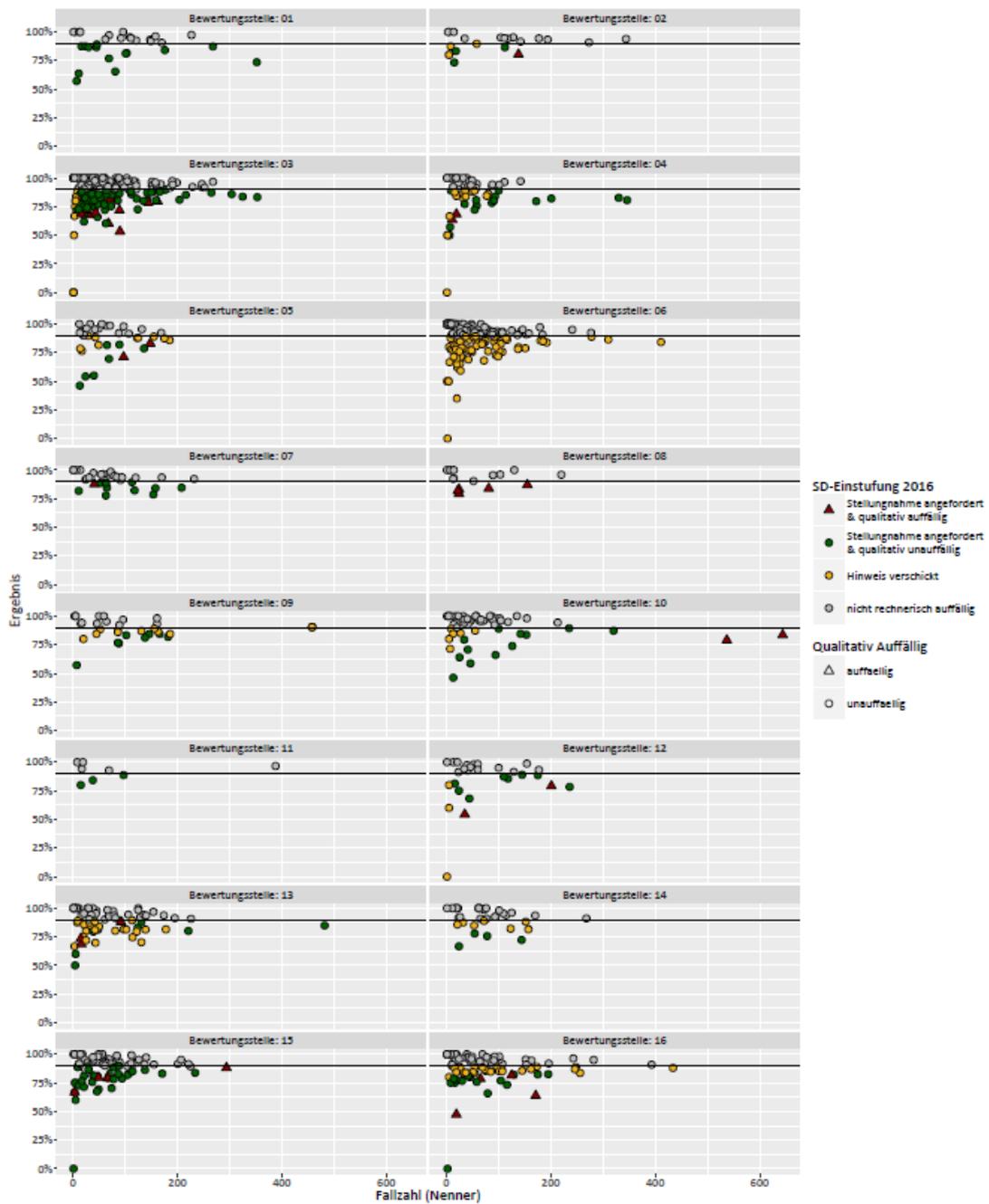


Abbildung 11: Ergebnisse des Strukturierten Dialogs zum Qualitätsindikator 09/1: 54134 „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“ für das Erfassungsjahr 2016 differenziert nach Fallzahl und Bewertungsstelle. Diese Daten sind die Grundlage für die modellbasierten Auswertungen in Abschnitt 2.2.

#### 4.5.3 DEK: QI 52010 Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4

##### Anteil an Stellungnahmen an allen rechnerisch auffälligen Standorten

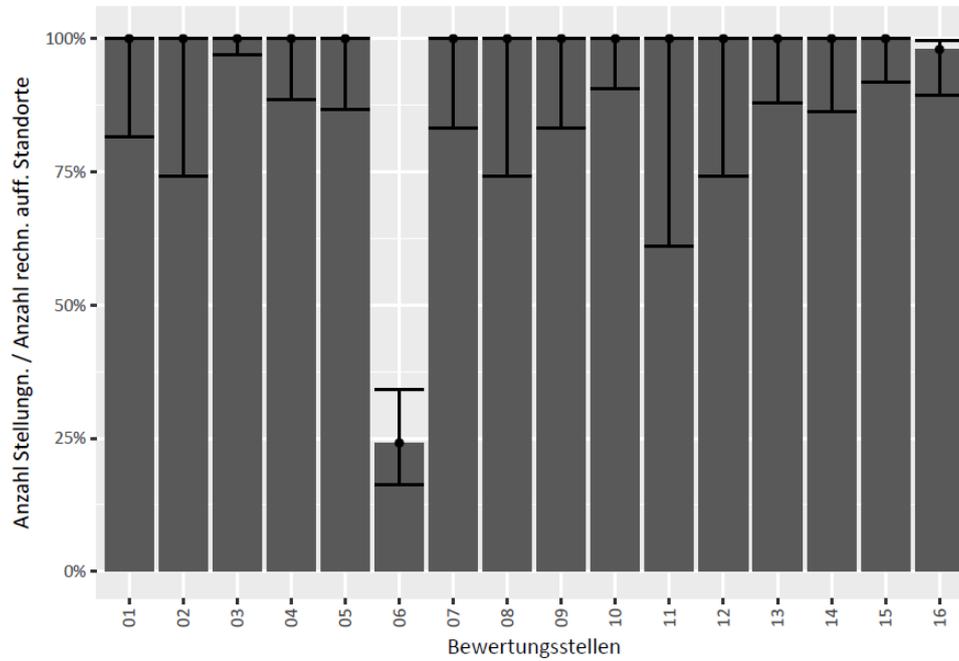


Abbildung 12: Qualitätsindikator DEK: 52010 „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“. Anfordern einer Stellungnahme nach rechnerischer Auffälligkeit

**Anteil an qualitativ auffälligen Bewertungen an allen Stellungnahmen**

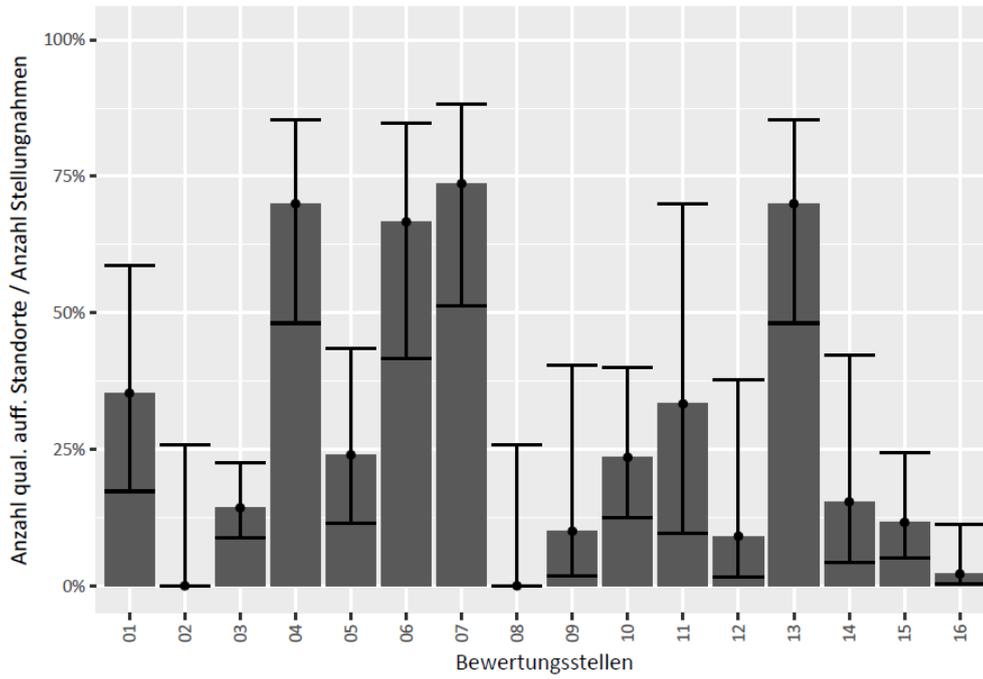


Abbildung 13: Qualitätsindikator DEK: 52010 „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“. Bewertungen nach Einleitung des Stellungnahmeverfahrens

**Darstellung der Standortergebnisse nach Bewertungsstelle und Fallzahl**

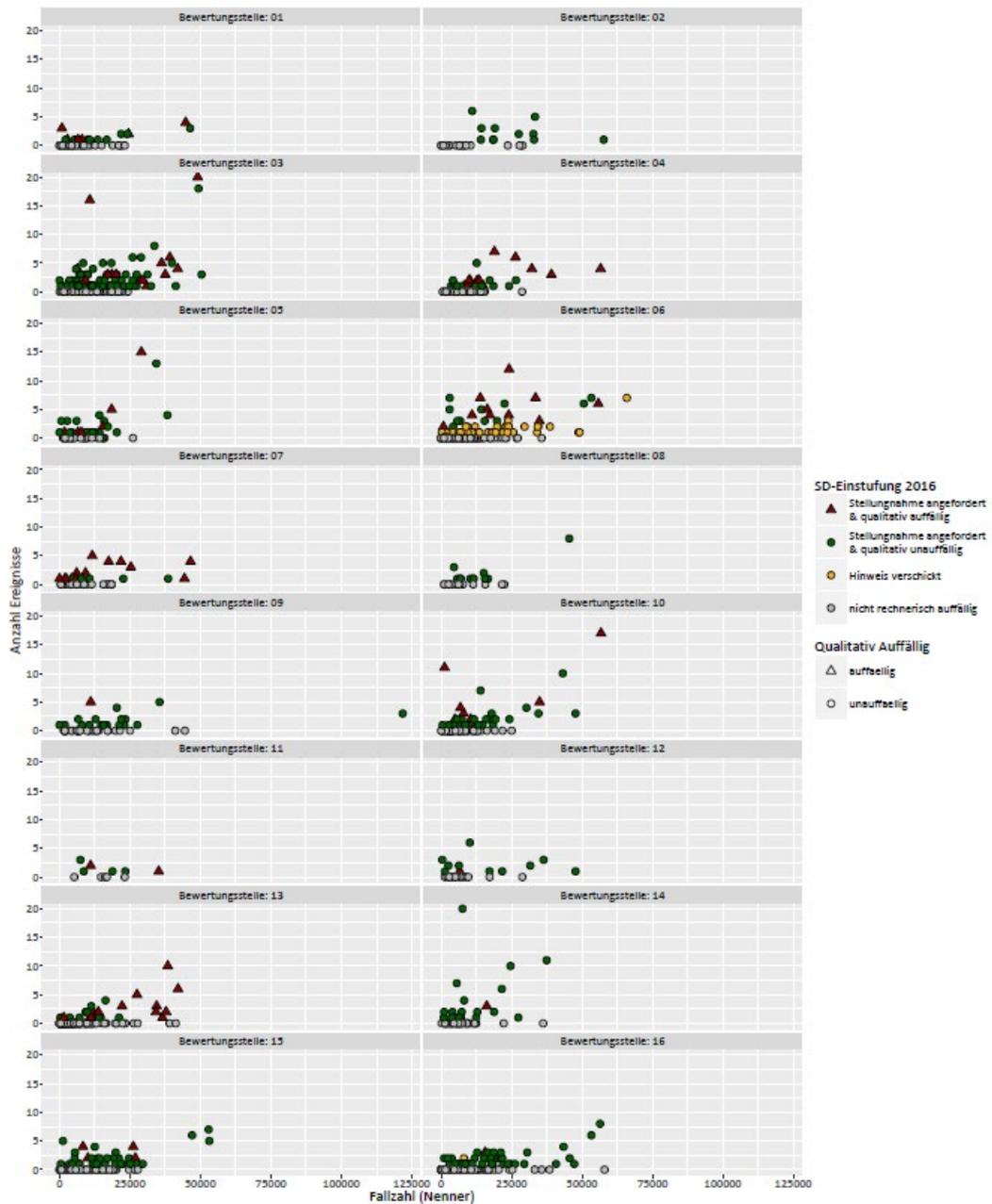


Abbildung 14: Ergebnisse des Strukturierten Dialogs zum Qualitätsindikator DEK: 52010 „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“ für das Erfassungsjahr 2016 differenziert nach Fallzahl und Bewertungsstelle. Diese Daten sind die Grundlage für die modellbasierten Auswertungen in Abschnitt 2.2.

## 5 Beispielfälle für die Qualitätsbewertung im Stellungnahmeverfahren

In seiner Stellungnahme kann ein Leistungserbringer Gründe anführen, die aus seiner Sicht das Verfehlen des Referenzbereichs erklären. Diese können sich auf einen oder mehrere Einflussfaktoren auf das Indikatorergebnis beziehen, die in der Berechnungsvorschrift des Indikators bisher nicht berücksichtigt sind. Im Folgenden werden fiktive Beispielfälle zum Stellungnahmeverfahren dargestellt, die veranschaulichen, wie diese Einflussfaktoren in der fachlichen Bewertung klassifiziert werden.

### 5.1 Fachlich-inhaltliche Einflussfaktoren

#### Beispiel 1: Kontraindikationen

Es wird ein Qualitätsindikator zur Thromboseprophylaxe genutzt, der abbilden soll, ob diese Behandlung in allen Fällen durchgeführt wurde, in denen sie angezeigt ist. In die Rechenregel des Indikators gehen Begleiterkrankungen der Patientinnen und Patienten nicht ein.

In einer Stellungnahme wird der Verzicht auf eine Thromboseprophylaxe bei einem Teil der Behandlungsfälle damit begründet, dass einige Patientinnen und Patienten ein erhöhtes Blutungsrisiko aufwiesen (z. B. wegen einer Thrombozytopenie).

#### *Analyse:*

Das Qualitätsziel des Indikators bezieht sich nicht auf Patientinnen und Patienten, die eine Kontraindikation zur Gabe von gerinnungshemmenden Medikamenten aufweisen. Die Verantwortung dafür, dass diese Kontraindikation gegeben ist, liegt auch nicht beim Leistungserbringer. Idealerweise wären Fälle mit erhöhtem Blutungsrisiko bei der Indikatorberechnung aus der Patientengrundgesamtheit des Indikators ausgeschlossen worden, da sie nicht zur Zielpopulation gehören. Der Einflussfaktor „Kontraindikation zur Thromboseprophylaxe“ kann bei der partiellen Nachberechnung in der fachlichen Bewertung berücksichtigt werden, indem geklärt wird, bei welchen der Behandlungsfälle dieser Einflussfaktor vorlag.

#### *Vorgehen:*

Diejenigen Fälle, bei denen die Fachkommission das Vorliegen einer „Kontraindikation zur Thromboseprophylaxe“ feststellt, werden bei der partiellen Nachberechnung aus der Grundgesamtheit ausgeschlossen.

#### Beispiel 2: Notfallbehandlung

Ein Qualitätsindikator beschreibt, ob bei einer diagnostischen Darmspiegelung (Koloskopie) der komplette Dickdarm beurteilt wurde (Erreichen des terminalen Ileums). In der Rechenregel des Indikators ist der Grund für die Untersuchung nicht berücksichtigt.

In einer Stellungnahme wird der Verzicht auf eine komplette Untersuchung des Dickdarms vom Leistungserbringer damit begründet, dass es sich bei einigen Patientinnen und Patienten um eine Notfalluntersuchung wegen einer akuten Blutung gehandelt habe.

*Analyse:*

Der Indikator ist nur auf elektive Untersuchungen anzuwenden. In akuten Notfallsituationen ist die Untersuchung des kompletten Dickdarms (in der Regel) kein Qualitätsmerkmal. Die Berücksichtigung bei der fachlichen Bewertung erfolgt dadurch, dass geklärt wird, bei welchen Fällen des Leistungserbringers der Einflussfaktor „akute Blutung als Indikation zur umgehenden Darmspiegelung“ Nachberechnung vorlag.

*Vorgehen:*

Diejenigen Fälle, bei denen LAG bzw. Bundesstelle und Fachkommission das Vorliegen einer „Notfalluntersuchung wegen akuter Blutung“ feststellen, werden bei der partiellen Nachberechnung aus der Grundgesamtheit ausgeschlossen.

**Beispiel 3: externe Verantwortlichkeit**

Ein Qualitätsindikator misst die In-Hospital-Sterblichkeit von Patientinnen und Patienten mit ambulant aufgetretenem akutem Herzinfarkt. Das Risikoadjustierungsmodell des Indikators berücksichtigt u. a. die wesentlichen Vorerkrankungen der Patientinnen und Patienten.

In einer Stellungnahme wird angeführt, dass aufgrund der regionalen Infrastruktur die Transportzeiten des Rettungsdienstes im Durchschnitt deutlich höher sind als bei anderen Einrichtungen. Als Beleg werden Studien zur mittleren Transportzeit angeführt und den eigenen, vom Leistungserbringer erfassten Daten gegenübergestellt.

*Analyse:*

Der Einflussfaktor „erhöhte Transportzeit des Rettungsdienstes“ ist vom Leistungserbringer nicht zu verantworten. Das Ausmaß, in dem die erhöhte Transportzeit zu schlechteren Ergebnissen (auch im Vergleich zu anderen Leistungserbringern) geführt hat, lässt sich nicht eindeutig angeben, wenn die Transportzeit nicht bei allen Leistungserbringern dokumentiert und für die Qualitätssicherung erfasst wird. Bei der fachlichen Bewertung wird dieser Einflussfaktor daher als „nicht für die partielle Nachberechnung geeignet“ klassifiziert.

*Vorgehen:*

Die „erhöhte Transportzeit des Rettungsdienstes“ wird nur im Rahmen der heuristischen Beurteilung durch die Fachkommission berücksichtigt. Kommt diese Beurteilung zum Schluss, dass bei den Fällen des betreffenden Leistungserbringers ein deutlich erhöhtes Risikoprofil in Bezug auf diesen Faktor im Vergleich zu den anderen Leistungserbringern vorliegt, kann dies als besondere Konstellation anerkannt werden.

**Beispiel 4: bekannter Risikofaktor**

Ein Qualitätsindikator beschreibt das Auftreten von Behandlungskomplikationen nach Implantation eines künstlichen Gelenks. Das Risikoadjustierungsmodell berücksichtigt die wichtigen Begleiterkrankungen, die zu einer erhöhten Häufigkeit von Komplikationen führen.

In seiner Stellungnahme weist ein Leistungserbringer auf ein erhöhtes Risikoprofil der von ihm behandelten Patientinnen und Patienten hin. In der Stellungnahme und den begleitenden Einzelfallbeschreibungen werden prolongierte, komplizierte Behandlungsverläufe geltend gemacht. Die darin genannten Risikofaktoren sind im Risikoadjustierungsmodell des Indikators berücksichtigt.

*Analyse:*

Es werden nur Einflussfaktoren genannt, die entweder vom Leistungserbringer zu verantworten sind oder bereits in der Berechnungsvorschrift des Indikators berücksichtigt sind. Der Einflussfaktor „erhöhtes Risikoprofil“ ist daher für eine partielle Nachberechnung nicht relevant. Im Rahmen der heuristischen Beurteilung durch die Fachkommission soll das erhöhte Risikoprofil ebenfalls nicht anerkannt werden, da im Vergleich zu den anderen Leistungserbringern nach Risikoadjustierung keine besondere Konstellation vorliegt.

*Vorgehen:*

Der Einflussfaktor „erhöhtes Risikoprofil“ wird nicht als Erklärung für das Verfehlen des Referenzbereichs anerkannt.

**Beispiel 5: strukturelle Sondersituation, planbar**

Ein Qualitätsindikator beschreibt Ergebnisse (z. B. perioperative Infektionen) oder durchzuführende Prozesse (z. B. postoperative Überwachung der Vitalfunktionen) in einem operativen Fachgebiet.

Die Stellungnahme eines Leistungserbringers verweist auf die Renovierung seines OP-Bereiches. Die operative Behandlung der Patientinnen und Patienten, die von dem Qualitätsindikator adressiert wird, fand vorübergehend in einem während der Renovierungsphase genutzten OP-Container statt.

*Analyse:*

Der Sachverhalt war planbar: Die Entscheidung, während der Renovierungsphase Patienten zu behandeln, impliziert die Verantwortlichkeit des Krankenhauses dafür, weiterhin eine ausreichende Qualität der Behandlung zu gewährleisten.

*Vorgehen:*

Der Einflussfaktor „Renovierung“ kann nicht als besondere Konstellation, die das Verfehlen des Referenzbereiches erklärt, geltend gemacht werden.

**Beispiel 6: unvorhergesehene strukturelle Situation**

Ein Qualitätsindikator beschreibt, ob eine gemäß Leitlinie geforderte radiologische CT-Untersuchung (z. B. zur Diagnostik einer Lungenembolie) durchgeführt wird.

In einer Stellungnahme wird geltend gemacht, dass die CT-Untersuchung bei einer relevanten Anzahl an Patientinnen und Patienten nicht durchgeführt werden konnte, da das Gerät unvorhersehbar ausgefallen war (z. B. durch einen technischen Defekt). Kurzfristige Versuche zur Behebung des Problems scheiterten.

*Analyse:*

Das Vorhalten eines zweiten Geräts kann nicht verlangt werden, ein Aufschub der Behandlung der Patientinnen und Patienten war wegen der Dringlichkeit der Versorgung nicht möglich. Für die Behandlungsfälle, die während des Ausfalls behandelt wurden, kann keine Qualitätsaussage getroffen werden, wenn die Situation nicht vom Leistungserbringer beeinflussbar war.

*Vorgehen:*

Der Einflussfaktor „unerwarteter Geräteausfall“ wird bei der partiellen Nachberechnung durch Ausschluss der betreffenden Behandlungsfälle berücksichtigt.

**Beispiel 7: Patientenpräferenz**

Ein Qualitätsindikator bildet ab, ob bei Patientinnen und Patienten mit Blutung aus dem Magen-Darm-Trakt die empfohlenen Untersuchungen zur Ermittlung der Blutungsquelle (endoskopische Diagnostik) durchgeführt werden.

In einer Stellungnahme wird der Verzicht auf die endoskopische Untersuchung bei einigen der Behandlungsfälle damit begründet, dass die Patientinnen und Patienten die Untersuchung abgelehnt haben. Es wird vom Leistungserbringer dargestellt, dass eine adäquate Aufklärung über Folgen der fehlenden Diagnostik erfolgt ist.

*Analyse:*

Die Verantwortung für den Einflussfaktor „Patientenpräferenz“ liegt nicht beim Leistungserbringer, sofern eine angemessene Aufklärung der Patientinnen und Patienten nachgewiesen ist. Im Rahmen der fachlichen Bewertung muss geklärt werden, für welche Behandlungsfälle eine entsprechende Aufklärung nachvollziehbar belegt ist.

*Vorgehen:*

Die Behandlungsfälle, bei denen LAG bzw. Bundesstelle und Fachkommission eine „Ablehnung der Behandlung durch die Patientinnen/Patienten trotz angemessener Aufklärung“ feststellen, werden bei der partiellen Nachberechnung berücksichtigt, indem sie aus der Grundgesamtheit des Indikators ausgeschlossen werden.

## 5.2 Datenfehler

### Beispiel 8: unentdeckte Softwarefehler

Ein Qualitätsindikator misst die In-Hospital-Sterblichkeit bei einer Herzklappenoperation. Das Risikoadjustierungsmodell berücksichtigt u. a. die Nierenfunktion, gemessen an der Kreatinin-Konzentration im Blutserum.

In einer Stellungnahme wird vorgebracht, dass durch einen Softwarefehler beim Export der QS-Daten die Kreatinin-Konzentrationen in einer falschen Einheit übermittelt wurden, ohne dass dies in der Benutzeroberfläche der QS-Software sichtbar war.

#### *Analyse:*

Wenn der Softwarefehler vom Leistungserbringer nicht bemerkt werden konnte, handelt es sich bei dem Datenfehler um einen nicht vom Leistungserbringer zu verantwortenden Faktor. Die Auswirkung auf das Indikatorergebnis lässt sich jedoch nicht durch einfachen Ausschluss der Behandlungsfälle im Rahmen der partiellen Nachberechnung angemessen beschreiben.

#### *Vorgehen:*

Ob die „Fehlmessung des Risikoprofils“ der betroffenen Behandlungsfälle so bedeutend war, dass sie das Verfehlen des Referenzbereichs erklärt, wird im heuristischen Beurteilungsschritt von LAG bzw. Bundesstelle und Fachkommission untersucht.

### Beispiel 9: Fehldokumentation

Ein Qualitätsindikator beschreibt das Auftreten von Komplikationen nach einem operativen Eingriff.

In seiner Stellungnahme macht der Leistungserbringer geltend, dass bei einem Teil der Behandlungsfälle tatsächlich keine Komplikationen vorgelegen hatten, sondern dass es sich um fehlerhafte Eingaben in der QS-Dokumentationssoftware handelt. Als Belege werden Auszüge der Behandlungsdokumentation (Patientenakte) beigefügt.

#### *Analyse:*

Soll das Ausmaß des vom Leistungserbringer zu verantwortenden Dokumentationsfehlers abgeschätzt werden, können bei einer Nachberechnung die betreffenden Behandlungsfälle als Fälle ohne Komplikation gezählt werden, z. B. durch Ausschluss aus dem Zähler des Indikators. Da die fehlerhaften Angaben in den QS-Daten jedoch durch Dokumentationsfehler des Leistungserbringers bedingt sind, handelt es sich um vom Leistungserbringer zu verantwortende Datenfehler. Eine Berücksichtigung im informativen Indikatorwert, der nur um Faktoren bereinigt ist, die nicht vom Leistungserbringer zu verantworten sind, ist nicht sachgemäß.

#### *Vorgehen:*

Bei der Ermittlung des informativen Indikatorwerts im Rahmen der partiellen Nachberechnung werden die Angaben zu den Komplikationen so berücksichtigt, wie sie ursprünglich dokumentiert waren. Die Dokumentationsfehler können nicht als besondere Konstellation, die den Hinweis auf ein Qualitätsdefizit entkräftet, geltend gemacht werden.

## 6 Textvorschläge Einholung Stellungnahmen

Für eine weitere Standardisierung im Vorgehen werden hier Textvorschläge für die Einholung einer Stellungnahme gemacht. Das IQTIG empfiehlt auch, aus Transparenzgründen die Methodik des Stellungnahmeverfahrens im Internet zu veröffentlichen und einen Link zu der Webseite in das Schreiben zur Einholung einer Stellungnahme einzufügen. Eine solche Transparenz ist einerseits aus wissenschaftlicher Sicht geboten und kann außerdem die Akzeptanz des Verfahrens steigern.

Wie aus dem Protokoll des Workshops mit den LAG- und LQS-Vertreterinnen und -Vertretern hervorgeht, nutzen einige Fachgruppen selbst entwickelte Fragebögen für die Durchführung des Stellungnahmeverfahrens oder des Strukturierten Dialogs (siehe Kapitel 1 in diesem Anhang). Da zum Zeitpunkt der Berichtslegung kein solcher Fragebogen vorliegt, kann nicht eingeschätzt werden, ob diese für die Prüfung auf das Vorliegen von nicht vom Leistungserbringer zu vertretene Einflussfaktoren geeignet sind.

Sehr geehrte Damen und Herren,

auf Basis der von Ihnen übermittelten Daten wurde für Ihre Einrichtung das Ergebnis für den Qualitätsindikator [NN] berechnet. Das Ergebnis Ihrer Einrichtung liegt statistisch signifikant außerhalb des Referenzbereichs. Dies bedeutet, dass der Indikator auf ein Qualitätsdefizit in der Versorgung hinweist. Sie erhalten hiermit die Möglichkeit, zu diesem Indikatorergebnis Stellung zu nehmen. Wenn Sie Gründe anführen möchten, die das Ergebnis erklären und nicht von Ihrer Einrichtung zu verantworten waren, können Sie diese innerhalb von fünf Wochen nach Zugang dieses Schreibens durch eine qualifizierte Stellungnahme nach untenstehenden Kriterien geltend machen. Ihre Stellungnahme wird dann einer Fachkommission vorgelegt, die bewerten wird, ob die von Ihnen angeführten Gründe anerkannt und das Indikatorergebnis mit „kein Hinweis auf Qualitätsdefizit“ eingestuft werden kann. Andernfalls wird das Indikatorergebnis als „Qualitätsdefizit“ bewertet. Erreicht uns innerhalb der Frist keine den untenstehenden Kriterien entsprechende Stellungnahme, wird das Ergebnis Ihrer Einrichtung als „Qualitätsdefizit“ bewertet.

### Neues Stellungnahmeverfahren

Bitte beachten Sie, dass sich das Vorgehen bei der Bewertung und im Stellungnahmeverfahren für diesen Qualitätsindikator aus der DeQS-RL von dem Ihnen bisher bekannten Vorgehen gemäß der QSKH-RL unterscheidet. Die Methodik für die Bewertung von Stellungnahmen und die abschließende Qualitätsbewertung können Sie hier nachlesen: [[Link zur Veröffentlichung der Methodik einfügen](#)]

### Formale Kriterien

Bitte beachten Sie beim Verfassen Ihrer Stellungnahme folgende Hinweise:

- Für jedes auffällige Indikatorergebnis ist eine separate Stellungnahme abzugeben.

- Die Stellungnahme ist einschließlich eventuell notwendiger Anhänge in Schriftform zu erstellen und einzureichen.
- Wenn Sie Belege haben, die Ihre Argumentation unterstützen, können Sie diese in angemessenem Umfang einreichen. Dazu zählen beispielsweise Einzelfalldokumente wie Arztbriefe, Epikrisen, Verlegungsberichte, Operationsberichte und Befunde. Belege können jedoch nur als Ergänzung des Erläuterungstextes berücksichtigt werden und ersetzen diesen nicht. Belege ohne erläuternden, übergreifenden Text (s. u.) werden nicht anerkannt.
- Die fachlich verantwortliche Instanz Ihrer Einrichtung ist über den indikatorbasierten Hinweis auf ein Qualitätsdefizit in Ihrer Einrichtung zu informieren und hat die angefertigte Stellungnahme zu autorisieren. Bitte bestätigen Sie in Ihrer Stellungnahme, dass dies geschehen ist.
- Bitte wahren Sie in Ihrer Stellungnahme die Anonymität sowohl Ihrer Patientinnen und Patienten (personenidentifizierende Daten) und Ihrer Einrichtung (einrichtungsidifizierende Daten) als auch ggf. weiterer Beteiligter. Liegt eine Verletzung des Datenschutzes vor, wird die Stellungnahme umgehend gelöscht und gilt als nicht abgegeben.
- Bitte beantworten Sie am Anfang Ihrer Stellungnahme folgende Frage: Sind Ihnen Fehler in den von Ihnen übermittelten Daten bekannt? Wenn ja, nennen Sie bitte den/die Fehler sowie – sofern zutreffend – die Gründe, warum der/die Fehler in der Datengrundlage nicht von Ihrer Einrichtung zu verantworten war/en.

#### **Inhaltliche Kriterien (für jeden Qualitätsindikator)**

Neben den obigen formalen Kriterien möchten wir Sie bitten, folgende inhaltliche Anforderung an Ihre Stellungnahme zu berücksichtigen. Wenn Sie eine Stellungnahme abgeben, soll diese eine Analyse enthalten, die erklärt, weshalb das Qualitätsindikatorergebnis vom Referenzbereich abweicht, obwohl Ihrer Meinung nach die Versorgungsqualität in Ihrer Einrichtung gut ist. In Ihrer Stellungnahme können zum einen Gründe angegeben werden, die sich nur auf einzelne oder weniger Behandlungsfälle beziehen (z. B. dass ein Notfall ein Abweichen vom üblichen Vorgehen erforderlich machte), und zum anderen Gründe, die alle Behandlungsfälle ihrer Einrichtung für diesen Indikator betreffen (z. B. dass wegen einer Spezialisierung Ihrer Einrichtung auf bestimmte Behandlungen die Patientinnen und Patienten ein besonderes Risikoprofil haben). Wenn möglich, sollte Ihre Stellungnahme jedoch auf eine behandlungsfallübergreifende Analyse fokussieren, um systemische Ursachen und ggf. systemische Verbesserungsansätze darzustellen.

- **Bei einem quantitativ auffälligen Ergebnis in einem Prozessindikator:**<sup>9</sup>  
Bitte schildern Sie in Ihrer Stellungnahme, warum die erforderlichen Prozessstandards (z. B. Durchführung, Anwendung oder Nicht-Durchführung einer Maßnahme, zeitliche Anforderungen, Dosierungen) nicht in dem Maße eingehalten wurden, wie durch den Referenzbereich des Qualitätsindikators gefordert wird, und warum dies dennoch nicht als Qualitätsproblem einzustufen ist.
- **Bei einem quantitativ auffälligen Ergebnis in einem risikoadjustierten Prozessindikator:**

---

<sup>9</sup> Bei der Erstellung der Anschreiben soll je nach betroffenem Qualitätsindikator nur der jeweils passende der hier aufgelisteten Textblöcke verwendet werden.

Da bei diesem Indikator patientenseitige Faktoren den Behandlungsprozess beeinflussen können, wird dieser Indikator risikoadjustiert. Bitte schildern Sie in Ihrer Stellungnahme, warum die erforderlichen Prozessstandards nicht eingehalten wurden und warum dies trotz Berücksichtigung patientenseitiger Faktoren in dem Indikator nicht als Qualitätsproblem einzustufen ist.

▪ **Bei einem quantitativ auffälligen Ergebnis in einem Indikationsindikator:**

Bitte schildern Sie in Ihrer Stellungnahme, warum die Prozedur häufiger durchgeführt wurde, als dies nach der Rechenvorschrift des Qualitätsindikators indiziert war, respektive warum die Prozedur trotz bestehender Indikation laut Qualitätsindikator nicht häufig genug durchgeführt wurde, und warum dies dennoch nicht als Qualitätsdefizit einzustufen ist.

▪ **Bei einem quantitativ auffälligen Ergebnis in einem risikoadjustierten Ergebnisindikator:**

Bei der Durchführung medizinischer Behandlungen sollen Komplikationen jeglicher Art und Schwere möglichst selten auftreten und Therapieziele möglichst häufig erreicht werden. Allerdings gibt es Umstände, in denen Komplikationen nicht gänzlich vermeidbar sind und Therapieziele nicht vollständig erreicht werden können. Um dies angemessen zu berücksichtigen, wird dieser Qualitätsindikator risikoadjustiert. Bitte schildern Sie in Ihrer Analyse, warum das Ergebnis Ihrer Einrichtung trotz Risikoadjustierung außerhalb des Referenzbereichs liegt und welche/r Einflussfaktor/en ggf. bei Ihren Patientinnen und Patienten so häufig auftraten, dass diese nicht mehr adäquat durch die Risikoadjustierung berücksichtigt wurde/n.

▪ **Bei einem quantitativ auffälligen Ergebnis in einem nicht risikoadjustierten Ergebnisindikator:**

Bei der Durchführung medizinischer Behandlungen sollen Komplikationen jeglicher Art und Schwere möglichst selten auftreten und Therapieziele möglichst häufig erreicht werden. Allerdings gibt es Umstände, in denen sie nicht gänzlich vermeidbar sind und Therapieziele nicht vollständig erreicht werden können. Bitte schildern Sie in Ihrer Analyse, welche Einflussfaktoren bei Ihren Patientinnen und Patienten häufiger auftraten als im Bundesdurchschnitt und warum das Verfehlen des Referenzbereichs daher nicht als Qualitätsproblem einzustufen ist.

Werden organisatorische Gründe für das quantitativ auffällige Indikatorergebnis benannt, ist nachzuweisen, dass Ihre Einrichtung mit der gebotenen Sorgfaltspflicht alles Zumutbare getan hat, um diese organisatorischen Mängel aufzufangen (z. B. die Einplanung von genügend Ressourcen und genügend Zeit).

**Angabe von Vorgangsnummern zu geltend gemachten Einflussfaktoren**

Wenn Sie in Ihrer Stellungnahme einen nicht von Ihrer Einrichtung zu verantwortenden Einflussfaktor geltend machen möchten, der in den Daten der Qualitätssicherung bisher nicht dokumentiert ist, geben Sie bitte in Ihrer Stellungnahme für alle Behandlungsfälle Ihrer Einrichtung zum jeweiligen Indikator an, ob dieser Einflussfaktor vorliegt oder nicht. Dazu sind die entsprechenden Vorgangsnummern mit der Stellungnahme zu übermitteln. Diese Angabe ist auch für diejenigen Behandlungsfälle zu machen, bei denen kein interessierendes Ereignis aufgetreten ist, d. h. die nicht im Zähler des Indikators erschienen sind. Eine detaillierte Einzelfalldarstellung aller Behandlungsfälle ist nicht erforderlich, sofern aus Ihrer Stellungnahme nachvollziehbar ist, warum der angeführte Einflussfaktor bei den entsprechenden Behandlungsfällen vorliegt.

Auf die Angabe der Vorgangsnummern für einen Einflussfaktor kann verzichtet werden, wenn anhand Ihrer Stellungnahme nachvollziehbar ist, dass dieser Einflussfaktor bei allen Fällen des Qualitätsindikators vorlag.

**Information über das Vorgehen bei Nichteinhaltung der formalen Kriterien**

Bei Nichteinhaltung der formalen Kriterien gilt die Stellungnahme als nicht abgegeben. Bei Nichteinhaltung der Vorgaben zur Wahrung der Anonymität Ihrer Patientinnen und Patienten, Ihrer Einrichtung oder weiterer Beteiligter wird die Stellungnahme darüber hinaus aus Datenschutzgründen vernichtet. Sie werden über diesen Vorgang benachrichtigt.

## 7 Beispiel zur Übermittlung von Vorgangsnummern in Stellungnahmen

Möchte ein Leistungserbringer in seiner Stellungnahme zu einem Indikatorergebnis geltend machen, dass ein oder mehrere nicht von ihm zu verantwortende Faktoren das Indikatorergebnis beeinflusst haben, so müssen die Vorgangsnummern der betreffenden Behandlungsfälle übermittelt werden (siehe Abschnitt 6.5.2 im Abschlussbericht). Dies wird im Folgenden an einem fiktiven Beispiel veranschaulicht<sup>10</sup>.

Als Beispiel wird ein Qualitätsindikator angenommen, der den Anteil der Patientinnen und Patienten misst, die bei einem ambulant aufgetretenen akuten Herzinfarkt innerhalb von 2 Stunden eine Koronarangiographie erhalten haben. Ein Leistungserbringer macht in seiner Stellungnahme folgende 4 Einflussfaktoren auf sein Indikatorergebnis geltend:

- Faktor 1: ein nicht vorhersehbarer Geräteausfall im einzigen Katheterlabor der Einrichtung
- Faktor 2: Ablehnung der Koronarangiographie durch die Patientin/den Patienten
- Faktor 3: ein fälschlich übermittelter Datensatz, der nicht zur Indikatorgrundgesamtheit gehört, aber wegen einer fehlerhaften Implementation des QS-Filters in der Software ausgelöst und dokumentiert wurde
- Faktor 4: lange Anfahrtszeit des Rettungsdienstes wegen geographischer Lage

Die folgende Tabelle 11 zeigt beispielhaft, in welcher Form die Übermittlung der Vorgangsnummern durch den Leistungserbringer erfolgen sollte. Dabei wird empfohlen, eine Übermittlung in digitaler Form zu etablieren, um Fehler bei der Weiterverarbeitung der Informationen zu vermeiden.

Tabelle 11: Fiktives Beispiel für die Angabe von Vorgangsnummern zur Stellungnahme zu einem Indikator

Vorgangsnummer des Behandlungsfalls	Faktor 1: Geräteausfall	Faktor 2: kein Therapiewunsch	Faktor 3: fälschlicher Datenexport wegen fehlerhafter QS-Filter-Implementation
12	1	1	
13	1		
14	1	1	
208			1
456		1	
77		1	

<sup>10</sup> Das Beispiel erhebt keinen Anspruch auf Korrektheit der medizinischen Inhalte oder der Beurteilung der Stellungnahme, sondern dient der Verdeutlichung des grundsätzlichen Vorgehens.

Vorgangsnummer des Behandlungsfalls	Faktor 1: Geräteausfall	Faktor 2: kein Therapiewunsch	Faktor 3: fälschlicher Datenexport wegen fehlerhafter QS-Filter-Implementation
79			1
558			1

In dem gezeigten Beispiel liegen aus Sicht des Leistungserbringers drei nicht von ihm zu verantwortende Einflussfaktoren vor, die sich nur auf einen Teil der Behandlungsfälle beziehen. Das Vorliegen des Faktors wird durch jeweils durch die 1 gekennzeichnet. Zu beachten ist, dass in einem Behandlungsfall ggf. mehrere Einflussfaktoren vorliegen können. Bei Vorgangsnummern, die in der Tabelle nicht aufgeführt sind, wird davon ausgegangen, dass keiner der in der Tabelle genannten Einflussfaktoren bei dem betreffenden Behandlungsfall vorlag. Zusätzlich sei für das Beispiel angenommen, dass der Leistungserbringer für den vierten Faktor „lange Anfahrtszeit des Rettungsdienstes wegen geographischer Lage“ in seiner Stellungnahme erläutert hat, dass dieser auf alle Behandlungsfälle zutrefte, und dass für diesen Faktor daher keine Vorgangsnummern übermittelt wurden.

## 8 Beispiel zur Berücksichtigung mehrerer Einflussfaktoren bei der partiellen Nachberechnung

Die partielle Nachberechnung im Rahmen der fachlichen Bewertung basiert auf der Kategorisierung der in einer Stellungnahme angeführten Einflussfaktoren hinsichtlich Verantwortlichkeit, Art des Einflussfaktors (fachlich-inhaltlich vs. Datenfehler) und Auswirkung auf das Indikatorergebnis (siehe Abschnitt 6.4.1 im Abschlussbericht). Am Beispiel der übermittelten Einflussfaktoren aus Kapitel 7 wird im Folgenden dargestellt, wie der informative Indikatorwert berechnet werden kann.

Für das Beispiel sei von folgenden Einschätzungen ausgegangen:

- Die Fachkommission erkennt alle vier in der Stellungnahme genannten Faktoren als nicht vom Leistungserbringer zu verantwortende Einflussfaktoren an.
- Die Faktoren 1, 2 und 4 werden als fachlich-inhaltliche Faktoren kategorisiert, der Faktoren 3 als Datenfehler.
- Die Auswirkung der Faktoren 1, 2 und 3 wird so eingeschätzt, dass die betreffenden Behandlungsfälle nicht in die Indikatorberechnung hätten eingehen sollen. Die Auswirkung von Faktor 4 auf das Indikatorergebnis wird als graduell bewertet, d. h. die betreffenden Fälle können nicht aus der Grundgesamtheit des Indikators ausgeschlossen werden.

Außerdem sei davon ausgegangen, dass bei der Berechnung des originalen Indikatorwerts die Behandlungsfälle mit den Vorgangsnummern 12, 13, 208, 77 und 79 ein interessierendes Ereignis aufgewiesen haben, sowie 5 weitere Fälle, die in der Tabelle der Stellungnahme (Tabelle 11) nicht aufgeführt sind (Vorgangsnummern 1 bis 5). Das bedeutet, dass der Leistungserbringer für letztere 5 Fälle mit interessierendem Ereignis keinen besonderen Einflussfaktor geltend gemacht, und des Weiteren, dass bei den Fällen mit den Vorgangsnummern 14, 456 und 558 zwar mindestens einer der geltend gemachten Einflussfaktoren vorgelegen hat, aber trotzdem kein interessierendes Ereignis aufgetreten ist. Tabelle 12 visualisiert die Berücksichtigung der Einflussfaktoren bei der partiellen Neuberechnung:

Tabelle 12: Fiktives Beispiel für die fallbezogene Berücksichtigung von Einflussfaktoren bei der partiellen Neuberechnung

Vorgangsnummer des Behandlungsfalls	Faktor 1: Geräteausfall	Faktor 2: kein Therapiewunsch	Faktor 3: fälschlicher Datenexport wegen fehlerhafter QS-Filter-Implementation	interessierendes Ereignis	partielle Nachberechnung	
					aus Nenner auszu-schließen	aus Zähler auszu-schließen
12	1	1		1	ja	ja

Vorgangsnummer des Behandlungsfalls	Faktor 1: Geräteausfall	Faktor 2: kein Therapiewunsch	Faktor 3: fälschlicher Datenexport wegen fehlerhafter QS-Filter-Implementation	interessierendes Ereignis	partielle Nachberechnung	
					aus Nenner auszuschließen	aus Zähler auszuschließen
13	1			1	ja	ja
14	1	1			ja	
208			1	1	ja	ja
456		1			ja	
77		1		1	ja	ja
79		1		1	ja	ja
558		1			ja	
1				1		
2				1		
3				1		
4				1		
5				1		

Aus diesen Beispielannahmen ergibt sich für die partielle Nachberechnung, dass 8 Fälle aus der Grundgesamtheit des Indikators (Nenner) ausgeschlossen werden, da  $a = 7$  Fälle aus fachlich-inhaltlichen Gründen (Faktoren 1 oder 2) und 1 Fall wegen eines Datenfehlers (Faktor 3) nicht für die Erfüllung des Qualitätsmerkmals des Indikators von Interesse sind. Der Zähler des Indikators wird um die Anzahl derjenigen aus dem Nenner auszuschließenden Behandlungsfälle vermindert, die auch ein Zählerereignis aufweisen ( $o_a = 5$  Fälle). Faktor 4 kann nicht in die partielle Nachberechnung eingehen, da sein Beitrag zum Indikatorergebnis nicht eindeutig anzugeben ist, und muss deshalb im heuristischen Beurteilungsschritt bewertet werden.

## 9 Berechnungsvorschrift für die partielle Nachberechnung

### 9.1 Berechnung des informativen Indikatorwerts

Im Folgenden wird die Berechnung des informativen Indikatorwerts für anteilsbasierte und risikoadjustierte Indikatoren beschrieben.

#### Anteilsbasierte Qualitätsindikatoren

Betrachtet man den einfachen Fall eines anteilsbasierten Qualitätsindikators, so liegt im betrachteten Zeitraum ein Indikatorergebnis vor mit  $J$  Fällen des Leistungserbringers in der Grundgesamtheit und  $o$  interessierenden Vorgängen im Zähler des Indikators. Ohne Beschränkung der Allgemeinheit sei hier angenommen, dass der optimale Indikatorwert bei  $o = 0$  also 0 % liegt.

Es bezeichne  $R$  den Referenzbereich des Indikators. Nach der quantitativen Bewertungsvorschrift  $s(o, J, R)$  (vgl. Abschlussbericht, Kapitel 5) wurde der Leistungserbringer als quantitativ auffällig bewertet.

Die fachliche Bewertung ergibt, dass eine gewisse Anzahl  $a$  von Fällen nicht in der Grundgesamtheit des Indikators hätte berücksichtigt werden sollen. Von diesen  $a$  Fällen hatten  $o_a$  Fälle ein interessierendes Ereignis. Unter den restlichen  $J - a$  Fällen wurde bei  $b$  Fällen festgestellt, dass für diese kein interessierendes Ereignis zu zählen ist (d. h. sie nicht in den Zähler des Indikators einfließen sollen).

Für die Beurteilung des partiell nachberechneten Indikatorwertes (informativer Indikatorwert) wird ein bereinigter Referenzbereich benötigt (siehe Abschlussbericht, Abschnitt 6.4). Dieser wird im Folgenden mit  $\tilde{R}$  bezeichnet.

Der Leistungserbringer wird nun nach der quantitativen Bewertungsvorschrift  $s(o - o_a - b, J - a, \tilde{R})$  erneut bewertet. In einem rein enumerativen Bewertungsschema (aktuelle „rechnerische Auffälligkeit“ der QSKH-Richtlinie) wäre die Entscheidungsregel

$$\frac{o - o_a - b}{J - a} > \tilde{R}.$$

#### Risikoadjustierte Qualitätsindikatoren

Im Falle risikoadjustierter Qualitätsindikatoren beinhaltet die quantitative Bewertung auch die Berücksichtigung der geschätzten Wahrscheinlichkeiten  $e_j$  für alle Patienten  $j = 1, \dots, J$ . Die quantitative Bewertung ist dann eine Funktion  $s(o, e, R)$ , wobei  $e = \sum_{j=1}^J e_j$ . Die fachliche Bewertung ergibt, dass eine gewisse Anzahl  $a$  von Fällen nicht in der Grundgesamtheit des Indikators hätte berücksichtigt werden sollen. Von diesen  $a$  Fällen hatten  $o_a$  Fälle ein interessierendes Ereignis. Unter den restlichen  $J - a$  Fällen wurde bei  $b$  Fällen festgestellt, dass für diese kein interessierendes Ereignis zu zählen ist. Die erneute quantitative Bewertung erfolgt dann nach

der Regel  $s(o - o_a - b, \sum_{j=1}^{J-a} e_j, \tilde{R})$ . An Stelle von  $e$  fließt somit nur noch die Summe der geschätzten Wahrscheinlichkeiten aller verbleibenden Fälle ein<sup>11</sup>. In einem rein enumerativen Bewertungsschema wäre die Entscheidungsregel dann:

$$\frac{o - o_a - b}{\sum_{j=1}^{J-a} e_j} > \tilde{R}.$$

Zu beachten ist, dass für diese Form der Nachberechnung die geschätzten Wahrscheinlichkeiten  $e_j$  für alle Behandlungsfälle des Leistungserbringers zum Zeitpunkt der Nachberechnung zur Verfügung gestellt werden müssen.

### **Abwägung von analytischer und enumerativer Herangehensweise der Nachberechnung**

Gemäß Abschlussbericht, Abschnitt 5.1.1 („Analytische Herangehensweise“) ist in der quantitativen Bewertung von Qualitätsindikatoren abzuwägen, in welchen Fällen eine enumerative bzw. eine analytische Herangehensweise an die Auswertung verfolgt wird. Für die quantitative Bewertung von anteilsbasierten und risikoadjustierten Indikatoren wird die analytische Herangehensweise empfohlen. Erstens wird diese durch in der Datenerhebung nicht quantifizierte Einflussfaktoren begründet, die eine Berücksichtigung statistischer Unsicherheit in der Einstufung erforderlich machen. Zweitens wird die analytische Herangehensweise gewählt, da im methodischen Verständnis des IQTIG ein Qualitätsindikator eine Messung der Kompetenz eines Leistungserbringers in einem speziellen Qualitätsmerkmal darstellt. Zur Messung dieser Kompetenz werden i. d. R. alle durch den QS-Filter spezifizierten Fälle eines Erfassungsjahres herangezogen. Zwar handelt es sich um eine Vollerhebung im Sinne des QS-Filters, jedoch im Sinne der Kompetenzmessung handelt es sich weiterhin um eine Stichprobe von Behandlungsfällen.

Unter der Annahme, dass durch die Stellungnahme und die fachliche Bewertung die nicht quantifizierten Einflussfaktoren sichtbar gemacht werden, stellt sich die Frage, inwiefern nach der Nachberechnung weiterhin eine analytische Herangehensweise gerechtfertigt ist. Stochastizität wird allerdings auch durch das Stellungnahmeverfahren nicht entfernt, da auch nach Entfernung der nun bekannten, gerichteten Einflussfaktoren noch unbekannte oder zufällige Effekte bleiben können. Des Weiteren beruht das Indikatorergebnis weiterhin nur auf einer Stichprobe von Behandlungsfällen, die zur Messung der Kompetenz des Leistungserbringers für das jeweilige Qualitätsmerkmal herangezogen werden. Für die Bewertung im Rahmen der partiellen Nachberechnung wird daher ebenfalls die analytische Herangehensweise gewählt.

## **9.2 Berechnung des bereinigten Referenzbereiches**

Es bezeichne  $\tilde{R}$  den bereinigten Referenzbereich des Indikators. Dieser wird berechnet, indem der originale Referenzbereich  $R$  in zwei Anteile aufgeteilt wird (siehe Abschlussbericht, Abschnitt 6.4.2): Dabei bezeichne  $R_1$  den Anteil gerichteter, nicht vom Leistungserbringer zu beeinflussender Faktoren auf die Indikatorwerte und  $R_2$  das tolerierte Qualitätsdefizit, wobei gilt:

<sup>11</sup> Es wird davon ausgegangen, dass die Wahrscheinlichkeiten  $e_j$  so nummeriert sind, dass die letzten  $J - a$  Fälle genau die aus der Grundgesamtheit auszuschließenden Fälle darstellen.

$$R = R_1 + R_2$$

Der bereinigte Referenzbereich berechnet sich dann als

$$\tilde{R} = \frac{R_2}{1 - R_1}$$

Die Korrektur im Nenner (gegenüber einer einfachen Setzung von  $\tilde{R} = R_2$ ) begründet sich dadurch, dass sich der Toleranzanteil  $R_2$  im originalen Referenzbereich nicht auf alle Behandlungsfälle der Grundgesamtheit des Indikators bezieht: Die Toleranz ist auf die Behandlungsfälle bezogen, deren Versorgungsqualität beurteilt werden soll, und damit nicht auf den Anteil  $R_1$  der Behandlungsfälle, der in der Grundgesamtheit des Indikators enthalten ist, weil ein Teil der vom Leistungserbringer nicht zu verantwortenden Faktoren nicht in der Berechnung berücksichtigt wird. Der Anteil der eigentlich für die Qualitätsbeurteilung interessierenden Behandlungsfälle beträgt somit  $1 - R_1$ . Durch den Korrekturfaktor wird sichergestellt, dass die relative Höhe des tolerierten Qualitätsdefizits auch nach der partiellen Nachberechnung gleichbleibt.

### 9.3 Retrospektive Schätzung der in Referenzbereichen enthaltenen Toleranz für Qualitätsdefizite

Für die Berechnung des bereinigten Referenzbereiches muss für den originalen Referenzbereich die Größe des Anteils gerichteter, nicht vom Leistungserbringer zu beeinflussender Faktoren einerseits und des Anteils tolerierter Qualitätsdefizite andererseits bekannt sein. Diese Größen wurden für die bisher in der externen Qualitätssicherung eingesetzten Referenzbereiche nicht gesondert ausgewiesen. Im Folgenden ist ein Vorschlag beschrieben, wie sie retrospektiv geschätzt werden können, damit sie für die Berechnung des bereinigten Referenzbereichs verwendet werden können.

Es wird empfohlen, die Differenz zwischen dem Idealwert des Indikators (0 % bzw. 100 %) und dem empirischen bundesweiten Wert (der Wert, der bei Berechnung des Indikators auf Grundlage aller Behandlungsfälle bundesweit resultiert) als Näherung für den Anteil gerichteter Einflussfaktoren zu verwenden (höchstens aber die Breite des Referenzbereichs). Dieses Vorgehen folgt der Überlegung, dass der Bundeswert einerseits per Definition auf dem durchschnittlichen Risikoprofil der Patientinnen und Patienten basiert, dass also alle gerichteten Einflussfaktoren mit ihrem mittleren Ausmaß in diesen Wert eingehen. Andererseits kann aus der Tatsache, dass der Indikator Qualitätsverbesserungen bewirken soll, abgeleitet werden, dass die bundesweite Versorgungsqualität nicht optimal ist. Für den Abstand des empirischen Bundeswertes vom Idealwert kann also angenommen werden, dass er mindestens so groß ist wie der mittlere Einfluss der gerichteten Faktoren auf den Indikatorwert<sup>12</sup>, und dass er zusätzlich einen Anteil enthält, der einer nicht optimalen bundesweiten Versorgungsqualität geschuldet ist

<sup>12</sup> Da es sich beim bundesweiten Wert um eine empirisch bestimmte Größe handelt, könnte es zufallsbedingt dazu kommen, dass diese Annahme nicht zutrifft. Die bundesweiten Fallzahlen werden jedoch als ausreichend groß angesehen, um diese Möglichkeit vernachlässigen zu können.

Es bezeichne  $R$  den originalen Referenzbereich des Indikators und  $\hat{\Pi}$  den empirischen Bundeswert des Indikators. Ohne Beschränkung der Allgemeinheit sei hier angenommen, dass der optimale Indikatorwert bei  $o = 0$  also 0 % liegt. Der Anteil gerichteter, nicht vom Leistungserbringer zu beeinflussender Faktoren  $R_1$  am originalen Referenzbereich wird dann geschätzt als

$$R_1 = \hat{\Pi}$$

falls  $R > \hat{\Pi}$ . Andernfalls wird  $R_1 = R$  gesetzt.

Der Anteil der tolerierten Qualitätsabweichung  $R_2$  am originalen Referenzbereich ergibt sich dann als Differenz aus dem originalen Referenzbereich und dem anhand des empirischen Bundeswerts geschätzten Anteils gerichteter Einflussfaktoren:

$$\begin{aligned} R_2 &= R - R_1 \\ &= R - \hat{\Pi} \end{aligned}$$

falls  $R > \hat{\Pi}$ . Andernfalls  $R_2 = 0$ .

Für risikoadjustierte Qualitätsindikatoren z. B. in O/E-Darstellung kann das beschriebene Vorgehen sinngemäß angewendet werden.

### **Auswirkungen des Schätzverfahrens auf die Qualitätsbewertung**

Bei der vorgeschlagenen Schätzung anhand des empirischen Bundeswertes wird der Anteil  $R_1$  gerichteter, nicht vom Leistungserbringer zu beeinflussender Faktoren an den Indikatorwerten tendenziell zu groß eingeschätzt. Dies führt umgekehrt dazu, dass der geschätzte Toleranzanteil  $R_2$  und damit der bereinigte Referenzbereich für die partielle Nachberechnung umso schmaler (strenger) sind, je weiter der empirische Bundeswert vom eigentlichen (unbekannten) erreichbaren Optimum entfernt ist und je näher er an der originalen Referenzbereichsgrenze liegt. Die in der Tendenz höhere Schätzung von  $R_1$  und niedrigere Schätzung von  $R_2$  führt dazu, dass Leistungserbringer in ihrer Stellungnahme einen größeren Anteil ihrer Abweichung vom Referenzbereich durch nicht von ihnen zu verantwortende Faktoren erklären müssen und ein geringerer Anteil als Qualitätsdefizit toleriert wird. Es kommt also in der fachlichen Bewertung umso leichter zu Bewertungen als „Qualitätsdefizit“, je näher empirischer Bundeswert und originale Referenzbereichsgrenze zusammenliegen.

Geht man davon aus, dass der Referenzbereich gezielt und unter Berücksichtigung der verfügbaren Informationen zum Ist-Zustand der Versorgung festgelegt wurde, so spiegelt sich in einem geringen Unterschied zwischen empirischem Bundeswert und Referenzbereichsgrenze die Absicht wieder, nur geringe Abweichungen vom Ist-Zustand zum Schlechteren hin zu tolerieren. Dabei bleibt zwar offen, ob bei der Festlegung des Referenzbereichs eine geringe Toleranz von Qualitätsdefiziten intendiert war oder ob davon ausgegangen wurde, dass bereits der Ist-Zustand, gemessen am empirischen Bundeswert, deutlich vom Optimum abweicht. Da jedoch bei den betreffenden Indikatoren offensichtlich von einem größeren Verbesserungspotenzial und damit Verbesserungsbedarf auszugehen ist, kann der oben beschriebene Effekt, nämlich eine eher niedrig geschätzte Toleranz, unter der Perspektive der Qualitätsverbesserung als gewünschter Effekt interpretiert werden.

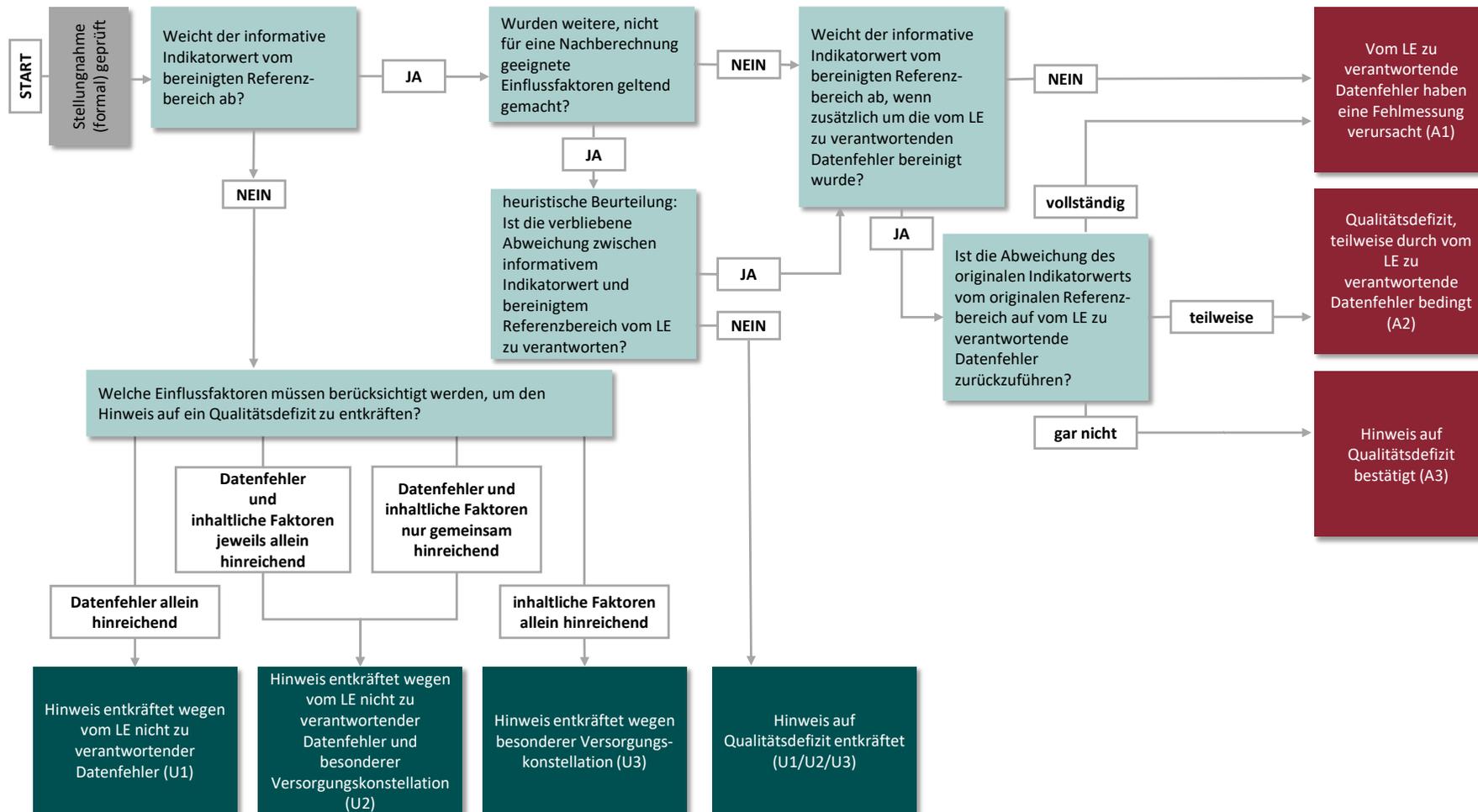
Ein Nachteil des vorgeschlagenen Vorgehens ist umgekehrt, dass – bei unverändertem Referenzbereich – umso weniger Abweichung vom Referenzbereich nach der partiellen Nachberechnung durch besondere Konstellationen erklärt werden muss, je besser der empirische Bundeswert ist. Dies könnte zu einer unangemessenen Toleranz von Qualitätsdefiziten bei einzelnen Leistungserbringern führen, wenn alle anderen Leistungserbringer deutlich bessere Versorgungsqualität aufweisen oder wenn es im Zeitverlauf zu deutlichen Verbesserungen im empirischen Bundeswert kommt. Wenn der empirische Bundeswert eines Qualitätsindikators weit innerhalb des Referenzbereichs liegt, sollte daher die Angemessenheit des Referenzbereichs überprüft werden.<sup>13</sup>

Grundsätzlich könnte der Toleranzanteil in den Referenzbereichen, die bisher in der externen Qualitätssicherung eingesetzt werden, auch auf andere Weise als hier vorgeschlagen geschätzt werden. Zum Zeitpunkt der Berichtslegung entwickelt das IQTIG eine Methodik, wie bei Festlegung von Referenzbereichen die Größe der beiden Anteile des Referenzbereichs angemessen und nachvollziehbar festgelegt werden kann.

---

<sup>13</sup> Die Prüfung kann auch ergeben, dass die Ergebnisse bei diesem Indikator bundesweit so gut sind, dass kein nennenswertes Verbesserungspotenzial mehr besteht. In diesem Fall muss die Option diskutiert werden, statt einer Verschärfung des Referenzbereichs die Verwendung des Indikators auszusetzen.

## 10 Gesamtalgorithmus für die fachliche Bewertung



# 11 Zielvereinbarungen

## 11.1 Formular zum Sachstandsbericht einzelner Zielvereinbarungen

Der Sachstandsbericht ist ein Controlling-Instrument zum Management von Zielvereinbarungen. Er dient der Kommunikation zwischen Geschäftsstelle<sup>14</sup> und Lenkungsgremium und stellt den Abschluss sowie den Erfüllungsgrad von Zielvereinbarungen dar.

### Mitteilung zu LE <Pseudonym>

Stand: *[Datum]*

Ansprechpartner Geschäftsstelle: *[Name]*

Anlass-Kategorie der Mitteilung: Wählen Sie ein Element aus.

*[Dropdown-Menü:*

- *Persistierendes Qualitätsproblem,*
- *Kritische Compliance bezüglich der Zusammenarbeit im Rahmen der Qualitätssicherung,*
- *mangelnde Einhaltung der formalen Kriterien der Zielvereinbarung,*
- *mangelnde Einhaltung der qualitativen Kriterien der Zielvereinbarung,*
- *Sonstiges]*

Betrifft das Erfassungsjahr	
QS-Verfahren	
Betroffene(r) QI / QIs	

### Begründete Einschätzung des Handlungsbedarfs

*[Begründung für die Involvierung des Lenkungsgremiums]*

....

### Quantitative Ergebnisse und fachliche Bewertung der Erfassungsjahre *[XY]* bis *[YZ]*

Erfassungsjahr	QI-Ergebnis (%)	Ergebnis außerhalb des Referenzbereichs	Bewertung nach Abschluss des Stellungnahmeverfahrens
<i>[QI-ID und QI-Bezeichnung]</i>			
...			
<i>[QI-ID und QI-Bezeichnung]</i>			

<sup>14</sup> Gemeint ist die Geschäftsstelle der LAG bzw. das IQTIG.

<b>Quantitative Ergebnisse und fachliche Bewertung der Erfassungsjahre [XY] bis [YZ]</b>			
...			
<i>[Erläuterung der dargestellten Ergebnisse und Entwicklungen]</i>			
....			
<b>Zielvereinbarung(en)</b>			
Frist für Bericht des LE zur Zielerreichung		<i>[Datum]</i>	
Unterschriebene Zusendung erfolgt		<i>[ja / nein]</i>	
<b><u>Vereinbartes Qualitätsziel 1: [Ziel]</u></b>			
Betrifft den QI		<i>[QI-ID und QI-Bezeichnung]</i>	
Begründung des Ziels		<i>[Text]</i>	
Frist zur Einhaltung des Ziels		<i>[Datum oder Erfassungsjahr]</i>	
Fristgerecht erfolgt		<i>[ja / nein]</i>	
<b><u>Vereinbartes Qualitätsziel 2 - n: [Ziel]</u></b>			
Betrifft den QI		<i>[QI-ID und QI-Bezeichnung]</i>	
Begründung des Ziels		<i>[Text]</i>	
Frist zur Einhaltung des Ziels		<i>[Datum oder Erfassungsjahr]</i>	
Fristgerecht erfolgt		<i>[ja / nein]</i>	
<b><u>Vereinbarte Maßnahme 1: [Maßnahme]</u></b>			
Betrifft den QI		<i>[QI-ID und QI-Bezeichnung]</i>	
Begründung der Maßnahme		<i>[Text]</i>	
Frist für den schriftlichen Nachweis		<i>[Datum]</i>	
Fristgerecht erfolgt		<i>[ja / nein]</i>	
<b><u>Vereinbarte Maßnahme 2 - n: [Maßnahme]</u></b>			
Betrifft den QI		<i>[QI-ID und QI-Bezeichnung]</i>	
Begründung der Maßnahme		<i>[Text]</i>	
Frist für den schriftlichen Nachweis		<i>[Datum]</i>	
Fristgerecht erfolgt		<i>[ja / nein]</i>	

<b>Zielvereinbarung(en)</b>	
<u>Empfohlene, nicht konsentierete Maßnahme 1: [Maßnahme]</u>	
Betrifft den QI	[QI-ID und QI-Bezeichnung]
Begründung der Maßnahme	[Text]
Begründung des Leistungserbringers für die Nicht-Konsentierung	[Text]
<u>Empfohlene, nicht konsentierete Maßnahme 2 - n: [Maßnahme]</u>	
Betrifft den QI	[QI-ID und QI-Bezeichnung]
Begründung der Maßnahme	[Text]
Begründung des Leistungserbringers für die Nicht-Konsentierung	[Text]
<b>Abschließende Einschätzung der Geschäftsstelle</b>	
[Text]	
...	
<b>Vorschlag der Geschäftsstelle zum weiteren Vorgehen</b>	
[Text]	
...	

<Ende des Formularinhalts>

## 11.2 Erläuterung zum Formular

### **Anlass-Kategorie der Mitteilung:**

#### Persistierendes Qualitätsproblem

Wurde bei einem Leistungserbringer in einem Indikator wiederholt ein Qualitätsdefizit festgestellt und zeigt sich trotz durchgeführter qualitätsfördernder Maßnahmen keine Verbesserung der Qualität, so stellt dies einen Anlass für eine Mitteilung an das Lenkungsgremium dar.

#### Kritische Compliance bezüglich der Zusammenarbeit im Rahmen der Qualitätssicherung

Eine mangelnde Compliance bezüglich der Zusammenarbeit im Rahmen der Qualitätssicherung ist zum Beispiel gegeben, wenn ein Leistungserbringer ohne Nennung eines nachvollziehbaren Grundes die Teilnahme an qualitätsfördernden Maßnahmen verweigert (z. B. Kollegiales Gespräch, Vor-Ort-Begehung oder Unterschrift einer Zielvereinbarung).

#### mangelnde Einhaltung der formalen Kriterien der Zielvereinbarung

Werden formale Kriterien (z. B. Einhaltung von Fristen) systematisch oder mehrfach nicht eingehalten, so stellt auch dies einen Anlass für eine Mitteilung an das Lenkungsgremium dar.

#### mangelnde Einhaltung der qualitativen Kriterien der Zielvereinbarung

Werden die vereinbarten Ziele und Maßnahmen nicht in ausreichender Qualität erfüllt und umgesetzt, so stellt dies einen Anlass für eine Mitteilung an das Lenkungsgremium dar.

Die Bewertung, ob die Kriterien qualitativ erfüllt werden, erfolgt durch die Geschäftsstelle nach fachlicher Beratung der Fachkommission.

#### Sonstiges

Da eine abschließende Auflistung der möglichen Gründe, die einen Anlass für eine Mitteilung an das Lenkungsgremium darstellen, nicht möglich ist, gibt es die Kategorie „Sonstiges“. Hierunter werden diejenigen Gründe gefasst, die über die bereits beschriebenen Anlass-Kategorien nicht abbildbar sind.

#### **Begründete Einschätzung des Handlungsbedarfs**

Zur weiteren Verdeutlichung der bereits genannten Anlass-Kategorie, wird an dieser Stelle näher beschrieben, welche Überlegungen oder Ereignisse zur Notwendigkeit einer Involvierung des Lenkungsgremiums geführt haben. Hier kann ebenso eine Einschätzung zur Dringlichkeit erfolgen.

#### **Ergebnisse der quantitativen Auffälligkeitseinstufung und fachliche Bewertung der Erfassungsjahre [XY] bis [YZ]**

Die tabellarische Zusammenstellung soll die quantitativen Ergebnisse zu den relevanten Qualitätsindikatoren, die Hinweise ob das Ergebnis entsprechend der Rechnungsvorschrift stat. signifikant vom Referenzbereich abweicht und ihre Bewertung nach Abschluss des Stellungnahmeverfahrens darstellen.

#### **Zielvereinbarung (-en)**

Es erfolgt eine detaillierte und zeitlich geordnete Auflistung aller Zielvereinbarungen, die den entsprechenden Indikator betreffen. Wurden in der Vergangenheit bereits Zielvereinbarungen getroffen, so werden auch diese hier dargestellt. Neben der Information, zu wann eine Unterschrift der gesamten Zielvereinbarung durch den Leistungserbringer erfolgen sollte und ob diese erfolgt ist, werden für jedes vereinbarte Ziel bzw. jede vereinbarte Maßnahme folgende Inhalte beschreiben:

1. Betrifft den QI:

Hier wird der QI (ID und Bezeichnung) genannt, für den das jeweilige Ziel / die jeweilige Maßnahme gilt.

2. Begründung des Ziels / der Maßnahme:

Es wird begründet, weshalb das Ziel / die Maßnahme aufgenommen wurde.

3. Frist für den schriftlichen Nachweis:

Für jedes Ziel / jede Maßnahme wird in der Zielvereinbarung eine Frist genannt, zu der die Erfüllung nachgewiesen werden muss. Diese wird hier aufgeführt.

4. Fristgerecht erfolgt:

Es erfolgt die Angabe (ja / nein), ob die geforderte Frist eingehalten wurde.

Wurden dem Leistungserbringer von der Geschäftsstelle weitere Maßnahmen vorgeschlagen, die Aufnahme dieser Maßnahmen in die Zielvereinbarung jedoch seitens des Leistungserbringers abgelehnt, werden diese ebenfalls aufgelistet. Dargestellt werden für diese Maßnahmen folgende Informationen:

5. Betrifft den QI:

Hier wird der QI (ID und Bezeichnung) genannt, für den das jeweilige Ziel / die jeweilige Maßnahme gelten soll.

6. Begründung des Ziels / der Maßnahme:

Es wird begründet, weshalb das Ziel / die Maßnahme empfohlen wurde.

7. Begründung des Leistungserbringers für die Nicht-Konsentierung:

Es wird die vom Leistungserbringer ggf. angeführte Begründung, weshalb einer Aufnahme in die Zielvereinbarung nicht zugestimmt wurde, genannt.

### **Abschließende Einschätzung der Geschäftsstelle**

Es erfolgt eine kurze Gesamteinschätzung des Sachverhalts aus Sicht der Geschäftsstelle.

### **Vorschlag der Geschäftsstelle zum weiteren Vorgehen (optional)**

Können dem Lenkungsgremium konkrete weitere Maßnahmen außerhalb der Handlungskompetenz der Geschäftsstelle empfohlen werden, werden diese hier benannt.

## Literatur

- Agresti, A (2013): *Categorical Data Analysis*. Third Edition. (Wiley Series in Probability and Statistics). Hoboken, US-NJ: Wiley. ISBN: 978-0-470-46363-5.
- Ash, AS; Fienberg, SE; Louis, TA; Normand, SLT; Stukel, TA; Utts, J (2012): *Statistical Issues in Assessing Hospital Performance*. Commissioned by the Committee of Presidents of Statistical Societies [*White paper*]. Revised: 27.01.2012. Baltimore, US-MD: Centers for Medicare and Medicaid Services; Committee of Presidents of Statistical Societies. URL: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf> (abgerufen am: 09.10.2019).
- Bates, D; Mächler, M; Bolker, B; Walker, S (2015): Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). DOI: 10.18637/jss.v067.i01.
- Gelman, A; Hill, J (2007): *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, GB: Cambridge University Press. ISBN: 978-0-521-86706-1.
- Hosmer, DW Jr.; Lemeshow, S; Sturdivant, RX (2013): *Applied Logistic Regression*. Third Edition. Hoboken, US-NJ: John Wiley & Sons. ISBN: 978-0-470-58247-3.
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2017): Bericht zum Strukturierten Dialog 2016. Erfassungsjahr 2015. Anhang. Stand: 15.05.2017. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2015/IQTIG\\_Bericht-zum-Strukturierten-Dialog-2016\\_Anhang.pdf](https://iqtig.org/downloads/berichte/2015/IQTIG_Bericht-zum-Strukturierten-Dialog-2016_Anhang.pdf) (abgerufen am: 27.02.2019).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018): IQTIG – R functions for hospital profiling. Package 'IQTIGpvc'. [Stand:] 06.03.2018. Berlin: IQTIG. URL: [https://iqtig.org/dateien/dasiqtig/grundlagen/biometrische\\_methoden/IQTIGpvc.pdf](https://iqtig.org/dateien/dasiqtig/grundlagen/biometrische_methoden/IQTIGpvc.pdf) (abgerufen am: 09.10.2019).
- Larsen, K; Petersen, JH; Budtz-Jørgensen, E; Endahl, L (2000): Interpreting Parameters in the Logistic Regression Model with Random Effects. *Biometrics* 56(3): 909-914. DOI: 10.1111/j.0006-341X.2000.00909.x.
- Säfken, B; Rügamer, D; Kneib, T; Greven, S (2018): Conditional Model Selection in Mixed-Effects Models with cAIC4. [Stand:] 17.03.2018. [Dokument Nr.:] arXiv:1803.05664v2 [stat.CO]. URL: <https://arxiv.org/pdf/1803.05664.pdf> (abgerufen am: 25.06.2018).
- Spiegelhalter, D; Sherlaw-Johnson, C; Bardsley, M; Blunt, I; Wood, C; Grigg, O (2012): Statistical methods for healthcare regulation: rating, screening and surveillance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 175(1): 1-47. DOI: 10.1111/j.1467-985X.2011.01010.x.
- Stroup, WW (2013): *Generalized Linear Mixed Models. Modern Concepts, Methods and Applications*. (Texts in Statistical Science). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC. ISBN: 978-1-4398-1512-0.

Winkler-Komp, G; Misselwitz, B; Kupfernagel, F; van Emmerich, C; Döbler, K (2014):

Maßnahmen zur Qualitätssicherung in Krankenhäusern: Strukturierter Dialog – Strukturen und Prozesse. Ergebnis einer Umfrage des Gemeinsamen Bundesausschusses bei den auf Landesebene beauftragten Stellen und der Institution nach § 137 a SGB V. *Das Krankenhaus* 2014(3): 198-205.

Wood, SN (2006): *Generalized Additive Models. An Introduction with R.* (Texts in Statistical Science). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC. ISBN: 978-1-58488-474-3.

Wood, SN (2013): A simple test for random effects in regression models. *Biometrika* 100(4): 1005-1010. DOI: 10.1093/biomet/ast038.



Institut für Qualitätssicherung und  
Transparenz im Gesundheitswesen

# Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern

Stellungnahmen zum Vorbericht

Erstellt im Auftrag des  
Gemeinsamen Bundesausschusses

Stand: 26. Mai 2020

---

# Impressum

**Thema:**

Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern. Stellungnahmen zum Vorbericht

**Ansprechpartnerin:**

Andrea Wolf

**Auftraggeber:**

Gemeinsamer Bundesausschuss

**Datum des Auftrags:**

18. Januar 2018

**Datum der Abgabe:**

26. Mai 2020

**Herausgeber:**

IQTIG – Institut für Qualitätssicherung  
und Transparenz im Gesundheitswesen

Katharina-Heinroth-Ufer 1  
10787 Berlin

Telefon: (030) 58 58 26-0  
Telefax: (030) 58 58 26-999

[info@iqtig.org](mailto:info@iqtig.org)

<https://www.iqtig.org>

# Inhaltsverzeichnis

## Organisationen nach § 137a Abs. 7 SGB V

- Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V. (AWMF)
- Bundesärztekammer (BÄK)
- Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS)
- Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e. V. (DGGG) und der Deutschen Gesellschaft für Senologie e. V. (DGS)
- Deutsche Gesellschaft für HNO-Heilkunde, Kopf- und Hals-Chirurgie e. V. (DGHNO-KHC)
- Deutsche Gesellschaft für Innere Medizin e. V. (DGIM)
- Deutsche Gesellschaft für Kardiologie – Herz- und Kreislaufforschung e. V. (DGK)
- Deutsche Gesellschaft für Kinder- und Jugendmedizin e. V. (DGKJ e. V.)
- Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS)
- Deutsche Gesellschaft für NeuroIntensiv- und Notfallmedizin (DGNI)
- Deutsche Gesellschaft für Neuroradiologie e. V. (DGNR)
- Deutsche Gesellschaft für Orthopädie und Unfallchirurgie e. V. (DGOU), Deutsche Gesellschaft für Orthopädie und Orthopädische Chirurgie e. V. (DGOOC) und Deutsche Gesellschaft für Unfallchirurgie e. V. (DGU)
- Deutsche Gesellschaft für Thorax-, Herz- und Gefäßchirurgie e. V. (DGTHG)
- Deutsche Krankenhausgesellschaft (DKG)
- Deutscher Pflegerat e. V. (DPR)
- Deutsches Netzwerk Versorgungsforschung e. V. (DNVF)
- Gesellschaft für Qualitätsmanagement in der Gesundheitsversorgung e. V. (GQMG)
- GKV-Spitzenverband (GKV-SV)
- Kassenärztliche Bundesvereinigung (KBV)
- Kassenzahnärztliche Bundesvereinigung (KZBV)
- Patientenvertretung: maßgebliche Organisationen nach § 140f SGB V (PatV)

## Organisationen auf Landesebene

- Ärztekammer Sachsen-Anhalt, Projektgeschäftsstelle Qualitätssicherung (AEKSA QS)
- Bayerische Arbeitsgemeinschaft für Qualitätssicherung in der stationären Versorgung (BAQ) bei der Bayerischen Krankenhausgesellschaft
- Geschäftsstelle Qualitätssicherung Hessen (GQH) bei der Hessischen Krankenhausgesellschaft
- Geschäftsstelle Qualitätssicherung Nordrhein-Westfalen bei den Ärztekammern Nordrhein und Westfalen-Lippe (QS-NRW)
- Landesgeschäftsstelle für einrichtungs- und sektorenübergreifende Qualitätssicherung der Sächsischen Landesärztekammer, Projektgeschäftsstelle Qualitätssicherung Sachsen (SLAEK QS)
- Landesgeschäftsstelle für Qualitätssicherung Berlin (LQS BE)
- Landesgeschäftsstelle für Qualitätssicherung Mecklenburg-Vorpommern (LQS MV)

- Landesgeschäftsstelle für Qualitätssicherung Rheinland-Pfalz (LQS RP)
- Landesgeschäftsstelle für Qualitätssicherung Saarland (LQS SL)
- Landesgeschäftsstelle für Qualitätssicherung Thüringen (LQS TH)
- Niedersächsische Krankenhausgesellschaft e. V., Projektgeschäftsstelle Qualitätssicherung (NKGeV QS)
- Qualitätsbüro Bremen (QB Bremen)
- Qualitätssicherung im Gesundheitswesen Baden-Württemberg (QiG BW GmbH)

<b>A</b> rbeitsgemeinschaft der	<i>Association of the</i>
<b>W</b> issenschaftlichen	<i>Scientific</i>
<b>M</b> edizinischen	<i>Medical</i>
<b>F</b> achgesellschaften e.V.	<i>Societies in Germany</i>



Berlin, 21.11.2019

**Stellungnahme**  
**der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften**  
**(AWMF)**  
**zum Vorbericht des IQTIG: Weiterentwicklung des Strukturierten Dialogs mit**  
**Krankenhäusern, Vorbericht zu Stufe 1 und Stufe 2**  
**Erstellt im Auftrag des Gemeinsamen Bundesausschusses**  
**Stand: 10. Oktober 2019**

Die AWMF wurde am 10. Oktober 2019 um eine Stellungnahme zu dem oben genannten Vorbericht gebeten. Die AWMF hat ihrerseits ihre Mitgliedsfachgesellschaften gebeten, bei Bedarf eine eigene Stellungnahme zu verfassen. Die bis einschließlich 21.11.2019 bei der AWMF eingegangenen Stellungnahmen von 12 Fachgesellschaften sind der Stellungnahme der AWMF als Anlage beigelegt (s. Anlage 1). Diese äußern sich auch zu wichtigen weiteren Aspekten, die wir ebenfalls zu berücksichtigen bitten. Die Deutsche Gesellschaft für Pneumologie und Beatmungsmedizin (DGP) schließt sich der Stellungnahme der AWMF an.

### 1) Hintergrund

Mit Beschluss vom 18. Januar 2018 hat der Gemeinsame Bundesausschuss (G-BA) das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) mit der Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern beauftragt .

Dieser Weiterentwicklungsauftrag ist in zwei Stufen gegliedert:

Der Abschlussbericht zu Stufe 1 wurde dem G-BA am 30. November 2018 übermittelt. Die AWMF und ihre Mitgliedsgesellschaften hatten auch hierzu im Vorfeld bereits Stellung genommen. Die darin formulierten Anregungen gelten weiterhin. Nun hat das IQTIG den Vorbericht zum zweiten Teil des Auftrags vorgelegt.

Die AWMF begrüßt den Auftrag des G-BA an das IQTIG und die entsprechenden Ausarbeitungen des IQTIG im vorliegenden Vorbericht zur Stufe 2. Im Kern liegt dadurch eine eindrucksvolle Bestandsaufnahme und Aufzeigung der Heterogenität und Verbesserungsfähigkeit der aktuellen Vorgehensweisen in den einzelnen Bundesländern in Bezug auf das Verfahren „Strukturierter Dialog“ vor.

Allerdings muss zur Vorgehensweise der Berichterstattung durch das IQTIG erneut kritisch angemerkt werden, dass auch dieser Vorbericht aus wissenschaftlicher Sicht an vielen Stellen von kritikwürdiger Qualität ist und das IQTIG von einer frühzeitigeren Einbindung der Expertise, die in den Wissenschaftlichen Medizinischen Fachgesellschaften zur Verfügung steht, profitieren könnte.

Die Stellungnahmeberechtigten, die Begutachtungen erst im Rahmen der Stellungnahmeverfahren im Ehrenamt bedienen dürfen, kommen an Grenzen der Machbarkeit. Wir gehen davon aus, dass

– wie bereits zuvor von der AWMF aufgezeigt - zahlreiche Fachgesellschaften die Möglichkeit einer eigenen Stellungnahme zu dem vorliegenden Vorbericht des IQTIG nicht wahrgenommen haben.

Dies betrifft auch die AWMF selbst. Aus pragmatischen Gründen bezieht sich daher auch die vorliegende Stellungnahme der AWMF nur auf wenige Aspekte des Methodenpapiers.

## **2) Auftrag des IQTIG**

Der G-BA hat das IQTIG beauftragt, den strukturierten Dialog mit den Krankenhäusern hinsichtlich der Ergebnisse der datengestützten einrichtungsübergreifenden Qualitätssicherung weiterzuentwickeln. Als Funktion des strukturierten Dialogs sind aufgeführt: Bewertung der Versorgungs- und Dokumentationsqualität und Qualitätsförderung. Der Auftrag umfasst:

- die Einheitlichkeit der Vorgehensweise sowie
- die Transparenz und Nachvollziehbarkeit der Entscheidungsfindung und
- die Effizienz im Sinne einer „Verschlankung“ des Verfahrens zu optimieren.

## **3) Punktuelle Anmerkungen zu den aktuellen Ausarbeitungen des IQTIG**

Wesentliche Ausarbeitungen zur Vereinheitlichung des Verfahrens werden von den stellungnehmenden Fachgesellschaften positiv beurteilt.

Kritisch zu sehen ist die hohe Bedeutung, die statistischen Methoden zur Quantifizierung von Ergebnissen beigemessen wird. Die Methoden erscheinen adäquat (s. Stellungnahme der GMDS und IBS-DR). Die Hinterfragung bezieht sich vielmehr auf die Aufwand- Nutzenabschätzung im Hinblick auf einen patientenrelevanten Nutzen der so „normierten“ Qualitätsauffälligkeiten (s.a. Stellungnahme der DGHNO-KHC).

Die folgenden Anmerkungen greifen einige wenige Aspekte heraus, die der AWMF zusätzlich zu den in den Stellungnahmen der Fachgesellschaften für besonders hervorhebenswert erscheinen, da hier noch eine Verbesserung möglich scheint.

### **Betroffene Leistungserbringer nicht im Vorfeld einbezogen**

Das IQTIG hat den Einbezug der Betroffenen für die Verbesserung des strukturierten Dialogs entsprechend der expliziten Beauftragung nur so umgesetzt, dass es einen Workshop mit Vertretern der Landesqualitätsstellen und Landesarbeitsgemeinschaften veranstaltet hat.

Erst nach der differenzierten Ausarbeitung des IQTIG werden nun die vom Verfahren betroffenen klinisch Tätigen einbezogen. Die AWMF hält dies für eine nicht optimale Vorgehensweise, wie bereits oben angemerkt. Die von den Maßnahmen Betroffenen sollten im Sinne der angestrebten Qualitätsförderung im Mittelpunkt stehen.

### **Datenübermittlung zu einem ungünstigen Zeitpunkt festgelegt/ Stellungnahmefristen zu kurz**

Nicht im Sinne der Betroffenen ist die praktische Organisation des strukturierten Dialogs geraten mit einer Datenübermittlung zu einem Zeitpunkt dünner Personaldecke aufgrund der Sommerferien (siehe u.a. Stellungnahme der DGGG und DGS sowie DGTHG in Bezug auf den ferienahen Zeitpunkt der Datenübermittlung) und einer mit vier Wochen zu kurzen Stellungnahmefrist (s. Stellungnahme der DGKJ).

### **Zeitnahe Datenübermittlung weiterhin nicht gewährleistet**

Auch die Intervalle der Datenübermittlung wurden im Hinblick auf ihre Nützlichkeit nicht mit den betroffenen Leistungserbringern abgestimmt (siehe Stellungnahme der GQMG in Bezug auf Zeitpunkte/Intervalle der Datenübermittlung).

**Klinische Fachexpertise weiterhin nicht ausreichend einbezogen**

Zu prüfen ist, inwieweit die Rolle von klinisch-wissenschaftlichen Fachexperten bei der Erarbeitung der Verfahren inklusive der Festlegung erforderlicher Datenfelder und der Beurteilung der Eignung von Referenzwerten weiter gestärkt werden kann (siehe u.a. Stellungnahme der DGVS, der DGIM, der DGNR und der DGTHG).

Dies betrifft auch die Einschätzung und Festlegung geeigneter Verfahren zur Qualitätsförderung.

In Bezug auf die Panelzusammensetzung der Fachgruppen hat das IQTIG nun die Fachexpertise sowie Kriterien für die Zusammensetzung konkretisiert. Es sollte jedoch für jedes Fachpanel dessen Repräsentativität geprüft werden z.B. im Abgleich mit vorliegenden interdisziplinären bzw. multiprofessionellen Leitlinien.

**Ziel: gelebtes Qualitätsmanagement, Feed-back: begleitende Evaluation erforderlich**

Qualitätsmanagement sollte hilfreich und nützlich sein. Dafür wäre es vorteilhaft, dass Rückmeldungen über die erzielten Leistungen nicht nur ein Mal pro Jahr erfolgen, sondern mindestens halbjährlich. Dies würde eine rasche Korrektur von fehlerhaften Prozessen erlauben, bzw. anderweitige rasche Verbesserungen.

Der Auftrag des IQTIG umfasste auch Empfehlungen zur Evaluation des neuen Konzepts. Diese werden im vergleichsweise kurz formulierten Kapitel 10 dargelegt. Erfreulicherweise schlägt das IQTIG als Bestandteil der formativen Evaluation eine prozessbegleitende Befragung auch der Leistungserbringer in Bezug auf Verständlichkeit und Nützlichkeit des neuen Konzepts vor. Darauf folgen ausführliche Darstellungen statistischer Verfahren zur Annäherung an eine summative Evaluation zu verschiedenen Zeitpunkten.

Leider wird die Befragung der Leistungserbringer in der Zusammenfassung nicht mehr erwähnt. Wir schlagen vor, für den Abschlussbericht die Vorschläge zur formativen Evaluation so zu konkretisieren, dass diese tatsächlich prozessbegleitend erfolgen kann. Nur so kann ermittelt werden, ob die Änderungen gewinnbringend sind.

Für Rückfragen stehen wir jederzeit gerne zur Verfügung.

Dr. med. Monika Nothacker, MPH [nothacker@awmf.org](mailto:nothacker@awmf.org)

Prof. Dr. med. Ina B. Kopp, [kopp@awmf.org](mailto:kopp@awmf.org)

Prof. Dr. med. Dr. med. dent. Wilfried Wagner [wagner@awmf.org](mailto:wagner@awmf.org)

**Anlage 1:** Stellungnahmen der Fachgesellschaften (in beigefügter Zip-Datei)  
Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS)  
Deutsche Gesellschaft für Gynäkologie und Geburtshilfe (DGGG) und Deutsche Gesellschaft für Senologie (DGS)  
Deutsche Gesellschaft für Hals-Nasen-Ohren-Heilkunde, Kopf- und Hals-Chirurgie e.V. (DGHNO-KHC)  
Deutsche Gesellschaft für Innere Medizin (DGIM)  
Deutsche Gesellschaft für Kardiologie - Herz- und Kreislaufforschung (DGK)  
Deutsche Gesellschaft für Kinder- und Jugendmedizin (DGKJ)  
Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) und Deutsche Sektion der der Internationalen Biometrischen Gesellschaft (IBS-DR)  
Deutsche Gesellschaft für Neuroradiologie (DGNR)  
Deutsche Gesellschaft für Thorax-, Herz- und Gefäßchirurgie (DGTHG)  
Deutsche Gesellschaft für Qualitätsmanagement im Gesundheitswesen (GQMG)



## **Stellungnahme der Bundesärztekammer**

zum IQTIG-Bericht

Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern

Vorbericht zu Stufe 1 und Stufe 2

Berlin, 21.11.2019

Korrespondenzadresse:

Bundesärztekammer  
Herbert-Lewin-Platz 1  
10623 Berlin

## 1. Vorbemerkung

Der Bericht beinhaltet gemäß Auftrag des G-BA vom 18.02.2018 Weiterentwicklungsvorschläge zum Strukturierten Dialog. Da mittelfristig alle Leistungsbereiche der externen stationären Qualitätssicherung (ESQS) aus der QSKH-RL in die DeQS-RL übergehen werden, sollen die Empfehlungen des Berichts trotz des anderslautenden Titels auch bzw. vor allem für zukünftige Stellungnahmeverfahren in der DeQS-RL gelten. Demzufolge gelten auch die Anmerkungen der Bundesärztekammer sowohl für den Strukturierten Dialog als auch für das Stellungnahmeverfahren.

## 2. Inhalt des Berichts

In einer Bestandsaufnahme wird die Ausgangssituation beim Strukturierten Dialog gemäß QSKH-RL analysiert (Kapitel 2). Die Heterogenität der Vorgehensweise der Landesgeschäftsstellen Qualitätssicherung (LQS) wird beleuchtet, z. B. bezüglich der Verwendung alleiniger Hinweise bei rechnerischen Auffälligkeiten. Betrachtungen zur Effizienz des Verfahrens zeigen, dass sich vor allem bei Einrichtungen mit niedrigen Fallzahlen aus einer großen Zahl von rechnerischen Auffälligkeiten häufig nur eine kleine Zahl von tatsächlichen qualitativen Auffälligkeiten ergibt.

In Kapitel 3 wird der gegenwärtig in den LQS durchgeführte Strukturierte Dialog unter methodischen Überlegungen zur Güte von Bewertungsprozessen betrachtet. Das Institut kritisiert einerseits, dass die auf Landesebene eingebundenen Expertinnen und Experten Entscheidungen zur qualitativen Auffälligkeit lediglich nach impliziten Bewertungsregeln vornehmen würden. Andererseits werde versucht, nicht nur die Qualitätsindikatorergebnisse, sondern auch die Behandlungsprozesse zu bewerten. Dies könne aber nicht die Aufgabe des Stellungnahmeverfahrens sein (Seite 56). Würden Qualitätsindikatoren lediglich als Aufgreifkriterium genutzt, sei zur Bewertung der Gesamtqualität ein separates von der indikatorbasierten Qualitätsmessung unabhängiges Verfahren notwendig (Seite 57). Zudem wird kritisiert, dass Elemente der Qualitätsbewertung und der Qualitätsförderung im Strukturierten Dialog vermischt würden.

Es wird konstatiert, dass nach Erkenntnissen der wissenschaftlichen Literatur zur Urteilsbildung die sogenannte klinische oder intuitive Urteilsbildung einer statistischen oder mechanischen Urteilsbildung unterlegen sei. Durch explizite Vorgaben an klinische Expertinnen und Experten sowie standardisierte Informationsgrundlagen könne die Güte von Entscheidungen erhöht werden (Seite 58). Vorrangige Ziele des Instituts seien daher die Formulierung expliziter Bewertungsregeln, die Standardisierung des Stellungnahmeverfahrens sowie die alleinige Beschränkung auf die Analyse der Indikatorergebnisse. Dadurch würden Stellungnahmeverfahren einerseits weniger heterogen und andererseits effizienter. Qualitätsbewertung und Qualitätsförderung seien zukünftig in Form von Modulen klar zu trennen (Kapitel 4).

Die Vorschläge zur statistischen Auswertungsmethodik nehmen einen Schwerpunkt des Berichts ein (Kapitel 5). Besonderes Augenmerk wird der sogenannten analytischen Herangehensweise an die statistische Datenauswertung gewidmet. Es wird postuliert, dass sich mit geeigneter mathematisch-statistischer Methodik über die Qualitätsindikatorergebnisse die ihnen zugrunde liegende „Kompetenz der Leistungserbringer“ abschätzen lasse. Damit seien nicht nur Rückschlüsse auf die Qualität der Behandlung bei Fällen eines zurückliegenden Erfassungsjahrs möglich, sondern auch „auf Fälle über das Erfassungsjahr hinaus“ (Seite 73).

Der Bericht enthält den Vorschlag, Stellungnahmeverfahren im Sinne eines Moduls „Qualitätsbewertung“ in einem zweischrittigen Verfahren neu zu konzeptionieren: ein erster

Schritt solle eine optimierte quantitative Auffälligkeitseinstufung beinhalten, sowohl für einen Ein-Jahreszeitraum als auch für einen Zwei-Jahreszeitraum, ein zweiter (qualitativer) Schritt soll einen neuen standardisierten Bewertungsprozess umfassen.

Als theoretische Basis dient dem Institut ein entscheidungstheoretisches Modell mittels bayesianischem Netzwerk. In diesem Modell hängen über Indikator gemessene Behandlungsergebnisse von einem „latenten Kompetenzparameter“ der Leistungserbringer ab. Indikatorergebnis, Fallzahl und Referenzbereich sind weitere Variablen, die in dem Modell zur Entscheidung über eine quantitative Auffälligkeit führen. Die Diskussion wird im Bericht mit dem Ziel geführt, eine möglichst effiziente, d. h. mit geringeren Aufwänden behaftete Entscheidung herbeizuführen. Es werden neben der aus der ESQS bekannten „rechnerischen Auffälligkeit“ und der aus dem PlanQI-Verfahren bekannten „statistisch signifikanten Auffälligkeit“ auch eine „statistisch relevante Auffälligkeit“ unterschieden, bei den die außerhalb des Referenzbereich liegenden maximal tolerierte Anzahl von Fällen im Fokus steht. Vor- und Nachteile dieser Auffälligkeitseinstufungen werden hinsichtlich unterschiedlicher Parameter wie z. B. Fallzahlabhängigkeit, Sensitivität und Spezifität miteinander verglichen. Im Weiteren werden verschiedene statistische Methoden zur Berechnung von Zwei-Jahresauffälligkeiten diskutiert.

Zusammenfassend schlägt das IQTIG vor, zukünftig bei der Berechnung und Bewertung der Indikatorergebnisse die statistische Unsicherheit zu berücksichtigen. Konkret wird die Anwendung der bayesianischen Version der statistisch signifikanten Auffälligkeit empfohlen. Für den Fall, dass viele Leistungserbringer nur kleine Fallzahlen aufweisen, sollten die Berechnungen auf den Daten von zwei Erfassungsjahren beruhen.

In Kapitel 6 werden Vorschläge für Schritt 2 des Stellungnahmeverfahrens gemacht. Liefert Schritt 1 bei einem Indikator einen „hinreichenden Hinweis auf ein Qualitätsdefizit“, so sollen die Expertinnen und Experten nach einem festen Bewertungsschema entscheiden, ob der Hinweis bestätigt wird oder nicht. Im Fokus des schriftlichen Stellungnahmeverfahrens soll stehen, ob in Schritt 1 bei einem Indikator auffällig gewordene Leistungserbringer besondere Einflussfaktoren geltend machen, die bei ihren Patientinnen und Patienten nicht berücksichtigt wurden. Dies soll für jeden einzelnen Fall eines Leistungserbringers erfolgen.

Die Aufgabe der Expertinnen und Experten soll es dann sein, die Angemessenheit dieser Einflussfaktoren zu prüfen. Werden Faktoren anerkannt, für die der Leistungserbringer nicht verantwortlich gemacht werden kann, so soll eine partielle Neuberechnung des Indikatorergebnisses erfolgen. Dabei können sowohl einzelne Fälle aus der Grundgesamtheit (Nenner) genommen werden, wenn sie zuvor fälschlicherweise eingeschlossen wurden, als auch aus dem Zähler, wenn fälschlicherweise ein Ereignis (z. B. Komplikation) gezählt wurde. Dokumentationsfehler („Datenfehler“) sollen dabei gesondert bewertet werden.

Das Ergebnis der Neuberechnung soll ein so genannter „informativer Indikatorwert“ sein. Auch der Referenzbereich ist neu zu berechnen, d. h. ein Anteil an „gerichteten Einflussfaktoren“ wird herausgerechnet. Liegt der informative Indikatorwert innerhalb des bereinigten Referenzbereichs, so ist für die Berichterstattung zum Stellungnahmeverfahren genau zu unterscheiden, ob dies aufgrund inhaltlicher Fehler, Datenfehler oder einer Mischung aus beiden geschehen ist. Liegt der informative Indikatorwert weiterhin außerhalb des bereinigten Referenzbereichs, so haben die Expertinnen und Experten zu entscheiden, ob berechtigterweise geltend gemachte Einflussfaktoren nicht berücksichtigt wurden. Dabei können sie entscheiden (so genannte heuristische Beurteilung), ob der „Hinweis auf ein Qualitätsdefizit“ bei einem Indikator trotzdem entkräftet werden kann. Andernfalls soll gemäß einem Bewertungsschema festgestellt werden, dass ein „Qualitätsdefizit“ vorliege.

Im Weiteren werden Vorschläge zu Mindestanforderungen des schriftlichen Stellungnahmeverfahrens und zur Zusammensetzung der Fachkommission gemacht. Das von der Qualitätsbewertung nach Auffassung des Instituts zu trennende „Modul“ der Qualitätsförderung wird kurz umrissen. So wird auch ein Peer Review als mögliche Maßnahme beschrieben. Für Zielvereinbarungen werden Mindestanforderungen dargelegt (Kapitel 7)

In Kapitel 8 wird eine zeitliche Verkürzung des Stellungnahmeverfahrens vorgeschlagen. Durch strikte Verfahrensvorgaben soll eine Verkürzung des neuen Moduls „Qualitätsbewertung“ gegenüber den Fristen gemäß DeQS-RL von vier auf drei Monate möglich sein.

Die Berichterstattung (Kapitel 9) soll nach Vorschlag des IQTIG zukünftig in getrennten Berichten zur Versorgungsqualität (Ergebnisse des Moduls Qualitätsbewertung) und zu Maßnahmen der Qualitätsförderung erfolgen.

Kapitel 10 enthält Empfehlungen zur Evaluation des neuen Konzepts. Neben einer formativen Evaluation während der Einführung des neuen Verfahrens wird eine summative Evaluation empfohlen, in der u. a. die Entwicklung der Heterogenität der Stellungnahmen und die Effizienz der Stellungnahmen überprüft werden sollen. Anhand der Daten der Erfassungsjahre 2016 und 2017 werden Beispielsberechnungen präsentiert.

Abschließend (Kapitel 11) werden die Empfehlungen zusammengefasst. Gemäß G-BA-Auftrag sollte das IQTIG außerdem prüfen, bei welchen Qualitätsindikatoren eine quantitative mit einer qualitativen Auffälligkeit gleichgesetzt werden kann. Das Institut kommt zu dem Schluss, dass trotz der Erhöhung der Spezifität durch neue statistische Verfahren vorerst nicht auf das Stellungnahmeverfahren verzichtet werden könne. In Beantwortung weiterer Einzelfragen des Auftrags spricht sich das Institut u.a. gegen den Einsatz von Indizes als Aufgreifkriterien aus und diskutiert den Einsatz leistungsbereichsübergreifender Qualitätsindikatoren.

### **3. Wesentliche Kritikpunkte**

#### *Konstrukt der Kompetenzkomponente der Leistungserbringer*

Mit der Feststellung „Bei der analytischen Herangehensweise liegt das Interesse am zugrunde liegenden Prozess. Im Rahmen der externen Qualitätssicherung bedeutet dies, dass man sich für die zugrunde liegende Kompetenz des Leistungserbringers bezüglich des im Indikator abgebildeten Qualitätsziels interessiert“ (Seite 73) wird das theoretische Konstrukt der „Kompetenz der Leistungserbringer“ postuliert.

Im Weiteren werden umfangreiche mathematische Modellierungen auf Basis dieses Konstrukts vorgenommen unter der Maßgabe, dass mit Qualitätsindikatoren die zugrunde liegende Kompetenz erfasst werde. Der Satz „Dies bedeutet, dass...bei der analytischen Herangehensweise nicht nur Schlüsse über die behandelten Fälle im betrachteten Erfassungsjahr mit den entsprechenden Ereignissen gezogen werden sollen, sondern das Interesse in der zugrunde liegenden Kompetenz des Leistungserbringers liegt, die sich auch auf die Fälle über das Ende des Erfassungsjahrs hinaus auswirkt“ (Seite 73) kommt der Behauptung gleich, dass sich mit Qualitätsindikatoren nicht nur vergangene, sondern auch zukünftig zu erwartende Qualität messen lasse.

Es handelt sich nach Ansicht der Bundesärztekammer um eine Neudefinition der Qualität bzw. der Funktion eines Qualitätsindikators, die durch keine Ableitung aus der wissenschaftlichen Literatur gedeckt ist. Die Befassung mit und die Messung von Kompetenzen ist der Domäne der Bildung und Weiterbildung vorbehalten. Instrumente zur Messung von Kompetenzen sind nicht mit den Indikatoren der externen Qualitätssicherung

vergleichbar. Unabhängig davon ist auch der Begriff des Leistungserbringers in diesem Kontext unbestimmt. Das Spektrum der denkbaren Konstellationen ist breit, vom fest zusammengesetzten oder wechselnden interdisziplinären Team bis hin zum einzelnen Arzt.

Die Bundesärztekammer rät von der Verwendung des Begriffs der „Kompetenz“ in diesem Zusammenhang dringend ab, insbesondere deshalb, weil er einen zentralen Bestandteil in der ärztlichen Weiterbildung darstellt und die Gefahr einer inhaltlichen und sprachlichen Uneindeutigkeit gegeben ist.

### *Fachliches Bewertungsverfahren*

Zentrales Element der vorgeschlagenen fachlichen Bewertung in Schritt 2 des Stellungnahmeverfahrens soll die Ergebnisbetrachtung aller Indikatoren auf Einzelfallebene sein. Das Ziel ist dabei nicht eine tatsächliche fachliche Erörterung des Einzelfalls mit den klinischen Experten. Vielmehr geht es um die fallweise Abklärung, ob sich der Indikatorwert und der Referenzbereich weiter „bereinigen“ lassen, indem vorher nicht berücksichtigte Risikofaktoren nachträglich berücksichtigt werden. Die Validität des Indikators soll hier quasi ex post rechnerisch optimiert werden.

Dieses Grundprinzip steht im Widerspruch zur Forderung an anderer Stelle im Bericht, dass „die Stellungnahme der Leistungserbringer möglichst auf eine Analyse von übergeordneten Gründen fokussieren soll“ (Seite 165). Dort heißt es (zu Recht) weiter: „Eine am Einzelfall orientierte Stellungnahme kann daher den Blick auf systemische Ursachen und notwendige systemische Verbesserungsansätze, wie z. B. die Verbesserung in der Einrichtungsorganisation oder der Patientensicherheitskultur, versperren“.

Der Aufwand, für jeden einzelnen Fall eines Erfassungsjahrs bei einem quantitativ auffälligen Indikator im Nachhinein mögliche Risikofaktoren zusammenzutragen, dürfte erheblich sein. Bei großen Fachabteilungen kann es notwendig werden, u.U. mehrere hundert Patientenakten für einen Indikator zu analysieren. Auch für die Expertinnen und Experten in den Fachgremien dürfte die Überprüfung der Angaben der Leistungserbringer aufwändig, wenn nicht sogar unpraktikabel sein.

Als Extrembeispiel sei im Leistungsbereich Hüftendoprothetik der QI 53004 (Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur) genannt: Im Jahre 2018 wurden 431 von 1.159 Krankenhäuser auffällig. Die Anzahl der auffälligen Fälle betrug 8.078 von 58.741 Fällen. Alleine für diesen Indikator müssten jedoch vermutlich mehrere Tausend Fälle betrachtet werden.

Das vorgeschlagene fachliche Verfahren will erklärtermaßen den Bewertungsspielraum der Expertinnen und Experten durch explizite Regeln einengen. „Es soll ein Maximum an Standardisierung und damit Objektivität der fachlichen Bewertung erreicht“ werden (Seite 163). Ein Dialog mit dem Leistungserbringer, wie er namensgebend für den Strukturierten Dialog war, ist nicht mehr vorgesehen.

Das Institut räumt selbst ein, dass durch die vorgeschlagenen expliziten Regeln das Verfahren deutlich komplexer werde und zudem für die klinischen Expertinnen und Experten ungewohnt sei (Seite 60). Tatsächlich dürfte das Verfahren inklusive der Neuberechnung der Indikatorwerte und Referenzbereiche für alle Beteiligten, also für die auf Landesebene beauftragten Stellen, die Expertinnen und Experten und die betroffenen Krankenhäuser und Vertragsärzte deutlich aufwändiger werden. Der im IQTIG-Konzept gewonnene Vorteil, dass durch die neue, auf statistischer Signifikanz beruhende Auswertungsmethodik weniger Auffälligkeiten im Stellungnahmeverfahren generiert werden, droht durch die Komplexität des nachfolgenden Bewertungsverfahrens

konterkariert zu werden. Es ist zudem zu befürchten, dass auch für die Adressaten der Qualitätsberichterstattung, die Versicherten und Patienten, die Ergebnisse des Stellungnahmeverfahrens nicht transparenter, sondern weniger anschaulich werden.

Im Bericht werden Komplikationsraten als Beispiele für das neue vorgeschlagene Bewertungsverfahren genannt. Das Spektrum der Qualitätsindikatoren in der ESQS reicht mit Prozessindikatoren, Indikatoren zur Indikation und komplexen Index-Indikatoren aber deutlich weiter. Hier sollte an weiteren Beispielen die Praktikabilität des neuen Verfahrens überprüft werden.

Die Ausführungen im Bericht beziehen sich zudem ausschließlich auf „negativ gepolte“ Indikatoren, d. h. Indikatoren, die ein negatives Ereignis, wie z. B. eine Komplikation, erfassen. Positiv gepolte Indikatoren, wie z. B. im Leistungsbereich Hüftendoprothetik QI 54002 (Indikation zum Hüftendoprothesen-Wechsel bzw. –Komponentenwechsel) sind umgekehrt zu behandeln. In diesen Fällen wird ein Referenzbereich bei Auffälligkeit nicht über- sondern unterschritten. Hier müsste von den betroffenen Krankenhäusern im Stellungnahmeverfahren für jeden nicht im Zähler mitgezählten (also primär nicht indizierten) Fall ex post nachgewiesen werden, dass er doch indiziert war. Diese Indikatorvarianten sollten im Bericht mit diskutiert werden.

Das Ergebnis des Bewertungsverfahrens soll ein neues Bewertungsschema sein. Statt dem bis dato verwendeten Begriff der „qualitativen Auffälligkeit“ soll zukünftig von einem „Qualitätsdefizit“ gesprochen werden. Es wird festgestellt, dass ersterer Begriff zu Missverständnissen führe, während letzterer „allgemeinverständlich das Nichterreichen der Anforderungen“ (Seite 144) repräsentiere. Tatsächlich ist dieses Bewertungsschema für die Außendarstellung von zentraler Bedeutung, da es die Bewertung der QS-Ergebnisse im Qualitätsbericht der Krankenhäuser wiedergibt. Es suggeriert jedoch die Sicherheit einer Tatbestandsfeststellung, die keineswegs gegeben ist. Dies wird schon alleine deshalb deutlich, da auch die Nichtwahrnehmung einer Möglichkeit zur Stellungnahme durch den Leistungserbringer und „teilweise durch Leistungserbringer zu verantwortende Datenfehler“ automatisch als „Qualitätsdefizit“ klassifiziert werden sollen.

#### *Nachberechnung der Indikatorergebnisse*

Im PlanQI-Verfahren wird gegenwärtig eine Neuberechnung der Indikatorergebnisse auf Basis von stichprobenhaften Datenvalidierungen durchgeführt. Das IQTIG hat in keinem seiner früheren Berichte wissenschaftlich dargelegt, dass die Vorgehensweise, das Ergebnis einer Vollerhebung auf Basis einer Stichprobenziehung zu korrigieren, methodisch angemessen bzw. statthaft ist.

Das Konzept des vorliegenden Berichts sieht nun vor, für bei einem Indikator auffällige Leistungserbringer partielle Nachberechnungen auf Basis der Rückmeldungen der Leistungserbringer im Stellungnahmeverfahren durchzuführen. Bei Ratenindikatoren können sich Zähler (interessierende Ereignisse, z. B. Komplikation) oder Nenner (eingeschlossene Grundgesamtheit des Indikators) ändern. Ergebnis soll ein so genannter informativer Indikatorwert sein. Da gleichzeitig auch der Referenzbereich individuell für jeden Leistungserbringer auf Basis der Ergebnisse „bereinigt“ werden soll, dürfte nicht selten das Ergebnis nach der „Bereinigung“ für das Krankenhaus noch ungünstiger sein als vorher (nachzuvollziehen anhand des fiktiven Beispiels in Abbildung 33 auf Seite 157). Die statistischen Auswertungen nach Neuberechnung eines Indikators und seines Referenzbereichs sind komplex und erfolgen unter Zeitdruck im Bewertungsverfahren. Bei komplexeren Indikatoren als einfachen Raten dürfte die Unanschaulichkeit der Ergebnisse zunehmen. Viele praktische Fragen bleiben offen. Kommt beispielsweise bei einem Indikator

mit einer statistischen Auffälligkeit über zwei Jahre auch eine Neuberechnung über zwei Jahre hinweg zur Anwendung?

In jedem Fall wäre für die Neuberechnung ein geeignetes Softwaretool unabdingbar, das den auf Landesebene beauftragten Stellen vom IQTIG zur Verfügung gestellt werden müsste.

#### *Peer Review*

An mehreren Stellen wird im Bericht auf das Ärztliche Peer Review Bezug genommen. Erfolgsfaktoren für dieses Instrument des kollegialen Dialogs wie Freiwilligkeit, Sanktionsfreiheit, fachliche Kompetenz der Peers und einheitliches Vorgehen werden zu Recht hervorgehoben (Kapitel 3.6). Allerdings ergeben sich Hinweise, dass bezüglich der Funktion eines Peer Reviews Missverständnisse bestehen. So wird auf Seite 23 eine Ähnlichkeit mit dem Strukturierten Dialog angedeutet, „auffällige Ergebnisse“ mit dem Leistungserbringer zu klären. Das Ärztliche Peer Review ist aber keinesfalls ein Qualitätsbewertungsinstrument, sondern ein Instrument des Qualitätsmanagements bzw. der Qualitätsförderung. Indikatorergebnisse können als Aufgreifkriterien dienen, müssen es aber nicht. Im Zentrum des kollegialen Dialogs steht nicht die Erörterung der Ergebnisse klinischer Einzelfälle, sondern die Betrachtung der Prozesse, die der Behandlung der Patientinnen und Patienten zugrunde liegt.

Auf Seite 174 wird vorgeschlagen, „diejenigen Leistungserbringer, bei denen mit hoher Sicherheit davon ausgegangen werden kann, dass sie die Qualitätsanforderungen erfüllt haben, zu identifizieren und als Peers für qualitätsfördernde Maßnahmen bei anderen Leistungserbringern zu gewinnen.“ Auch diese Überlegung, dass im Sinn eines „Lernen von den Besseren“ die besuchenden Peers „die Besseren“ seien, entspricht nicht dem Grundgedanken des Ärztlichen Peer Reviews. „Peers“ sind schon dem Wort nach Kolleginnen und Kolleginnen, die sich auf Augenhöhe begegnen. Im Peer Review kann auch der besuchende Peer vom besuchten Peer lernen. Zudem kann er oder sie selbst bei einem anderen Peer Review zum besuchten Peer werden.

#### 4. Fazit

Der Auftrag des G-BA an das IQTIG vom 18.02.2018 zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern enthielt einen umfangreichen Katalog von Aufgaben zur Optimierung der Einheitlichkeit der Vorgehensweise, der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung und der Effizienz des Verfahrens. Zudem sollte ein Rahmenkonzept für ein „Verfahren zur Feststellung von Auffälligkeiten in Krankenhäusern, zu deren Bewertung sowie zur Initiierung von Maßnahmen“ entwickelt werden. Die meisten dieser Aufgaben wurden im vorliegenden Bericht abgearbeitet, einige allerdings nur oberflächlich gestreift. Der Schwerpunkt des Berichts liegt auf der Qualitätsbewertung. Zur Qualitätsförderung hätte man sich substantziellere Vorschläge gewünscht.

Die Analyse des IQTIG zum Ist-Zustand des Strukturierten Dialogs gemäß QSKH-RL ist weitgehend nachvollziehbar. Betont werden die Heterogenität der Verfahren in den einzelnen Bundesländern und der Aufwand für die beteiligten Krankenhäuser und LQS. Letzterer ist durch eine niedrige Spezifität des Verfahrens bedingt.

Zur Erhöhung der Spezifität wird vom Institut eine statistische Methodik abgeleitet, nach der nicht mehr nur rechnerische, sondern statistisch signifikante Auffälligkeiten ausgewiesen werden sollen, bei Bedarf über zwei Erfassungsjahre statt über nur ein Erfassungsjahr. Die Ableitung der statistischen Methodik nimmt im Bericht einen dominanten Teil ein. Hier wäre die Auslagerung der mathematisch vertiefenden Ausführungen in eine Anlage zielführender gewesen. Die vorgeschlagene Berechnungsmethode erscheint - zumindest für die Beispielindikatoren und- auf theoretischer Ebene - plausibel. Eine Erhöhung der Spezifität der Indikatorergebnisse würde eine Reduzierung des Aufwands für alle Beteiligten bedeuten. Allerdings wird die vom Institut postulierte Grundannahme für die mathematische Modellierung, mit Qualitätsindikatoren ließe sich die „Kompetenz“ der Leistungserbringer messen, von der Bundesärztekammer nicht geteilt. Es existiert auch keine Grundlage in der Literatur dafür.

Da im Bericht nur wenige Beispiele für die neue Methodik der Indikatoreauswertung benannt werden, lässt sich noch keine Aussage darüber treffen, ob wirklich die gesamte datengestützte Qualitätssicherung von den neuen Methoden profitieren würde. Die Anschaulichkeit und Nachvollziehbarkeit der ermittelten Werte sollte Priorität haben.

Als methodische Grundlage für eine Neukonzeption des Strukturierten Dialogs gemäß QSKH-RL bzw. des Stellungnahmeverfahrens gemäß DeQS-RL sollen Forschungsergebnisse zur ärztlichen Urteilsbildung dienen. Eine statistische bzw. mechanistische Urteilsbildung sei der intuitiven Urteilsbildung hinsichtlich ihrer Treffgenauigkeit überlegen. Ob sich die zitierten Forschungsergebnisse wirklich 1:1 auf die spezielle Situation des Strukturierten Dialogs übertragen lassen, muss jedoch kritisch hinterfragt werden.

Das bis heute gültige Konzept, dass Qualitätsindikatoren zunächst als Aufgreifkriterien dienen und erst im kollegialen, strukturierten Dialog der Expertinnen und Experten mit den Leistungserbringern die wahren Qualitätsprobleme identifiziert werden, wird im Bericht quasi als obsolet erklärt. Stellungnahmeverfahren sollen nach Vorstellung des Instituts zukünftig nach expliziten, d. h. detailliert vorgegebenen Regeln erfolgen. Inhalt des Stellungnahmeverfahrens sollen alleine die Indikatorergebnisse sein. Dabei soll der Kontext, in dem diese Ergebnisse entstanden sind, nicht mehr in die Bewertung einfließen. Die Leistungserbringer sollen für jeden ihrer behandelten Fälle angeben, ob individuelle Risikofaktoren vorgelegen haben. Die Expertinnen und Experten sollen dann nach detaillierten Vorgaben für jeden Indikator und jede Vorgangsnummer überprüfen, ob eine Neuberechnung des Indikators gerechtfertigt ist. Lediglich für eine Restkategorie nicht in

Rechenregeln umsetzbarer Begründungen wird den Expertinnen und Experten ein gewisser heuristischer Beurteilungsspielraum eingeräumt.

Dieser Paradigmenwechsel im Strukturierten Dialog bzw. im Stellungnahmeverfahren wird von der Bundesärztekammer –zumindest in diesem Ausmaß – abgelehnt. Ein solcher mechanistischer Ansatz würde das Bewertungsverfahren auf einen abzuarbeitenden Algorithmus reduzieren. Ein eigentlicher Dialog mit den betroffenen Einrichtungen soll nicht mehr stattfinden. Zum Ausdruck kommt ein erschreckend reduktionistisches Qualitätsverständnis, bei dem nur noch Indikatorergebnisse, aber nicht mehr die zugrunde liegenden Prozesse und Strukturen betrachtet werden sollen. Letzteres angemessen zu berücksichtigen, ist jedoch genau die Stärke des Strukturierten Dialoges und sollte auch nicht mit „Heterogenität“ verwechselt werden, die es um jeden Preis zu beseitigen gelte.

Die vorgeschlagene Trennung des Strukturierten Dialogs in zwei unterschiedliche, nacheinander abzuarbeitende Module Qualitätsbewertung und Qualitätsförderung scheint vordergründig im Sinne einer klaren Aufgabentrennung vorteilhaft zu sein. Da dadurch allerdings in der Qualitätsbewertung keinerlei Raum mehr für einen kollegialen Dialog vor der abschließenden Bewertung gelassen wird, wird diese strikte Trennung von der Bundesärztekammer abgelehnt.

Ein weiteres und dabei grundlegendes Problem der Anwendbarkeit des neuen Konzepts besteht durch die Historie des Auftrags. Das Konzept ist formuliert in Anlehnung an den Strukturierten Dialog für den Krankenhausbereich. Inwieweit die Vorschläge für Bewertungsverfahren auch im vertragsärztlichen Bereich gelten können, wird gar nicht thematisiert.

Zusammenfassend enthält der IQTIG-Vorbericht einige bedenkenswerte Vorschläge v. a. zur methodischen Weiterentwicklung des Auswertungsverfahrens. Ein erster Schritt sollte allerdings die Überprüfung der bestehenden Indikatoren auf ihre Güte und ggf. die Streichung von Indikatoren bzw. die Neuentwicklung geeigneter valider Indikatoren sein. Das im Bericht vorgelegte Konzept betont jedoch einseitig die nachgelagerten Schritte der Auswertung und Bewertung mit dem offensichtlichen Ziel, mehr Validität ex post zu erreichen. Folgt man den vorgeschlagenen Änderungen im Bewertungsverfahren, würde am Ende das Kernstück des Strukturierten Dialogs – nämlich der Dialog – weitgehend abgeschafft sein.

Die Bundesärztekammer rät von einer Umsetzung der Änderungsvorschläge zum Bewertungsverfahren aufgrund der vielen offenen Fragen derzeit ab.

## **IQTIG Projekt: Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern (Stufe 1 und 2)**

### **Stellungnahme DGVS-11-2019**

Thomas Frieling

Ludger Leifeld

Petra Lynen

### **Hintergrund**

Mit Beschluss vom 18. Januar 2018 hat der Gemeinsame Bundesausschuss (G-BA) das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) mit der *Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern* beauftragt. Dieser Weiterentwicklungsauftrag ist in zwei Stufen mit unterschiedlichen Inhalten und Abgabefristen gegliedert. Der Abschlussbericht zu Stufe 1 wurde dem G-BA am 30. November 2018 übermittelt. Der vorliegende Bericht beinhaltet sowohl die weiterentwickelten Empfehlungen des Berichts zu Stufe 1 als auch die für Stufe 2 beauftragten Inhalte.

Als Strukturierter Dialog werden diejenigen Prozesse bezeichnet, die gemäß §§ 11 bis 14 QSKH-RL eingeleitet werden sollen, wenn ein Leistungserbringer in einem Qualitätsindikator oder einem Auffälligkeitskriterium der QSKH-RL vom Referenzwert abweicht. Laut § 12 QSKH-RL soll dadurch geprüft werden, ob die dokumentierten Leistungen in der fachlich gebotenen Qualität erbracht und korrekt dokumentiert wurden.

Die Verantwortung für die Durchführung des Strukturierten Dialogs liegt bei der jeweils mit der Durchführung des Verfahrens beauftragten Stelle. Dies sind der *Unterausschuss Qualitätssicherung des G-BA* (direkte Verfahren) und für den Großteil der QS-Verfahren (indirekte Verfahren) die *Landesgeschäftsstellen für Qualitätssicherung (LQS)* bzw. die *Landesarbeitsgemeinschaften (LAG)* in den Bundesländern.

Analysen der Ergebnisse des Strukturierten Dialogs zeigen

- eine deutliche Heterogenität im Vorgehen und in den Ergebnissen zwischen den beauftragten Stellen und
- einen erheblichen Aufwand der beauftragten Stellen mit steigender Anzahl von QS-Verfahren und Indikatoren

### **Zielauftrag**

Das IQTIG wurde daher beauftragt, Empfehlungen für die Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern zu entwickeln:

1. Optimierung der Einheitlichkeit der Vorgehensweise
2. Optimierung der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung
3. Optimierung der Effizienz des Verfahrens
4. Entwicklung eines Rahmenkonzepts, das die Elemente des Strukturierten Dialogs weiterentwickelt und mit anderen relevanten Richtlinien abstimmt

Die Beauftragung verlangt auch, die auf Landesebene mit der Durchführung des Strukturierten Dialogs nach QSKH-RL beauftragten Stellen (LQS) und soweit wie möglich die Landesarbeitsgemeinschaften (LAG) in die Weiterentwicklung einzubeziehen. **Der Workshop fand am 27. Mai 2019 in den Räumlichkeiten des IQTIG statt. Insgesamt nahmen 19 LAG- und LQS-Vertreterinnen und -vertreter aus 14 Bundesländern sowie die beim IQTIG für die direkten Verfahren zuständigen Personen am Workshop teil.**

Gemäß § 137a Abs. 7 SGB V sind bei Entwicklungsarbeiten des IQTIG verschiedene Organisationen, Verbände und Interessenvertretungen des Gesundheitswesens zu beteiligen. Um dieser gesetzlichen Pflicht nachzukommen, führt das IQTIG ein **Stellungnahmeverfahren** für den gemeinsamen Abschlussbericht zu Stufe 1 und Stufe 2 der Beauftragung durch. Die folgenden Organisationen, Verbände und Interessenvertretungen erhielten vom 10. Oktober 2019 bis zum 21. November 2019 die Gelegenheit, schriftlich Stellung zu den Inhalten dieses Berichts zu nehmen:

- Kassenärztliche Bundesvereinigung
- Deutsche Krankenhausgesellschaft
- Spitzenverband Bund der Krankenkassen
- Verband der Privaten Krankenversicherung
- Bundesärztekammer
- Bundeszahnärztekammer
- Bundespsychotherapeutenkammer
- Berufsorganisationen der Krankenpflegeberufe
- **die wissenschaftlichen medizinischen Fachgesellschaften**
- Deutsches Netzwerk Versorgungsforschung
- die für die Wahrnehmung der Interessen der Patientinnen und Patienten und der Selbsthilfe chronisch kranker und behinderter Menschen maßgeblichen Organisationen auf Bundesebene
- Beauftragter der Bundesregierung für die Belange der Patientinnen und Patienten
- zwei von der Gesundheitsministerkonferenz der Länder zu bestimmende Vertreter
- Bundesoberbehörden im Geschäftsbereich des Bundesministeriums für Gesundheit, soweit ihre Aufgabenbereiche berührt sind
- die mit der Durchführung der QS-Verfahren gemäß QSKH-RL sowie DeQS-RL beauftragten Stellen auf Landesebene

Die eingegangenen Stellungnahmen werden vom IQTIG schriftlich gewürdigt, und der Bericht wird auf Basis der Stellungnahmen auf Anpassungsbedarf hin geprüft.

## **Empfehlungen des IQTIG**

### **Ermittlung einer statistischen Auffälligkeit**

- Die fallzahlabhängigen Qualitätsindikatoren der DeQS-RL werden nach einer analytischen Herangehensweise berechnet und ausgewertet, d. h. durch stochastische Einflüsse entstehende Unsicherheit ist bei der Berechnung der Indikatorergebnisse und beim Abgleich mit dem Referenzbereich zu berücksichtigen.
- Der Bewertungsprozess der QI-Ergebnisse erfolgt jährlich. Dabei ist zwischen der Berichterstattung der Ergebnisse eines Erfassungsjahres und dem Verfahren zur Bestimmung des hinreichenden Hinweises für eine Abweichung vom Referenzbereich zu unterscheiden.
- Ergibt sich anhand der vorliegenden Daten eines Leistungserbringers kein hinreichender Hinweis auf ein Qualitätsdefizit, wird das Indikatorergebnis mit „kein hinreichender Hinweis auf ein Qualitätsdefizit“ (U0) bewertet.
- Der hinreichende Hinweis auf ein Qualitätsdefizit wird als vorgegebener Schwellenwert (entsprechend dem Signifikanzniveau) für die Posteriori-Verteilung der Null-Hypothese im Rahmen eines bayesianischen Hypothesentests statistisch operationalisiert.
- Ergibt sich anhand der vorliegenden Daten eines Leistungserbringers ein hinreichender Hinweis auf ein Qualitätsdefizit, erhält der Leistungserbringer die Möglichkeit zur Stellungnahme.
- Die Funktion des Stellungnahmeverfahrens ist es, zu prüfen, ob Gründe für die statistische Abweichung vom Referenzbereich vorliegen, die nicht vom Leistungserbringer zu verantworten sind. Bei der Prüfung sind in der Regel nur Gründe, die nicht bereits in der Rechenregel des Indikators oder in dem Risikoadjustierungsmodell des Indikators berücksichtigt sind, zu beachten.
- Um eine höhere Treffsicherheit bei der statistischen Operationalisierung des hinreichenden Hinweises auf ein Qualitätsdefizit zu gewährleisten, soll zukünftig eine Mehrjahresbetrachtung bei der Klassifikation der Leistungserbringerergebnisse stattfinden.

### **Formaler Ablauf des Stellungnahmeverfahrens**

- Das Anforderungsschreiben für eine Stellungnahme wird am 30. Juni von der zuständigen Stelle an die Leistungserbringer versendet.
- Der Leistungserbringer erhält ab Eingang des Anforderungsschreibens vier Wochen Zeit, um eine qualifizierte Stellungnahme bei der zuständigen Stelle einzureichen.
- Belege, die die Argumentation des Leistungserbringers unterstützen, können in angemessenem Umfang ergänzend zur Stellungnahme eingereicht werden, ersetzen diese jedoch nicht. Belege, die ohne erläuternden Text übermittelt werden, sollen nicht anerkannt werden.

- Die fachlich verantwortliche Instanz des Leistungserbringers (z. B. die Chefärztin/der Chefarzt, Praxisinhaberin/Praxisinhaber) ist über den Hinweis auf ein Qualitätsdefizit zu informieren und sie soll die angefertigte Stellungnahme autorisieren.
- Die Anonymität sowohl der Patientinnen und Patienten (personenidentifizierende Daten) als auch ggf. weiterer Beteiligter (z. B. Zuweiser) ist zu wahren. Bei Nichteinhaltung der Vorgaben zur Wahrung der Anonymität soll die Stellungnahme aus Datenschutzgründen vernichtet werden und der Leistungserbringer über diesen Vorgang benachrichtigt werden.
- Der Leistungserbringer bestätigt die Korrektheit der übermittelten Daten oder erklärt, welche Fehler vorliegen und inwieweit diese von ihm zu vertreten sind.
- Der Leistungserbringer soll sowohl für einzelfallbezogene als auch für einzelfallübergreifende Gründe angeben, auf welche Behandlungsfälle diese Gründe zutreffen.
- Für die Zuordnung der Gründe zu den Behandlungsfällen sind die Vorgangsnummern der entsprechenden Behandlungsfälle mit der Stellungnahme zu übermitteln. Gleiches gilt auch für diejenigen Behandlungsfälle, bei denen kein interessierendes Ereignis (Behandlungsfälle, die nur im Nenner des Indikators vorkommen) aufgetreten ist.
- Die Erfüllung der formalen Kriterien für Stellungnahmen prüft und dokumentiert die zuständige Stelle bevor sie die Stellungnahme an die Fachkommission weiterleitet.

### **Fachliche Bewertung**

- Die Prüfung, ob Gründe für die statistische Auffälligkeit des Leistungserbringers vorliegen, die nicht vom Leistungserbringer zu vertreten sind und nicht schon in der Rechenregel des Indikators oder in dem Risikoadjustierungsmodell des Indikators berücksichtigt sind, heißt „fachliche Bewertung“ und wird von einer Fachkommission bei der zuständigen Stelle auf Basis der schriftlichen Stellungnahme des Leistungserbringers durchgeführt.
- Begehungen und Besprechungen bzw. kollegiale Gespräche werden ausschließlich im Rahmen der Qualitätsförderung eingesetzt.
- Die Fachkommission klassifiziert die angeführten Gründe dahingehend, ob sie vom Leistungserbringer zu verantworten sind oder nicht.
- In der Stellungnahme angeführte Gründe, bei denen aus der Stellungnahme nicht erkenntlich ist, ob sie nur bei den Behandlungsfällen mit einem interessierenden Ereignis oder auch bei Fällen ohne interessierendes Ereignis vorgelegen haben, werden in der fachlichen Bewertung nicht berücksichtigt.
- In der Stellungnahme angeführte Gründe, die aus Sicht des Leistungserbringers nicht von ihm zu verantworten waren, sollen (mittels schriftlichem Beleg) nachgewiesen werden.
- Die Fachkommission führt auf Basis der Klassifikation der angeführten Gründe mittels technischer Unterstützung eine Nachberechnung des Indikatorergebnisses für

den Leistungserbringer durch. Dabei werden die Fälle, bei denen die Fachkommission zu der Einschätzung gelangt, dass die angeführten Gründe nicht vom Leistungserbringer zu verantworten sind, aus der Grundgesamtheit des Indikators ausgeschlossen.

- Liegt nach der Nachberechnung kein hinreichender statistischer Hinweis auf ein Qualitätsdefizit vor, dann erfolgt eine Bewertung des Indikatorergebnisses mit „Kein Hinweis auf Qualitätsdefizit“ (U).
- Liegt nach der Nachberechnung ein hinreichender statistischer Hinweis auf ein Qualitätsdefizit vor, prüft die Fachkommission, ob weitere, nicht eindeutig klassifizierbare Gründe vom Leistungserbringer geltend gemacht wurden.
- Wurden vom Leistungserbringer weitere, nicht eindeutig klassifizierbare Gründe geltend gemacht, beurteilt die Fachkommission auf Basis ihrer fachlichen Expertise, ob diese Gründe hinreichend sind, um die statistische Auffälligkeit nach der Nachberechnung zu erklären.
- Wenn die Fachkommission zu der Einschätzung gelangt, dass die angeführten Gründe hinreichend sind, um die statistische Auffälligkeit zu erklären, empfiehlt sie das Indikatorergebnis als „Kein Hinweis auf Qualitätsdefizit“ (U) zu bewerten. Die Empfehlung zur Einstufung in die Ziffern 1 bis 3 erfolgt auf Basis der fachlichen Expertise der Fachkommission.
- Wenn die Fachkommission zu der Einschätzung gelangt, dass die angeführten Gründe nicht hinreichend sind, um die statistische Auffälligkeit nach der Nachberechnung zu erklären, empfiehlt sie das Indikatorergebnis als „Qualitätsdefizit“ (A) zu bewerten. Die Empfehlung zur Einstufung in die Ziffern 1 bis 3 erfolgt auf Basis der fachlichen Expertise der Fachkommission.
- Die Fachkommission empfiehlt der zuständigen Stelle eine abschließende Bewertung des Indikatorergebnisses gemäß dem Bewertungsschema einschließlich der Einstufung in die Ziffern.
- Die zuständige Stelle trifft die abschließende Entscheidung in Bezug auf die Einordnung des Indikatorergebnisses in das Bewertungsschema.
- Weicht die zuständige Stelle von der Einstufungsempfehlung der Fachkommission ab, begründet und dokumentiert sie diese Abweichung.
- Mit der abschließenden Entscheidung in Bezug auf die Einordnung des Indikatorergebnisses in das Bewertungsschema ist die Qualitätsbewertung abgeschlossen.
- In Ausnahmefällen kann die zuständige Stelle das Indikatorergebnis eines Leistungserbringers mit „Sonstiges – ohne Bewertung“ (S) bewerten. Der Sachverhalt, der eine Qualitätsbewertung unmöglich machte, ist durch die zuständige Stelle anzugeben.

### **Zusammensetzung der Fachkommissionen**

- Die zuständige Stelle trägt die Verantwortung für die Auswahl der Expertinnen und Experten sowie die Zusammensetzung der Fachkommission.
- Die Fachkommissionen sollen interdisziplinär zusammengesetzt werden.
- In Abhängigkeit vom jeweiligen QS-Verfahren sind folgende Berufsgruppen einzubinden: Fachärztinnen und Fachärzte, Pflegefachkräfte, Klinische Qualitätsmanagerinnen und Qualitätsmanager, Patientenvertreterinnen und -vertreter, weitere Angehörige der Gesundheitsberufe, Krankenhaushygienikerinnen und -hygieniker oder Hygienefachkräfte

### **Qualitätsfördernde Maßnahmen**

- Qualitätsfördernde Maßnahmen schließen sich nur an eine abgeschlossene Qualitätsbewertung an.
- Die Fachkommission empfiehlt unter Berücksichtigung des Indikatorergebnisses und seiner abschließenden Bewertung den Einsatz von Fördermaßnahmen.
- Die zuständige Stelle leitet ab dem zweiten mit „Qualitätsdefizit“ (A) bewerteten Indikatorergebnis bei dem gleichen oder einem inhaltlich ähnlichen Indikator innerhalb von drei Jahren immer qualitätsfördernde Maßnahmen ein.
- Es liegt im Ermessen der zuständigen Stelle unter Angabe von Gründen bereits bei einem mit „Qualitätsdefizit“ (A) bewerteten Indikatorergebnis qualitätsfördernde Maßnahmen bei einem Leistungserbringer einzuleiten.
- Die zuständige Stelle entscheidet über die Art und Einleitung von notwendigen Maßnahmen gemäß DeQS-RL und QFD-RL.
- Begehungen und Besprechungen bzw. kollegiale Gespräche werden der Maßnahmenstufe 1 zugeordnet.
- Der Einsatz und Umsetzungsgrad der qualitätsfördernden Maßnahmen ist zu dokumentieren.
- Die zuständige Stelle hat dem Lenkungsgremium regelmäßig über den Stand der qualitätsfördernden Maßnahmen zu berichten.

### **Stellungnahme der DGVS (November 2019)**

Die Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS) unterstützt grundsätzlich die Aktivitäten des IQTIG zur Entwicklung von Empfehlungen für die Weiterentwicklung und Straffung des Strukturierten Dialogs mit Krankenhäusern. Hierbei stimmt die DGVS mit den formalen Inhalten der Empfehlungen bezüglich der *Ermittlung einer statistischen Auffälligkeit* und des *Formalen Ablaufs des Stellungnahmeverfahrens* überein. Wir verweisen aber inhaltlich auf unsere Stellungnahme zum IQTIG Projekt Qualitätsorientierte Vergütung vom 04. Dezember 2018. So sieht die DGVS es weiterhin als kritisch an, mit Bündeln von unterschiedlichen Qualitätsindikatoren die Qualität einer Abteilung, Klinik bzw. Krankenhaus sachgerecht und umfassend zu beschreiben, und sie hat anhand der Literaturlage eine gesunde Skepsis in der Einschätzung

der Potenz von Qualitätsindikatoren. Ebenfalls erscheint der administrative Aufwand des Ansatzes, auch nach Strukturierung durch das IQTIG, weiterhin sehr hoch, sowohl für die Leistungserbringer als auch für die Planungsbehörden.

Die DGVS stimmt mit den im Protokoll zum *Workshop zur Weiterentwicklung des Strukturierten Dialogs* aufgeführten Kritikpunkten bzw. Fragen grundsätzlich überein. Darüber hinaus sieht die DGVS, dass im Ablauf des Strukturierten Dialogs die Fachkommissionen bei der *Fachlichen Bewertung* und den *Qualitätsfördernden Maßnahmen* eine besondere Stellung einnehmen, da sie die Auswahl der Indikatoren und die Bewertung der Qualitätsindikatorergebnisse vornehmen und Qualitätssicherungsmaßnahmen anstoßen. Bisher werden die fachkundigen Mitglieder in Bezug auf ihre Qualifikation ausschließlich nach Kenntnissen aus dem Bereich Qualitätsmanagement und Qualitätssicherung gewählt. Auf welchem Gebiet die Mitglieder darüber hinaus fachkundig sein sollten, ist bisher nicht näher beschrieben. Folglich ist unklar, über welche genaue Qualifikation die einzelnen Mitglieder verfügen sollen. Hier muss eine Definition der Qualitätskriterien unter Einbindung der wissenschaftlich medizinischen Fachgesellschaften entwickelt werden. Zudem ist die Zusammensetzung der Fachgruppen und die Einbindung in Entscheidungsprozesse regional sehr heterogen. Hier müssen verbindliche Rahmenbedingungen und Standards erarbeitet werden. Die DGVS unterstützt aus diesem Grunde insbesondere die Empfehlung des IQTIG, nach der die Zusammensetzung der Fachkommissionen strukturiert und interdisziplinär erfolgen und Berufsgruppen wie Fachärztinnen und Fachärzte, Pflegefachkräfte, klinische Qualitätsmanagerinnen und Qualitätsmanager und Patientenvertreterinnen und -vertreter eingebunden werden sollen. Hier ist die Beteiligung der jeweiligen Fachgesellschaften aufgrund ihrer Expertise ebenfalls notwendig.

Insbesondere ist bei den *Qualitätsfördernden Maßnahmen* die Expertise der Fachgesellschaften gefordert. Hierbei unterstützt die DGVS die vom IQTIG vorgeschlagenen Instrumente wie die Teilnahme an Fortbildungen, Fachgesprächen und Kolloquien, die Teilnahme an Qualitätszirkeln, die Implementierung von Behandlungspfaden, die Durchführung von Audits und insbesondere die Durchführung von Peer Reviews bzw. die Implementierung von Handlungsempfehlungen anhand von Leitlinien. Wichtig ist in diesem Zusammenhang auch die Definition der Qualifikation der Auditoren bzw. Peers und die verwendeten Auswahlkriterien.

Für die Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS)

Prof. Dr. med. Thomas Frieling

Prof. Dr. med. Ludger Leifeld

Priv. Doz. Dr. med. Petra Lynen Jansen



Repräsentanz der DGGG und Fachgesellschaften · Hausvogteiplatz 12 · 10117 Berlin

Repräsentanz der DGGG  
und Fachgesellschaften  
Hausvogteiplatz 12  
10117 Berlin  
Telefon: 030 / 514 88 33 33  
Telefax: 030 / 514 88 344  
E-Mail: [stellungnahmen@dggg.de](mailto:stellungnahmen@dggg.de)

12. November 2019

## **Stellungnahme der Deutschen Gesellschaft für Gynäkologie und Geburtshilfe e.V. (DGGG) und der Deutschen Gesellschaft für Senologie e.V. (DGS)**

zum IQTIG-Vorbericht

### **„Weiterentwicklung des strukturierten Dialogs mit Krankenhäusern – Vorbericht zur Stufe 1 und Stufe 2“ (Stand: 10.10.2019)**

#### **Für den Fachbereich Mammachirurgie**

Sehr geehrte Damen und Herren,

wir bedanken uns für die Zusendung des Vorberichts „Weiterentwicklung des strukturierten Dialogs mit Krankenhäusern – Vorbericht zur Stufe 1 und Stufe 2“, zu dem wir sehr gerne Stellung nehmen und einige Aspekte fokussieren möchten.

Zunächst einmal ist es beeindruckend, wie heterogen der Strukturierte Dialog in den einzelnen Bundesländern geregelt ist. Dementsprechend ist eine klare Struktur mit Transparenz essentiell. Die Weiterentwicklung wird aus Sicht der Fachgesellschaften dementsprechend begrüßt. Es müssen klare und gleiche Kriterien für Stellungnahmen, Besprechungen und Begehungen aller Bundesländer vorliegen, so dass auch für den Leistungserbringer eine Transparenz besteht. Zudem ist es aus Sicht der Fachgesellschaften wesentlich, bei der Weiterentwicklung die Ressourcen der Leistungserbringer im Auge zu behalten. Ein ressourcensparendes Vorgehen ist für die Kliniken essentiell.

Es wird so auch auf Seite 66 des Vorberichts (oben) genannt, dass eine Optimierung der Objektivität der Qualitätsbewertung zur Folge hat, dass in der Regel eine Aufwandsreduktion einhergeht. Dieses ist durchweg zu begrüßen.

Allerdings möchten wir noch auf einzelne Aspekte näher eingehen:

*Vorbericht Seite 69 oben: „...wird seitens des Leistungserbringers auf eine Stellungnahme verzichtet, gilt gemäß des Verständnisses von Qualitätsindikatoren (s. Abschnitt 3.1), dass ein hinreichender Hinweis für ein Qualitätsdefizit vorliegt, der Indikator in seiner Validität nicht angezweifelt wird und somit ein Qualitätsdefizit vorliegt (s. Abbildung 9)...“*

Diesen Schritt können wir aus Sicht der Fachgesellschaften nicht mitgehen. Es gab zahlreiche Rückmeldungen von Leistungserbringern, dass sie nicht über die Notwendigkeit einer Stellungnahme informiert worden sind. Dieses lag z.T. daran, dass die entsprechenden Beauftragten in den Kliniken im Urlaub oder krank waren. So wurden einzelne Chefärztinnen und Chefarzte nicht über die Notwendigkeit einer Abgabe einer



Stellungnahme informiert. Dieses ist somit nicht gleichbedeutend mit einem Hinweis auf ein Qualitätsdefizit. Es können durchaus strukturelle oder personelle Probleme in einzelnen Kliniken vorliegen, die dazu führen, dass der Leistungserbringer nicht über die Notwendigkeit der Stellungnahme informiert wurde. Somit ist es aus Sicht der Fachgesellschaften notwendig, zunächst zu prüfen, wieso auf eine Stellungnahme verzichtet wird. Zumindest eine Bestätigung des Leistungserbringers, dass bewusst auf eine Stellungnahme verzichtet wird, ist essentiell.

*Vorbericht Seite 145 - Tabelle 13: Ergebnis der Qualitätsbewertung A Qualitätsdefizit Ziffer 0: „Verzicht auf eine Einreichung einer Stellungnahme oder Stellungnahme nicht fristgerecht eingereicht oder Stellungnahme entsprach nicht den Anforderungen.“*

Hier sollte der Verzicht auf die Einreichung einer Stellungnahme separat von der Beurteilung, dass eine Stellungnahme nicht den Anforderungen entsprach, geführt werden. Beides in der gleichen Kategorie zu führen, ermöglicht keine getrennte Auswertung und schafft in diesem Fall keine entsprechende Transparenz.

*Vorbericht Seite 145: Ergebnis der Qualitätsbewertung A Qualitätsdefizit, Ziffer 1: „Vom Leistungserbringer zu verantwortende Datenfehler haben eine Fehlmessung verursacht.“*

Die Ziffer ist selbstverständlich adäquat, es fehlt aber eine Ziffer für die Fälle, bei denen der Datenfehler nicht vom Leistungserbringer zu verantworten ist, sondern durch das System bedingt ist. Dieses tritt immer wieder auf und sollte adäquat abgebildet sein.

*Vorbericht Seite 163: „... wie in Abschnitt 3.5 angedeutet, sollte für die Umsetzung der Empfehlungen für die fachliche Bewertung eine Kosten-Nutzen-Abwägung durchgeführt werden. Der Nutzen eines möglichst expliziten Vorgehens in Form größtmöglicher Objektivität und Transparenz muss gegenüber den Kosten, die die Umsetzung des Vorgehens bedeuten würde, abgewogen werden...“*

Die grundsätzliche Kosten-Nutzen-Abwägung wird von den Fachgesellschaften begrüßt. Allerdings sollte die Kosten-Nutzen-Abwägung auf jeden Fall auch die Ressourcen des Leistungserbringers berücksichtigen. Nur wenn alle Partner des Strukturierten Dialogs berücksichtigt werden, ist eine entsprechende Bewertung der Kosten gegenüber des Nutzens möglich.

*Vorbericht Seite 165: „... die Stellungnahme ist innerhalb von 4 Wochen nach Anforderung der Stellungnahme abzugeben...“*

Wie bereits vorne aufgeführt, ist aus unserer Sicht in dem Fall eine Empfangsbestätigung über die Aufforderung zur Stellungnahme notwendig. Sollte der Leistungserbringer nicht bestätigen, dass er die Aufforderung für eine Stellungnahme erhalten hat, muss dem nachgegangen werden. Ggf. sollte in dem Fall die Chefärztin bzw. der Chefarzt der Klinik direkt informiert werden.

*Vorbericht Seite 167: „... diese sollten insbesondere die in Abschnitt 3.6 beschriebenen Erfolgsfaktoren (fachliche Kompetenz, Unabhängigkeit und Interdisziplinarität) bei der Zusammensetzung der Fachkommission berücksichtigen. Daher wird empfohlen, eine fachlich qualifizierte, unabhängige und interdisziplinäre Fachkommission zusammenzusetzen, um die Akzeptanz des Gesamtverfahrens sicher zu stellen und Qualitätsdefizite erfolversprechend abzubauen. Die Verantwortung für die*



*Zusammensetzung einer fachlich qualifizierten, unabhängigen und interdisziplinären Fachkommission soll zukünftig bei der zuständigen Stelle liegen...“*

Grundsätzlich sind die Kriterien für die Besetzung der Fachkommissionen aus Sicht der Fachgesellschaften adäquat. Allerdings wäre es zu empfehlen, hier die Fachgesellschaften bei der Auswahl der entsprechenden Mitglieder der Fachkommissionen zu berücksichtigen. Zahlreiche Mitglieder der Fachgesellschaften erfüllen die entsprechenden Kriterien. So können die Fachgesellschaften bei der Auswahl der Mitglieder unterstützen.

*Vorbericht Seite 168: „...daher sollen die zu berufenen ärztlichen Fachexpertinnen und Fachexperten nach der Facharztausbildung die letzten 3 Jahre vor ihrem Mitwirken in der Fachkommission im entsprechenden Versorgungsbereich praktisch tätig gewesen sein...“*

Im Bereich der Senologie (Mammachirurgie) wirken drei Jahre sehr lang. Die Änderungen in der Diagnostik und Therapie sind rasend schnell. Somit empfehlen die Fachgesellschaften eine Zeitspanne von 2 Jahren zu nehmen, so dass die Fachexpertinnen und -experten noch ausreichend berufliche Expertise besitzen und die aktuellen Prozesse in den Kliniken kennen.

*Vorbericht Seite 180: „...eine weitere Komponente zur Förderung der Homogenität und der Effizienz des Bewertungsprozesses wäre eine IT-technische Unterstützung ... Dieses kann z.B. eine dem Bewertungsprozess angepasste Softwarelösung sein, welche einfache und komplexe Schritte, wie z.B. den Zugriff auf die Stellungnahmen und die Nachberechnung von Ergebnissen nach Berücksichtigung von besonderer Versorgungskonstellation für die Anwender erleichtert...“*

Aus Sicht der Fachgesellschaften ist dieses keine „Kann-“ sondern eine „Mussoption“. Die entsprechenden Softwarelösungen müssen vorhanden sein, um den Strukturierten Dialog zu homogenisieren und ressourcensparend durchzuführen. Zudem sollten diese Softwarelösungen für den Leistungserbringer kostenneutral sein.

*Vorbericht Seite 181: „...Nach Eingang der Rückmeldeberichte bei den zuständigen Stellen benötigen diese Zeit, die Aufforderungsschreiben zu erstellen und zu versenden, so dass die Aufforderung zur Stellungnahme eine Woche später (am 07. Juli) beim Leistungserbringer eingehen könnten...“*

Es ist zu berücksichtigen, dass in zahlreichen Bundesländern im Juli bereits die Sommerferien begonnen haben. Die Leistungserbringer haben in diesem Zeitraum eine reduzierte personelle Decke. Zudem sind manche ärztlichen Kolleginnen und Kollegen aufgrund der Ferien nicht verfügbar und ansprechbar (z.B. Leiterin/ Leiter des Brustzentrums, Leiterin/ Leiter des Perinatalzentrums, Chefärztin/ Chefarzt selbst). In diesem Fall ist eine Frist von 4 Wochen für die Erstellung der Stellungnahme äußerst knapp bemessen. Sollte somit die Stellungnahmezeit in die Ferienzeit fallen, wäre es zu begrüßen, eine Frist von 6 Wochen anstatt von 4 Wochen vorzugeben. Ansonsten würden erneut Limitationen für einen regelhaften Ablauf des Systems bestehen.

*Vorbericht Seite 192: „...als Ziele der Weiterentwicklung des Strukturierten Dialogs wurden definiert: Die Einheitlichkeit der Vorgehensweise, die Transparenz und Nachvollziehbarkeit der Entscheidungsfindung, die Effizienz im Sinne einer Verschlankung des Verfahrens...“*



Diese Ziele werden von den Fachgesellschaften begrüßt. Alle Ziele könnten zur Reduktion des Ressourcenaufwandes des Leistungserbringers führen und gleichzeitig die Qualität verbessern. Allerdings sollten die Evaluation und Reduktion der Kosten auch als direktes Ziel der Weiterentwicklung des Strukturierten Dialogs genannt werden. Die Verhältnismäßigkeit zwischen Aufwand und Nutzen ist wesentliche Grundlage der gesamten Weiterentwicklung des Strukturierten Dialogs.

Die Stellungnahme wurde durch Herrn Prof. Dr. Michael P. Lux verfasst.

Herzliche kollegiale Grüße

A handwritten signature in blue ink, appearing to read 'A. Scharl'.

Prof. Dr. Anton Scharl  
Präsident der DGGG e.V.

A handwritten signature in blue ink, appearing to read 'S. Brucker'.

Prof. Dr. Sara Yvonne Brucker  
Vorsitzende der DGS e.V.



Geschäftsstelle:  
 Friedrich-Wilhelm-Str. 2 · 53113 Bonn  
 Telefon +49 (0) 2 28 / 923 922-0  
 Telefax +49 (0) 2 28 / 923 922-10  
 Frau Ulrike Fischer  
 E-Mail: [info@hno.org](mailto:info@hno.org)  
 Homepage: <http://www.hno.org>

Deutsche Gesellschaft für HNO-Heilkunde, Kopf- und Hals-Chirurgie e.V.  
 Friedrich-Wilhelm-Str. 2 · 53113 Bonn

per E-Mail: [weiterentwicklung-sd@iqtig.org](mailto:weiterentwicklung-sd@iqtig.org)

**IQTIG**  
 Institut für Qualitätssicherung und  
 Transparenz im Gesundheitswesen  
 Katharina-Heinroth-Ufer 1  
 10787 Berlin

21. November 2019

**Stellungnahme der Deutschen Gesellschaft für HNO-Heilkunde, Kopf- und Hals-Chirurgie  
 an IQTIG und AWMF wegen Qualitätsparametern und Strukturiertem Dialog  
 „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“**

Sehr geehrte Damen und Herren,

das Präsidium der Deutschen Gesellschaft für HNO-Heilkunde, Kopf- und Hals-Chirurgie bedankt sich für die Möglichkeit zu dem Vorbericht „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“ wie folgt Stellung nehmen zu können:

Der Entwurf macht zunächst einmal gut deutlich, wie schwierig es ist (und auch weiterhin sein wird), valide Qualitätsindikatoren zu etablieren, die einen wirklichen Informationswert besitzen. Immerhin existiert der strukturierte Dialog bereits seit einigen Jahren und scheint nur wenig aussagekräftige Ergebnisse zutage gefördert zu haben. Die Gründe (Heterogenität des Patientenguts, Fallzahlabhängigkeit, fehlerhafte Daten) sind evident. Ein weiterer Faktor ist die uneinheitliche Methodik der statistischen Auswertung für die einzelnen Parameter, für den in dem Entwurf nun ein neues Rechenmodell vorgestellt wird, welches sich wird bewähren müssen. Es bleibt die Frage offen, ob das IQTiG sich der Erfahrungen der Initiative Qualitätsmedizin bedienen konnte oder hier zumindest einen Dialog gefunden hat. Hier konnten wir lediglich eine vermutliche Auseinandersetzung mit Literatur aus diesem Bereich finden.

Durch die Verbesserung der Methode sollen in Zukunft weniger falsch-positive Resultate auftreten, wodurch sich dann das anschließende Stellungnahmeverfahren verschlankt. Nach einer Phase der Praxiserprobung soll das System in der Weise scharf gestellt werden, dass statistisch-signifikante Schwellenwerte definiert werden, die ein Qualitätsdefizit ‚eindeutig‘ anzeigen und dann kein Stellungnahmeverfahren mehr erfordern, d.h. das Qualitätsdefizit eines Leistungserbringers kann dann unmittelbar und ohne Rücksprache veröffentlicht werden. Diese Methode der sozusagen computergenerierten Veröffentlichung von Qualitätsdefiziten halten wir für höchst problematisch, da dann ohne Plausibilisierung eine Krankenhausabteilung in den möglicherweise unberechtigten Vorwurf eines Qualitätsdefizites kommen würde. Hier muss unseres Erachtens eine Dialogmöglichkeit erhalten werden, auch wenn dieses Personal und Zeit erfordert. Auch ein Gericht stellt keine Urteile nur aufgrund einer operationalisierten Computeranalyse aus, hier sind Ansätze der künstlichen Intelligenz fraglos überfordert. Die Tragweite einer solchen Publikation von vermeintlichen Qualitätsdefiziten ist erheblich und kann eine Krankenhausabteilung existentiell bedrohen.

b. w.

Präsident:  
 Prof. Dr. A. Dietz  
 Univ. HNO-Klinik  
 Liebigstr. 10-14  
 04103 Leipzig

Stellv. Präsident:  
 Prof. Dr. S. Plontke  
 Univ. HNO-Klinik  
 Ernst-Grube-Str. 40  
 06120 Halle/S.

Generalsekretär:  
 Prof. Dr. Th. Deitmer  
 DGHNO-KHC  
 Friedrich-Wilhelm-Str. 2  
 53113 Bonn

Schriftführer:  
 Prof. Dr. Dr. H.-J. Welkoborsky  
 Klinikum Nordstadt, HNO-Klinik  
 Haltenhoffstr. 41  
 30167 Hannover

Schatzmeister:  
 Prof. Dr. K.-W. Delank  
 HNO-Klinik, Klinikum der  
 Stadt Ludwigshafen  
 Bremser Str. 79  
 67063 Ludwigshafen

Telefon: 03 41 / 9 72 17 00  
 Telefax: 03 41 / 9 72 17 09  
[andreas.dietz@medizin.uni-leipzig.de](mailto:andreas.dietz@medizin.uni-leipzig.de)

Telefon: 03 45 / 5 57 17 84  
 Telefax: 03 45 / 5 57 18 59  
[stefan.plontke@uk-halle.de](mailto:stefan.plontke@uk-halle.de)

Telefon: 02 28 / 9 23 92 20  
 Telefax: 02 28 / 9 23 92 210  
[thomas.deitmer@hno.org](mailto:thomas.deitmer@hno.org)

Telefon: 05 11 / 9 70 43 77  
 Telefax: 05 11 / 9 70 46 42  
[hans-juergen.welkoborsky@krh.de](mailto:hans-juergen.welkoborsky@krh.de)

Telefon: 06 21 / 5 03 34 01  
 Telefax: 06 21 / 5 03 34 03  
[delankw@klllu.de](mailto:delankw@klllu.de)

Aus unserer Sicht wird es vor allem die Aufgabe der betroffenen Fachgesellschaften sein, in der Phase der Erprobung (d.h. solange jedes Qualitätsdefizit noch durch eine Stellungnahme entkräftet werden kann), die eventuellen Mängel der statistischen Auswertung aufzudecken und zu beanstanden. Der Praxistest wird zeigen, ob das System valide Aussagen liefert oder nicht. Insofern ist eine hinreichend lange und nachhaltige Beteiligung der wissenschaftlichen Fachgesellschaften an der Erprobung des Verfahrens entscheidend und wird dann letztlich auch die Akzeptanz erhöhen.

Die Auswahl von Experten für den strukturierten Dialog dürfte nach vorgesehenen Regeln sehr problematisch sein. Für einen solchen Dialog wird auch im Sinne einer Verbesserung der Ergebnisse nur ein Peer tätig werden können, der authentisch und aus eigener praktischer Erfahrung die Probleme sehen und beurteilen kann.

Den straffen Zeitplan in Bezug auf die Stellungnahmen halten wir ebenfalls für überzogen und kaum realistisch. Die Platzierung einer so entscheidenden Stellungnahmefrist mitten in die Urlaubszeit ist unverständlich. Auch außerhalb der Urlaubszeit sind 4 Wochen definitiv zu knapp und die Frist müsste mindestens mit 2 Monaten bemessen werden.

Mit freundlichen Grüßen



Prof. Dr. med. Thomas Deitmer  
Generalsekretär

Deutsche Gesellschaft für Innere Medizin e.V. · Postfach 2170 · 65011 Wiesbaden

IQTIG  
Institut für Qualitätssicherung und  
Transparenz im Gesundheitswesen  
Katharina-Heinroth-Ufer 1  
10787 Berlin

**Generalsekretär a. D.**  
Prof. Dr. med. Dr. h.c. Ulrich R. Fölsch

Geschäftsstelle Wiesbaden  
Irenenstraße 1 · 65189 Wiesbaden

Tel.: 0611 2058040-0 · Fax: 0611 2058040-46  
E-Mail: [urfoelsch@1med.uni-kiel.de](mailto:urfoelsch@1med.uni-kiel.de)  
Internet: [www.dgim.de](http://www.dgim.de)

Dependance Berlin  
Oranienburger Str. 22 · 10178 Berlin  
Tel.: 030 24625900 · Fax 030 24625905

19. November 2019 UF/er

## **Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern, Vorbericht zur Stufe 1 und 2, 10. Oktober 2019**

Sehr geehrte Damen und Herren,

sehr bedanken wir uns für die Möglichkeit, Ihre umfassende Arbeit zur Weiterentwicklung des Strukturierten Dialogs im Rahmen des Bestrebens zur Förderung der medizinischen Versorgungsqualität in Deutschland. Auch wenn wir von einer guten Versorgungsqualität in Deutschland ausgehen, ist es unbestritten, dass diese regelmäßig überprüft werden sollte, möglicherweise mehr durch Etablierung und Kontrolle struktureller Kriterien als durch die hier aufgeführten vielen Qualitätsindikatoren.

Wir erlauben uns, einige Themen zu kommentieren verbunden mit der Hoffnung, die Inhalte auch richtig verstanden zu haben. Zum Schluss erlauben Sie uns noch auf einige offene Punkte hinzuweisen.

### Übergeordnetes Ziel

Wir begrüßen das Bestreben, den strukturierten Dialog durch die Entwicklung eines verbindlichen Rahmenkonzeptes unter Wahrung einer hohen Transparenz zu vereinheitlichen. Wir hoffen, dass ein solches Vorgehen nicht durch zu viele Akteure innerhalb des G-BA und durch unsere föderale Verwaltung außerhalb des G-BA zu sehr verwässert wird.

### Gremien

Das Papier geht auf die unterschiedlichen Zuständigkeiten ein. Bei der Bildung von Fachkommissionen stellt sich hier die Frage, wie sehr Vorschläge(-gaben) in praxi dann auf Landesebene eingehalten werden. Wir möchten darauf hinweisen, dass bei der sicher notwendigen Einbeziehung wissenschaftlicher Fachgesellschaften bedacht werden sollte, dass diese in der Regel bundesweit organisiert sind. Die alleinige Berücksichtigung auf Landesebene, würde hier zu kurz greifen. Bisherige Erfahrung

Mit Interesse haben wir festgestellt, welche Datenflut schon jetzt bei Berücksichtigung eines relativ begrenzten Krankheitsspektrums anfällt und welche Probleme die Filterung hin zu einem strukturierten Dialog aufwirft. Wir fragen uns, ob das wirklich weiter ausgedehnt werden sollte und

ob es nicht viel sinnvoller wäre, im Dialog mit den Leistungserbringern Strukturvorgaben (z.B. Mindestzahl an Operationen, Personalbesetzung, bauliche und apparativ-technische Voraussetzungen, Dokumentationsqualität etc.) zu entwickeln und festzulegen, deren Einhaltung dann entsprechend überprüft wird.

#### Methodik

Die Ausführungen zur Methodik des Strukturierten Dialogs haben wir mit Interesse gelesen, jedoch einige der ausgedehnten statistisch- methodischen Ausführungen nur teilweise bzw. nicht verstanden. Das Problem der Aussagekraft bei kleiner Fallzahl wird mehrfach angesprochen. Hierbei erscheint uns die „Aggregation“ mehrerer Standorte problematisch. Hier sollte man diskutieren, ob Mindestzahlen nicht eine bessere Lösung darstellen.

Das einheitlich standardisierte und transparente Vorgehen bei der Qualitätsbewertung – auch in Bezug auf Fristen - befürworten wir.

Der Hinweis auf die Qualität der Datenquellen ist wichtig. Allerdings ist die Diskussion über die Asymmetrie des Aufwands in Bezug auf rechnerisch unauffällige Ergebnisse in hohem Maße bedenkenswert. Die Auswertung über einen Zeitraum von zwei Jahren erscheint plausibel. Bei allen diesen Betrachtungen und Analysen sollte jedoch immer darauf geachtet werden, den bürokratischen Aufwand für die Leistungserbringer nicht „überborden“ zu lassen. Gerade von dieser Seite ist das Personal in den Kliniken schon jetzt massiv belastet.

#### Referenzwerte

Auf das Problem der Referenzwerte und der Auffälligkeitseinstufung wird in dem Bericht mehrfach eingegangen. Das erscheint uns eine sehr wichtige Frage, in die die wissenschaftlichen Schwerpunktgesellschaften auch bei den vorbereitenden Beratungen (nicht nur in der Diskussion zu weitgehend abgeschlossenen Vorlagen wie dieser) mit einbezogen werden sollten. Eine Festlegung dieser Referenzwerte hat entscheidenden Einfluss auf alle nachfolgenden Schritte. Hier sollte man sich über Deutschland hinaus auch international orientieren. Sind die Kriterien zur nachträglichen Risikoadjustierung einheitlich?

#### Konsequenzen des Verfahrens

Wenn das Verfahren des strukturierten Dialogs zur einer sanktionsfreien konstruktiven Qualitätsförderung führt, ist das ein positives Ergebnis. Auf der anderen Seite hat das Verfahren natürlich auch schon jetzt über die öffentliche Berichterstattung und auch über vergütungsrelevante Konsequenzen indirekten Einfluss auf die Planung. Dies sollte bei der Weiterentwicklung der eigentlichen PlanQis bedacht und analysiert werden.

#### Detailfragen/-Hinweise

S. 21.: hier werden Sensitivität und Spezifität der Qualitätsbewertung diskutiert. Bezogen auf welches Außenkriterium?

S. 30: Datenqualität und Dokumentation ist offenbar doch noch ein vordringliches Problem. Kann man das zeitnah lösen?

S. 36: Priorisierung von Verfahren: Gibt es hier feste Vorgaben?

S. 54: Lassen sich die statistischen Unsicherheitsfaktoren überhaupt bereinigen/beheben?

S. 55: Hier wird ein wichtiger Punkt angeschnitten. Bilden die Qualitätsindikatoren, die Versorgung wirklich angemessen ab?

S. 63: Wird hier für ein primär sanktionsfreies Verfahren plädiert? Dies stünde im Widerspruch zur bisherigen Entwicklung.

S. 72 ff: Diese Ausführungen sind schwer verständlich

S. 79ff: Entscheidungstheoretische Modellierung des Bewertungsprozesses: Sehr komplex, schwer zu verstehen

S. 113ff: „Zusammenhang zwischen Strukturbruchproblemen, statistischer Prozesskontrolle und sequentiellen Entscheidungsproblemen“ diese Überschrift ist ebenso wie der dann folgende Text für den „Laien“ schwer zu durchdringen

S. 118 ff: „Laufregel“ auch hier ist die zugehörige methodische Abhandlung ohne Zuhilfenahme eines Statistikers nicht beurteilbar.

S. 192ff: Wie viel Zeit muss für die Evaluierung eines neuen Verfahrens veranschlagt werden? Die Ausführungen zum Evaluierungsdesign und dessen Limitationen erscheinen wiederum sehr komplex und schwer verständlich.

Wir danken Ihnen nochmals für die Möglichkeit zur Einsicht in diesen detaillierten und umfassenden Vorbericht und sind

mit freundlichen Grüßen

Ihre

Deutsche Gesellschaft für Innere Medizin e.V.



Prof. Dr. Dr. h.c. Ulrich R. Fölsch  
Generalsekretär a.D.



Prof. Dr. Tilman Sauerbruch  
Beauftragter der DGIM



**DGK.**

Deutsche Gesellschaft für Kardiologie  
– Herz- und Kreislaufforschung e.V.

Grafenberger Allee 100  
40237 Düsseldorf  
Telefon: +49 (0) 211 / 600 692-0  
Fax: +49 (0) 211 / 600 692-10  
E-Mail: info@dgk.org  
Web: www.dgk.org

IQTIG

Institut für Qualitätssicherung und Transparenz

im Gesundheitswesen

- Stellungnahme zum Vorbericht „Weiterentwicklung des  
Strukturierten Dialogs mit KH“ -

Katharina-Heinroth-Ufer 1

10787 Berlin

Per E-Mail: [weiterentwicklung-sd@iqtig.org](mailto:weiterentwicklung-sd@iqtig.org)

Nachrichtlich: [s-iqtig@awmf.org](mailto:s-iqtig@awmf.org)

Düsseldorf, den 20. November 2019

DGK\_V2019\_065 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern/06

## Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern

### Vorbericht zu Stufe 1 und Stufe 2

Erstellt vom IQTIG im Auftrag des Gemeinsamen Bundesausschusses

Stand: 10. Oktober 2019

Sehr geehrte Damen und Herren,

Ziel des vom IQTIG vorgelegten Vorberichts ist die Weiterentwicklung des „Strukturierten Dialogs“ (SD) mit Krankenhäusern im Rahmen der Qualitätssicherung (QS). Eine Vereinheitlichung und Straffung des SD ist grundsätzlich zu befürworten.

Im Beschluss zur Beauftragung des IQTIG wird der SD beschrieben als Verfahren, „in dem unter Einbeziehung von Experten auffällige Ergebnisse von Einrichtungen bewertet und erforderlichenfalls geeignete Maßnahmen zur Qualitätsverbesserung und -förderung eingeleitet werden“. Dies entspricht der bisherigen Vorgehensweise. Anders als in der Industrie treten in der Medizin Komplikationen jedoch meist individuell und nicht systematisch auf, so dass einer Standardisierung Grenzen gesetzt sind.

Durch die vorgelegten, äußerst umfangreichen statistischen Modelle wird suggeriert, dass sich medizinische Vorgänge komplett operationalisieren ließen. Dies mag partiell auf Indikations- und Prozessindikatoren zutreffen, verfehlt jedoch im Besonderen das Ziel bei der Beurteilung von Komplikationen sowie der Indikationsqualität. Nur die fachspezifische Beurteilung des einzelnen Falles (unter Einbeziehung aller verfügbaren Qualitätsindikatoren) kann eine sachgerechte Bewertung vornehmen. Dieser Aspekt deckt sich mit der Aussage des IQTIG auf **Seite 61**: „Für ein explizites Vorgehen müssten bereits A-priori-Bewertungsregeln für alle möglichen Einflussfaktoren vorliegen.“

Es ist aber charakteristisch für die zu bewertenden besonderen Konstellationen, dass eben nicht alle möglichen Einflussfaktoren vor Beginn des Stellungnahmeverfahrens bekannt sind. Es gibt Konstellationen, in denen die verfügbaren Informationen nicht ausreichen, um Entscheidungsregeln und Berechnungsvorschriften festzulegen oder diese sinnvoll anzuwenden.“

Zudem erschweren komplexe statistische Verfahren die Nachvollziehbarkeit einer Auffälligkeit beim Leistungserbringer im Sinne einer „Black Box“. Nach der Literatur ist die Überzeugung der Teilnehmer vom Nutzen des Verfahrens aber ein wichtiger Erfolgsfaktor, wie das IQTIG selbst anführt.

Ein angestrebter, automatisierter Entscheidungsprozess über Auffälligkeiten als Qualitätsmängeln widerspricht dem Wesen und der Funktion eines SD, dem genau dieses Urteil bisher vorbehalten war. Er soll ja bei rechnerischen Auffälligkeiten bewerten, ob eine Stellungnahme erforderlich ist, ob unter Würdigung der Stellungnahme Qualitätsmängel vorliegen und ob eine Begehung, Peer Review oder ähnliche Maßnahmen vorzusehen sind. Hinzu kommt, dass Qualitätsmängel in Zukunft „strafbewehrt“ sind und zu Abteilungs- oder Klinik-Schließungen usw. führen können. Die vorgeschlagenen Verfahren stellen damit keine Verbesserung des Strukturierten Dialoges dar, sondern seine partielle Abschaffung.

Eine medizinischzentrierte Qualitätssicherung mit der Zielsetzung durch Edukation (da, wo es notwendig ist) Verfahren für Patient\*innen "besser" zu machen, kann durch einfachere Vorgehensweisen erzielt werden, z.B. durch ein Auditing-Verfahren, wie es z.B. das QulK-Register seit vielen Jahren erfolgreich betreibt: Auswahl von Fällen (und damit von Ärzten) nach dem Zufallsverfahren, NACH Abgabe der Datensätze mit Vor-Ort-Besuch und Einsichtnahme in die gesamte Krankengeschichte der "gezogenen" Patient\*innen. Bei Auffälligkeiten gibt es die Möglichkeit des Ziehens zusätzlicher Fälle.

#### **Zu S. 50:**

Es wird vorgeschlagen, durch eine Aggregation aller Einzelfälle, die in Standorten gleicher Fallzahl behandelt wurden, eine durchschnittliche Tendenz des tatsächlichen Zusammenhangs zwischen Fallzahl und Versorgungsqualität zu ermitteln. Dieser Ansatz erscheint sinnvoll, da das Problem kleiner Fallzahlen in der QS immer wieder zu erheblichen Problemen führt.

#### **Zu S. 53:**

Die derzeit verwendete Methode der rechnerischen Auffälligkeitseinstufung würde tendenziell zu viele Standorte mit kleiner Fallzahl als auffällig einstufen, die sich dann im SD nicht bestätigen. Die relativ hohe Zahl rechnerischer Auffälligkeiten bei kleiner Fallzahl beruht darauf, dass bereits ein Einzelfall zu einer Auffälligkeit führen kann. Dieser bedarf dann der medizinischen Bewertung im SD, ggf. unter Hinzuziehung weiterer Indikatoren. Es ist nicht nachvollziehbar, wie dieses Dilemma durch komplexe statistische Verfahren beseitigt werden könnte, ohne hier mögliche medizinische Probleme zu übersehen.

#### **In Kapitel 5**

werden umfangreiche statistische Auswertemethoden vorgestellt. Zwei dieser Methoden seien bereits in der QS eingesetzt worden, es fehlen jedoch Aussagen, wie diese auf ihre Funktionsfähigkeit validiert wurden.

Analog zur empfohlenen Qualifikation der Fachkommissionsmitglieder bitten wir um Offenlegung, welche persönlichen Erfahrungen die Autoren dieser Auswertemethoden mit dem SD haben.

#### **Zu S. 109:**

Hier wird ein perzentilbasierter Referenzwert vorgeschlagen, bei welchem ein vorher definierter Anteil an Leistungserbringern quantitativ auffällig wird. Dies bedeutet, dass bewusst in Kauf genommen wird, dass Leistungserbringer auffällig werden, selbst wenn theoretisch alle eine gute Qualität erbringen. Hier muss über zusätzliche Referenzwerte diskutiert werden, welche primär verfehlt werden müssen, bevor ein perzentilbasierter Wert greift. Ansonsten liegt eine fortwährende, systembedingte Diskriminierung von Leistungserbringern vor.

#### **Zu 5.4 Berücksichtigung der Daten mehrerer Erfassungsjahre:**

Die Sammlung zusätzlicher Informationen über die vergangenen Erfassungsjahre zur Vergrößerung der Datengrundlage bei kleinen Fallzahlen erscheint sinnvoll und ist zu begrüßen.

#### **Zu Laufregel S. 119 unten:**

Das hier angeführte Beispiel zeigt, dass das vorgeschlagene Verfahren sehr problematisch ist, da ein Leistungserbringer selbst mit 5 Ereignissen bei einer Grundgesamtheit von 25 Fällen nicht auffällig wird. In einem solchen Fall ist ein Qualitätsproblem sehr wahrscheinlich. Das Verfahren ist in dieser Form daher abzulehnen.

#### **Zu S. 139 oben:**

Hier wird festgestellt, dass statistisch relevante Einstufungen erst ab einer gewissen Fallzahl möglich sind, da diese Vorgehensweise danach geht, wie viele Patient\*innen von einem möglichen Qualitätsdefizit betroffen sind. Dies bedeutet im Umkehrschluss, dass man auf der einen Seite bereit ist, Qualitätsdefizite zu tolerieren, wenn nur wenige Patient\*innen betroffen sind. Dies ist ethisch kaum zu tolerieren, da in der Medizin das Wohl des Einzelnen im Vordergrund steht.

#### **Zu S. 140:**

„Insgesamt zeigen die Analysen, dass die korrekte Klassifikation kleiner Leistungserbringer eine besondere Herausforderung darstellt, die entweder nur über eine Erweiterung der Datenbasis oder mittels nicht statistischer Methoden zu lösen ist.“ Hierzu sei aus kardiologischer Erfahrung angemerkt, dass z.B. bei der Schrittmacherimplantation etwa die Hälfte der Kliniken durchschnittlich 25 Implantationen jährlich durchführt. Bei ICD-Implantationen ist die Zahl noch geringer. Aus dem zitierten Statement ist damit zu folgern, dass die vorgestellten statistischen Methoden hier überhaupt nicht anwendbar sind!

#### **Zu 6 Fachliche Bewertung:**

Die beste erzielbare Bewertung lautet: „Kein Hinweis auf Qualitätsdefizit“. Dies erinnert an „aus Mangel an Beweisen“ und schließt sich an die unsägliche Diktion der „Patientengefährdung“ an. Die Bewertung könnte neutraler formuliert sein: „Kein Qualitätsdefizit“. Genauso bei Bewertungen U0: „Kein hinreichender Hinweis auf ein Qualitätsdefizit“.

**Zu S. 165 unten:**

„Qualitätsindikatoren treffen eine aggregierte Qualitätsaussage auf Ebene des Leistungserbringers und dienen nicht dazu, Einzelfälle zu bewerten.“

Dieser Aussage kann nicht gefolgt werden. Gerade die langjährige Erfahrung aus dem SD zeigt, dass insbesondere bei kleinen Fallzahlen häufig nur wenige Fälle analysiert und beurteilt werden müssen. Dass dies einen hohen Aufwand zur Folge hat, ist korrekt, dient aber zweifelsohne der Qualitätsverbesserung.

Aus klinischer Sicht sind hier wenige, auch zusammenfassende Informationen, nicht ausreichend und die Prüfung der betreffenden Behandlungsfälle im Detail wird erforderlich.

**Zu S. 183 Sommerurlaubszeit:**

Diese soll richtigerweise für die Erarbeitung der Stellungnahmen berücksichtigt werden, dies muss aber genauso für die weitere Planung und Sitzungen der Fachkommissionen berücksichtigt werden.

**Zu S. 211 feste Referenzbereiche:**

In Anbetracht der o.a. Kommentare sollte, wie vorgeschlagen, eine Überführung perzentilbasierter Referenzbereiche in feste Referenzbereiche durch die Fachgruppen für jeden Indikator überprüft werden.

**Zu S. 215-216:**

Das IQTIG empfiehlt für die meisten fallbasierten Qualitätsindikatoren eine statistische Operationalisierung des hinreichenden Hinweises auf ein Qualitätsdefizit auf Basis des quantitativen Indikatorergebnisses. Das Signifikanzniveau wird als zusätzliche Stellschraube für die Entscheidung, ob eine Stellungnahme eingeholt werden soll, bezeichnet. Dies ist abzulehnen. Grundsätzlich muss eine Einstufung als Qualitätsdefizit zur Stellungnahme führen.

**Zu S. 217 Fachliche Bewertung von Stellungnahmen:**

Explizite Regeln für die Bewertungen sind unter klinischen Gesichtspunkten praktisch nicht umsetzbar, da in der Mehrzahl aller Fälle der individuelle Krankheitsverlauf betrachtet werden muss. Ein Peer Review muss grundsätzlich erfolgen, bevor gravierende Konsequenzen eingeleitet werden. Kollegiale Gespräche und Begehungen sollen reduziert werden und nicht mehr in die Beurteilung einfließen. Hier verkennt der Entwurf die Abläufe des SD, da diese Mittel erst bei wiederholten Hinweisen auf Qualitätsprobleme zum Einsatz kommen und nicht in der primären Bewertung von Stellungnahmen.

**Zu S. 217 im 1.Abschnitt unten:**

„Dies steht im Einklang mit der Empfehlung, zukünftig nur noch beim Vorliegen überzeugender Gründe Indikatorergebnisse, die hinreichende Evidenz für einen Qualitätsdefizit nahelegen, in der fachlichen Bewertung mit „kein Hinweis auf Qualitätsdefizit“ zu bewerten.“ Das würde bedeuten, dies wäre bisher nicht der Fall gewesen und gehört als tendenziöse Aussage hier nicht hin.

**Zu S. 220**

**Abschließende Entscheidung der zuständigen Stelle, ggf. Begründung einer Abweichung von der Fachkommission:** Dies würde eine fachliche Aufsicht über die jeweilige unabhängige Fachkommission begründen und ist folglich abzulehnen.

Für die Qualität maßgeblich entscheidend ist natürlich die Qualität der Indikationsstellung. Die Bewertung der Indikationsqualität lässt sich sicherlich am besten durch ein Auditingverfahren erfassen und nicht durch die Bestimmung einzelner Indikatoren wie z.B. dem objektiven Ischämienachweis.

**Fazit:**

Die Weiterentwicklung des Strukturierten Dialogs wird das Niveau des zugrunde liegenden Verfahrens nicht verbessern.

Für Rückfragen stehen wir selbstverständlich gerne zur Verfügung.

Mit freundlichen Grüßen

Prof. Dr. Andreas M. Zeiher  
Präsident  
Deutsche Gesellschaft für Kardiologie  
Herz- und Kreislaufforschung e.V.

Prof. Dr. Bernd Nowak  
Autor und Vorsitzender  
*Ausschuss Bewertungsverfahren*  
Ständiger Ausschuss  
Qualität- / Leistungsbewertung in der Kardiologie

Prof. Dr. Johann Bauersachs  
Vorsitzender  
Kommission für Klinische Kardiovaskuläre Medizin

**Autoren**

Diese Stellungnahme wurde für die Deutsche Gesellschaft für Kardiologie - Herz- und Kreislaufforschung (**DGK**) erarbeitet von

Prof. Dr. Bernd Nowak (Frankfurt) federführend  
Prof. Dr. Albrecht Elsässer (Oldenburg)  
Prof. Dr. Karl-Heinz Kuck (Hamburg)  
Prof. Dr. Matthias Pauschinger (Nürnberg)  
Prof. Dr. Bernhard Schieffer (Marburg)  
Prof. Dr. Christoph Stellbrink (Bielefeld)  
PD Dr. Michael A. Weber (Dachau)



# DEUTSCHE GESELLSCHAFT FÜR KINDER- UND JUGENDMEDIZIN e.V.

DGKJ e.V. | Geschäftsstelle | Chausseestr. 128/129 | 10115 Berlin

IQTIG  
Institut für Qualitätssicherung und Transparenz im  
Gesundheitswesen  
Katharina-Heinroth-Ufer 1  
10787 Berlin

**Die Präsidentin**  
Prof. Dr. Ingeborg Krägeloh-Mann

**Geschäftsstelle**  
Chausseestr. 128/129  
10115 Berlin  
Tel. +49 30 3087779-0  
Fax: +49 30 3087779-99  
info@dgkj.de | www.dgkj.de

**Hausadresse:**  
Universitätsklinik für Kinder- und  
Jugendmedizin Tübingen  
Abt. Neuropädiatrie, Entwicklungs-  
neurologie, Sozialpädiatrie  
Hoppe-Seyler-Str. 1  
72076 Tübingen  
Tel. +49 7071 29-84735  
Fax: +49 7071 29-5473  
kraegeloh-mann@dgkj.de

Tübingen, 18.11.2019

## **Stellungnahme der DGKJ zum IQTIG-Vorbericht „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“**

Sehr geehrte Damen und Herren,

für die Zusendung der Unterlagen zum Vorbericht „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“ und die Möglichkeit zur Stellungnahme danken wir. Die nachfolgende Stellungnahme würde in Zusammenarbeit mit Herrn Prof. Dr. Andreas Trotter, Verband leitender Kinder- und Jugendärzte und Kinderchirurgen Deutschlands e. V. (VLKKD), und Herrn Jochen Scheel, Gesellschaft der Kinderkrankenhäuser und Kinderabteilungen in Deutschland e. V. (GKinD) erarbeitet.

Die Weiterentwicklung des strukturierten Dialogs hat zum Ziel, die derzeit bestehende Heterogenität zwischen den Bundesländern in der Umsetzung des strukturierten Dialogs zu vereinheitlichen. Im Sinne der besseren Vergleichbarkeit begrüßen wir dies, da bisher ein als auffällig eingestuftes Indikatorergebnis in verschiedenen Bundesländern unterschiedlich häufig zur Einforderung einer Stellungnahme der leistungserbringenden Klinik geführt hat. Damit werden auch Zielvereinbarungen zur Verbesserung der Qualität unterschiedlich häufig abgeschlossen. Ein weiteres aus unserer Sicht positiv zu bewertendes Ziel ist die Verringerung des zum Teil sehr hohen vor allem personellen Aufwands für strukturierte Dialoge zum Beispiel durch Verringerung der zu prüfenden Qualitätsindikatoren. Ob dies bei eher steigender Zahl von Qualitätsindikatoren wirklich gelingt, ist jedoch zu bezweifeln.

Wir unterstützen die Zielsetzung, die zuständigen Landesgremien (Landesgeschäftsstellen für Qualitätssicherung (LQS) bzw. Landesarbeitsgemeinschaften (LAG)) in ihrer Struktur und Qualifikation zu vereinheitlichen. So sollen Fachexpertinnen/-experten über die für das jeweilige QS-Verfahren relevante klinisch-praktische Erfahrung verfügen.

Einzelvertretungsberechtigt i. S. d. § 26 BGB:  
Prof. Dr. Ingeborg Krägeloh-Mann, Präsidentin  
Prof. Dr. Christian von Schnakenburg, Schatzmeister  
5707 00)

Eingetragen unter VR 26463 B  
Sitz des Vereins: Berlin  
USt.-IdNr. 27/663/60401

Bank für Sozialwirtschaft  
IBAN: DE26 1002 0500 0001 5707 01  
(für Spenden/Mitgliedsbeitrag: IBAN: DE53 1002 0500 0001

BIC: BFSWDE33BER

Außerdem sehen wir die geplante Einbindung von Vertretern des Pflegeberufs in der entsprechenden Fachkommission als Gewinn (s. S. 170).

Bei den Mindestanforderungen an vom Leistungserbringer zu formulierende Stellungnahmen (s. S. 165) wird unter anderem eine Einreichfrist innerhalb von 4 Wochen angegeben. Diese Frist ist unseres Erachtens zu kurz gefasst, da die verantwortliche Instanz des Leistungserbringers (z. B. die Chefärztin/der Chef-  
arzt) eine angefertigte Stellungnahme immer autorisieren soll. Eine entsprechende Zusammenstellung und Abstimmung der Daten ist in der Regel zeitaufwändiger. Wir plädieren, die regulär zu gewährende Bearbeitungszeit zu verlängern.

Eine konstante Sitzungsbeteiligung in den Gremien ist wünschens- und unterstützenswert. Um dies zu garantieren, soll die zeitliche Verfügbarkeit von sich bewerbenden Experten/innen im Vorfeld schriftlich bestätigt werden (s. S. 169). Bei der derzeitigen Arbeitsbelastung nahezu aller praktisch in der Klinik tätigen Berufsgruppen kann dies unserer Einschätzung nach eine Hürde darstellen.

Für Rückfragen stehen wir gern zur Verfügung.

Mit freundlichen Grüßen



Prof. Dr. Ingeborg Krägeloh-Mann



GMDS e.V. · Industriestraße 154 · D-50996 Köln

IQTIG - Institut für Qualitätssicherung und Transparenz  
im Gesundheitswesen  
Dr. med. Christof Veit  
Katharina-Heinroth-Ufer 1  
D-10787 Berlin

Per E-Mail: [weiterentwicklung-sd@iqtig.org](mailto:weiterentwicklung-sd@iqtig.org)  
[s-iqtig@awmf.org](mailto:s-iqtig@awmf.org)

20. November 2019

**Stellungnahme zum Beteiligungsverfahren des IQTIG nach § 137a Abs. 7 SGB V zum Abschlussbericht "Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern"**

Sehr geehrter Herr Dr. Veit,

wir bedanken uns für die Möglichkeit über die AWMF (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V.) eine Stellungnahme zum Beteiligungsverfahren des IQTIG nach § 137a Abs. 7 SGB V zum Abschlussbericht "Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern" abgeben zu können.

Die Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR) und die Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) e.V. haben die Stellungnahme im Rahmen der gemeinsamen Präsidiumskommission "Methodenaspekte in der Arbeit des IQWiG und IQTIG" angefertigt.

Mit freundlichen Grüßen

Prof. Dr. Alfred Winter  
- Präsident der GMDS -

Prof. Dr. Werner Brannath  
- Präsident der IBS-DR -

Prof. Dr. Tim Friede  
- Leiter der PK Methodenaspekte in  
der Arbeit des IQWiG und IQTIG -

**GMDS Geschäftsstelle**

Industriestraße 154  
D-50996 Köln

Telefon: +49 (0)2236 33 19 958  
Telefax: +49 (0)2236 33 19 959

E-Mail: [info@gmDS.de](mailto:info@gmDS.de)  
Internet: [www.gmDS.de](http://www.gmDS.de)

**Geschäftsführung**

Beatrix Behrendt

**Präsident und Vize-Präsidenten**

Prof. Dr. Alfred Winter  
(Leipzig), Präsident

Prof. Dr. Harald Binder  
(Freiburg), 1. Vizepräsident

Prof. Dr. Andreas Stang, MPH  
(Essen), 2. Vizepräsident

**IBS-DR Geschäftsstelle**

Heike Krubert  
c/o Institut für Biometrie, Epidemiologie und  
Informationsverarbeitung  
Tierärztliche Hochschule  
Hannover  
Bünteweg 2  
D-30559 Hannover

Telefon: +49 (0) 511 953 79 51  
Telefax: +49 (0) 511 953 79 74

E-Mail: [biometrische-gesellschaft@tih-hannover.de](mailto:biometrische-gesellschaft@tih-hannover.de)

**IBS-DR - Präsident und Vizepräsident**

Prof. Dr. Werner Brannath  
(Bremen), Präsident

Prof. Dr. Andreas Faldum  
(Münster), Vizepräsident



**Stellungnahme zu „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern:  
Vorbericht zu Stufe 1 und Stufe 2“ des IQTIG durch die gemeinsame Präsidiumskommission  
„Methodenaspekte in der Arbeit des IQWiG und IQTIG“ der GMDS und IBS-DR**

20. November 2019

Das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) hat am 10. Oktober 2019 einen Bericht zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern zu den Stufen 1 und Stufe 2 vorgelegt. Den Bericht zur Stufe 1 hatte das IQTIG bereits am 30. November 2018 dem G-BA übermittelt. Der nun vom IQTIG vorgelegte Bericht beinhaltet zum einen einige Weiterentwicklungen zur Stufe 1 wie auch Ausführungen zur Stufe 2.

In dieser Stellungnahme gehen wir insbesondere auf die biometrischen Aspekte des Berichts des IQTIG ein. Es sei vorweg bemerkt, dass die Qualitätssicherung im Gesundheitswesen aus statistischer Sicht ein komplexes und dynamisches Anwendungsfeld ist, das von der fortlaufenden Entwicklung adäquater statistischer Methoden profitiert. Wir begrüßen deshalb ausdrücklich die ausführliche und sorgfältige Beschreibung und Behandlung der statistisch-methodischen Aspekte durch das IQTIG. Im Folgenden möchten wir einige Punkte aufgreifen, die als Anstoß für weitere Entwicklungen gedacht sind. Unsere Fachgesellschaften und insbesondere die gemeinsame Präsidiumskommission sind offen für einen Dialog mit dem IQTIG in Bezug auf die biometrischen Aspekte und würden zukünftige Interaktionen begrüßen.

Im Abschnitt 2.4.3 Bewertung der Dokumentationsqualität von Qualitätsindikatoren wird ausgeführt, dass rechnerisch und potenziell qualitative auffällige Ergebnis in die Kategorie „D“ („Bewertung nicht möglich wegen fehlerhafter Dokumentation“) eingestuft werden. Weiter wird berichtet, dass „[b]undesweit [...] für das Erfassungsjahr 2015 durchschnittlich 10,4% der rechnerisch auffälligen Indikatorergebnisse aufgrund nachträglicher Angaben von Dokumentationsfehlern nicht bewertet [wurden]“ (Seite 30). Dieser Umgang mit fehlenden oder fehlerhaften Dokumentationen birgt aus unserer Sicht ein hohes Potential für Verzerrungen (Bias) und systematischen Unterschätzungen der Indikatorraten. Zudem ist diese Selektion in den statistischen Methoden nicht angemessen reflektiert. Wir sehen hier ein Handlungsfeld für weitere Entwicklungen im Bereich der statistischen Methodik.

Die Risikoadjustierung ist ein wichtiges Instrument der epidemiologischen Forschung. Entsprechend wird diese auch an verschiedenen Stellen des vorliegenden Berichts wiederholt genannt. So heißt es z.B. im Abschnitt 3.1.1 Einflussfaktoren auf das Indikatorergebnis: „Bei der Konstruktion von Qualitätsindikatoren wird durch eine angemessene Festlegung der Berechnungsvorschrift und durch Maßnahmen der Risikoadjustierung der Einfluss bekannter, nicht qualitätsbezogener Faktoren auf das Indikatorergebnis so weit wie möglich reduziert. Für die Berücksichtigung von stochastischen Einflüssen auf die Indikatorergebnisse stehen beispielsweise verschiedene statistische Methoden zur Verfügung (siehe Abschnitt 5.1).“ (Seite 54 - 55). Hier ist es uns wichtig, darauf hinzuweisen, dass

diese Adjustierungen nur unter bestimmten Voraussetzungen überhaupt funktionieren können. Zum einen müssen relevante Risikofaktoren erhoben und mit hinreichender Qualität dokumentiert worden sein. Zum anderen müssen auch gewisse statistische Anforderungen je nach verwendeter Methode erfüllt sein. Hierzu kann z.B. zählen, dass sich die Verteilungen der Risikofaktoren in den verschiedenen Krankenhäusern zumindest überlappen müssen, so dass keine Separierungen auftreten. Dieser Probleme sind typisch für viele Anwendungsfelder der Epidemiologie. In der Pharmakoepidemiologie beispielsweise wird vom ‚confounding by indication‘ gesprochen und ein großer Aufwand betrieben, um trotzdem valide Ergebnisse zu erzielen. Hier verweisen wir beispielhaft auf Brookhart et al (2010) und McMahon et al (2003).

Wie z.B. in Abbildung 3 auf Seite 40 deutlich wird, hängt die Variabilität in der Bestimmung eines Qualitätsindikators entscheidend von der Fallzahl ab. Kleine Fallzahlen führen zu einer hohen Variabilität. Uns ist es wichtig, hier auf die Unsicherheit in der Schätzung eines Indikators und der Notwendigkeit der Berücksichtigung dieser Unsicherheit in Evaluationen hinzuweisen. Naive Verwendung von Qualitätsindikatoren ohne Berücksichtigung der Unsicherheit in der Schätzung z.B. im Zusammenhang mit der Perzentilmethode, kann in irreführenden Ergebnissen resultieren. Wir begrüßen daher ausdrücklich die aufgeführten Ansätze zur angemessenen statistischen Behandlung des Problems (siehe Abschnitt 5). Hierzu zählt z.B. die ‚statistisch relevante Einstufungsmethode‘, die in Abschnitt 5.3.1.3 Statistisch relevante Auffälligkeit behandelt wird. Dieses Verfahren benutzt eine Verlustfunktion, ein aus statistischer Sicht sinnvolles Konzept zur Behandlung des Problems. Allerdings ist die Definition solcher Verlustfunktionen (siehe z.B. Tabelle 6 auf Seite 92) immer abhängig von der jeweiligen Anwendung, was auch die Tür zu möglicher Kritik öffnet.

Der Bericht ist geprägt durch einen Fokus auf einzelne Qualitätsindikatoren. Wenn Qualitätsindikatoren isoliert betrachtet werden, kann dies zu einem Informationsverlust führen, da sowohl inhaltliche Zusammenhänge wie statistische Korrelationen ungenutzt bleiben. Aus statistischer Sicht wäre es sinnvoll, zu einer Gesamtschau verschiedener Indikatoren zu gelangen, bei der dann die gemeinsamen Verteilungen betrachtet oder Kombinationen von Indikatoren gebildet werden und nicht nur marginale Verteilungen der einzelnen Qualitätsindikatoren.

Abschließend möchten wir noch einmal die ausführliche und sorgfältige Beschreibung und Behandlung der zum Teil recht schwierigen statistisch-methodischen Aspekte durch das IQTIG würdigen. Das IQTIG hat nach unserer Einschätzung hier einen wichtigen Schritt in die richtige Richtung getan. In den nächsten Jahren wird es darauf ankommen, dass die hier begonnenen Entwicklungen konsequent weiterverfolgt werden. Unsere Fachgesellschaften und insbesondere die gemeinsame Präsidiumskommission stehen zum methodisch-fachlichen Austausch bereit.

### Referenzen

Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. *Med Care*. 2010; 48(6 Suppl): S114-S120.

McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf*. 2003; 12(7): 551-558.



Geschäftsstelle DGNI – c/o Conventus Congressmanagement - Carl-Pulfrich-Str. 1 - 07745 Jena

Frau Sabine Lehman  
AWMF

Deutsche Gesellschaft für  
Neurointensiv- und Notfallmedizin  
Carl-Pulfrich-Str. 1  
07745 Jena  
phone: 03641 31 16 450  
fax: 03641 31 16 244  
mail: gs@dgni.de  
www.dgni.de

4 May 2020

**Betreff: Weiterentwicklung des strukturierten Dialogs mit  
Krankenhäusern Stand 10.Oktober 2019**

Sehr geehrte Damen und Herren,

von Seiten der DGNI sehen wir keine Notwendigkeit zur separaten  
Stellungnahme zu dem vorliegenden Papier und würden uns den  
Ausführungen der AWMF und anderen größeren Fachgesellschaften  
anschießen.

Mit freundlichen Grüßen

Dr. med. Sylvia Bele  
AWMF Leitliniendelegierte der DGNI

Präsident:  
Prof. Dr. med. Oliver Sakowitz  
RKH Klinikum Ludwigsburg  
Neurochirurgisches Zentrum  
Ludwigsburg-Heilbronn  
Postlpostr. 4  
71640 Ludwigsburg  
phone: 07141 996 71 01  
mail: oliver.sakowitz@kliniken-lb.de

1. Vizepräsident:  
Prof. Dr. med. Georg Gahn, M.B.A., FEAN  
Städtisches Klinikum Karlsruhe gGmbH  
Neurologische Klinik  
Moltkestr. 90  
76133 Karlsruhe  
phone: 0721 974 36 01  
mail: georg.gahn@klinikum-karlsruhe.de

2. Vizepräsident:  
Prof. Dr. med. Julian Bösel, FNCS  
Klinikum Kassel GmbH  
Klinik für Neurologie  
Mönchebergstr. 41-43  
34125 Kassel  
phone: 0561 980 40 51  
mail: julian.boesel@klinikum-kassel.de

Schatzmeister:  
Prof. Dr. med. Wolfgang Müllges  
Universitätsklinikum Würzburg  
Neurologische Klinik und Poliklinik  
Josef-Schneider-Str. 11  
97080 Würzburg  
phone: 0931 20 12 37 64  
mail: muellges\_w@ukw.de

Schriftführerin:  
Dr. med. Katja Wartenberg  
Universitätsklinikum Leipzig  
Klinik und Poliklinik für Neurologie  
Liebigstr. 20  
04103 Leipzig  
phone: 0341 972 00 72  
mail: katja.wartenberg@medizin.uni-leipzig.de

Beisitzer:  
Dr. Sylvia Bele, IFAANS  
Universitätsklinikum Regensburg  
Klinik und Poliklinik für Neurochirurgie  
Franz-Josef-Strauß-Allee 11  
93093 Regensburg  
phone: 0941 944 90 55  
mail: sylvia.bele@ukr.de

Prof. Dr. Thomas Westermaier, MHBA  
Universitätsklinikum Würzburg  
Neurochirurgische Klinik und Poliklinik  
Josef-Schneider-Str. 11  
97080 Würzburg  
phone: 0931 20 12 45 27  
mail: Westermaie\_T@ukw.de

Nichtärztlicher Beisitzer:  
Peter Nydahl, MScN, BScN  
Universitätsklinikum Schleswig-Holstein  
Klinik für Anästhesie und operative Intensivmedizin  
Brunswiker Str. 10  
24105 Kiel  
phone: 0431 50 01 28 13  
mail: Peter.Nydahl@uksh.de



Dieses Schreiben wurde maschinell erstellt und ist ohne Unterschrift gültig.



Institut für Qualitätssicherung und Transparenz im Gesundheitswesen

**Herrn Christof Veit**

**Frau Andrea Wolf**

Katharina-Heinroth-Ufer 1

10787 Berlin

Per E-Mail: [weiterentwicklung-sd@iqtig.org](mailto:weiterentwicklung-sd@iqtig.org)

VORSTAND

Präsident:

Prof. Dr. Horst Urbach

Weitere Mitglieder

Prof. Dr. Arnd Dörfler

Prof. Dr. Marc Brockmann

Prof. Dr. Mirko Pham

Prof. Dr. Werner Weber

KONTAKT

DGNR-Geschäftsstelle

Florian Schneider

Melek Mirzanli

Tel.: 030 916 070-70

Fax: 030 916070-22

Mail: [dgnr@neuroradiologie.de](mailto:dgnr@neuroradiologie.de)

Berlin, 20. November 2019

**Weiterentwicklung des Strukturierten Dialogs mit  
Krankenhäusern/Beteiligungsverfahren: Stellungnahme der Deutschen Gesellschaft  
für Neuroradiologie e.V.**

Sehr geehrter Herr Dr. Veit, sehr geehrte Frau Wolf,  
sehr geehrte Damen und Herren,

vielen Dank für die Zusendung der Unterlagen zur Weiterentwicklung des Strukturierten Dialogs, zu denen wir im Auftrag der Deutschen Gesellschaft für Neuroradiologie (DGNR) gern Stellung nehmen:

Im Auftrag des GBA wurde ein umfangreiches Methodenpapier entwickelt mit dem Ziel, den Strukturierten Dialog als wesentliches Instrument der gesetzlich vorgeschriebenen Qualitätssicherung bundesweit zu standardisieren und bezüglich seiner Regeln transparenter zu gestalten.

Durch die Streichung von Datenfeldern in den Erhebungsbögen und die Beschränkungen auf wenige möglichst Evidenz-basierte Qualitätsindikatoren kommt dem Strukturierten Dialog eine zunehmende Bedeutung zu. Die bislang ehrenamtlich tätigen Kommissionen bei den Landesgeschäftsstellen müssen zunehmend erst feststellen, warum eine Klinik rechnerisch auffällig geworden ist, weil Details der Eingriffe wie z. B. die Art von Komplikationen nicht mehr spezifisch erhoben werden. Den Datenfeldern für die invasive Behandlung von Karotisstenosen kann man beispielsweise nicht mehr entnehmen, welche Techniken bei der chirurgischen oder interventionellen Behandlung angewandt wurden und was die Auslöser für als Komplikation aufgetretene neurologische Defizite waren.

Dies im Strukturierten Dialog zu ermitteln, wäre auch in einem besser strukturierten Prozess aufwändig. Um Qualitätsverbesserungen zu erreichen, müssten detaillierte Daten aus dem Strukturierten Dialog wieder in das Verfahren einfließen. Sinnvoller wäre es aus unserer Sicht, das Prinzip der Datensparsamkeit etwas zu lockern, damit es leichter gelingt, Praktiken, die zu erhöhten Komplikationsraten führen, zu identifizieren. Eine solche Lockerung wäre günstiger als – wie es jetzt für die Karotis geplant ist – Angaben zu Körpergröße oder Gewicht zu erheben, deren Bedeutung als Risikofaktoren für Karotis-Prozeduren durch wissenschaftliche Studien kaum belegbar ist.

Die früheren Freiheiten der Bundesfachgruppen, gemeinsam mit den Institutionen der Qualitätssicherung die Inhalte des Verfahrens festzulegen, wurden auf eine relativ unverbindliche Beratungstätigkeit reduziert. Die Möglichkeiten eines verpflichtenden Registers, für die Beurteilung von Prozeduren relevante Daten zu erheben, bleiben ungenutzt. Inhaltlich generiert die Minimalisierung der Datensätze undifferenzierte rechnerische Auffälligkeiten, von denen im Strukturierten Dialog mit großem Aufwand der Anteil der Einrichtungen mit wirklich schlechter Qualität herausgefiltert werden muss.

Die geplante Standardisierung des Strukturellen Dialogs mag im Sinne der Transparenz sinnvoll sein und die Beurteilung durch die Kommissionen erleichtern. Die zahlreichen vorgeschlagenen Regelungen beinhalten jedoch auch das Risiko einer zunehmenden Bürokratisierung ohne Zugewinn an Qualität.

Bislang sind die Landesfachgruppen durch ehrenamtlich tätige FachärztInnen und VertreterInnen des MDK besetzt. Es ist nicht leicht, für diese Tätigkeit ausreichend fachlich und bezüglich ihrer Reputation qualifizierte ExpertInnen zu gewinnen. Die Kriterien für Bewerbungsverfahren, wie sie beispielsweise beim IQWiG üblich sind, sind nicht besonders transparent und nachvollziehbar. Bei der Einbindung von PatientenvertreterInnen wäre nach deren Motivation und Kompetenz in der Beurteilung oft komplexer medizinischer Sachverhalte zu fragen.

Die geplante Standardisierung der Kriterien für die Besetzung der Expertenkommissionen nimmt den Landesgeschäftsstellen und Landesfachgruppen einen Teil ihrer Freiheit, funktionierende und in der inhaltlichen Arbeit erprobte Gremien zusammenzustellen, die in der Lage sind, im Konsensus verbindliche Entscheidungen zu treffen.

Inhaltliche und methodische Freiheiten auf Landesebene müssen sich nicht nachteilig auswirken. Wie die landesspezifische Qualitätssicherung der Arbeitsgemeinschaft Deutscher Schlaganfallregister zeigt, ermöglichen die Freiheiten eine Entwicklung suffizienter Verfahren mit einem gewissen Wettbewerb zwischen den beteiligten Bundesländern. Die notwendige Standardisierung von Qualitätsindikatoren und Referenzbereiche für rechnerische Auffälligkeiten ergeben sich aus dem Prozess und aus den Vorgaben aus Studien und Leitlinien.

Die standardisierte Qualitätsbewertung erfolgt mit dem Ziel, qualitativ auffällige Kliniken „methodisch korrekt“ zu identifizieren, sie zu veröffentlichen und ökonomisch zu sanktionieren. Neben der Durchsetzung klarer Strukturvorgaben wie Personal-, Geräteausstattung und ggf. Mindestfallzahlen sollte das Ziel des Verfahrens – wie dies in Teilen des Papiers anklingt – in erster Linie eine nachhaltige Qualitätsverbesserung sein.

Aus Sicht der DGNR reicht es dazu nicht aus, Kliniken mit Qualitätsmängeln zu identifizieren und schlimmstenfalls aus der Versorgung auszuschließen. Derartige Schritte benötigen eine wissenschaftliche Grundlage und einen Expertenkonsensus eines für den Eingriff repräsentativen Gremiums mit einwandfreier Reputation. Eine stetige Qualitätsverbesserung sollte vor möglichen Sanktionen im schriftlichen und mündlichen Dialog mit den von Mängeln betroffenen Kliniken erreicht werden. Dazu sind insbesondere für wiederholt auffällige Einrichtungen die geplanten Peer-Review-Verfahren mit Analysen und Zielvereinbarungen zu begrüßen. Eine derart aufwändige Tätigkeit durch die eingebundenen Experten kann nicht mehr ausschließlich ehrenamtlich erfolgen. Eine entsprechende Aufwandsentschädigung muss – ähnlich wie beispielsweise für Gerichtsgutachten – gewährleistet sein.

Wir hoffen, dass in den unter politischem Einfluss und mit methodischen Vorgaben nicht immer sachgerecht weiter entwickelten Qualitätssicherungsverfahren genügend Freiheiten für fachliche Expertise und die Beurteilung komplexer Einzelfälle erhalten bleiben. Gewachsene Strukturen wie die Landesfachgruppen sollten nicht zerschlagen werden. Bei Neuerungen im Ablauf des Strukturierten Dialogs und in der Zusammensetzung der Gremien empfehlen wir, der generellen Umsetzung eine Erprobungsphase mit Modellversuchen und anschließender Evaluierung der endgültigen Regelung vorzuschalten.

Mit freundlichen Grüßen



Prof. Dr. Horst Urbach  
Präsident der DGNR



Prof. Dr. Joachim Berkefeld  
Autor der Stellungnahme



Prof. Dr. Werner Weber  
Vorstandsmitglied

DGOU/DGOOC/DGU Geschäftsstelle · Str. des 17. Juni 106-108 · 10623 Berlin

Institut für Qualitätssicherung und Transparenz  
im Gesundheitswesen (IQTiG)  
Katharina-Heinroth-Ufer 1

10787 Berlin

Per E-Mail: [weiterentwicklung-sd@iqtig.org](mailto:weiterentwicklung-sd@iqtig.org)

## GESCHÄFTSSTELLE

DGOU e. V. / DGOOC e. V. / DGU e. V.

Straße des 17. Juni 106 – 108  
10623 Berlin

Tel. +49 - (0)30 - 340 60 36 00

Fax +49 - (0)30 - 340 60 36 01

[office@dgou.de](mailto:office@dgou.de)

[www.dgou.de](http://www.dgou.de)

Berlin, 21.11.2019

**Gemeinsame Stellungnahme  
der Deutschen Gesellschaft für Orthopädie und Unfallchirurgie (DGOU),  
der Deutschen Gesellschaft für Orthopädie und Orthopädische Chirurgie (DGOOC) und  
der Deutschen Gesellschaft für Unfallchirurgie (DGU)  
zum Vorbericht des IQTiG „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“**

Sehr geehrte Damen und Herren,

in der Anlage übersenden wir Ihnen die gemeinsame Stellungnahme der DGOU, der DGOOC und der DGU zum Vorbericht des IQTiG „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“, welche von Herrn Prof. Dr. Christoph Lohmann, Leiter des DGOU-Ausschusses Versorgung, Qualität und Sicherheit, erstellt wurde.

Mit freundlichen Grüßen

Prof. Dr. med. Bernd Kladny  
Generalsekretär der DGOU  
Generalsekretär der DGOOC

Prof. Dr. Dietmar Pennig  
Stellv. Generalsekretär der DGOU  
Generalsekretär der DGU

### Vorstand (gemäß §26 BGB Abs. 1)

Präsident: Prof. Dr. Paul A. Grützner, Stellvertretender Präsident: Prof. Dr. Carsten Perka

Generalsekretär: Prof. Dr. Bernd Kladny, Stellvertretender Generalsekretär: Prof. Dr. Dietmar Pennig

DGOU-Bankverbindung: APO-Bank München, IBAN: DE34 3006 0601 0007 4267 39, SWIFT-BIC: DAAEDED3

DGOU-Steuer-Nr. 27/640/53836, Amtsgericht Bochum, VR 3953

**Gemeinsame Stellungnahme  
der Deutschen Gesellschaft für Orthopädie und Unfallchirurgie (DGOU),  
der Deutschen Gesellschaft für Orthopädie und Orthopädische Chirurgie (DGOOC) und  
der Deutschen Gesellschaft für Unfallchirurgie (DGU)  
zum Vorbericht des IQTiG „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“**

Die Weiterentwicklung des strukturierten Dialogs bietet verschiedene Vorteile, insbesondere die Möglichkeit der Objektivierung des Verfahrens auf den verschiedenen Ebenen und eine höchst mögliche Transparenz für alle Beteiligten.

Hervorzuheben ist, dass Möglichkeiten zur Qualitätsförderung geschaffen werden. Ebenfalls ist wichtig, dass eine sehr viel feinere Betrachtungsweise der einzelnen Indikatoren durchgeführt wird, zum Beispiel durch eine Morbiditätsadjustierung.

Erfreulich ist, dass das Projekt EndoCert der DGOOC als Modellprojekt für die verschiedenen Facetten der Verbesserung und Weiterentwicklung des strukturierten Dialogs angesehen wird.

Insgesamt empfinde ich das vorgelegte Papier als eine sehr gelungene, fein graduierte Weiterentwicklung des strukturierten Dialogs.

Gez. Prof. Dr. Christoph Lohmann  
Leiter des DGOU-Ausschusses Versorgung, Qualität und Sicherheit

**Vorstand (gemäß §26 BGB Abs. 1)**

Präsident: Prof. Dr. Paul A. Grützner, Stellvertretender Präsident: Prof. Dr. Carsten Perka

Generalsekretär: Prof. Dr. Bernd Kladny, Stellvertretender Generalsekretär: Prof. Dr. Dietmar Pennig

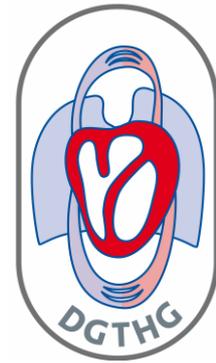
**DGOU-Bankverbindung:** APO-Bank München, IBAN: DE34 3006 0601 0007 4267 39, SWIFT-BIC: DAAEDED

**DGOU-Steuer-Nr.** 27/640/53836, **Amtsgericht** Bochum, VR 3953

# DEUTSCHE GESELLSCHAFT FÜR THORAX-, HERZ- UND GEFÄSSCHIRURGIE e.V.

DGTHG • Langenbeck-Virchow-Haus • Luisenstraße 58/59 • 10117 Berlin

An das  
IQTIG  
Institut für Qualitätssicherung und Transparenz im Gesundheitswesen  
Katharina-Heinroth-Ufer 1  
10787 Berlin



**SEKRETÄR**  
*Prof. Dr. Andreas Markewitz*

Geschäftsstelle:  
Langenbeck-Virchow-Haus  
Luisenstraße 58/59  
10117 Berlin  
Tel.: +49 (0) 30 28004 370  
Fax: +49 (0) 30 28004 379  
sekretariat@dgthg.de  
www.dgthg.de

---

sekretaer@dgthg.de

Nur per Mail an: [weiterentwicklung-sd@iqtig.org](mailto:weiterentwicklung-sd@iqtig.org)

20.11.2019

**Betr.:** Weiterentwicklung Strukturierter Dialog

**Hier:** Stellungnahme der Deutschen Gesellschaft für Thorax-, Herz- und Gefäßchirurgie (DGTHG)

**Bezug:** E-Mail von [weiterentwicklung-sd@iqtig.org](mailto:weiterentwicklung-sd@iqtig.org) vom 10.10.2019

Sehr geehrte Damen und Herren,

zunächst möchten wir uns im Namen der DGTHG bedanken, dass unsere Fachgesellschaft zu dem o.g. Projekt Stellung beziehen kann.

Unsere schriftliche Stellungnahme übersenden wir als Anlage

Mit freundlichen Grüßen

Prof. Dr. Andreas Markewitz  
*Sekretär*



## **Stellungnahme der DGTHG zum Vorbericht des IQTIG zu Stufe 1 und Stufe 2 der Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern**

Das IQTIG legt in dem o.a., 230 Seiten umfassenden Bericht seine Empfehlungen für die Gestaltung eines Stellungnahmeverfahrens für die DeQS-RL vor, wobei Stellungnahmeverfahren und Strukturierter Dialog als inhaltsgleich verstanden werden.

Er soll konkrete **Empfehlungen** beinhalten für

1. die Auswertung und
2. Bewertung von leistungserbringerbezogenen Qualitätsindikatorergebnissen
3. auf Basis von
  - 3.1. Dokumentationsdaten der Leistungserbringer,
  - 3.2. Sozialdaten sowie
  - 3.3. durch Patientenbefragungen erhobener Daten.

Im weiteren Text werden die Punkte 3.2 und 3.3 ohne weitere differenzierte Ausführungen nur noch erwähnt.

Durch die Empfehlungen des Berichts sollen **3 Ziele** des GBA verwirklicht werden, nämlich

1. Reduktion der Heterogenität in der Vorgehensweise zwischen den beauftragten Stellen,
2. Erhöhung der Transparenz und Nachvollziehbarkeit der Entscheidungen sowie
3. Steigerung der Effizienz des Strukturierten Dialogs.

Das IQTIG hat die vom GBA erteilten **Aufträge** für seinen Bericht wie folgt verstanden:

1. Optimierung der Einheitlichkeit der Vorgehensweise
2. Optimierung der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung
3. Optimierung der Effizienz des Verfahrens
4. Entwicklung eines Rahmenkonzepts zur Weiterentwicklung des Strukturierten Dialogs

Im Bericht sind die **Kapitel 2 und 3** (S. 20 – 65) einer ausführlichen **Ist-Analyse** z.T. mit einigen Hinweisen zu Möglichkeiten einer Modifikation des bisherigen Vorgehens gewidmet, in den **Kapiteln 4 und 5** (S. 66 – 142) werden statistische Methoden angesprochen, auf die in dieser Stellungnahme nicht eingegangen wird.

Ab S. 143 bzw. **Kapitel 6** erfolgen Vorschläge zur **Fachlichen Bewertung** und den damit verbundenen Änderungen des bisherigen Verfahrens.

### **1. Neues Bewertungsschema zur Einstufung der Indikatorergebnisse**

Die bisherigen Begriffe „qualitativ auffällig“ bzw. „qualitativ unauffällig“ sollen durch eine detailliertere und damit auch komplexere, in Tab.13 des Berichts dargestellte Matrix ersetzt werden.

Dagegen gibt es **fachlich keine Einwände**.



In diesem Zusammenhang wird von IQTIG zurecht angemahnt, dass die Bezeichnung „kein Hinweis auf Qualitätsdefizit“ offen lässt, ob zureichende Qualität vorliegt oder nur Unsicherheit über die Qualitätsbewertung besteht. **Konkrete Hinweise des IQTIG; wie dies geändert werden kann, sind nicht zu erkennen, wären aber wünschenswert.**

## 2. Methodik der fachlichen Bewertung

Das IQTIG empfiehlt hier ein kombiniertes Vorgehen, bei dem Einflussfaktoren, deren Beitrag zum Indikatorergebnis eindeutig angegeben werden kann, in Form einer partiellen Nachberechnung des Indikatorwertes und die übrigen angeführten Einflussfaktoren durch heuristische Beurteilung mit Unterstützung der Expertinnen und Experten berücksichtigt werden.

Dagegen gibt es **fachlich keine Einwände**, das **Vorgehen kompliziert jedoch das Verfahren.**

## 3. Ablauf der fachlichen Bewertung

Hier wird die Beurteilung anhand einer Kombination aus Ergebnissen einer partiellen Nachberechnung und der heuristischen Beurteilung vorgeschlagen. Auf die entscheidende Bedeutung der fachlichen Expertise von Experten bei diesen Schritten wird explizit hingewiesen.

Dagegen gibt es grundsätzlich keine **fachlich keine Einwände**, das **Vorgehen kompliziert jedoch das Verfahren.**

Der letzte Absatz dieses Kapitels (6.3) ist allerdings **aus fachlicher Sicht nicht akzeptabel**. Nachdem das Verfahren der fachlichen Bewertung zuvor sehr detailliert und transparent dargestellt wurde, heißt es am Ende:

*Die zuständige Stelle trifft die abschließende Entscheidung in Bezug auf die Einordnung des Indikatorergebnisses in das Bewertungsschema. **Weicht die zuständige Stelle von den Empfehlungen der Fachkommission ab, begründet und dokumentiert sie diese Abweichung.** Mit der abschließenden Entscheidung in Bezug auf die Einordnung des Indikatorergebnisses in das Bewertungsschema ist die Qualitätsbewertung abgeschlossen.*

Es bleibt unklar,

1. wer in der zuständigen Stelle
2. mit welcher Qualifikation
3. nach welchen Kriterien die abschließende Entscheidung trifft und
4. in welcher Form dieses Vorgehen nach außen transparent gemacht werden soll.
5. Weiter bleibt unklar, warum die fachliche Expertise bei einer abweichenden Entscheidung nicht erneut eingeholt wird.

Insgesamt kann man dieses Vorgehen aus fachlicher Sicht nur als „Degradierung“ der Fachkommission verstehen.

## 4. Bewertungsalgorithmus zur Beurteilung von Stellungnahmen

In diesem in mehrere Unterkapitel aufgeteilten Abschnitt werden die methodischen Schritte der Punkte 2. und 3. detailliert ausgeführt. Im Zusammenhang mit der partiellen Nachberechnung des Indikatorwertes werden die Begriffe „informativer Indikatorwert“ und



„bereinigter Referenzwert“ neu eingeführt, erklärt und ihre Bedeutung für die Neubewertung der Ergebnisse eines Qualitätsindikators (QI) dargestellt. Weiter wird darauf hingewiesen, dass weitere Einflussfaktoren, die einer Nachberechnung nicht zugänglich sind, durch Fachexperten abschließend auf ihre Relevanz geprüft werden sollen.

Weitere Hinweise zum Vorgehen mit Beispielen finden sich in Kapitel 5 und 6 des Anhangs zu dem Bericht.

Bis **hierhin können die Inhalte fachlich nachvollzogen** werden.

Im **Unterkapitel 6.4.3.** wird dann zurecht darauf hingewiesen, dass „verteilungsbezogene Referenzbereiche keine Aussagen über die Erfüllung definierter Standards erlauben, sondern lediglich vergleichende Aussagen relativ zu den Ergebnissen aller Leistungserbringer.“ Es wird logisch gefolgert, dass „strenggenommen daher aus methodischer Sicht für Indikatoren mit einem verteilungsbezogenen Referenzbereich das Einstufungsergebnis „keine Qualitätsbewertung“ (S) lauten müsste“.

Dem kann man sich **fachlich anschließen**, es ist allerdings ein **irritierendes Ergebnis für viele seit mehr als 10 Jahren bestehende QS-Verfahren.**

Das dann als „**Interimslösung**“ vorgeschlagene Vorgehen „perzentilbasierte Referenzbereiche im Stellungnahmeverfahren wie feste Referenzbereiche zu behandeln“ kann **fachlich nicht mitgetragen werden.** Es sollte im Gegenteil umgehend durch die als „perspektivisch“ bezeichnete „Festlegung fester, kriteriumsbezogener Referenzbereiche für diese Indikatoren Klarheit geschaffen werden, welche Qualitätsstandards von den Leistungserbringern gefordert werden.“ Die für diese Aufgabe notwendige Expertise ist mit den diversen Expertengremien auf Bundes- und Landesebene vorhanden.

Die im weiteren vorgestellten **Unterkategorien für die Einstufung „Qualitätsdefizit“** erfordern einen **zusätzlichen, nicht unerheblichen Rechenaufwand**, sind vermutlich in der Praxis **schwer vermittelbar** und haben aus fachlicher Sicht **keinen sinnvollen Vorteil** für das Verfahren, sondern führen **eher dazu Qualitätsdefizite zu verschleiern.**

Auch die **Unterkategorien bei Entkräftung des Hinweises auf ein Qualitätsdefizit** erfordern einen **zusätzlichen, nicht unerheblichen Rechenaufwand**, sind vermutlich in der Praxis **schwer vermittelbar** und haben aus fachlicher Sicht **keinen sinnvollen Vorteil** für das Verfahren.

## 5. Mindestanforderungen an Stellungnahmen

In diesem Kapitel (6.5) werden formale Kriterien vorgestellt, die bundeseinheitlich als Mindestanforderung für jede Stellungnahme gelten sollen und von der zuständigen Stelle geprüft werden. Es wird vom IQTIG weiter empfohlen, bundesweit einheitlich die Textvorschläge aus Kapitel 7 des Anhangs für die Einholung von Stellungnahmen zu verwenden.

Dies kann **fachlich nachvollzogen** werden.

## 6. Zusammensetzung von Fachkommissionen

In diesem Kapitel (6.6) werden orientiert an den Zielen und Aufgaben der Fachkommission bestimmte Anforderungen sowohl an die einzelnen Expertinnen und Experten als auch an die Zusammensetzung der Gruppe als Ganzes formuliert.



Diese sollten bestimmte **Erfolgsfaktoren** (fachliche Kompetenz, Unabhängigkeit und Interdisziplinarität) bei der Zusammensetzung der Fachkommission berücksichtigen. Es wird empfohlen, eine **fachlich qualifizierte, unabhängige und interdisziplinäre Fachkommission** zusammenzusetzen, um die Akzeptanz des Gesamtverfahrens sicherzustellen und Qualitätsdefizite erfolgversprechend abzubauen. Die **Verantwortung** für die Zusammensetzung einer fachlich qualifizierten, unabhängigen und interdisziplinären Fachkommission soll zukünftig **bei der zuständigen Stelle** liegen. Zur Steigerung der Akzeptanz des Verfahrens gegenüber den Expertinnen und Experten der Fachkommission wird eine **bundesweit einheitliche Regelung zur Aufwandsentschädigung**, die gleichermaßen für alle Expertinnen und Experten gilt, empfohlen. Die Aufwandsentschädigung wird im weiteren Text nicht erneut thematisiert oder konkretisiert.

Bei den **Experten** wird zwischen **fachlicher Eignung** (Qualifikation und berufliche Erfahrung) und **formalen Anforderungskriterien** (z. B. Zusicherung der Vertraulichkeit, Angaben zu Interessenkonflikten) unterschieden.

Um Experten mit möglichst aktuellem Wissen und Erfahrungen zu gewinnen, wird **fachlich inhaltlich** gefordert, dass die Experten in den letzten 3 Jahren vor ihrem Mitwirken in der Fachkommission in dem entsprechenden Versorgungsgebiet tätig gewesen sein müssen. Experten, die in den Ruhestand wechseln, können bis zu 4 Jahre nach ihrem Ausscheiden aus dem aktiven Berufsleben mitwirken.

Die Forderung nach Experten mit möglichst aktuellem Wissen und Erfahrungen kann **fachlich nachvollzogen** werden, die Details, wie das Vorhandensein bzw. das Nicht-Vorhandensein dieser Expertise definiert wird, werden aus folgenden Gründen **fachlich nicht mitgetragen**:

1. *Fachexpertinnen und -experten sollen nach der Facharztausbildung die letzten drei Jahre vor ihrem Mitwirken in der Fachkommission im entsprechenden Versorgungsbereich praktisch tätig gewesen sein.*

Es ist nicht erkennbar, welche Evidenz bzw. welche Kriterien zu der Zeitgrenze von 3 Jahren geführt haben. Aus praktischer Sicht reichen 3 Jahre nicht aus, um bei einem Verfahren der externen Qualitätssicherung als Experte fungieren zu können.

2. *Expertinnen und Experten, die während ihrer Mitgliedschaft in der Fachkommission in den Ruhestand wechseln, können bis zu vier Jahre nach ihrem Ausscheiden aus dem aktiven Berufsleben mitwirken.*

Es ist nicht erkennbar, welche Evidenz bzw. welche Kriterien zu der Zeitgrenze von 4 Jahren geführt haben. Bei Verfahren der externen Qualitätssicherung ist neben der praktischen Erfahrung, die am Ende eines Berufslebens in ausreichendem Maße vorhanden sein sollte, insbesondere die Kenntnisse der aktuellen Leitlinien und der themenbezogenen Literatur von Bedeutung. Es ist nicht nachvollziehbar, warum diese Kenntnisse nach 4 Jahren nicht mehr vorhanden sein sollten bzw. nicht weiter vervollständigt werden können.

3. *Expertinnen und Experten, die im Ausland tätig sind oder waren, sollten mit den nationalen fachlichen Standards vertraut sein und mindestens fünf Jahre als Klinikerinnen/Kliniker in Deutschland tätig gewesen sein.*

Es ist nicht erkennbar, welche Evidenz bzw. welche Kriterien zu der Zeitgrenze von 5 Jahren geführt haben.



Die **formalen Anforderungskriterien** sind **fachlich nachvollziehbar**, allerdings in der detaillierten Beschreibung auch als formalistisch zu bezeichnen, was an ihrer Berechtigung nichts ändert.

Bei den Kriterien für die Zusammensetzung der Fachkommission steht die Interdisziplinarität im Vordergrund. Das IQTIG macht dabei **konkrete Vorschläge über Details der Zusammensetzung der Fachkommissionen**, die möglicherweise im Fall der Krankenhaushygieniker bzw. Hygienefachkräfte an die **Grenzen der realen Verfügbarkeit dieser Berufsgruppen** stoßen wird, im Übrigen aber **fachlich nachvollzogen** werden kann.

**Unbeantwortet bleibt die Frage**, inwieweit die in Kapitel 6.6.2 aufgeführten Mitglieder der Fachkommission die in Kapitel 6.6.1 definierten Kriterien für die Auswahl der Expertinnen und Experten erfüllen müssen oder können.

Im sich anschließenden **Kapitel 7** wird das Thema **Qualitätsförderung** unter Betonung der Maßnahme Zielvereinbarung kurz vorgestellt, ohne in die Tiefe zu gehen. Die Inhalte des Kapitels sind **fachlich nachvollziehbar**.

**Kapitel 8** behandelt die **Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens**. Die damit verbundenen Vorteile liegen auf der Hand und werden im Bericht kurz zusammengefasst.

Eine Verkürzung des Stellungnahmeverfahrens kann grundsätzlich auf **drei unterschiedliche Weisen** erreicht werden:

1. Eine Möglichkeit ist, den Aufwand für die zuständigen Stellen zu reduzieren, indem z. B. die **Anzahl zu prüfender Qualitätsindikatorergebnisse verringert** wird. Dies kann über die Anzahl an ausgewerteten Qualitätsindikatoren sowie über die Methodik zur Berechnung der quantitativen Auffälligkeitseinstufung für diese Qualitätsindikatoren gesteuert werden. Dadurch stünden mehr Ressourcen für die Bearbeitung der verbleibenden Indikatorergebnisse zur Verfügung und die Bearbeitung könnte unter Umständen beschleunigt werden.

Diese Möglichkeit wird insbesondere von den LQS und LAG favorisiert, kann bei fachlich begründetem Verzicht auf bestimmte QI auch umgesetzt werden, wird jedoch **aus fachlicher Sicht nicht empfohlen**.

2. Eine zweite Möglichkeit besteht darin, die **Prozesse des Stellungnahmeverfahrens effizienter** zu gestalten.

Dabei soll bei Nicht-Einhaltung der formalen und inhaltlichen Anforderungen an die Stellungnahmen der Leistungserbringer direkt eine abschließende Bewertung ohne weitere Einholung von Präzisierungen erfolgen.

Dies wird **aus fachlicher Sicht abgelehnt**.

Eine Vorgabe von einzuhaltenden **Fristen zur Verkürzung des Verfahrens** wird grundsätzlich begrüßt, allerdings ist der für die Stellungnahme durch die Leistungserbringer vorgesehene Monat Juli einer der Haupturlaubsmonate in Deutschland, und dementsprechend ungeeignet, um eine 4-Wochenfrist anwenderfreundlich für die zuständigen Stellen



auf Landesebene bzw. die Leistungserbringer zu realisieren. Dies wurde im Übrigen bereits im Vorfeld durch die zuständigen Stellen auf Landesebene adressiert.

Inwieweit die in **Kapitel 8.3** aufgeführten Möglichkeiten zur Verkürzung des Verfahrens realistisch oder gar zielführend sind, kann **nicht abschließend beurteilt werden**.

3. Drittens besteht die Möglichkeit, **mehr Ressourcen zur Verfügung zu stellen**, z. B. indem mehr Personal bei den LQS oder LAG eingesetzt wird, um die Bearbeitung zu beschleunigen.

Dies wird **aus fachlicher Sicht mit Nachdruck befürwortet**, da dies die wesentliche Voraussetzung für eine mögliche Verkürzung des Stellungnahmeverfahrens darstellt.

Demgegenüber werden die in **Kapitel 8.2.4 vorgestellten Empfehlungen** für einen beschleunigten Abschluss des Stellungnahmeverfahrens **aus fachlicher Sicht als unrealistisch** eingeschätzt.

In **Kapitel 9** werden **Empfehlungen für eine optimierte Berichterstattung** zu den Ergebnissen des Strukturierten Dialogs vorgestellt.

Dabei wird zunächst empfohlen, dass die Berichterstattung zukünftig aus 2 Berichtsarten besteht, die beide zu unterschiedlichen Zeiten in den Bundesqualitätsbericht (BQB) einfließen:

1. Berichterstattung zur Versorgungsqualität: Qualitätsergebnisse am Ende des Moduls „Qualitätsbewertung“
2. Berichterstattung über die Maßnahmen der Qualitätsförderung mit Empfehlungen für die einzuleitenden qualitätsfördernden Maßnahmen und die Ergebnisse aus der Umsetzung dieser Maßnahmen.

Hier ergeben sich **aus fachlicher Sicht keine grundsätzlichen Einwände**, da keine belastbaren Daten für eine Beurteilung dieser Vorschläge vorhanden sind.

In **Kapitel 10** werden **Empfehlungen für die Evaluation** des neuen Konzepts vorgestellt.

Dabei werden ein mögliches Evaluationskonzept, die Datenquellen und die Zeitschiene sowie der Evaluationsgegenstand, die quantitative Qualitätsbewertung, beschrieben und im Anschluss durch Ausführungen zum statistischen Vorgehen ergänzt. Auf eine Bewertung der überwiegend aus der Beschreibung statistischer Methoden aufgebauten Aussagen wird in dieser Stellungnahme bewusst verzichtet.

Insofern bleibt abzuwarten ob die Empfehlung, sowohl eine formative als auch eine summative Evaluation des Konzepts durchzuführen, sich als zielführend erweisen wird.

In **Kapitel 11** werden die **Empfehlungen und Konsequenzen zusammengefasst**.

Dabei wird zunächst auf die **Weiterentwicklung der Qualitätsindikatoren** fokussiert und nachvollziehbar gefordert, im Rahmen der Weiterentwicklung der Indikatoren für alle Indikatoren mit verteilungsbezogenen (perzentilbasierten) Referenzbereichen zu prüfen, ob diese in feste Referenzbereiche überführt werden können.

Diese Forderung wird **fachlich mit Nachdruck unterstützt**.



Im Anschluss werden weitere Möglichkeiten der Weiterentwicklung diskutiert, wie die Eignungsprüfung anhand von Eignungskriterien.

Die **Gleichsetzung quantitativer Auffälligkeit mit qualitativer Auffälligkeit** im Sinne eines Verzichts auf das Stellungnahmeverfahren wird vor dem Hintergrund des Weiterentwicklungsbedarfs einiger bestehender Qualitätsindikatoren und dem damit einhergehenden Bedarf an einer (unbekannten) Steigerung der Spezifität des Vorgehens nicht empfohlen.

Auch von der Verwendung von **aggregierten Qualitätsindikatorergebnissen (Indices) als Aufgreifkriterien** wird dringend abgeraten.

Hinsichtlich **leistungsbereichübergreifender Indikatorensets** wird darauf hingewiesen, dass im Rahmen einer Konzeptentwicklung die Bedingungen für einen sinnvollen Einsatz dieser Indikatorensets beschrieben werden können, was für den Leser nur einen sehr limitierten Informationsgewinn darstellt.

Den genannten Empfehlungen kann man sich jedoch **aus fachlicher Sicht anschließen**

Der vorletzte Satz des Berichts ist dann allerdings **irritierend**, da am Ende des 230 Seiten umfassenden Berichts **vom IQTIG die Entwicklung eines Umsetzungskonzepts** empfohlen wird. Dies führt unweigerlich zu der Frage, warum dies im vorliegenden Bericht nicht gelungen ist und nicht als Bestandteil des Auftrags verstanden wurde.

## Fazit:

1. Der Bericht enthält eine Reihe von fachlich nachvollziehbaren Empfehlungen, ohne dass sich ein zusätzlicher Nutzen für den Strukturierten Dialogs bzw. das Stellungnahmeverfahren direkt erschließt.
2. Die Empfehlungen werden jedoch die Anforderungen an die Leistungserbringer, die Fachkommissionen und die zuständigen Stellen auf Landesebene deutlich erhöhen.
3. Das Potential, das Verfahren bei Umsetzung der Empfehlungen zu verkürzen und die Aussagen zur Versorgungsqualität zu schärfen, ist erkennbar, geht aber mit einer deutlichen Steigerung der Komplexität des Verfahrens einher, so dass die Frage der Umsetzbarkeit vorerst unbeantwortet bleibt.
4. Die Kriterien für die Zusammensetzung der Fachkommissionen sollten überarbeitet, korrigiert bzw. ergänzt werden
5. Inakzeptabel ist die Intransparenz zur abschließenden Entscheidung in Bezug auf die Einordnung des Indikatorergebnisses in das Bewertungsschema und die damit einhergehende Relativierung der Bedeutung der Fachkommission.

---

**Stellungnahme  
der Deutschen Krankenhausgesellschaft**

vom 21. November 2019

---

**zum IQTIG-Bericht  
„Weiterentwicklung des Strukturierten Dialogs mit  
Krankenhäusern - Vorbericht zu Stufe 1 und Stufe 2“  
(Stand 10. Oktober 2019)**

## Inhalt

Executive Summary .....	3
Einleitung.....	5
Zur Erfüllung der Beauftragung .....	7
Allgemeine Stellungnahme .....	11
Kapitelbezogene Stellungnahme .....	13
1.    Einleitung.....	13
2.    Analyse der Ausgangssituation .....	15
3.    Methodischer Hintergrund .....	15
4.    Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und -förderung ..	17
5.    Statistische Auswertungsmethodik .....	18
6.    Fachliche Bewertung .....	19
7.    Modul Qualitätsförderung .....	25
8.    Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens .....	25
9.    Berichterstattung .....	26
10.   Empfehlungen zur Evaluation des neuen Konzepts .....	26
11.   Zusammenfassung der Empfehlungen und Konsequenzen .....	27
12.   Literatur .....	28

Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten gleichwohl für alle Geschlechter.

## Executive Summary

---

Das Institut nach § 137a SGB V (IQTIG) hat am 10. Oktober 2019 den Bericht „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern - Vorbericht zu Stufe 1 und Stufe 2“ vorgelegt. Der wissenschaftlich gut aufbereitete Bericht enthält sehr weitreichende Vorschläge zur Optimierung der Einheitlichkeit der Vorgehensweise und der Effizienz des Verfahrens, die gut geeignet erscheinen, der Kritik an der Heterogenität des bisherigen Strukturierten Dialoges zu begegnen. Das IQTIG legt nachvollziehbar die Notwendigkeit einer maximalen methodischen Güte der Qualitätsindikatoren dar, verzichtet aber leider auf die beauftragte und notwendige Weiterentwicklung der Indikatoren. Stattdessen soll ein Bewertungsalgorithmus aus expliziten Entscheidungsregeln und einer heuristischen Komponente die derzeitige methodische Schwäche der Qualitätsindikatoren ausgleichen. Es verbleiben Zweifel, dass dies mit dem vorgeschlagenen Konzept gelingt. Daher sind Modifikationen erforderlich. Nachfolgend eine Zusammenfassung der relevanten Punkte:

- Die Trennung von Qualitätsbewertung und Qualitätsförderung kann die Effizienz des Verfahrens steigern und wird insofern mitgetragen, allerdings sind sowohl Qualitätsbewertung als auch Qualitätsförderung Teilaspekte *eines* Verfahrens – mit dem Ziel der Qualitätsverbesserung. Qualitätsfördernde Maßnahmen können hoch effektiv sein und müssen zunächst im Vordergrund stehen (DKG 2019).
- Der Schwerpunkt des vorliegenden Berichtes liegt auf der Qualitätsbewertung; Ausführungen zur Qualitätsförderung beschränken sich weitestgehend auf Verweise zur DeQS-RL und QFD-RL und sollen zukünftig bei wiederholten Auffälligkeiten innerhalb von 3 Jahren durch die zuständige Stelle eingeleitet werden. Positiv zu werten ist das im Anhang aufgenommene Muster für eine Zielvereinbarung.
- Die standardisierte Einholung schriftlicher Stellungnahmen erscheint im Sinne eines einheitlichen Vorgehens und zur Verkürzung des Verfahrens sinnvoll. Allerdings können eine ausschließlich schriftliche Darlegung von Ursachen im Rahmen des Stellungnahmeverfahrens und eine „Bewertung nach Aktenlage“ durch ein Expertengremium nicht gänzlich den kollegialen fachlichen Dialog auf Augenhöhe ersetzen.
- Der Bewertungsalgorithmus muss - auch aufgrund der methodischen Schwächen der Qualitätsindikatoren - um zwei Ablaufschritte ergänzt werden:
  1. Sofern die schriftliche Stellungnahme nicht ausreicht, muss die Fachkommission vor der abschließenden Bewertung weiterhin Zusatzinformationen aus Begehungen oder kollegialen Dialogen einholen können.
  2. Eine negative Bewertung ist dem Krankenhaus als Zwischenergebnis mit der befristeten Möglichkeit einer Erwiderung vorab mitzuteilen.
- Die Terminologie der Beurteilung und das Bewertungsschema müssen überarbeitet werden.

- Die Evaluation mit einem prozessbegleitenden und retrospektiven Ansatz einschließlich Wirkmodell ist überzeugend konzipiert, sollte aber um eine Pilotierung mit ausgewählten Indikatoren ergänzt werden.
- Die neue statistische Auswertungsmethode erscheint vielversprechend und wurde sorgfältig mittels externer Expertise entwickelt. Der Nachweis, dass diese treffsicherer als die bisher verwendete Methode ist und die Objektivität und Effizienz des Verfahrens steigert, muss nicht nur theoretisch anhand von Modellrechnungen, sondern auch in der Praxis im Rahmen der begleitenden Evaluation erbracht werden.

## Einleitung

---

Die Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern (QSKH-RL) definiert den Strukturierten Dialog wie folgt: Der Strukturierte Dialog ist eine Maßnahme, bei der unter Einbeziehung von Expertinnen und Experten auffällige Ergebnisse der Krankenhäuser im Dialog mit dem Krankenhaus bewertet und erforderlichenfalls weitere geeignete Maßnahmen zur Qualitätsverbesserung und –förderung (...) eingeleitet werden (§ 11 Abs. 1 QSKH-RL) (G-BA 2019a).

Der Strukturierte Dialog ist bisher von einer starken Heterogenität bei der Umsetzung und gleichzeitig hohem Aufwand geprägt. Die Berichte zum Strukturierten Dialog lassen ein unterschiedliches Vorgehen der Länder sowohl in den eingesetzten Maßnahmen – hierfür stehen gemäß Richtlinie Hinweise, Stellungnahmen, kollegiale Gespräche, Begehungen und Zielvereinbarungen zur Auswahl – als auch in der Ergebnisbewertung erkennen. Die Richtlinie lässt diese Heterogenität allerdings zu, da sie lediglich Eckpunkte für die Durchführung des Strukturierten Dialogs vorgibt. Zu nennen sind hier unter anderem:

- In Ausnahmefällen, z.B. bei der Verwendung von Qualitätsindikatoren-Sets oder bei Vorliegen landesindividueller Vorgaben, kann auf die Einleitung von Maßnahmen verzichtet werden (§10 Abs. 2 Satz 3 QSKH-RL). Gleiches gilt, sofern die Auffälligkeiten nur durch einen Fall pro Qualitätsindikator begründet ist und es sich nicht um einen Sentinel-Event-Indikator handelt.
- Die Richtlinie gibt keine Kriterien vor, wann die Versendung von Hinweisen ausreicht und wann Stellungnahmen angefordert werden sollen.
- In der Richtlinie sind keine Kriterien für die Veranlassung der weiterführenden Maßnahmen (kollegiale Gespräche, Begehung, Zielvereinbarung) festgelegt und auch keine Kriterien zur Nachverfolgbarkeit der Umsetzung von Zielvereinbarungen.

Vor diesem Hintergrund wurde bereits 2012 im Gemeinsamen Bundesausschuss (G-BA) aufgrund der anhaltenden Kritik über eine Weiterentwicklung beraten. Der G-BA hat 2013 eine Ist-Analyse und einen Workshop mit den Landesgeschäftsstellen für Qualitätssicherung durchgeführt und Projektgruppen zur weiteren Konkretisierung des Weiterentwicklungsbedarfs zum Strukturierten Dialog eingesetzt. Aufgrund des Wechsels des Instituts nach § 137a SGB V ruhten zunächst die Beratungen. Auf Antrag der Deutschen Krankenhausgesellschaft (DKG) wurden die Beratungen im Mai 2017 durch den Unterausschuss Qualitätssicherung des G-BA wieder aufgenommen und das IQTIG am 18. Januar 2018 im Plenum des G-BA mit der Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern beauftragt (G-BA 2018).

Ziel der Weiterentwicklung des Strukturierten Dialogs ist es

- die Einheitlichkeit der Vorgehensweise sowie
- die Transparenz und Nachvollziehbarkeit der Entscheidungsfindung und
- die Effizienz im Sinne einer Verschlankung des Verfahrens

zu optimieren.“

Beauftragt wurde ein zweistufiges Vorgehen. Grundsätzlich sollte bei allen Schritten der Beauftragung ein zu entwickelndes Rahmenkonzept und eine Abstimmung der verschiedenen Richtlinien mitberücksichtigt werden.

Das IQTIG hat am 30. November 2018 den Bericht zu Stufe 1 des Auftrags vom 18. Januar 2018 zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern vorgelegt. Dieser stellte den vom G-BA beauftragten ersten Bearbeitungsschritt zur Umsetzung des Auftrags vom 18. Januar 2018 dar, in einer ersten Entwicklungsstufe Vorschläge zur Optimierung der Einheitlichkeit der Vorgehensweise und zur Optimierung der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung vorzulegen. Ein Beteiligungsverfahren gemäß § 137a Abs. 7 SGB V zu diesem Bericht wurde vom IQTIG nicht durchgeführt. Ursprünglich sollten gemäß Beauftragung bis zum 31. Juli 2019 (mit Zwischenbericht bis 31. Januar 2019), die Entwicklungsarbeiten zu Stufe 2 folgen.

Am 10. Oktober 2019 legte das IQTIG den Bericht „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern - Vorbericht zu Stufe 1 und Stufe 2“ vor und eröffnete das Beteiligungsverfahren gemäß § 137a Abs. 7 SGB V. Die DKG bedankt sich für die Möglichkeit zur Stellungnahme. Im Folgenden wird zur sachgerechten Umsetzung des Auftrags Stellung genommen. Die Stellungnahme folgt im Wesentlichen der inhaltlichen Gliederung des Vorberichts. Hierdurch kann es zu mehrfacher Erwähnung ähnlicher Inhalte kommen.

## Zur Erfüllung der Beauftragung

---

Das IQTIG hat am 10. Oktober den Bericht „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern - Vorbericht zu Stufe 1 und Stufe 2“ zur Beauftragung des G-BA vom 18. Januar 2018 vorgelegt. Die Beauftragung umfasste folgende Auftragsteile:

### 1. Optimierung der Einheitlichkeit der Vorgehensweise

Beauftragte Entwicklungsleistungen:

- a) *Definition von Kriterien wann Hinweise (in besonderen Einzelfällen) und wann Stellungnahmen eingesetzt werden sollen*
  - b) *Erarbeitung von Empfehlungen zu Form und Inhalt der Einholung von Stellungnahmen*
  - c) *Erarbeitung von Arbeitshilfen (z.B. Beispielfälle) sowie Kriterien für die Bewertung der Arbeitsgruppen*
  - d) *Erarbeitung von standardisierten Bewertungskriterien zur Einstufung qualitativer Auffälligkeiten*
  - e) *Empfehlungen zur Evaluation des neuen Konzepts*
  - f) *Entwicklung von Empfehlungen in Bezug auf Anforderungen an die Zusammensetzung der Arbeitsgruppen und an die Experten (z.B. Qualifikation, Erfassung und Umgang mit Interessenkonflikten und Befangenheiten)*
  - g) *Erarbeitung einer Vorgehensweise, um eine möglichst unabhängig von Interessenkonflikten vorgenommene Bewertung zu erhalten, z.B. sicherstellen, dass die zu bewertenden Krankenhäuser den Bewertenden nicht bekannt sind, „Ringtausch“ der Bewertungen über die Bundesländer hinaus.*
- Insgesamt ist Nr. 1 der Beauftragung bis auf die Erarbeitung von Arbeitshilfen für die Bewertung der Arbeitsgruppen (z.B. Beispielfälle) im Wesentlichen erfüllt: Es werden ein standardisiertes Vorgehen zur Einholung von Stellungnahmen, ein neuer Bewertungsalgorithmus zur Vereinheitlichung der Vorgehensweise und Vorgaben für die Zusammensetzung der Fachkommissionen vorgeschlagen. Die in Kapitel 10 enthaltenen Empfehlungen zur Evaluation des vorliegenden Konzepts liefern eine gute Basis, wie prozessbegleitend und retrospektiv der Umsetzungsgrad und Nutzen sowie die Zielerreichung untersucht werden könnte. Das vorgeschlagene Wirkmodell muss um eine Pilotierung mit ausgewählten Qualitätsindikatoren vor der Einführung ergänzt werden.

## 2. Optimierung der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung

Beauftragte Entwicklungsleistungen:

- a) *Entwicklung eines Konzeptes zur Gewährleistung einer einheitlichen Datengrundlage von Qualitätsreport und Bericht zum Strukturierten Dialog*
  - b) *Optimierung der Berichterstattung zum Strukturierten Dialog (z.B. über die eingeleiteten Maßnahmen bei auffälligen Ergebnissen) und zur Datenvalidierung, dabei insbesondere Prüfung der Möglichkeit zur besseren Darstellung und Quantifizierung von Ursachen qualitativer Auffälligkeiten und eingeleiteter Verbesserungsmaßnahmen*
  - c) *Festlegung von Mindestanforderungen für die Länderberichte zum Strukturierten Dialog, dabei insbesondere Erstellung von Mustertabellen (Bereitstellung von Templates) für eine lückenlose Darstellung der durchgeführten Maßnahmen und der Bewertung sowie des Vorgehens bei fehlenden oder unzureichenden Stellungnahmen*
  - d) *Entwicklung eines Konzeptes zur ergänzenden Berichterstattung für die Öffentlichkeit und für teilnehmende Krankenhäuser (z.B. periodischer Newsletter) unter Nutzung von Informationen aus den Analysen auffälliger Ergebnisse, die bislang nur auffälligen Krankenhäusern zur Verfügung standen und/oder unter Nutzung von weiterführenden Auswertungen, die im Rahmen der Weiterentwicklung der Verfahren durchgeführt werden und modellhafte Erarbeitung an mindestens einem Beispiel*
  - e) *Entwicklung eines Konzeptes, wie identifizierte Verbesserungspotentiale und beispielhaft positive Versorgungsmodelle breit kommuniziert und ggf. für ein Fortbildungsinstrument nutzbar gemacht werden können. Ziel ist es, den Nutzen des Strukturierten Dialogs und damit des Verfahrens der externen stationären Qualitätssicherung für alle Beteiligten zu steigern.*
- Die Mindestanforderungen für die Länderberichte fehlen, ebenso Konzepte zur Publikation, Kommunikation und Fortbildung, um den Nutzen des Strukturierten Dialogs für alle Beteiligten zu steigern. Diese Beauftragungspunkte werden vom IQTIG zwar adressiert (vgl. Anmerkungen zu Kapitel 7 und Kapitel 9), aber unter Verweis auf das Auslaufen der QSKH-RL und die neuen Regelungen der DeQS-RL als derzeit noch nicht bearbeitbar bezeichnet. Es verbleiben daher Zweifel, ob weiterhin ein Benchmark im Sinne eines „Lernen von den Besten“ im neuen Konzept erfolgen kann.

## 3. Optimierung der Effizienz des Verfahrens

Beauftragte Entwicklungsleistungen:

- a) *Optimierung der Diskriminationsfähigkeit der Indikatoren unter Prüfung verschiedener (auch derzeit noch nicht genutzter) statistischer Verfahren, Abwägung von Vor- und Nachteilen, z. B. in Anlehnung an das Verfahren Planungsrelevante Qualitätsindikatoren („statistische Signifikanz“)*

*b) Prüfung der Verwendbarkeit von Indizes für derzeitige Leistungsbereiche im Sinne von Aufgreifkriterien für die Bewertung ganzer Leistungsbereiche*

*c) Prüfung von Optionen für leistungsbereichsübergreifende Indikatorensets,*

*d) Ggf. Identifikation von Indikatoren, für die eine rechnerische Auffälligkeit mit einer qualitativen Auffälligkeit gleichgesetzt werden kann*

*e) Methodische Weiterentwicklung der Indikatoren unter Nutzung der Erfahrungen aus der Datenbank zum Strukturierten Dialog und einer Umfrage bei allen auf Landesebene beauftragten Stellen und soweit möglich auch bei den LAGen*

*f) Notwendige Anpassungen der Berichterstattung (Bundes- und Landesauswertungen, Qualitätsbericht)*

*g) Prüfung von Optionen zur Verkürzung des Verfahrens Es ist darzustellen, welcher Ressourcenbedarf sich aus den Empfehlungen ergibt. Insbesondere falls sich die Empfehlungen als nicht effizient erweisen, sollten auch alternative Optionen zur Optimierung der Effizienz dargestellt werden.*

- Die Beauftragungspunkte zur Optimierung der Effizienz des Verfahrens sind nur als teilweise erfüllt anzusehen. Leider verzichtet das IQTIG aus Ressourcengründen auf die methodische Weiterentwicklung der Indikatoren und schlägt stattdessen eine Beurteilung der Qualitätsindikatoren durch die Bundesfachgruppen vor. Diese methodische Weiterentwicklung der Qualitätsindikatoren ist eine Aufgabe des wissenschaftlichen Instituts. Gerade im Hinblick auf die zukünftige Fokussierung auf ein weitgehend schriftliches Stellungnahmeverfahren sind hohe Anforderungen an die Güte der Qualitätsindikatoren zu stellen. Wäre eine solche Beurteilung der methodischen Güte der Indikatoren durch die Bundesfachgruppen leistbar, würde man eine Empfehlung zur Abschaltung oder Beibehaltung der Qualitätsindikatoren auch bereits im Rahmen der Systempflege erwarten.
- Das IQTIG legt nachvollziehbar und gut begründet dar, warum das Stellungnahmeverfahren als Prüfschritt des Verfahrens beizubehalten ist. Gleiches gilt für die Ablehnung einer Verwendung von Indizes. Leider verzichtet das IQTIG auch auf die nähere Untersuchung leistungsbereichsübergreifender Indikatorensets, obwohl diesbezüglich bereits positive Ergebnisse des nordrhein-westfälischen Pilotprojekts zur Weiterentwicklung des Strukturierten Dialogs (qs-nrw 2016) vorliegen.
- Mit einer Trennung von Qualitätsbewertung, und Qualitätsförderung und einer Standardisierung der Einholung von Stellungnahmen macht das IQTIG weitreichende Vorschläge zur Optimierung der Effizienz des Verfahrens.
- Unter Einholung externer Expertise wird sehr ausführlich eine neue Methodik der statistischen Auffälligkeit zur Optimierung der Diskriminationsfähigkeit vorgeschlagen, die treffsicherer sein soll als die der rechnerischen Auffälligkeit und insofern die Objektivität und Effizienz des Verfahrens steigern soll.

- Das Konzept wurde vom IQTIG „mit Blick auf die Implementation in der DeQS-RL entwickelt, durch die modulare Betrachtungsweise lasse sich das Modul Qualitätsbewertung jedoch prinzipiell auch in andere Richtlinien integrieren“ (S. 210). Diese Feststellung kann das beauftragte Rahmenkonzept nicht ersetzen. Zudem erschließt sich nicht ohne weiteres, dass auch die Qualitätsindikatoren anderer zukünftiger Verfahren generell mit dieser Methode auswertbar sein sollen, da Qualitätsindikatoren jeweils für einen vorgesehenen Anwendungsbereich entwickelt werden (IQTIG 2019a).

### **Fazit:**

Die Vorschläge des IQTIG zur Weiterentwicklung des Strukturierten Dialogs beinhalten gute Ansätze (Einheitlichkeit der Vorgehensweise, Vorgaben für die Bewertung und Zusammensetzung der Fachkommissionen), müssen aber noch weiterentwickelt und ergänzt werden (Prüfung der Verwendung von Qualitätsindikatorensets, Weiterentwicklung der Qualitätsindikatoren, Stellungnahmeverfahren mit Zwischenergebnis und erneuter Positionierung, Beibehaltung von Kollegialem Dialog und Begehung als Zusatzinformation, falls aus Sicht der Fachkommission erforderlich).

Daher sind Nachbesserungen am vorliegenden Bericht erforderlich, der in sich schlüssig und (bis auf den statistischen Teil in Kap. 5) nachvollziehbar geschrieben ist und gut aufbereitet die Problemstellungen der Beauftragung adressiert.

## Allgemeine Stellungnahme

---

Der Strukturierte Dialog ist ein wesentliches Verfahrenselement der externen stationären Qualitätssicherung. Er wird nach Maßgabe der Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern (QSKH-RL) für die 11 indirekten Verfahren auf der Landesebene und für die 10 direkten Verfahren auf der Bundesebene durchgeführt und geht mit einem extrem hohen Aufwand für alle Verfahrensbeteiligten einher. Wenngleich die zahlenmäßige Verteilung der Leistungsbereiche eine Gleichverteilung zwischen Bundes- und Landesebene vermuten lässt, ist festzuhalten, dass der Aufwand auf Landesebene ungleich höher ist. Direkte Verfahren weisen vergleichsweise geringe Fallzahlen und/oder eine geringe Anzahl beteiligter Standorte auf (§5 Abs. 2 QSKH-RL).

Gleichzeitig ist das Verfahren von hohem Aufwand und geringer Effizienz geprägt: Zum Erfassungsjahr 2017 waren von insgesamt 110.662 Indikatorergebnissen 11.413 Ergebnisse (10,3%) rechnerisch auffällig. Hierzu wurden im Rahmen des Strukturierten Dialogs 7.229 Stellungnahmen eingefordert, 4094 Hinweise verschickt, 314 kollegiale Gespräche und 17 Begehungen durchgeführt sowie 668 Zielvereinbarungen getroffen. Letztlich wurden nur 1,3% aller Indikatorergebnisse des Erfassungsjahres 2017 als qualitativ auffällig bewertet (IQTIG 2019)

Die datengestützte, einrichtungsübergreifende Qualitätssicherung muss relevanter, effizienter und fairer werden. Die Berücksichtigung der unterschiedlichen Risikoprofile der behandelten Patienten sowie das Verfahren zur abschließenden Bewertung der Ergebnisse durch neutrale praktisch tätige Fachkollegen sind essentieller Bestandteil fairer Qualitätsvergleiche (DKG 2019). Die Beurteilung der rechnerischen Auffälligkeit in Form eines Peer Reviews im fachlichen Gespräch ist derzeit das Kernstück des Strukturierten Dialogs. Die Klärung, ob der rechnerischen Auffälligkeit eines Qualitätsindikators tatsächlich qualitative Mängel zugrunde liegen, kann nicht gänzlich durch die Einholung einer schriftlichen Stellungnahme ersetzt werden. Dies lassen die Diskriminationsfähigkeit der Qualitätsindikatoren, ihre Risikoadjustierung und die Methodik zur Festlegung der Referenzbereiche nicht zu.

Daher ist zunächst eine grundlegende Prüfung der Evidenz und Aktualität der verwendeten Qualitätsindikatoren erforderlich. Vor diesem Hintergrund ist der Verzicht des IQTIG auf die beauftragte Weiterentwicklung der Qualitätsindikatoren nicht nachvollziehbar. Erschwerend kommt hinzu, dass auch die Nichteinhaltung formaler Kriterien oder eine nicht fristgerechte Übermittlung der Stellungnahme als „Qualitätsdefizit“ bewertet werden soll. Damit würde den Patienten in irreführender Weise ein Rückschluss auf eine defizitäre Qualität der Leistungserbringung suggeriert, obwohl lediglich Formfehler vorliegen. Erforderlich ist daher eine konsequente Korrektur der Terminologie des Bewertungssystems. Zudem ist fraglich, ob die Nichteinhaltung formaler Kriterien und deren Beurteilung als Qualitätsdefizit einer juristischen Prüfung standhalten. Zur Erhöhung der Justitiabilität des Verfahrens ist zumindest ein Zwischenschritt im Stellungnahmeverfahren erforderlich, der das Krankenhaus über das Zwischenergebnis einer absehbaren Bewertung als „qualitativ auffällig“ informiert und eine Möglichkeit zur Positionierung innerhalb einer vorgegebenen Frist einräumt.

Vor dem Hintergrund, dass die Stellungnahme darlegen soll, ob Gründe für die statistische Abweichung vom Referenzbereich vorliegen, die nicht vom Leistungserbringer zu verantworten sind und dass bei der Prüfung nur Gründe zu beachten sind, die nicht bereits in der Rechenregel des Indikators oder in dem Risikoadjustierungsmodell des Indikators berücksichtigt sind, setzt dies hohes Fachwissen der Experten in der Anwendung der neuen Methodik voraus. Es sollte der Fachkommission daher auch weiterhin stets möglich sein, ergänzende Informationen aus einem Gespräch mit dem Krankenhaus und/oder einer Begehung einzuholen, sofern aus ihrer Sicht die schriftliche Stellungnahme für eine Beurteilung nicht ausreicht. Ferner sollte auch ein Zeitfenster für ein Widerspruchsverfahren implementiert werden, wenn die Beurteilung der qualitativen Auffälligkeit aus Sicht des Krankenhauses nicht gerechtfertigt erscheint.

Der Strukturierte Dialog muss weiter mit Fachwissen durchgeführt werden, aber auch mit entsprechenden Konsequenzen. Der Gesetzgeber hat ein Verfahren gemäß § 137 Absatz 1 SGB V vorgesehen, wonach die Nichteinhaltung von Qualitätsanforderungen mit Konsequenzen verknüpft wird. Diese Vorgabe wird mit der Qualitätsförderungs- und Durchsetzungsrichtlinie (QFD-RL) umgesetzt. Bei der Durchsetzung von Qualitätsanforderungen sind die Maßnahmen verhältnismäßig und in einem gestuften System auszugestalten. Qualitätsfördernde und unterstützende Maßnahmen (z.B. Zielvereinbarungen, verpflichtende Teilnahme an geeigneten Fortbildungen oder Qualitätszirkeln, Begehungen, Teilnahme an Peer Reviews oder Implementierung von Standard Operating Procedures) können hoch effektiv sein und müssen zunächst im Vordergrund stehen. (DKG 2019). Die Qualitätsfördermaßnahmen finden im vorliegenden Konzept nicht genügend Raum.

## Kapitelbezogene Stellungnahme

---

### 1. Einleitung

#### *Kapitel 1.1 Hintergrund*

- Das IQTIG stellt fest, dass die Verantwortung für die Durchführung des Strukturierten Dialogs bei der jeweils mit der Durchführung des Verfahrens beauftragten Stelle liegt. Dies sei bei den direkten Verfahren der Unterausschuss Qualitätssicherung und bei den indirekten Verfahren die LQS bzw. LAG. Dies ist missverständlich: Das Pendant zum Unterausschuss Qualitätssicherung ist das nach § 14 QSKH-RL verantwortliche Gremium auf Landesebene (z.B. das Lenkungsgremium) (G-BA 2019). Die Durchführung des Strukturierten Dialogs für die direkten Verfahren obliegt dem IQTIG.

#### *Kapitel 1.2 Schnittstellen zu anderen Beauftragungen und Richtlinien*

- Die Schnittstellen zu anderen Beauftragungen und Richtlinien sind korrekt dargestellt.
- Den Begriff „Strukturierter Dialog“ der QSKH-RL durch „Stellungnahmeverfahren“ gemäß DeQS-RL zu ersetzen ist im Hinblick auf das Auslaufen der QSKH-RL ebenfalls nachvollziehbar; gleichwohl fällt auf, dass im Begriff „Dialog“ ein Austausch in beide Richtungen mit Nachfragen und Antworten verankert ist, während die Abgabe einer Stellungnahme dies nicht impliziert. Hierin besteht eine Schwäche dieses Begriffs im Hinblick auf seine Verwendung im vorliegenden Konzept.

#### *Kapitel 1.3 Auftragsverständnis*

- Das Verständnis des Auftrags ist folgerichtig abgeleitet. Allerdings fällt auf, dass die unter Nr. 3 ebenfalls beauftragte Optimierung der „Diskriminationsfähigkeit“ der Indikatoren aus dem Auftrag ausgeklammert und stattdessen die Entwicklungsleistungen zur Effizienz in den Vordergrund gerückt werden.
- Die Effizienz des Verfahrens wird allerdings nur in Bezug auf den Aufwand bei den zuständigen Stellen (LQS/LAG und IQTIG) interpretiert und nicht anwender- bzw. leistungserbringerorientiert. Dies ist eine Schwäche des vorliegenden Konzepts.
- Zu der ebenfalls beauftragten methodischen Weiterentwicklung der Indikatoren wird lediglich auf erste Empfehlungen, wie diese Weiterentwicklung durchgeführt werden könnte, verwiesen. Diese beziehen sich auf eine Befragung der Bundesfachgruppen und der auf Landesebene beauftragten Stellen, deren Ergebnisse dann für die Weiterentwicklung genutzt werden sollen. Nähere Ausführungen in Form wissenschaftlicher Ansätze zur Weiterentwicklung der Qualitätsindikatoren fehlen im Bericht.

### *Kapitel 1.4 Einbeziehung der auf Landesebene beauftragten Stellen*

- Die ebenfalls beauftragte Einbeziehung der auf Landesebene beauftragten Stellen wurde in Form eines Workshops mit 19 Teilnehmern aus 14 Bundesländern realisiert, deren Ergebnisse protokolliert und im Anhang zum Bericht ausführlich dargelegt werden. Dies ist außerordentlich positiv zu bewerten.

Folgende Anmerkungen der Teilnehmer sind für das vorliegende Konzept besonders relevant:

- Bezüglich des Konzepts wird Praxistauglichkeit und Entbürokratisierung gewünscht
- „Hinreichender Verdacht auf ein Qualitätsdefizit“ könnte verstanden werden als „Generalverdacht, der entkräftet werden müsse“
- Die aktuelle Anzahl der Qualitätsindikatoren wird als zu groß eingeschätzt, einige QI bilden die Qualität nicht ab, viele QI sind nicht geeignet zwischen guter und schlechter Qualität zu unterscheiden, QI sind teilweise veraltet
- Deutlichere Differenzierung des Bewertungsschemas in Bezug auf Fehldokumentationen als gesonderte Bewertungskategorie
- Die Nichteinhaltung formaler Kriterien darf nicht zur Bewertung „qualitativ auffällig“ führen; auch im Hinblick auf die Akzeptanz des Verfahrens
- Berücksichtigung der Haupturlaubszeit für das Stellungnahmeverfahren
- Nicht genügend Fachexperten vorhanden, die die hohen Anforderungen erfüllen und bereit sind mitzuwirken, insbesondere in kleinen Ländern.
- Die Bewertung soll zwingend von der Fachgruppe durchgeführt werden; die LQS/LAG soll nicht die abschließende Entscheidung treffen und diese verantworten müssen
- Erforderlich sind Kriterien zur Vorgabe, wann welche Fördermaßnahmen einzuleiten sind

### *Kapitel 1.5 Einbeziehung weiterer Expertise*

- Das IQTIG verweist darauf, dass die Entwicklung der statistischen Methoden (Kap. 5) mit Unterstützung eines externen Experten für statistische Prozesskontrolle erfolgte. Die dargestellte statistische Auswertungsmethodik ist für den Laien nicht nachvollziehbar. Die Auswertungsmethodik muss zukünftig für alle Beteiligten (LQS/LAG, Fachkommission, Leistungserbringer) transparent und nachvollziehbar sein.

## 2. Analyse der Ausgangssituation

### *Kapitel 2.1 Funktionen des Strukturierten Dialogs*

Das IQTIG definiert folgende Kernfunktionen des SD: Bewertung der Versorgungsqualität – Bewertung der Dokumentationsqualität – Qualitätsförderung.

- Den hier genannten Kernfunktionen des SD ist insgesamt zuzustimmen; allerdings erfolgt die Bewertung der Dokumentationsqualität primär im Rahmen des Datenvalidierungsverfahrens gemäß § 9 QSKH-RL und stellt - anders als hier dargestellt – keine *Kernfunktion* des Strukturierten Dialogs dar, auch wenn im Rahmen des Strukturierten Dialogs die valide Dokumentation der Daten ebenfalls zu beachten ist. Sowohl der Strukturierte Dialog als auch die Überprüfung der Dokumentationsqualität im Rahmen der Datenvalidierung sind gemäß § 15 QSKH-RL demzufolge auch getrennte Aspekte der Berichterstattung.
- Gleichwohl ist die Akzeptanz oder Nicht-Akzeptanz von Datenfehlern bei der Ergebnisbewertung im Rahmen des Strukturierten Dialogs von Bedeutung. (vgl. Anmerkungen zu Kap. 3 und Kap. 6)

### *Kapitel 2.6 Biometrische Analyse zu Aufwand und Effizienz des Strukturierten Dialogs*

- Das IQTIG stellt unter anderem nachvollziehbar dar, dass die bisher verwendete Methode der Auffälligkeitseinstufung ein Problem bei kleinen Fallzahlen erzeugt.

## 3. Methodischer Hintergrund

### *Kapitel 3.1 Die Methodische Funktion von Stellungnahmen*

- Das IQTIG stellt hier ausführlich und gut nachvollziehbar dar, dass und warum das Ergebnis von Qualitätsindikatoren nicht nur von der Versorgungsqualität der Leistungserbringer bestimmt wird, sondern auch von weiteren nicht von diesem beeinflussbaren Faktoren abhängt (z.B. Risikoprofil der Patienten, stochastische Einflüsse, im Indikator unberücksichtigte Versorgungssituationen).
- Der Referenzbereich soll diese unbekannt oder nicht gemessenen Einflussfaktoren berücksichtigen und das Stellungnahmeverfahren soll die Validität des Indikatorergebnisses für den jeweiligen Leistungserbringer prüfen, d.h. ob alle wichtigen Einflussfaktoren, die nicht vom Leistungserbringer beeinflussbar sind, berücksichtigt wurden.
- Das Stellungnahmeverfahren dient somit dazu, die Abweichung als gerechtfertigt (nicht beeinflussbarer Faktor) oder ungerechtfertigt (beeinflussbarer Faktor) zu klären.

- Der Referenzbereich an sich wird gar nicht mehr in Frage gestellt, lediglich Einflussfaktoren, die im Einzelfall eine Abweichung hiervon bedingen können. Diese Interpretation lässt unberücksichtigt, dass die grundsätzliche Schwierigkeit von Referenzbereichen darin besteht, eine scharfe Grenze zwischen zureichender oder unzureichender Qualität zu ziehen.

### *Kapitel 3.2 Reichweite der Qualitätsaussage von Stellungnahmen*

- Die fachliche Bewertung im Stellungnahmeverfahren soll sich nur noch auf das vom jeweiligen Qualitätsindikator adressierte Qualitätsmerkmal beziehen, auch wenn im Rahmen der Bewertung andere Auffälligkeiten – bis hin zu groben Behandlungsfehlern – zu Tage treten. Diese Reduzierung der Beurteilungsperspektive ist höchst problematisch, auch wenn sie der Homogenität der Vorgehensweise dient und der im Konzept verfolgten strikten Trennung von Qualitätsbewertung und Qualitätsförderung geschuldet ist.
- Es sollte zumindest möglich bleiben, im Rahmen des Stellungnahmeverfahrens erkannte Verbesserungspotenziale zu kommunizieren, damit diese in das Modul der Qualitätsförderung einfließen können.
- Auch im Stellungnahmeverfahren zutage getretene methodische Schwächen der Qualitätsindikatoren sollten breit kommuniziert und zur Weiterentwicklung der Qualitätsindikatoren genutzt werden.

### *Kapitel 3.3 Gütekriterien für Bewertungsprozesse*

- In diesem Kapitel wird die „intuitiv-klinische“ Urteilsbildung der „statistisch-mechanischen“ Urteilsbildung gegenüber gestellt. Das IQTIG konstatiert, dass die Urteilsforschung eine Überlegenheit der Güte der mechanischen Urteilsbildung zeige.
- Es gibt auch Untersuchungen, die zum gegenteiligen Schluss kommen und in einer intuitiven gegenüber einer rein kriteriengeleiteten Urteilsbildung Vorteile sehen (Gigerenzer 2007). Diese Diskussion sollte hier ergänzt werden.

### *Kapitel 3.4 Ableitung von Anforderungen an das Stellungnahmeverfahren und Kapitel 3.5 Vor- und Nachteile expliziter Beurteilungsregeln im Stellungnahmeverfahren*

- Ein möglichst standardisiertes, explizites Vorgehen bei der Einholung von Stellungnahmen, bzgl. ihres Inhaltes und diesbezüglicher Bewertungsregeln ist anzustreben. Die Standardisierung stößt jedoch an ihre Grenzen, wenn für jeden Einflussfaktor vorab eine Rechenvorschrift und Gewichtung festgelegt werden müssten. Dies ist bei unbekanntem Einflussfaktoren im Vorhinein des Bewertungsprozesses ohnehin nicht möglich.

- Die diesbezüglichen Limitationen expliziter Beurteilungsregeln werden vom IQTIG sehr gut beschrieben und münden in der Empfehlung technischer Unterstützung und einer grundsätzlichen Abwägung von Aufwand und Nutzen. Beidem ist uneingeschränkt zuzustimmen.
- Die Vorgaben zur Einholung der Stellungnahmen und die Vorgaben zur Zusammensetzung der Fachkommissionen werden bereits zur Objektivität der Qualitätsbewertung beitragen. Die heuristische Komponente in der Qualitätsbewertung durch die Fachexperten muss auch weiterhin ein wesentlicher Bestandteil der Qualitätsbewertung bleiben.

#### 4. Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und -förderung

##### *zur modularen Betrachtungsweise*

- Eine Trennung von Qualitätsbewertung und Qualitätsförderung ist grundsätzlich sinnvoll. Sie eröffnet nachvollziehbar Rationalisierungspotenziale.
- Allerdings muss den Experten auch die Einholung von Zusatzinformationen aus Gesprächen – gegebenenfalls auch telefonisch (qs-nrw 2016) - oder Begehungen weiterhin möglich sein, sofern aus Sicht der Experten die Beurteilung des Sachverhalts durch eine schriftliche Stellungnahme allein nicht möglich ist.

##### *zum grundlegenden Ablauf des Moduls der Qualitätsbewertung und -förderung*

- Es besteht Ergänzungs- und Änderungsbedarf hinsichtlich der Ablaufschritte, der Terminologie und des Bewertungsschemas, die in den nachfolgenden Kapiteln angemerkt wird; unter anderem hinsichtlich folgender Punkte:
  - Die Nichterfüllung formaler Anforderungen darf nicht mit „Qualitätsdefizit“ (der Leistungserbringung) beurteilt werden.
  - Die statistische Operationalisierung des hinreichenden Hinweises auf ein Qualitätsdefizit bei den verwendeten Qualitätsindikatoren muss für alle Beteiligten (Leistungserbringer, zuständige Stelle, Fachkommission) transparent und nachvollziehbar sein.
  - Der Leistungserbringer muss zur Abgabe einer Stellungnahme innerhalb einer angemessenen Frist aktiv von der zuständigen Stelle aufgefordert werden.
  - Der Leistungserbringer muss über das Zwischenergebnis der Bewertung informiert werden und die Möglichkeit erhalten, sich hierzu erneut innerhalb einer vorgegebenen Frist zu äußern.
  - Die Fachkommission muss vor der abschließenden Bewertung bei Bedarf Zusatzinformationen erheben können, z.B. in Form eines Gesprächs (auch telefonisch) oder einer Begehung.

## 5. Statistische Auswertungsmethodik

Die statistische Auswertungsmethodik (S. 71 bis 142) bildet einen Schwerpunkt des vorliegenden Berichts und ist mittels externer Expertise eines Fachexperten für statistische Prozesskontrolle erstellt worden. Die Ausführungen in diesem Kapitel sind für Laien größtenteils nicht verständlich.

### *Kapitel 5.1 Rahmenkonzept der Auswertungsmethodik*

- Durch die Wahl des Signifikanzniveaus (d.h. mit welcher Wahrscheinlichkeit liegt das Ergebnis - „der latente Kompetenzparameter“ - des Leistungserbringers innerhalb oder außerhalb des Referenzbereichs) dieser Auffälligkeitseinstufungsmethode kann die Anzahl der einzuholenden Stellungnahmen beeinflusst werden. Das klingt nach einer Variabilität, die für den vom IQTIG antizipierten Einsatz dieser Methode für die unterschiedlichen QS-Verfahren sicher von Vorteil ist, aber gleichzeitig auch ein fundiertes Verständnis für die Methodik voraussetzt, das weder im IQTIG (da externe Expertise) noch in den zuständigen G-BA Gremien vorhanden sein wird.
- Die für dieses Verfahren (noch) auszuwählenden Qualitätsindikatoren müssen so gewählt werden, dass sie eindeutig ein Qualitätsdefizit anzeigen und damit eindeutig vorgeben, ab wann Stellungnahmen einzuholen sind.
- Es ist fraglich, wie aufgrund dieser Methode eine Vergleichbarkeit im Benchmark für den Leistungserbringer selbst und für die Patienten bei Veröffentlichung der Ergebnisse möglich sein wird.
- Wenn zukünftig überwiegend „nach Aktenlage“ entschieden werden soll, ist die Nachvollziehbarkeit der Auswertungsergebnisse für die Leistungserbringer umso wichtiger. Die neue Komplexität der QI-Berechnung wird zu Lasten der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung gehen. Diese ist auch für die Justitiabilität des Verfahrens notwendig.
- Schon die derzeitigen Rechenregeln und Risikoadjustierung der Qualitätsindikatoren stellt für die meisten der Stellungnehmenden eine Herausforderung dar. Die nun nicht mehr linear berechneten Qualitätsindikatoren erfordern höchste statistische Expertise.

### *Kapitel 5.4 Berücksichtigung der Daten mehrerer Erfassungsjahre*

- Die vom IQTIG vorgeschlagene Berücksichtigung der Daten mehrerer Erfassungsjahre soll die Güte der Methodik erhöhen. Dies ist insbesondere bei kleinen Fallzahlen zu begrüßen. Hier ergibt sich jedoch das Dilemma, den Bezug zu dem die Auffälligkeit auslösenden Ereignis zu erhalten. Dieser muss für das Stellungnahmeverfahren zwingend möglich bleiben. Auch dies muss in der Evaluation untersucht werden (vgl. Anmerkungen zu Kap. 10). Eine Mehrjahresbetrachtung setzt zudem konstante Rechenregeln voraus.

## 6. Fachliche Bewertung

### Kapitel 6.1 Bewertungsschema zur Einstufung der Indikatorergebnisse

Das Bewertungsschema zur Einstufung der Indikatorergebnisse ist im Ungleichgewicht zu Ungunsten der Leistungserbringer und muss anders definiert werden:

- Kritisch gesehen wird, dass durch die verwendeten Begrifflichkeiten „statistische Auffälligkeit = hinreichender Hinweis auf Qualitätsdefizit“ und „qualitativ unauffällig = kein Hinweis auf Qualitätsdefizit“ die Leistungserbringung im Krankenhaus unter den Generalverdacht defizitärer Qualität gestellt wird.
- Das Indikatorergebnis wird im negativen Fall mit „Qualitätsdefizit“ und im positiven Fall lediglich mit „Kein Hinweis auf ein Qualitätsdefizit“ bewertet. Entweder müssten die Bewertungspaare „Qualitätsdefizit - kein Qualitätsdefizit festgestellt“ oder aber „Hinweis auf Qualitätsdefizit - kein Hinweis auf Qualitätsdefizit festgestellt“ heißen.
- Das IQTIG weist darauf hin, dass die Abwesenheit eines hinreichenden Hinweises auf ein Qualitätsdefizit nicht gleichzusetzen ist mit einem Hinweis auf zureichende Qualität und erklärt dies mit einer stets verbleibenden statistischen Unsicherheit, da ein Indikatorwert immer unbekanntem oder nicht gemessenen Faktoren ausgesetzt sei. Diese Unsicherheit verbleibt allerdings auch im negativen Fall, da Risikoadjustierungsmodelle niemals perfekt sein können - dennoch bezeichnet das IQTIG die Qualität dann als defizitär.
- Dies widerspricht auch der eindeutigen Interpretation von Bewertungsergebnissen, die das IQTIG nach eigenen Aussagen mit dieser Skala verfolgt, denn „Qualitätsdefizit“ ist stets negativ konnotiert und „kein hinreichender Hinweis auf ein Qualitätsdefizit“ ebenfalls.
- Positiv zu werten ist, dass zukünftig der Begriff „Datenfehler“ anstatt „Dokumentationsfehler“ verwendet wird, um klarzustellen, dass es auch Datenfehler geben kann, die ihre Ursache nicht in fehlerhaften Dokumentationsprozessen haben.
- Das IQTIG verweist darauf, die Bezeichnungen „hinreichender Hinweis auf ein Qualitätsdefizit“ und „Verfehlen des Referenzbereichs“ synonym zu verwenden. Inwieweit die Berechnung des indikatorbasierten *hinreichenden Hinweises auf ein Qualitätsdefizit* (Kap. 5) eine geeignetere Methode zur Unterscheidung zwischen auffälliger und unauffälliger Qualität ist, bleibt abzuwarten.
- Die mit Kategorie A 0 („Verzicht auf Einreichung einer Stellungnahme oder Stellungnahme nicht fristgerecht eingereicht oder Stellungnahme entsprach nicht den Anforderungen“) aufgenommenen formalen Gründe mit einem *Qualitätsdefizit* (der Leistungserbringung) zu bewerten, ist in der Außerdarstellung nicht sachgerecht, da hiermit ein Defizit der

Leistungserbringung mit einem Defizit der Einhaltung formaler Anforderungen gleichgesetzt wird, was zu Fehlinformationen der Patienten führt. Darüber hinaus erhalten Leistungserbringer, deren Leistung qualitativ einwandfrei ist und solche, deren Leistungsqualität nicht einwandfrei ist, ungerechtfertigter Weise dieselbe Beurteilung.

- Die Beurteilung formaler Fehler bei der Abgabe der Stellungnahme ist zwingend außerhalb des Bewertungsschemas für die Qualitätsindikatoren zu etablieren, z.B. in einer eigenen Bewertungskategorie. Gerade zu Beginn der Umstellung des Strukturierten Dialogs auf ein nahezu ausschließlich schriftliches Stellungnahmeverfahren wird die Einhaltung der formalen Anforderungen für alle Beteiligten eine große Herausforderung, die nicht mit der Bewertung der Leistungsqualität vermischt werden sollte.
- Auch die Gleichsetzung von Dokumentationsfehlern und fehlerhafter Qualität der Leistungserbringung innerhalb ein und derselben Bewertungskategorie (A) ist problematisch und irreführend in der Außendarstellung und muss geändert werden.
- Was sind vom Leistungserbringer zu verantwortende Datenfehler (Kategorien A1) bzw. *teilweise* zu verantwortende Datenfehler (Kategorie A2) und was sind vom Leistungserbringer nicht zu verantwortende Datenfehler (Kategorien U1 und U2)? Auf welchen Kriterien basiert die Beurteilung der Datenfehler? Inwieweit auf diesen Begriffsdefinitionen ein standardisiertes Vorgehen in der Bewertung etabliert werden kann, erschließt sich nicht ohne weiteres.

### *Kapitel 6.2 Methodik der fachlichen Bewertung*

- Das IQTIG verweist darauf, dass die fachliche Bewertung zu beurteilen habe, ob Einflussfaktoren vorliegen, die „nicht vom Leistungserbringer zu verantworten sind und die in ihrer Gesamtheit das Verfehlen des Referenzbereiches bei diesem Indikator erklären“. Die Stellungnahme müsse das Vorliegen solcher Einflussfaktoren „nachvollziehbar darstellen“. Hinzu kommt, dass nicht nur das Vorliegen, sondern auch die Wirkung der Einflussfaktoren auf das Indikatorergebnis in der fachlichen Bewertung berücksichtigt werden müsse. Hier ergeben sich Interpretations- und damit Beurteilungsspielräume, die ein standardisiertes Vorgehen erschweren.
- Die Beurteilung, ob die Einflussfaktoren nicht vom Leistungserbringer zu verantworten sind, erfolgt derzeit mittels heuristischer Beurteilung durch die zuständige Stelle und die beratenden Experten auf der Grundlage individueller fachlicher Erfahrungen. Demgegenüber soll nunmehr, um eine möglichst hohe Objektivität und Reliabilität der Bewertung zu erreichen, mittels statistischer Methoden (quantitative Gewichtung der genannten Einflussfaktoren, Algorithmus) eine Neubewertung des Indikatorergebnisses eines Leistungserbringers unter Berücksichtigung der Einflussfaktoren erfolgen. Das IQTIG legt nachvollziehbar dar, dass dies nur teilweise möglich ist und schlägt daher eine Kombination beider Ansätze vor. (Partielle Nachberechnung für

geeignete Einflussfaktoren und heuristische Beurteilung für sonstige Einflussfaktoren). Auch die Einschätzung, für welche Einflussfaktoren eine Nachberechnung möglich ist, soll von der Fachkommission entschieden werden. Im Sinne eines homogenen Vorgehens muss zukünftig gewährleistet sein, dass die Fachkommissionen länderübergreifend dieselben Einflussfaktoren akzeptieren.

### *Kapitel 6.3 Ablauf der fachlichen Bewertung*

- Der Gesamtalgorithmus für die fachliche Bewertung ist sehr ausführlich und auch mit der im Anhang dargelegten Abbildung schlüssig und gut nachvollziehbar dargestellt. Es stellt sich allerdings die Frage, ob diese sehr aufwändige Nachberechnung des Indikatorwertes auf Basis der Stellungnahmen den Nutzen rechtfertigt, wenn am Ende eine Variante der Bewertungskategorie A das Ergebnis ist. Anstelle des sehr detaillierten Bewertungsalgorithmus sollte es in vielen Fällen ausreichen, wenn eine Expertengruppe die Stellungnahme als hinreichend interpretiert und beurteilt, ob die Einflussfaktoren alle Fälle des Indikators betreffen oder nicht und entsprechend bewertet.
- Die auf Landesebene beauftragten Stellen müssen befähigt werden, die Extraktion der adressierten Einflussfaktoren und Klassifikation als Datenfehler oder fachlicher Faktor vorzunehmen. Außerdem müssen sie Datenfehler, die nicht vom Leistungserbringer zu verantworten sind, trennen von solchen, die dem Leistungserbringer zuzuschreiben sind. Die Diskussionen im Zusammenhang mit der 100% Dokumentationsrate haben gezeigt, wie schwierig eine solche Abgrenzung sein kann. Zudem müssen sie beurteilen, welche fachlich-inhaltlichen Einflussfaktoren nicht für eine Nachberechnung geeignet sind. Die Evaluation muss daher auch zeigen, inwieweit mit einem solchen Vorgehen tatsächlich eine Standardisierung des Verfahrens und Gleichbehandlung der Leistungserbringer und damit das angestrebte homogene Vorgehen der Landesebene erzielt werden kann.
- Die für die Durchführung des Stellungnahmeverfahrens vom IQTIG vorgeschlagene angemessene technische Unterstützung der auf Landesebene beauftragten Stellen wird ausdrücklich begrüßt. Dies erfordert eine entsprechende Ausstattung und Finanzierung der beauftragten Stellen, damit nicht aus einer ungleichen Ressourcenverteilung eine Verzerrung der Ergebnisse resultiert.
- Die Schwerpunktsetzung auf schriftliche Stellungnahmen bei der Qualitätsbewertung darf nicht dazu führen, dass sich der ehemals medizinische fachlich-kollegiale Dialog in eine Bewertung nach rein formalen Kriterien auf „Sachbearbeiterebene“ der zuständigen Stellen verwandelt. Die Experten sollten daher auch zukünftig die Möglichkeit haben, zusätzliche weiterführende Informationen auf Basis einer Begehung oder eines Gespräches mit dem Krankenhaus einzuholen, sofern sie der Auffassung sind, Zusatzinformationen für den Bewertungsprozess zu benötigen - auch wenn sich hierdurch der Bewertungsprozess verlängert.

- Es ist folgerichtig, dass an der Pseudonymisierung der Leistungserbringer grundsätzlich festgehalten wird, um Verzerrungen im Beurteilungsprozess zu vermeiden. Sofern Zusatzinformationen aus Begehungen oder kollegialen Gesprächen aus Sicht der Fachkommission im Einzelfall für die weitere Beurteilung erforderlich sind, muss auch eine De-Pseudonymisierung mit Zustimmung des Krankenhauses möglich sein. Dies erscheint gerechtfertigt, da es darum geht, das negative Beurteilungsergebnis zu hinterfragen.
- Eine Prüfung der vom IQTIG gelieferten Auswertungen kann durch die zuständigen Stellen nicht mehr in derselben Durchdringungstiefe erfolgen wie bisher, da die Landesebene zukünftig keine Einsicht in Rohdaten mehr haben wird. Daher ist auch die Informationsgrundlage der zuständigen Stellen und Fachkommissionen als Basis für die fachliche Bewertung zukünftig eingeschränkt. Es muss gewährleistet sein, dass die zuständigen Stellen weiterhin fachliche Ansprechpartner der Krankenhäuser bleiben. In der Abbildung der fachlichen Bewertung fehlt daher eine Rückkoppelung der zuständigen Stellen zum IQTIG, falls sich Rückfragen zu den Auswertungen und Indikatorergebnissen ergeben und die Informationsgrundlage nicht ausreicht für die fachliche Bewertung.

#### *Kapitel 6.4 Bewertungsalgorithmus*

Das IQTIG empfiehlt eine Kombination aus impliziter (heuristische Beurteilung durch die Fachkommission) und expliziter (quantitative Gewichtung der als Grund anerkannten Einflussfaktoren) Beurteilung. Begründet wird dies damit, dass eine vollständige explizite und quantitative Bewertung der in den Stellungnahmen geltend gemachten Einflussfaktoren nicht möglich sei, da z.B.

- nicht für alle vorgebrachten Einflussfaktoren bestimmbar ist, welches Ausmaß sie auf das Indikatorergebnis haben
- die Gewichtung der möglichen Einflussfaktoren nicht im Vorhinein standardisiert festlegbar (da die Einflussfaktoren vorher nicht bekannt oder zu selten sind)

Folgende Aspekte sind anzumerken:

- Abweichende Voten innerhalb der Fachkommissionen sollten transparent gemacht werden.
- Es sollten zwei Jahre Vorlaufzeit für den Katalog an akzeptierten Einflussfaktoren auf das Indikatorergebnis (=Einflussfaktoren die zu einer Nachberechnung des Indikatorwertes für den Leistungserbringer führen können) eingeräumt werden, damit die Krankenhäuser und alle am Bewertungsprozess Beteiligten sich darauf einstellen können.
- Zu diesem Katalog der anzuerkennenden fachlich-inhaltlichen Einflussfaktoren sollte auch die Stellungnahme der jeweiligen Fachgesellschaften eingeholt werden.

## Kapitel 6.5 Mindestanforderungen an Stellungnahmen

### Formale Kriterien

- Die im Anhang enthaltenen Textvorschläge zur Einholung von Stellungnahmen sind als gelungener Ansatz zu werten, die Stellungnahmen zu standardisieren und die Akzeptanz des Verfahrens zu erhöhen.
- Das IQTIG konstatiert: „*Das Ergebnis der fachlichen Bewertung soll keine umfassende Qualitätsaussage treffen, sondern nur die Angemessenheit der Indikатораussage untersuchen*“ (S. 166). Diese Auffassung würde einen Paradigmenwechsel der Qualitätsbewertung bedeuten. Bisher untersucht der Strukturierte Dialog, ob die Leistung in der fachlich gebotenen Qualität erbracht wurde. Der Fokus der Qualitätsbewertung muss auch weiterhin auf der Untersuchung der Qualität der Leistungserbringung liegen und nicht auf einer Untersuchung der methodischen Güte von Qualitätsindikatoren. Es dürfen daher nur Qualitätsindikatoren verwendet werden, die für die Beurteilung der Leistungsqualität hinreichend valide sind. Gleichwohl sollten Informationen aus dem Stellungnahmeverfahren für die Weiterentwicklung der Qualitätsindikatoren genutzt werden.
- In der Stellungnahme soll lt. IQTIG nachgewiesen werden, dass mit der „*gebotenen Sorgfaltspflicht alles Zumutbare getan wurde*“. Auch hier zeigen sich wieder definitorische Schwächen im Bericht, die ein standardisiertes Vorgehen erschweren: Was umfasst die *gebotene Sorgfaltspflicht* und was ist *zumutbar*?
- In der Stellungnahme angeführte Gründe, die aus Sicht des Leistungserbringers nicht von ihm zu verantworten waren, sollen zukünftig mittels schriftlichen Belegs nachgewiesen werden. Diesbezüglich sind daher genaue Vorgaben erforderlich, was als schriftlicher Beleg gilt.
- Zukünftig soll *die fachlich verantwortliche Instanz des Leistungserbringers (z.B. Chefärztin/Chefarzt über den Hinweis auf ein Qualitätsdefizit informiert werden und die angefertigte Stellungnahme autorisieren*. Es ist darauf zu achten, dass diese Autorisierung möglichst einfach erfolgen kann, z.B. über eine entsprechende Erklärung des/der benannten Arztes/Ärztin, dass die Abstimmung mit dem Chefarzt erfolgt ist.
- Entspricht eine Stellungnahme nicht den formalen Kriterien – auch nicht im Rahmen der Korrekturfrist von 4 Wochen – soll die Stellungnahme als nicht abgegeben gelten: Die Ablehnung von Stellungnahmen aufgrund von Formfehlern ist nicht rechtskonform, da Stellungnahmen daraufhin zu prüfen sind, ob die im überprüften Jahr dokumentierten Leistungen in der fachlich gebotenen Qualität erbracht und valide dokumentiert wurden (§12 Abs. 1 Satz 1 QSKH-RL). Bei einer Ablehnung aufgrund von Formfehlern findet keine derartige Prüfung statt. Formale Fehler dürfen daher niemals zu einer automatischen Bewertung „Qualitätsdefizit“ führen.

- Eine lediglich mit Formfehlern behaftete Stellungnahme darf im Ergebnis nicht mit einer gänzlich fehlenden Stellungnahme gleichgesetzt werden.
- Die Frist zur Erst-Abgabe einer Stellungnahme ist mit 4 Wochen zu kurz. Hier sollten – vor allem in der Urlaubszeit – Verlängerungsmöglichkeiten eingeräumt werden. 6 Wochen erscheinen realistischer. Kritisch zu sehen ist der fehlende Austausch zwischen Sender und Empfänger der Stellungnahmen.
- Hinweise an die Krankenhäuser können als Frühwarnsystem im zukünftigen Modul der Qualitätsförderung weiterhin nützlich sein.
- Eine Mehrjahresbetrachtung, auch unter Berücksichtigung der Ergebnisse von Stellungnahmen aus dem Vorjahr wird begrüßt, denn anhaltende Auffälligkeiten sollten anders beurteilt werden als erstmalige Auffälligkeiten.

### *Kapitel 6.6 Zusammensetzung von Fachkommissionen*

- Aufgrund der hohen Bedeutung der Fachkommissionen im Rahmen des Strukturierten Dialogs sowie im Sinne der erforderlichen bundesweiten Homogenisierung des Vorgehens und der Entscheidungsfindung gehen die Vorschläge des IQTIG zur Zusammensetzung der Fachkommissionen in die richtige Richtung. Fachkommissionen müssen über Expertise jeweils aus dem ambulanten und stationären Bereich verfügen, entsprechend der jeweiligen sektorspezifischen oder sektorenübergreifenden Ausrichtung der Verfahren.
- Hinsichtlich der Zusammensetzung der Fachkommission fehlen Vorgaben zur Transparenz von Interessenkonflikten und einer Geschäftsordnung. So muss z.B. vermieden werden, dass Fachkommissionsmitglieder konkurrierende Einrichtungen bewerten.
- Die Funktionsweise und die Rolle der Fachkommissionen muss durch entsprechende Regelungen für Fachkunde, Anzahl, Auswahl, Besetzung und Besetzungsdauer, Stimmrechten, Befangenheiten und Kosten klar und eindeutig bestimmt sein und richtet sich nach den diesbezüglichen Vorgaben in der DeQS-RL, über die derzeit im G-BA beraten wird.
- Die Regelungen müssen auch Vorgaben zu Tagungsfrequenzen, Protokollführung und Vertraulichkeit enthalten, die bundesweit einheitlich sind.
- Die Fachkommissionen sollten insbesondere folgende Kernfunktionen erfüllen:
  - Fachliche Bewertung der von der Bundesauswertungsstelle übermittelten Auswertungen
  - Feststellung der Notwendigkeit eines Stellungnahmeverfahrens
  - Bewertung der Ergebnisse des Stellungnahmeverfahrens

- Empfehlung über den Abschluss des Stellungnahmeverfahrens oder die Einleitung von weiterführenden Maßnahmen sowie den Zeitrahmen innerhalb dessen die Maßnahmen umgesetzt werden sollen
- Empfehlung eines geeigneten Verfahrens zur Überprüfung des Erfolgs der weiterführenden Maßnahmen gegenüber der LAG bzw. der Bundesstelle

## 7. Modul Qualitätsförderung

Folgende Aspekte sind anzumerken:

### *Kapitel 7.1 Empfehlungen zur Einleitung qualitätsfördernder Maßnahmen*

- Als ein wichtiger Beitrag zur Qualitätsförderung sollte die Bewertung (insbesondere eine negative Bewertung) mit einer ausführlichen und strukturierten Begründung mit Nennung möglicher Verbesserungsmaßnahmen abschließen.
- Der fachlich-kollegiale Dialog muss – sofern er nicht zur Entscheidungsfindung für Zusatzinformationen erforderlich ist (vgl. Anmerkungen zu Kap. 6) – nunmehr in der Qualitätsförderung verortet werden, damit die Qualitätssicherung im Krankenhaus durch die Fokussierung auf schriftliche Stellungnahmen keinen Rückschritt erleidet. Hinsichtlich der Etablierung von Qualitätsfördermaßnahmen wäre z.B. eine Verbindung zum Peer-Review-Verfahren der Ärztekammern oder Initiativen der Fachgesellschaften denkbar.

## 8. Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens

Folgende Aspekte sind anzumerken:

- Die Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens sind nachvollziehbar dargestellt. Allerdings muss eine verkürzte Stellungnahmefrist zukünftig auch für Module funktionieren, die Sozialdaten (z.B. wegen nachfolgender Komplikationen) berücksichtigen sollen. Dieser Aspekt fehlt in der Gesamtbetrachtung.
- Die vierwöchige Frist zur Abgabe der Stellungnahme dürfte erst ab Eingang der Aufforderung beginnen und müsste eine Nachlaufzeit bei Nachfragen der LAG haben. Vor diesem Hintergrund erscheint eine sechswöchige Abgabefrist realistischer.
- Die neuen verkürzten Fristen setzen zudem voraus, dass das Verfahren sehr routiniert abläuft. Gerade zu Beginn sollte von einer Verkürzung der Fristen zunächst abgesehen werden, um für die Umstellung auf das neue Verfahren und die neue Methodik, die zwangsläufig zu erhöhtem Informationsbedarf und Nachfragen auf allen Ebenen (zwischen Krankenhäusern, Fachkommissionen, zuständigen Stellen und IQTIG) führen wird, genügend Zeit zur Verfügung zu stellen.

- Wie bereits oben erwähnt, ist in die Fristen auch die Zusatzinformation der Fachkommissionen aus Begehungen und/oder Gesprächen einzuplanen, sofern aus ihrer Sicht die Information der Stellungnahme nicht ausreicht.
- Die Berücksichtigung der Sommerurlaubszeit bei der Erarbeitung der Stellungnahmen ist ein wichtiger Punkt. Dieser Aspekt ist ebenso bei den Sitzungen der Fachkommissionen zu berücksichtigen. Das Stellungnahmeverfahren im Juli starten zu lassen führt dazu, dass es sich in die Urlaubszeit verschiebt. Ggf. könnte die Aufforderung zur Stellungnahme beim Leistungserbringer am 15. Juni eingehen.
- Eine Verkürzung der Datenannahmefrist auf den 31. Januar ist für die Krankenhäuser schwierig umzusetzen und wäre auch ein logischer Bruch, da aktuell immer am 15. des zweiten Monats im Quartal für entlassene Fälle des Folgequartals geliefert wird. Folglich sollte als Datenannahmefrist der 15. Februar gelten.

## 9. Berichterstattung

Folgende Aspekte sind anzumerken:

- Die Erläuterungen zur zukünftigen Berichterstattung sind unter Verweis auf das Auslaufen der QSKH-RL und die zukünftige Umsetzung im Rahmen der DeQS-RL nachvollziehbar und schlüssig dargelegt.
- Insgesamt reicht die beschriebene Berichterstattung und Abfrage der Qualitätsfördermaßnahmen für eine Analyse und ein Konzept im Sinne eines „Lernen von den Besten“ nicht aus.
- Der Vorschlag des IQTIG, zu einem späteren Zeitpunkt ein Konzept zu entwickeln, wie die Analyseergebnisse und Verbesserungsvorschläge aus der Qualitätsförderung allen beteiligten Einrichtungen zur Verfügung gestellt und z.B. für Fortbildungszwecke genutzt werden können, ist sehr zu begrüßen.

## 10. Empfehlungen zur Evaluation des neuen Konzepts

Folgende Aspekte sind anzumerken:

### *Kapitel 10.2 Datenquellen und Zeitschiene*

- Die vorgeschlagene Befragung zur Evaluation des Konzeptes bei den Leistungserbringern und der mit der Durchführung beauftragten Stellen wird begrüßt.
- Es sollte auch eine Befragung der Fachkommissionen zur Nachvollziehbarkeit und Anwendbarkeit der neuen Methodik und zum Aufwand und Nutzen erfolgen.

## Kapitel 10.4 Zusammenfassung der Empfehlungen zur Evaluation des vorliegenden Konzepts

- Die Evaluationsvorschläge sind umfangreich und ausführlich beschrieben.
- Die vom IQTIG vorgeschlagene sowohl prozessbegleitende als auch summative Evaluation (retrospektive Kohortenstudie) wird begrüßt. Sie sollte auch die Themen „Aufwand und Nutzen“ und „Zeiträumen“ beinhalten.
- Auch der Vorschlag, für die Evaluation zunächst ein detailliertes Modell zur Erfassung aller Auswirkungen (Wirkmodell) zu erarbeiten, sollte umgesetzt werden. Hierbei sind auch die Bewertungsalgorithmen (im Hinblick auf Verständlichkeit und Nachvollziehbarkeit für alle Beteiligten) zu prüfen.
- Die prozessbegleitende Evaluation muss zwingend auch die Eignung der ausgewählten Qualitätsindikatoren für das neue Verfahren mit einschließen, da die Güte der Qualitätsindikatoren in erheblichem Maß die möglichen Fragestellungen zur Heterogenität a) hinsichtlich der Aufnahme des Stellungnahmeverfahrens und b) der Bewertung als qualitativ auffällig nach Durchführung des Stellungnahmeverfahrens beeinflussen werden.
- Die Evaluation muss auch mögliche strukturelle Unterschiede der mit der Durchführung beauftragten Stellen einschließen, die ebenfalls eine Ursache für heterogenes Vorgehen sein können. Hierzu gehören die Anzahl und die Zusammensetzung der Fachkommission und die personelle und technische Ausstattung der Geschäftsstellen.
- Die Fachkommissionen haben zwar weiterhin eine wichtige Funktion im Stellungnahmeverfahren bezüglich der Bewertung der Einflussfaktoren, allerdings ist deren Beurteilungsexpertise durch die formalen Vorgaben vergleichsweise wenig gefragt. Dies dürfte die Gewinnung von Experten für das Verfahren zukünftig erschweren. Auch dieser Aspekt sollte evaluiert werden.
- Die hier vorgeschlagene *prozessbegleitende* und *retrospektive* Evaluation reicht nicht aus. Dies wird auch anhand der in diesem Kapitel aufgezeigten Limitationen (Kap. 10.3.1) deutlich. Die Evaluation sollte um eine Pilotierung vor der Umsetzung mit für die Methodik geeigneten ausgewählten Qualitätsindikatoren (von hoher Güte und insbesondere guter Diskriminationsfähigkeit) zur Erprobung des Verfahrens ergänzt werden.

## 11. Zusammenfassung der Empfehlungen und Konsequenzen

### Kapitel 11.5 Zusammenfassung der zentralen Empfehlungen

- Vgl. Anmerkungen zu den jeweiligen Kapiteln

## 12. Literatur

DKG [Deutsche Krankenhausgesellschaft] (2019) Qualität und Patientensicherheit-Positionen der Deutschen Krankenhausgesellschaft. Berlin : DKG  
URL:[https://www.dkgev.de/fileadmin/default/Mediapool/2\\_Themen/2.6. Qualitaet Hygiene und Sicherheit/DKG Qualitaet-Sicherheit Positionen.pdf](https://www.dkgev.de/fileadmin/default/Mediapool/2_Themen/2.6_Qualitaet_Hygiene_und_Sicherheit/DKG_Qualitaet-Sicherheit_Positionen.pdf)  
(abgerufen am: 12.11.2019)

G-BA [Gemeinsamer Bundesausschuss] (2018) Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des IQTIG: Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern vom 18. Januar 2018. Berlin: G-BA.  
URL: [https://www.g-ba.de/downloads/39-261-3196/2018-01-18\\_IQTIG-Beauftragung\\_Weiterentwicklung-strukturierter-Dialog.pdf](https://www.g-ba.de/downloads/39-261-3196/2018-01-18_IQTIG-Beauftragung_Weiterentwicklung-strukturierter-Dialog.pdf)  
(abgerufen am: 12.11.2019)

G-BA [Gemeinsamer Bundesausschuss] (2019a) Richtlinie des Gemeinsamen Bundesausschusses gemäß § 136 Abs. 1 SGB V i. V. m. § 135a SGB V über Maßnahmen der Qualitätssicherung für nach § 108 SGB V zugelassene Krankenhäuser (Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern/QSKHRL)  
Berlin: G-BA.  
URL: [https://www.g-ba.de/downloads/62-492-1834/QSKH-RL\\_2019-03-22\\_iK-2019-06-01.pdf](https://www.g-ba.de/downloads/62-492-1834/QSKH-RL_2019-03-22_iK-2019-06-01.pdf)  
(abgerufen am: 12.11.2019)

G-BA [Gemeinsamer Bundesausschuss] (2019b) Richtlinie des Gemeinsamen Bundesausschusses zur datengestützten einrichtungsübergreifenden Qualitätssicherung (DeQS-RL) Berlin: G-BA.  
[https://www.g-ba.de/downloads/62-492-1960/DeQS-RL\\_2019-01-17\\_iK-2019-11-05\\_AT-04-11-2019-B1.pdf](https://www.g-ba.de/downloads/62-492-1960/DeQS-RL_2019-01-17_iK-2019-11-05_AT-04-11-2019-B1.pdf)  
(abgerufen am: 17.11.2019)

Gigerenzer, G (2007): Gut Feelings. The Intelligence of the Unconscious. New York: Viking. ISBN: 13 978-0-67003-863-3

IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2019b) Qualitätsreport. Berlin: IQTIG  
URL: <https://iqtig.org/gs-berichte/qualitaetsreport/>  
(abgerufen am: 12.11.2019)

IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2019a) Methodische Grundlagen Version 1.1 Stand: 15.04.2019. Berlin: IQTIG. URL: <https://iqtig.org/das-iqtig/grundlagen/methodische-grundlagen/> Berlin: IQTIG  
(abgerufen am: 15.11.2019)

QS-NRW [Qualitätssicherung Nordrhein-Westfalen] (2016) Weiterentwicklung des Strukturierten Dialogs – Bericht zum Pilotprojekt im Auftrag des Lenkungsausschusses QS NRW. Düsseldorf/Münster: QS-NRW. URL: [https://www.qs-nrw.org/app/main/doc/Projektbericht\\_Weiterentwicklung\\_strukturierter\\_Dialog.pdf](https://www.qs-nrw.org/app/main/doc/Projektbericht_Weiterentwicklung_strukturierter_Dialog.pdf)  
(abgerufen am: 12.11.2019)

**Stellungnahme des Deutschen Pflegerates e.V. (DPR) zur  
Weiterentwicklung des Strukturierten Dialogs mit  
Krankenhäusern**

Vorbericht zu Stufe 1 und Stufe 2

Erstellt im Auftrag des Gemeinsamen Bundesausschusses

Stand: 10. Oktober 2019

Der Deutsche Pflegerat (DPR) vertritt als Dachverband der bedeutendsten Berufsverbände des deutschen Pflege- und Hebammenwesens die Positionen der Pflegeorganisationen und ist primärer Ansprechpartner für die Politik. Im Nachfolgenden nimmt der DPR Stellung zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern.

**Zu Punkt 2.1 Funktionen des Strukturierten Dialogs**

S. 20 Absatz 3 **Qualitätsförderung**

*Stellungnahme*

Inhaltlich geht es um Qualitätssicherung unter der Nutzung von Instrumenten zur Qualitätsförderung, daher ist der o.g. Begriff „Qualitätsförderung“ hier verwirrend. Dies gilt auch bzgl. der Bezeichnung im Folgenden.

Dementsprechend regen wir eine Anpassung an.

**Zu Punkt 3.6 Methodische Einordnung der Qualitätsförderung**

S. 63 Absatz 2

*Stellungnahme*

Sicherung und Förderung der Qualität sind zwei verschiedene Begriffe mit unterschiedlichen Instrumenten, die hier vermischt werden.

Vor diesem Hintergrund regen wir eine Prüfung des Absatzes 2 an.

### **Zu Punkt 5.4.5.2 Probleme bei zeitlichen Veränderungen in Daten und Indikatoren**

Seite 136 Absatz 2

#### *Stellungnahme*

Hier sollte klarstellend aufgenommen werden, ob die dort ausgesprochenen „Empfehlungen“ Empfehlungen des IQTIG sind.

### **Zu Punkt 6.6.1.1 Fachlich-inhaltliche Anforderungskriterien**

S. 167/168

#### *Stellungnahme*

Bei den fachlich-inhaltlichen Anforderungen sollte Folgendes berücksichtigt werden: anders als bei Chefärzten/innen und ärztlichen Professoren/innen sind Pflegende mit einer hohen Expertise nicht selten nicht mehr in der Praxis tätig. Auf diese Expertise darf aus unserer Sicht für die professionelle Pflege nicht verzichtet werden, da viele akademisch tätige Pflegende aktuell in den Fachgruppen aktiv sind. Auch Pflegeexperten/innen aus den Kammern sollten die Möglichkeit haben Mitglieder der Fachkommission zu sein.

Entsprechend regen wir hier dringend eine Anpassung an.

Grundsätzlich wollen wir noch Folgendes anmerken:

Nutzen der vorgeschlagenen neuen Verfahren ist womöglich eine höhere Wahrscheinlichkeit für ein auffälliges Ergebnis. Dies erscheint jedoch nicht nur verbunden mit einem hohen Aufwand, sondern auch notwendigerweise verknüpft mit einer erhöhten erforderlichen statistischen Expertise in den Landesgeschäftsstellen. Wenn der Aufwand im Verfahren steigt, haben wir die Sorge, dass das Verständnis und die Nachvollziehbarkeit unangemessen erschwert werden. Aus diesem Grunde ist der hohe statistische Aufwand und der Nutzen bzw. das verstärkte „Unverständnis“ des Verfahrens abzuwägen. Aus unserer Sicht sollte das Verfahren der Sektorenübergreifenden Qualitätssicherung für alle verständlich und nachvollziehbar bleiben.

Berlin, 21. Oktober 2019

Deutscher Pflegerat e.V. – DPR  
Alt- Moabit 91  
10559 Berlin  
Tel.: + 49 30 / 398 77 303  
Fax: + 49 30 / 398 77 304  
E-Mail: [info@deutscher-pflegerat.de](mailto:info@deutscher-pflegerat.de)  
[www.deutscher-pflegerat.de](http://www.deutscher-pflegerat.de)

Deutsches Netzwerk für Versorgungsforschung

## **Stellungnahme zum Vorbericht des IQTIG zur „Weiterentwicklung des strukturierten Dialogs mit Krankenhäusern“**

Der G-BA hat das IQTIG am 18. Januar 2018 mit der Weiterentwicklung des strukturierten Dialogs mit Krankenhäusern beauftragt. Das IQTIG legt nun den Abschlussbericht mit den weiterentwickelten Empfehlungen zu beiden Stufen der Beauftragung vor. Der Vorbericht umfasst 230 Seiten plus einem Anhang von 71 Seiten.

In Anbetracht des Umfangs des Vorberichts verzichtet das DNVF auf eine zusammenfassende Darstellung aller Inhalte des Vorberichts, sondern nimmt nur zu einzelnen Punkten der Kapitel 1-11 Stellung.

### Ad 1) Einleitung

Ausgangspunkt der Beauftragung ist die Beobachtung einer hohen Heterogenität im Vorgehen der auf Landesebene bzw. auf Bundesebene zuständigen Stellen für den strukturierten Dialog sowie der große Aufwand, der für die zuständigen Stellen mit der steigenden Anzahl von QS-Verfahren und Indikatoren einhergeht. Ziel der Weiterentwicklung soll die Reduktion der Heterogenität der Vorgehensweise der beauftragten Stellen sein, die Erhöhung der Transparenz und Nachvollziehbarkeit der Entscheidungen und die Steigerung der Effizienz des strukturierten Dialogs; zusätzlich soll eine Vereinheitlichung der Bewertung von Stellungnahmen, der Anforderungen an Fachexperten und des Umgang mit Interessenkonflikten erzielt werden.

Das dargelegte Auftragsverständnis des IQTIG begrenzt die Beauftragung durch den G-BA in der Form, dass zwar ein Rahmenkonzept erstellt wird und die Qualitätsbewertung auf der Basis einzelner Indikatoren sehr ausführlich weiterentwickelt wird, jedoch kaum die weiteren Punkte bearbeitet werden, nämlich die Prüfung von Indizes als Aufgreifkriterien, die Prüfung leistungsbereichübergreifender Indikatorensets, die methodische Weiterentwicklung aller Indikatoren sowie das beauftragte Konzept zur Evaluation des Rahmenkonzepts. Dadurch, dass nicht explizit ein Konzept zur Bewertung der Daten- und Dokumentationsqualität beauftragt wurde, wird auch hierzu nicht Stellung genommen. Zu einzelnen dieser Punkte wird im Bericht erwähnt, dass Beauftragungen erwartet werden, die in Zukunft bearbeitet werden könnten. Inwieweit dieses Auftragsverständnis im Sinne des G-BA ist, ist fraglich.

### Ad 2) Analyse der Ausgangssituation

Als Funktionen des strukturierten Dialogs sieht das IQTIG A) die Bewertung der Versorgungsqualität, B) die Bewertung der Dokumentationsqualität und C) die Qualitätsförderung zum Beispiel durch Zielvereinbarungen.

Das IQTIG beschreibt eindrucksvoll die derzeitigen Unzulänglichkeiten des strukturierten Dialogs und führt empirische Belege für die Heterogenität und Ineffizienz des derzeitigen Verfahrens an. In Anbetracht der neuen Anforderungen an Qualitätsindikatoren, die zunehmend zu Steuerungszwecken genutzt werden, wird vom IQTIG eine Abkehr von der bisherigen Präferenzierung hoher Sensitivität bei niedriger Spezifität hin zu einer hohen Spezifität bei eventuell geringerer Sensitivität bevorzugt.

Die Analysen zum Zusammenhang zwischen der Fallzahl eines Standortes und rechnerischen Auffälligkeiten zeigen, dass die Wahrscheinlichkeit für eine rechnerische Auffälligkeit mit sinkender

Fallzahl steigt. Das IQTIG bemängelt, dass die statistische Unsicherheit bisher nicht adäquat berücksichtigt wird, wodurch viele Standorte mit kleineren Fallzahlen auffällig werden und die Arbeit für die beauftragten Stellen überproportional ansteigt. Unseres Erachtens kann eine Schlussfolgerung aus den Analysen jedoch nicht unterstützt werden: aus den modellierten Ergebnissen folgert das IQTIG, dass qualitativ auffällige Ergebnisse in Wahrheit Fallzahl-unabhängig sind, die Einstufung als rechnerisch auffällig aufgrund der Nichtberücksichtigung der statistischen Unsicherheit jedoch falsch wäre, weil sich anhand der empirischen Ergebnisse des strukturierten Dialoges zeigen würde, dass keine erhöhte Anzahl qualitativer Auffälligkeiten bei kleiner Fallzahl festzustellen sei. Hierbei wird nicht berücksichtigt, dass die beauftragten Stellen bzw. die jeweiligen Fachkommissionen in ihrer Beurteilung der qualitativen Auffälligkeit mit hoher Wahrscheinlichkeit die jeweilige Fallzahl der Einrichtung mit berücksichtigen werden, sodass diese Berücksichtigung dazu führt, dass keine höheren Anteile qualitativer Auffälligkeiten bei rechnerischen Auffälligkeiten in kleineren Einrichtungen empirisch festzustellen sind. Dies sollte das IQTIG nochmals auf der Basis der vorliegenden Befragungsergebnisse der beauftragten Stellen überprüfen.

### Ad 3) Methodischer Hintergrund

Die Funktion des Stellungnahmeverfahrens besteht laut IQTIG zunächst darin, für jeden Einzelfall zu prüfen, ob das Indikatorergebnis in diesem Einzelfall valide ist, d. h. ob alle nicht vom Leistungserbringer zu verantwortenden Einflussfaktoren berücksichtigt sind. Das Problem besteht darin, dass bei der Entwicklung von Indikatoren oftmals nicht alle Einflussfaktoren auf Seiten der Patienten bekannt sind, diese zum Beispiel wegen eines zu hohen Aufwands nicht erhoben werden oder eine unzureichende Objektivität und Reliabilität bestimmter Daten vorliegt. Ein weiterentwickeltes Stellungnahmeverfahren kann die Spezifität des Gesamtverfahrens durch Erhöhung der Validität steigern. Die Stellungnahmen sollen sich laut IQTIG nur auf das jeweils beobachtete Qualitätsmerkmal beziehen und nicht die Gesamtqualität der Behandlung bewerten, da diese andere Bewertungsverfahren voraussetzen. Die bisherigen Stellungnahmeverfahren seien nicht objektiv und sollen durch Standardisierung d. h. eine einheitliche Informationsgrundlage für die Bewertung der Stellungnahmen und explizite Entscheidungsregeln verbessert werden. Dabei müsste die Erhöhung der Komplexität des Verfahrens durch die Beachtung methodischer Anforderungen und Fragen der Umsetzbarkeit / des Aufwands gegeneinander abgewogen werden.

Zudem wird eine Bewertungssystematik mit klaren Festlegungen zu den Bewertungsinhalten und zum Bewertungsvorgehen angestrebt. Der Bewertungsvorgang soll abgeschlossen sein, bevor qualitätsfördernde Maßnahmen eingeleitet werden. Regeln zur Vertraulichkeit und zum Datenschutz sollten für die Qualitätsbewertung und die Maßnahmen der Qualitätsförderung jeweils einzeln verabschiedet werden. Wichtig ist die Überzeugung der Teilnehmer vom Nutzen des Verfahrens und die Kenntnis und Einhaltung der Verfahrensregeln durch alle Teilnehmer.

Die in diesem Kapitel genannten Anmerkungen sind aus Sicht des DNVF gut nachvollziehbar und methodisch nicht zu beanstanden.

### Ad 4) Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und -förderung

In diesem Kapitel werden Eckpunkte des Rahmenkonzepts aufgeführt. Nur die Abbildung ist anschaulich, die restlichen Ausführungen sind größtenteils redundant.

### Ad 5) Statistische Auswertungsmethodik

Dieses alleine 72 Seiten umfassende Kapitel erweitert die bisherigen methodischen Ansätze zur Bewertung der Indikatorergebnisse einzelner Leistungserbringer. Das Kapitel wird sicher Einzug in die nächste Überarbeitung der „Methodischen Grundlagen“ finden. Zur Qualitätsbewertung schlägt das

IQTIG nach ausführlicher mathematischer Herleitung für die Zukunft folgende Eckpunkte vor: Berücksichtigung der statistischen Unsicherheit auf der Basis bayesianischer Inferenz-Tests, Beurteilungen auf der Basis der statistischen Signifikanz und Berücksichtigung von zwei Erfassungsjahren zur Prävalenzanreicherung bei geringen Fallzahlen.

Die Überlegungen zur Abwägung von Aufwand und Nutzen der Verfahren, insbesondere zur Festlegung des richtigen Signifikanzniveaus, um eine hohe Sensitivität und zugleich Spezifität zu erreichen, jedoch nicht zu viele bzw. die richtige Anzahl an Stellungnahmeverfahren auszulösen, sind gut nachvollziehbar. Kritisch zu bewerten sind die Überlegungen zur statistischen Relevanz, die am Ende dazu führen könnten, dass Leistungserbringer mit kleinen Fallzahlen aufgrund der geringen Anzahl von schlechter Qualität betroffener Patientinnen und Patienten nicht als defizitär im Hinblick auf ihre Qualität bewertet würden.

Insgesamt zeugen die in diesem Kapitel beschriebenen Konzepte von einer methodisch aktuellen Durchdringung der Problematik der Bewertung von Indikatoregebnissen.

#### Ad 6) Fachliche Bewertung

Das IQTIG schlägt in diesem Kapitel für die einzelnen Teilschritte der fachlichen Bewertung jeweils Vorgehensweisen vor, die im Großen und Ganzen methodisch nicht zu beanstanden sind. Größtenteils geht es um eine Standardisierung der Vorgehensweise bei den zuständigen Stellen und in den Fachkommissionen; zum Teil werden neue Begrifflichkeiten vorgeschlagen, wie zum Beispiel ein neuer Indikatorwert auf der Basis einer Nachberechnung, der als „informativer Indikatorwert“ bezeichnet wird und ein neu berechneter Referenzbereich, der „bereinigte Referenzbereich“. Der Nachberechnung liegt ein Ausschluss derjenigen Fälle zu Grunde, bei denen die Fachkommissionen entschieden haben, dass diese Auffälligkeiten nicht in der Verantwortung der Leistungserbringer liegen. Für die Nachberechnung stellt das IQTIG eine Softwarelösung für die zuständigen Stellen in Aussicht, die noch zu erarbeiten wäre.

Hilfreich wäre gewesen, wenn im Bericht Angaben dazu gemacht worden wären, wie oft bisher in den Stellungnahmen Gründe angeführt werden, die zu einer Nachberechnung führen müssten und welcher Anteil nur qualitativ zu beurteilen ist. Für eine Abschätzung des zu erwartenden Aufwands wäre eine solche empirische Basis wünschenswert. Unklar ist weiterhin, warum der Vorschlag des IQTIG für die retrospektive Abschätzung der Anteile der Referenzbereiche, die dem Bereich „gerichtete Einflüsse“ bzw. dem Bereich „nicht von jedem ist optimale Qualität erreichbar“ zuzuordnen sind, in den Anhang verschoben wurde. Darüber hinaus schlägt das IQTIG Unterkategorien für die Einstufung „Qualitätsdefizit“ vor (ganz oder teilweise oder gar nicht durch Datenfehler verursacht, die vom Leistungserbringer zu verantworten sind), bei denen fraglich ist, ob diese vom Publikum verstanden werden.

Für eine fachliche Bewertung ist die Einbeziehung der Anwendersicht wichtig. Dies könnte z.B. schon bei der Zusammensetzung von Fachkommissionen hilfreich sein. Hier bieten sich ÄrztInnen, Dokumentationsverantwortliche, Qualitätsmanagementbeauftragte etc. als wertvolle Hinweisgeber an. Weiterhin ist es für den Erfolg des Verfahrens essentiell, rechtzeitig den konkreten Zeitstrahl des strukturierten Dialoges bekannt zu geben, damit sich die Kliniken darauf einrichten können. So wird seitens des IQTIG gefordert, dass alle Behandler zum strukturierten Dialog aussagefähig sind. Dies bedeutet auch, dass die Aufforderungen zur Stellungnahme seitens der Fachgruppen präzise und eindeutig formuliert sein müssen einschließlich Angaben zum Umfang der Stellungnahme. Aus der Praxis wird angemahnt, dass insbesondere bei den direkten Verfahren wiederholte Stellungnahmeverfahren zu denselben Fällen problematisch sind. Hier muss aus der Aufforderung klar hervorgehen, welche neuen Sachverhalte für die Fachgruppe relevant sind.

Ad 7) Modul Qualitätsverbesserung

Das IQTIG empfiehlt nachvollziehbar, dass ab dem zweiten Qualitätsdefizit bei einem Indikatorergebnis oder bei inhaltlich ähnlichen Indikatoren innerhalb von drei Jahren immer qualitätsfördernde Maßnahmen durch die zuständigen Stellen eingeleitet sollen; sie könnten dies auch schon bei erstmaligem Auftreten eines Qualitätsdefizits.

In diesem Kapitel werden Maßnahmen der Qualitätsförderung nur benannt, die als Stufe eins bzw. Stufe zwei kategorisiert werden, wobei die Stufe zwei aus denjenigen Maßnahmen besteht, wie der G-BA in der Qualitätsförderungs- und Durchsetzung-Richtlinie, QFD-RL, genannt hat. Eine genaue Zuordnung, welche der Maßnahmen bei welchen Kategorien von Indikatoren bzw. welchen Qualitätsdefiziten hilfreich sind, fehlt bisher. Für eine der möglichen Maßnahmen, das Peer Reviewing, wird nur angemerkt, dass dafür nur Leistungsbereiche infrage kommen, bei denen einzelnen Leistungserbringern eine zureichende Qualität bescheinigt werden kann, wofür noch zusätzliche methodische Arbeiten notwendig wären. Für die bisher auch schon genutzten Zielvereinbarungen wird eine Standardisierung vorgeschlagen, für die im Anhang nützliche Formulare dargestellt werden.

Letztlich bleibt dieses Kapitel jedoch zum großen Teil auf einem oberflächlichen Niveau, das ausbaufähig ist.

Ad 8) Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens

Das IQTIG beschränkt die Vorschläge in diesem Kapitel auf die Daten der Einrichtungs- und fallbezogenen QS-Dokumentation, der Umgang mit dem zukünftigen Sozialdaten und Patientenbefragungen wird noch nicht behandelt. Der Aufwand soll einerseits durch Volumenreduktion erzielt werden, d. h. Reduktion der Anzahl an Qualitätsindikatorergebnissen, für die ein Stellungnahmeverfahren einzuleiten ist und durch Verringerung der Anzahl Qualitätsindikatoren nach Prüfung der fachlichen Eignung. An dieser Stelle wäre es hilfreich gewesen, eine empirische Abschätzung der potenziell zu entfernenden Qualitätsindikator anzuführen. Nicht wirklich klar bleibt, ob der G-BA schon mit diesem Auftrag gemeint hatte, dass das Indikatorenset insgesamt schon hätte einer Prüfung unterzogen werden müssen. Dies wäre dann auf jeden Fall bisher noch nicht erledigt worden.

Weiterhin wird die Effizienzsteigerung durch die in den vorherigen Ausführungen benannter Standardisierung im Vorgehen erzielt. Zudem ließen sich die bisher geplanten Fristen für die einzelnen Teilschritte des Bewertungsverfahrens wesentlich verkürzen. Die hier angegebenen Zeiten erscheinen sehr ambitioniert und müssten den vom IQTIG selber genannten umfangreichen Prüfungen der Umsetzbarkeit und Konsequenzen unterzogen werden. Zudem ist unklar, inwieweit diese kürzeren Fristen bereits mit den zuständigen Stellen besprochen wurden, die hier eine Einschätzung der Umsetzbarkeit abgeben müssten.

Ad 9) Berichterstattung

Dadurch, dass alle wesentlichen Auswertungen in Zukunft durch das IQTIG erfolgen, wird in Zukunft eine einheitliche Datengrundlage gegeben sein. Die Qualitätssicherungsergebnisberichte der Länder sollen maschinenlesbar eingereicht werden. Es wird empfohlen, einen Bericht zur Versorgungsqualität und einen anderen Bericht zu den Maßnahmen der Qualitätsförderung zu verlangen. Die Spezifikation soll der neuen DeQS-RL folgen. Die konkrete Ausarbeitung will das IQTIG erst dann vornehmen, wenn das vorgelegte Rahmenkonzept zur Qualitätsbewertung verabschiedet ist. Gleiches gilt für die Berichterstattung über die Maßnahmen der Qualitätsförderung, für die

darauf verwiesen wird, dass eine Spezifikation nach QFD-RL vorliegt, die nur angepasst werden müsste. Der Auftrag des G-BA scheint insoweit nur zum Teil umgesetzt worden zu sein.

#### Ad 10) Empfehlungen zur Evaluation des neuen Konzepts

Auch in diesem Kapitel werden nur grundsätzliche erste Überlegungen angestellt. In Grundsätzlich seien sowohl formative als auch summative Evaluationen möglich und nützlich. Das IQTIG scheint nur vorläufige Angaben zu machen und die Beauftragung eines Evaluationsplans zu erwarten, statt schon jetzt ein Konzept vorzulegen, wie die Evaluation wirklich gestaltet werden soll.

#### Ad 11) Zusammenfassung der Empfehlungen und Konsequenzen

In diesem Kapitel werden zum einen nochmals das Ziel der Beauftragung benannt und die Empfehlungen der einzelnen Kapitel stichpunktartig aufgeführt, zum anderen werden aber grundsätzliche Einschränkungen erwähnt, die darauf hindeuten, dass nur Teile der Beauftragung erfüllt wurden.

Besonders zu nennen, ist, dass das IQTIG zunächst den Bedarf sieht, die methodische Güte jedes Qualitätsindikator zu prüfen, umso eine Reduktion der Anzahl Qualitätsindikatoren zu erzielen. Danach erst sollen die Erfahrungen der zuständigen Stellen mit den Indikatoren für die Weiterentwicklung genutzt werden. Zudem seien quantitative Auffälligkeiten aus Sicht des IQTIG noch nicht mit qualitativen Auffälligkeiten gleichzusetzen, da die Qualitätsindikatoren noch nicht gut genug sind, um auf den Prozess des Stellungnahmeverfahrens als Prüfschritt zu verzichten. Bei diesem Prüfschritt werden ja diejenigen Konstellationen identifiziert, die zu einer fälschlichen Auffälligkeit eines Leistungserbringers geführt haben. Die Quantifizierung des Aufwands für das Stellungnahmeverfahren (bei den beauftragten Stellen und bei den Leistungserbringern) wurde im Rahmen der Beauftragung noch nicht durchgeführt, weil dies nach Sicht des IQTIG den Rahmen der Beauftragung gesprengt hätte.

Ebenso werden zu leistungsbereichübergreifenden Qualitätsindikatorensets nur grundsätzliche Überlegungen angestellt, die später im Rahmen einer Konzeptentwicklung beschrieben werden könnten. Im Fazit empfiehlt das IQTIG auch die Erstellung eines Umsetzungskonzepts, wobei die Frage ist, ob dafür wieder eine Beauftragung erfolgen soll und welchen Zeitrahmen das IQTIG dafür ansetzt.

#### **Fazit**

Das IQTIG legt ein insgesamt durchdachtes Konzept zur Weiterentwicklung des strukturierten Dialogs vor. Insbesondere die Weiterentwicklungen im Bereich der Methodik der Qualitätsbewertung sind positiv hervorzuheben. Das Konzept liefert wichtige Hintergrundinformationen zum strukturierten Dialog, die bislang in dieser Form nicht transparent waren. Die Ansätze sind als gut zu betrachten, allerdings muss darauf geachtet werden, dass zusätzliche Formalien dem Ziel der Prozessverschlinkung nicht entgegenstehen.

Unklar ist jedoch, inwieweit das Auftragsverständnis des IQTIG dem des G-BA folgt. Viele der als Punkte der Beauftragung genannten Aufgaben sind in dem vorgelegten Konzept nur angedacht und nicht so weit ausgeführt, dass sie in naher Zukunft umgesetzt werden könnten.

---

Die DNVF-Stellungnahme wurde im Auftrag des DNVF-Vorstands von Mitgliedern der Arbeitsgruppe Qualitäts- und Patientensicherheitsforschung unter Mitarbeit von Prof. M. Geraedts, Prof. S Kuske,

Dr. K. Dödler, Dr. B. Misselwitz und Prof. M. Eberlein-Gonska angefertigt. Mit Beschluss des Vorstands vom 18.11.2019 wurde die vorliegende Stellungnahme verabschiedet.

**Offenlegung von Interessenkonflikten:**

Herr Prof. Max Geraedts ist Mitglied des Wissenschaftlichen Beirats des IQTIG. Die weiteren Autoren geben keine Interessenkonflikte an.

**Kontakt:**

Deutsches Netzwerk Versorgungsforschung (DNVF) e.V.

Prof. Dr. Max Geraedts (Sprecher der AG Qualitäts- und Patientensicherheitsforschung, Mitglied im Vorstand)

Prof. Dr. Monika Klinkhammer-Schalke (Vorsitzende)

c/o DNVF-Geschäftsstelle

Kuno-Fischer-Straße 8

14057 Berlin

E-Mail: [info@dnvf.de](mailto:info@dnvf.de)

Tel.: 030 1388 7070

Köln, den 21.11.2019

**Stellungnahme zum wissenschaftlichen Vorbericht des IQTIG „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“**

Das IQTIG ist mit Beschluss vom 18. Januar 2018 vom Gemeinsamen Bundesausschuss (G-BA) mit der Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern beauftragt worden. Der jetzt vorgelegte Bericht beinhaltet sowohl die weiterentwickelten Empfehlungen des Berichts zu Stufe 1 als auch die für Stufe 2 beauftragten Inhalte.

Im Rahmen des Auftrags sollten drei Aspekte bearbeitet werden:

1. Optimierung der Einheitlichkeit der Vorgehensweise
2. Optimierung der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung
3. Optimierung der Effizienz des Verfahrens

Der letzte Punkt beinhaltet sowohl die Optimierung der „Diskriminationsfähigkeit“ der Indikatoren unter Prüfung verschiedener statistischer Verfahren als auch die Prüfung der Möglichkeiten zur Verkürzung des Verfahrens.

Im Beschluss zur Beauftragung des IQTIG wird der Strukturierte Dialog beschrieben als Verfahren, „in dem unter Einbeziehung von Experten auffällige Ergebnisse von Einrichtungen bewertet und erforderlichenfalls geeignete Maßnahmen zur Qualitätsverbesserung und –förderung eingeleitet werden“. Im Strukturierten Dialog sind die Bewertung der Versorgungsqualität, die Bewertung der Dokumentationsqualität und die Qualitätsförderung abzugrenzen. Weiterhin sieht das IQTIG im Strukturierten Dialog auch die implizite Funktion, die Spezifität der Qualitätsmessung zu steigern.

Zur Ausgangslage konstatiert das IQTIG nachvollziehbar, dass es eine derzeit eine hohe Heterogenität in der Durchführung und im Ergebnis des Strukturierten Dialogs über alle Bundesländer hinweg besteht. Weiterhin wird die überproportionale Häufung von rechnerischen Auffälligkeiten bei kleiner Fallzahl konstatiert, die im Ergebnis oft zu unauffälligen Bewertungen führen. Hier wird ein Potenzial zur Reduzierung von Heterogenität und Aufwand gesehen.

Aussagekräftige Qualitätsbewertungen sind nach Ansicht des IQTIG die zwingende Voraussetzung für die Qualitätsförderung. Bei reduzierter Aussagekraft bestünde die Gefahr, dass darauf aufbauende Maßnahmen fehlschlügen. Dies kann sicherlich zutreffend sein, jedoch sollten Möglichkeiten einer Früherkennung und Frühförderung nicht ungenutzt bleiben, um Qualitätsdefizite nicht erst statistisch belastbar manifest werden zu lassen.

Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und –förderung sind die strikte Trennung von Bewertung und Förderung und die objektive hinreichende aussagekräftige Qualitätsbewertung mit Beweislastumkehr (Leistungserbringer muss nachweisen, dass kein Qualitätsdefizit besteht). Hierzu wird eingeräumt, dass für die Zuschreibbarkeit höhere methodische Anforderungen an die indikatorbasierte Qualitätsmessung zu stellen sind, die über die Funktion eines Indikators als Aufgreifkriterium hinausgehen. Bezüglich der Auswertungsmethodik von Indikatoren wird ein umfangreiches Konzept zur statistischen Bewertung von Indikatorergebnissen vorgestellt. Inwieweit dies in kritischen Situationen, z.B. kleine Fallzahlen, hinreichend für die erhöhten Anforderungen die Qualitätsmessung sind, bleibt offen. Hier sind umfangreichere Vergleichsmessungen aus vorliegenden Datensätzen der vergangenen Jahre sinnvoll.

Die Abgabe von Stellungnahmen durch den Leistungserbringer wird notwendig, um bei einem indikatorbasiert festgestelltem hinreichenden Qualitätsdefizit den Gegenbeweis der Unauffälligkeit anzutreten. Hierbei können formale Fehler zu Ablehnung der Stellungnahme führen. Inhaltlich können so Falsch-positive Bewertungen bestehen bleiben. Im Rahmen des Stellungnahmeverfahrens sind ggf. Nachberechnungen erforderlich, die wiederum nicht zu einer Verkürzung beitragen.

Die Qualitätsförderung setzt auf der Feststellung eines Qualitätsdefizits nachgelagert auf und stützt sich auf bekannte Vorgehensweisen und Instrumente und sieht ein Potenzial für ein „Lernen von den Besseren“. Die Einbeziehung einer Evaluation des Erfolgs der Qualitätsförderung wird nicht adressiert.

Die Empfehlungen zur Verkürzung des Stellungnahmeverfahrens konzentrieren sich auf eine Straffung des Vorgehens ohne neue inhaltliche Aspekte zu benennen, die zu einer Verkürzung beitragen könnten.

**Aus Sicht der GQMG sind nachstehende Aspekte auffällig bzw. zu adressieren:**

1. Die Güte von vielen aktuell genutzten Qualitätsindikatoren ist für die hinreichende Feststellung eines Qualitätsdefizits problematisch. Auch musste mehr auf das Versorgungsgeschehen bzw. die Fachabteilung und nicht auf einzelne Indikatoren geschaut werden.
2. Die formale Rigidität des Stellungnahmeprozesses kann sich potenziell negativ auf die Feststellung von Q-Defiziten auswirken (Falsch-Positive!)
3. Die Qualitätsförderung als zentraler Gedanke des Strukturierten Dialogs kommt zu kurz. Warum nicht in eine rollierende Betrachtung von Indikatoren, Auffälligkeiten und Verbesserungsmöglichkeiten gehen, um zeitnah zu reagieren (PDCA-Gedanke)? Die Betrachtung des Datenjahres wäre dann eigentlich nur eine Zusammenfassung der Zwischenergebnisse. Damit wäre wirklich eine Verkürzung erreichbar. Das Jahresergebnis ist Ausdruck der Summe von Qualitätsergebnis und Wirksamkeit der Qualitätsförderung.
4. Letztlich sollte eine Evaluation des Gesamtverfahrens unter dem Aspekt des Kosten-/Nutzen-Aufwands und der erreichten Qualitätsverbesserung in regelmäßigen Abständen durchgeführt werden.

Insgesamt leistet das vorgelegte Konzept einen wesentlichen Beitrag zur Verbesserung des Strukturierten Dialogs hinsichtlich der Optimierung der Einheitlichkeit der Vorgehensweise und der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung. Inwieweit die hierfür notwendige Verbesserung der Qualitätsmessung erzielt werden kann, bleibt offen. Für die Optimierung der Effizienz und die Qualitätsförderung sollte der Qualitätsförderung mehr Raum gegeben werden und durch ein beispielsweise rollierendes Bewertungsverfahren mehr Wert auf Früherkennung potenzieller Qualitätsdefizite im Verbund mit frühzeitiger Qualitätsförderung gelegt werden. Dies schließt eine jährliche robuste und aussagekräftige Qualitätsbewertung nicht aus.

Für den Vorstand

Prof. Dr. med. Ralf Waßmuth

(Mitglied des GQMG Vorstands)



**Stellungnahme  
des GKV-Spitzenverbandes  
vom 21.11.2019**

**zu der Weiterentwicklung des Strukturierten Dialogs mit  
Krankenhäusern  
Vorbericht zu Stufe 1 und 2 vom 10. Oktober 2019**

**GKV-Spitzenverband**  
Reinhardtstraße 28, 10117 Berlin  
Telefon 030 206288-0  
[www.gkv-spitzenverband.de](http://www.gkv-spitzenverband.de)



Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäu-  
sern des IQTIG vom 10. Oktober 2019  
Seite 2 von 28

## Inhaltsverzeichnis

### Inhaltsverzeichnis

<b>Hintergrund .....</b>	<b>3</b>
<b>I. Zusammenfassende Stellungnahme.....</b>	<b>4</b>
1 Weiterentwicklung der Qualitätsindikatoren .....	6
2 Umgang mit kleiner Fallzahl.....	7
3 Technische (IT-) Lösungen zur Unterstützung der Qualitätsbewertung .....	7
4 Im Vorbericht bisher nicht oder nicht vollständig dargestellt wurden folgende Anforderungen der Beauftragung: .....	9
<b>II. Stellungnahme zu einzelnen Aspekten des Vorberichts .....</b>	<b>13</b>
Kapitel 1: Einleitung.....	13
Kapitel 2: Analyse der Ausgangssituation.....	14
Kapitel 3: Methodischer Hintergrund.....	16
Kapitel 4: Eckpunkte des Rahmenkonzeptes für die Qualitätsbewertung und -förderung“ ...	17
Kapitel 6: Fachliche Bewertung .....	19
Kapitel 7: Modul Qualitätsförderung .....	25
Kapitel 8: Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens .....	26
Kapitel 9: Berichterstattung .....	27
Kapitel 10: Empfehlungen zur Evaluation .....	27
Kapitel 11: Zusammenfassung der Empfehlungen und Konsequenzen .....	27

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 3 von 28

## Hintergrund

Der „Strukturierte Dialog“ in Krankenhäusern ist neben der Datenvalidierung eines der Kernelemente der „Externen Stationären Qualitätssicherung“ auf Basis der QSKH-RL (ESQS). Er dient der Klärung rechnerisch auffälliger Indikatorergebnisse und ist Ausgangspunkt für Maßnahmen der Qualitätsverbesserung und -förderung. Am 18.01.2018 beauftragte der Gemeinsame Bundesausschuss das IQTIG, den Strukturierten Dialog im Rahmen der externen Qualitätssicherung mit Krankenhäusern in zwei Stufen weiterzuentwickeln. Ziel der Beauftragung ist zum einen die Effizienzsteigerung des Verfahrens, im Sinne einer „Verschlankung“, zum anderen die Entwicklung eines richtlinienübergreifenden „Rahmenkonzeptes für ein Verfahren zur Feststellung von Auffälligkeiten in Krankenhäusern, zu deren Bewertung sowie zur Initiierung von Maßnahmen“.

Mit dieser Beauftragung inhaltlich unmittelbar verknüpft, ist ein weiterer Auftrag an das IQTIG zur „Entwicklung von Kriterien für die Datenbewertung und die Einleitung und Durchführung von Qualitätssicherungsmaßnahmen“ im Rahmen der Richtlinie zur sektorenübergreifenden Qualitätssicherung (Qesü-RL; vom G-BA beauftragt am 17.05.2018, Bericht des IQTIG vom 15.03.2019).

Mit Inkrafttreten der Richtlinie zur „Datengestützten einrichtungsübergreifenden Qualitätssicherung“ (DeQS-RL) am 01.01.2019 musste die Konzeptentwicklung auf die geänderten Anforderungen angepasst werden. Der G-BA beabsichtigt, alle Verfahren der ESQS sukzessive bis zum Jahr 2021 in die DeQS-RL zu überführen. Das bedeutet für den Weiterentwicklungsauftrag an das IQTIG eine Anpassung der Konzepte und Vorschläge unter Berücksichtigung des jetzigen DeQS-Kontextes bei zunehmender Komplexität der Anforderungen.

Das IQTIG stellt vor diesem Hintergrund sein Auftragsverständnis in dem Sinne dar, dass es den „Strukturierten Dialog“ der QSKH-RL und das „Stellungnahmeverfahren“ der mittlerweile außer Kraft gesetzten Qesü-RL und der die Qesü-RL ersetzenden Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung (DeQS-RL) als analoge Konzepte zur quantitativen und qualitativen Bewertung der Ergebnisse von Qualitätsindikatoren ansieht.

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 4 von 28

## I. Zusammenfassende Stellungnahme

Das IQTIG legte am 10.10.2019 den Vorbericht zu Stufe 1 und 2 „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“ vor. Der Bericht führt die grundsätzlichen Überlegungen und Analysen des Berichts zu Stufe 1 fort und leitet schwerpunktmäßig auf hohem wissenschaftlichen Niveau eine statistische Auswertungsmethodik ab, die sich nach Interpretation des GKV-Spitzenverbandes maßgeblich an der plan.QI-Richtlinie orientiert. Wissenschaftlich begründet wird die Abkehr von der rein rechnerischen Auffälligkeit zur statistisch signifikanten Abweichung vom Referenzbereich empfohlen. Die quantitative und qualitative Bewertung der Indikatorergebnisse soll stärker formalisiert und standardisiert erfolgen, um die bekannten Probleme der zwischen Bundesländern und QS-Verfahren stark unterschiedlichen Vorgehensweisen zu adressieren. Ein Zuwachs an Homogenität und nicht zuletzt auch an Effizienz soll damit in der Folge zu erzielen sein. Der Prozess der Qualitätsbewertung wird klar von den Maßnahmen der Qualitätsförderung getrennt.

Der GKV-Spitzenverband würdigt das Konzept als wissenschaftlich fundiert und begründet. Die Darstellung des statistischen Hintergrundes für die vorgeschlagenen Änderungen der quantitativen Auffälligkeitseinstufung ist ausführlich und erscheint überzeugend. Der Bericht setzt sich zudem in bislang für die datengestützte externe Qualitätssicherung nicht erreichter Systematik und Tiefe mit dem Hintergrund und der Methodik der qualitativen Bewertung von quantitativen Ergebnissen auseinander. Beide Punkte führen jedoch insgesamt dazu, dass konkrete Empfehlungen und Umsetzungshinweise relativ zu den methodischen Teilen des Berichts zu kurz kommen. Hier muss der Abschlussbericht unbedingt konkretisiert und ergänzt werden.

Den Ausführungen des IQTIG nach sollen entsprechend der neuen statistischen Methodik nicht mehr rechnerische bzw. numerische Abweichungen vom Referenzbereich, sondern statistisch signifikante Abweichungen zum Auslöser des Stellungnahmeverfahrens werden. Der Qualitätsindikator wird nicht mehr als Aufgreifkriterium für weitere Klärungen, sondern an und für sich als ausreichender Hinweis auf ein Qualitätsdefizit betrachtet, der lediglich durch die Geltendmachung „besonderer Versorgungssituationen“ durch den Leistungserbringer entkräftet werden kann. Das Verfahren ist maximal auf die Spezifität ausgerichtet, d.h. auf das Herausfiltern von falsch-positiven Ergebnissen im ersten quantitativen Schritt der Bewertung, und somit letztlich auch auf eine

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 5 von 28

gesteigerte Effizienz des Stellungnahmeverfahrens. Sofern jedoch eine solche quantitative Auffälligkeit nach der neuen Definition vorliegt, soll auch immer ein Stellungnahmeverfahren ausgelöst werden, d.h. reine Hinweisschreiben soll es dann nicht mehr geben.

Neu ist die Nachberechnung des Ergebnisses anhand der vom Krankenhaus angeführten Gründe für eine besondere Versorgungssituation (die letztlich analog zu den Ausnahmetatbeständen in der plan.QI-RL zu verstehen sind). Diese müssen außerhalb der Verantwortung des Krankenhauses liegen und können dann unter Umständen zur Korrektur des Rechenergebnisses benutzt werden.

Der Begriff „Strukturierter Dialog“ wird vom Institut verlassen, stattdessen der Begriff des „Stellungnahmeverfahrens“ aus der DeQS-RL eingeführt: Ein Paradigmenwechsel, der sich schon in den Begriffen widerspiegelt. Die Dialogebene in der Bewertung, die mit hohem Ressourcenaufwand und auch bisher durch die große Rolle, die implizite Bewertungskriterien spielen, mit Heterogenität der Bewertungsergebnisse verbunden ist, soll durch ein weitestgehend standardisiertes Verfahren ersetzt. Das dialogische Element wird eher auf den der Bewertung nachfolgenden Schritt der Festlegung von qualitätsfördernden Maßnahmen verlegt.

Beide Vorschläge – die Einführung der statistischen Signifikanz als Kriterium für Auffälligkeit und die weitgehende Standardisierung der Bewertung – sind aus Sicht des GKV-Spitzenverbandes geeignet, die Effizienz und Objektivität des Verfahrens erheblich zu steigern und sollten daher umgesetzt werden.

Die Einteilung des Konzeptes in die Module Qualitätsbewertung, Dokumentationsqualität und Qualitätsförderung ist nachvollziehbar. Die auf die Qualitätsbewertung aufbauenden Maßnahmen der Qualitätsförderung sollen im Sinne der Trennung von Qualitätsbewertung und Qualitätsförderung nicht als Teil des Stellungnahmeverfahrens verstanden werden. Zusätzlich kann diese Trennung auch dazu beitragen, dass die Qualitätsergebnisse früher verfügbar sein könnten als bisher und auch der Öffentlichkeit früher zur Verfügung stehen (Qualitätsberichte). Insbesondere für die Datenvalidierung, die im Bericht überhaupt nicht adressiert wird, müssen die Prozesse unbedingt zeitlich und inhaltlich aufeinander abgestimmt sein.

Unter den Themen der Beauftragung, die bisher nicht oder nicht ausreichend klar und präzise bearbeitet wurden, befindet sich auch die Forderung nach einem übergreifenden Rahmenkonzept für die qualitative Bewertung. Dieses wurde für den vorliegenden Bericht in Aussicht gestellt und sollte auch wesentliche Fragen vor allem zur Konsistenz zwischen den verschiedenen datengestützten Verfahren bzw. Richtlinien (plan.QI) beantworten.

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019

zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019

Seite 6 von 28

Zusammengefasst begrüßt der GKV-Spitzenverband die genannten Schwerpunkte der Vorschläge des IQTIG zur statistischen Methodik und zur Standardisierung der Bewertung. Die Umsetzung dieser Vorschläge muss vor dem Hintergrund der Anforderungen der DeQS-Richtlinie und auf Grund der Vorgaben des G-BA auch zeitnah erfolgen. Entsprechend den Formulierungen im §12 zur Themenspezifischen Bestimmung PCI heißt es unter „Bewertungen der Auffälligkeiten“: *„Der G-BA wird innerhalb von drei Jahren nach Beginn der Datenerhebung bundesweit einheitliche Kriterien für die Datenbewertung und die Einleitung und Durchführung von Maßnahmen gegenüber den Leistungserbringerinnen und Leistungserbringern gemäß Teil 1 §3 Satz 2 Nr. 9 der Richtlinie festlegen“*. Für diese Anforderungen muss das IQTIG die erforderlichen Voraussetzungen zur konkreten Umsetzung im Rahmen des Abschlussberichts zur Verfügung stellen. Der GKV-Spitzenverband erwartet daher für den Abschlussbericht eine Klärung der zum vorliegenden Vorbericht verbleibenden Fragen zur Umsetzung, wie sie untenstehend angesprochen werden, sowie darüber hinaus die abschließende Bearbeitung der Beauftragung im Hinblick auf Kriterien zur Bewertung wie vereinbart und letztmals im Unterausschuss QS am 06.11.2019 dargestellt.

Von besonderer Bedeutung ist aus Sicht des GKV-Spitzenverbandes die Klärung und Berücksichtigung folgender drei grundsätzlicher Punkte als essentielle, unabdingbare Voraussetzungen für eine praktische Umsetzung des vorgelegten Konzeptes.

## **1 Weiterentwicklung der Qualitätsindikatoren**

Ein entscheidender Faktor im neuen Prozess des Stellungnahmeverfahrens ist die Güte der Qualitätsindikatoren. Die vom IQTIG vorgeschlagene statistische Methodik setzt valide und hochwertige Qualitätsindikatoren voraus, die auch über feste Referenzbereiche verfügen sollten.

Die mehr als 200 Indikatoren der ESQS entsprechen nicht generell diesen hohen Anforderungen und haben teilweise erheblichem Anpassungs- und Überarbeitungsbedarf. Das IQTIG schlägt hier eine Überprüfung und Überarbeitung der Indikatoren auf Grundlage von Eignungskriterien gemäß den „Methodischen Grundlagen V1.1 (IQTIG 2019a)“ vor.

Der GKV-Spitzenverband erwartet vom IQTIG hierzu eine konkrete Angabe zur Zeitplanung. Die Überarbeitung der Indikatoren hat nach Verständnis des GKV-Spitzenverbandes im Rahmen der Verfahrenspflege kontinuierlich im Routinebetrieb zu erfolgen (siehe Kapitel 7.2 Methodenpapier des IQTIG).

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 7 von 28

## 2 Umgang mit kleiner Fallzahl

Die vorgeschlagene statistische Methodik bietet nach Darstellung des IQTIG eine wesentliche Unsicherheit: Die Bewertung von Leistungserbringern mit kleineren Fallzahlen könnte eine besondere Herausforderung darstellen. Das IQTIG schreibt im Vorbericht

*S. 112: Ein kritischer Punkt bei sowohl der statistisch signifikanten als auch der statistisch relevanten Auffälligkeitseinstufung ist, dass es, je nach Wahl der A-priori-Verteilung und der Tuning-Parameter der Methoden, Fallzahlen geben kann, bei denen Leistungserbringer unabhängig vom Ergebnis nicht auffällig werden können.*

*S. 141: Für belastbare Aussagen bei Leistungserbringern mit kleinen Fallzahlen ist die Datengrundlage eines Erfassungsjahres häufig nicht ausreichend. Es besteht somit die Gefahr, dass mögliche Qualitätsdefizite nicht detektiert werden können. In diesen Fällen empfiehlt das IQTIG, die Datengrundlage zu erweitern und beispielsweise zwei Erfassungsjahre gemeinsam für die Bewertung heranzuziehen. Konkret empfiehlt das IQTIG die bayesianische Version der statistisch signifikanten Auffälligkeitseinstufung basierend auf den Daten zweier Erfassungsjahre.*

Diese Ausführungen müssen nach Ansicht des GKV-Spitzenverbandes weiter präzisiert werden. Ab wann spricht das IQTIG von einer „kleinen Fallzahl“? Wie ist das konkrete Vorgehen vom IQTIG geplant? Welche konkrete Methodik sollte ab welcher Fallzahl angewandt werden?

Nach Vorstellung des GKV-Spitzenverbandes könnte es insbesondere in Leistungsbereichen mit kleinen Fallzahlen oder Grundgesamtheiten und seltenen Ereignissen Probleme bezüglich der Auswertungsmethodik geben. Der GKV-Spitzenverband bittet das IQTIG um Stellungnahme und konkrete Vorschläge zum Umgang am Beispiel eines Transplantationsverfahrens. Die neue statistische Methodik darf keinesfalls dazu führen, unzureichende Behandlungsqualität lediglich aufgrund der geringen Fallzahl nicht mehr zu erkennen.

## 3 Technische (IT-) Lösungen zur Unterstützung der Qualitätsbewertung

Die Nachberechnung des quantitativen Qualitätsindikatorergebnisses und das Führen durch die Schritte und Entscheidungen des Bewertungsprozesses sollen nach Darstellung des IQTIG automatisiert erfolgen. Das IQTIG schreibt in dem Vorbericht:

*Für die Umsetzung des vorgeschlagenen Bewertungsprozesses wird außerdem empfohlen, geeignete technische Unterstützung für die Durchführung der fachlichen Bewertung zur Verfügung zu*

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 8 von 28

*stellen. Dies könnte beispielsweise in Form eines für alle Bewertungsstellen einheitlichen Web-basierten Portals erfolgen, welches Funktionalitäten für die Berechnung bereinigter Ergebnisse bietet und die einzelnen Schritte des Bewertungsprozesses sequentiell abfragt. Eine einheitliche Softwarelösung für alle Bewertungsstellen würde zudem Heterogenität in der Bewertung aufgrund unterschiedlicher Softwarelösungen bei den Landesstellen vermeiden.*

*S 180: Eine weitere Komponente zur Förderung der Homogenität und der Effizienz des Bewertungsprozesses wäre eine IT-technische Unterstützung der hier vorliegenden Empfehlungen. Dies kann z. B. eine dem Bewertungsprozess angepasste Softwarelösung sein, welche einfache und komplexe Schritte, wie z. B. den Zugriff auf die Stellungnahmen und die Nachberechnung von Ergebnissen nach Berücksichtigung von besonderen Versorgungskonstellationen für die Anwender erleichtert.*

Der GKV-Spitzenverband begrüßt grundsätzlich die dargestellte Softwareunterstützung im Sinne der Effizienzsteigerung und Homogenisierung des Verfahrens. Jedoch erscheint insbesondere die praktische Umsetzung der „Nachberechnung“ von Ergebnissen und Referenzbereichen während des Stellungnahmeverfahrens an vielen Stellen noch unklar sowie äußerst aufwändig und komplex und lässt beim GKV-Spitzenverband viele Fragen offen. Wie z.B. wird entschieden und quantifiziert, welche der angegebenen Einflussfaktoren wie in eine Nachberechnung eingehen?

Der GKV-Spitzenverband bittet das Institut um konkrete Darstellung, wie, durch wen und in welchem Zeitraum ein solches web-basiertes Portal zur Verfügung gestellt werden kann. Darüber hinaus sollte unbedingt dargestellt werden, in welchem Verhältnis dieses Portal zu der mandantenfähigen Datenbank gemäß DeQS-RL steht.

Die Zusammenfassung der Empfehlungen in Kapitel 11 des Berichts ist bereits hilfreich in der Darstellung, sollte aus unserer Sicht jedoch noch wesentlich stringenter, konkreter und mit Blick auf die tatsächliche Umsetzung auf Landesebene und über mögliche Richtlinienänderungen abgefasst werden.

Für den GKV-Spitzenverband ist neben der theoretischen Grundlage die konkrete Umsetzung des Konzeptes von besonderer Relevanz. Es werden daher insbesondere zu den drei großen, oben genannten Themenbereichen im Abschlussbericht konkrete Antworten und Lösungsvorschläge erwartet. Dazu gehört eine realistische Einschätzung in Bezug auf Zeit- und Ressourcenbedarf für die Anpassungen der Qualitätsindikatoren wie auch der Einführung neuer technischen Lösungen. Der GKV-Spitzenverband bittet das IQTIG, im Abschlussbericht einen klaren Zeitplan aufzustellen.

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 9 von 28

Aus Sicht des GKV-Spitzenverbandes ist es zudem erforderlich zu Fragen der weiteren Präzisierungen und für die Umsetzungsvorgaben des vorgelegten Konzeptes die dafür auf Landesebene Verantwortlichen (LAG) miteinzubeziehen.

Im Rahmen der Umsetzungsplanung sollte über eine Übergangsregelung nachgedacht werden, denn die Erfahrungen mit der plan.QI-Richtlinie zeigen, dass eine gewisse Umstellungsphase notwendig sein könnte. Der GKV-Spitzenverband empfiehlt in diesem Zusammenhang auch die Durchführung einer frühen, begleitenden Evaluation der Einführungsphase. Ergänzend sollte vor Umsetzung ein Kommunikations- und Schulungskonzept erstellt werden, um alle Beteiligten in dem Veränderungsprozess mitnehmen zu können.

Der Empfehlung des IQTIG, zunächst ein Umsetzungskonzept zu beauftragen, kann der GKV-Spitzenverband jedoch nicht folgen. Das IQTIG schreibt auf Seite 222: *Aufgrund der Vielzahl der Änderungen, die die Umsetzung des empfohlenen Vorgehens mit sich bringen würde und aufgrund der Vielzahl der beteiligten Akteure, deren Handeln aufeinander abgestimmt werden müsste, empfiehlt das IQTIG die Entwicklung eines Umsetzungskonzepts. Im Rahmen eines solchen Umsetzungskonzepts sollte dann auch die Entscheidung über die konkrete Ausgestaltung des Stellungnahmeverfahrens unter Abwägung der Möglichkeiten zur Standardisierung des Verfahrens getroffen werden.* Die Ausgestaltung und Standardisierung und Klärung der Praktikabilität war ausdrücklich Inhalt der Beauftragung zur Weiterentwicklung des Strukturierten Dialogs. Ein vorangehendes Umsetzungskonzept wird daher für nicht erforderlich erachtet und würde zudem nur einen weiteren Zeitverzug bei der Umsetzung selbst bedeuten. Vielmehr müssen – entsprechend der Beauftragung – die Empfehlungen und Hinweise zur Umsetzung hinreichend konkret bereits in dem jetzt abzugebenden Abschlussbericht enthalten sein.

#### **4 Im Vorbericht bisher nicht oder nicht vollständig dargestellt wurden folgende Anforderungen der Beauftragung:**

##### **1. Optimierung der Einheitlichkeit der Vorgehensweise**

g) Erarbeitung einer Vorgehensweise, um eine möglichst unabhängig von Interessenkonflikten vorgenommene Bewertung zu erhalten, z.B. sicherstellen, dass die zu bewertenden Krankenhäuser den Bewertenden nicht bekannt sind, „Ringtausch“ der Bewertungen über die Bundesländer hinaus

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 10 von 28

## 2. Optimierung der Transparenz und Nachvollziehbarkeit der Entscheidungsfindung

b) Optimierung der Berichterstattung zum Strukturierten Dialog (z.B. über die eingeleiteten Maßnahmen bei auffälligen Ergebnissen) und zur Datenvalidierung, dabei insbesondere Prüfung der Möglichkeit zur besseren Darstellung und Quantifizierung von Ursachen qualitativer Auffälligkeiten und eingeleiteter Verbesserungsmaßnahmen

d) Entwicklung eines Konzepts zur ergänzenden Berichterstattung für die Öffentlichkeit und für teilnehmende Krankenhäuser (z.B. periodischer Newsletter) unter Nutzung von Informationen aus den Analysen auffälliger Ergebnisse, die bislang nur auffälligen Krankenhäusern zur Verfügung standen und/oder unter Nutzung von weiterführenden Auswertungen, die im Rahmen der Weiterentwicklung der Verfahren durchgeführt werden und modellhafte Erarbeitung an mindestens einem Beispiel

e) Entwicklung eines Konzepts, wie identifizierte Verbesserungspotentiale und beispielhaft positive Versorgungsmodelle breit kommuniziert und ggf. für ein Fortbildungsinstrument nutzbar gemacht werden können.

## 3. Optimierung der Effizienz des Verfahrens

e) Methodische Weiterentwicklung der Indikatoren unter Nutzung der Erfahrungen aus der Datenbank zum Strukturierten Dialog und einer Umfrage bei allen auf Landesebene beauftragten Stellen und soweit möglich auch bei den LAGen.

## 4. Entwicklung eines Rahmenkonzepts für ein Verfahren zur Feststellung von Auffälligkeiten in Krankenhäusern, zu deren Bewertung sowie zur Initiierung von Maßnahmen

Der GKV-Spitzenverband bittet das IQTIG die entsprechenden Ergänzungen im Abschlussbericht aufzuführen. Weitere unten genannten Themenbereiche sind in die Überarbeitung miteinzubeziehen.

### Methodik zur qualitativen Bewertung von quantitativen Ergebnissen

Die differenzierte Abwägung von Vor- und Nachteilen einer impliziten und einer expliziten Bewertung in Kapitel 3.5 begründet die vom IQTIG letztlich abgeleiteten Schlussfolgerungen und Empfehlungen. Die Darstellung macht aber auch die Limitationen jedes Verfahrens von qualitativen Urteilen durch Experten deutlich und stellt heraus, dass zwar eine möglichst hohe Objektivität angestrebt wird, jedoch aufgrund der Komplexität der zu bewertenden Prozesse letztlich immer implizite Elemente in der Urteilsfindung in Kauf genommen werden müssen.

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 11 von 28

Eine wesentliche Limitation ist darin zu sehen, dass die qualitative Bewertung ausschließlich darauf ausgerichtet ist, Qualitätsdefizite zu identifizieren bzw. verifizieren. Eine differenzierte, abgestufte Bewertung der Qualität verschiedener Einrichtungen ist somit nicht möglich. Das IQTIG stellt diese Limitation ausdrücklich dar. Es weist in diesem Zusammenhang auch darauf hin, dass die vorliegenden Empfehlungen auch keine Identifikation „zureichender“ Qualität ermöglichen und „kein Hinweis auf Qualitätsdefizit“ nicht mit „zureichender Qualität“ gleichgesetzt werden kann, begründet dies und verweist auf grundlegenden methodischen Entwicklungsbedarf, um differenziertere Bewertungen zu ermöglichen.

Das IQTIG empfiehlt *„die Methodik dahingehend weiterzuentwickeln, dass hinreichende Sicherheit über das Erreichen des Solls in einem Indikator (z. B. des Referenzbereichs für erwartbare Qualität) mit einer entsprechenden Bewertungskategorie (z. B. „zureichende Qualität“) bezeichnet werden kann.“*

Eine solche Entwicklung sollte möglichst kurzfristig erfolgen.

In diesem Zusammenhang erscheint dem GKV-Spitzenverband auch der Aspekt der bisher fehlenden Kriterien für gute oder exzellente Qualität besonders wichtig, die entsprechend den Anforderungen der DeQS-Richtlinie vorliegen sollten:

*§17 DeQS-RL (2) Das Verfahren nach Satz 1 und 2 kann darüber hinaus auch dann durchgeführt werden, wenn eine Leistungserbringerin oder ein Leistungserbringer auffällig gute Ergebnisse hat oder in Vorjahren wiederholt auffällig war.*

Der GKV-Spitzenverband bittet das IQTIG um eine konkrete Empfehlung.

### Rahmenkonzept

Ein Unterpunkt in der Beauftragung des IQTIG zur Weiterentwicklung des Strukturierten Dialogs nimmt die „Entwicklung eines Rahmenkonzepts für ein Verfahren zur Feststellung von Auffälligkeiten in Krankenhäusern, zu deren Bewertung sowie zur Initiierung von Maßnahmen“ ein.

Dieser Auftrag ist nach Meinung des GKV-Spitzenverbandes noch nicht abschließend erfüllt, insbesondere in Bezug auf die Integration der Bewertung, wie sie in anderen Richtlinien des G-BA zur datengestützten Qualitätssicherung vorgenommen wird. Im Abschlussbericht sollte im Sinne eines Rahmenkonzepts das hier vorgeschlagene Konzept mit dem Verfahren zu planungsrelevanten Indikatoren im Einzelnen in Beziehung gesetzt werden. Es sollte dargestellt werden, ob das Verfahren zu planungsrelevanten Indikatoren in der bisher angewendeten Form beibehalten oder angepasst werden soll, denn die vorgelegten Empfehlungen weisen viele Ähnlichkeiten aber auch Unterschiede auf und zu einigen Aspekten bleibt unklar, ob Abweichungen vorliegen. Zu folgenden Aspekten erscheint nicht klar, ob Unterschiede zum Verfahren der Plan-QI-RL bestehen:

- Ist die Bewertung „unzureichend“ (Plan-QI) gleichzusetzen mit „Qualitätsdefizit“?

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 12 von 28

- Sind „Ausnahmetatbestände“ (Plan-QI) gleichzusetzen mit den „Einflussfaktoren“ in den vorliegenden Empfehlungen?
- Ist die Feststellung von statistischen Abweichungen im einen Verfahren (Plan-QI) nach „frequentistischer“ und im anderen Verfahren nach „bayesianischer“ Methodik erforderlich?
- Können die Fristen innerhalb der Verfahren angeglichen werden?

Zu den offensichtlichen Unterschieden sollte das IQTIG darstellen, ob Unterschiede unvermeidlich sind und dies ggf. begründen. Es sollte ein Vorschlag gemacht werden, wie ggf. eine Vereinheitlichung möglich gemacht werden kann.

Es bleibt unklar, wie eine Bewertung als „unzureichend“ (bzw. „in erheblichem Maße unzureichend“) im Verfahren zu planungsrelevanten Indikatoren im Verhältnis zum „Qualitätsdefizit“ in den vorliegenden Empfehlungen einzuordnen ist. Die Darstellung macht ebenfalls nicht deutlich, wie andere „Kategorisierungen“ von Qualität (z.B. „Exzellenzqualität“ im Sinne der Regelungen zu qualitätsbasierten Zu- und Abschlägen) in dieses „Rahmenkonzept“ integriert werden können. Für den GKV-Spitzenverband ist es auch nicht nachvollziehbar, dass das IQTIG ausdrücklich die Bewertung der Daten- und Dokumentationsqualität aus der Konzeptentwicklung ausschließt. In der QSKH-RL ist eine solche Bewertung Bestandteil des Strukturierten Dialogs und sollte somit auch Bestandteil einer Weiterentwicklung des Strukturierten Dialogs sein. Der GKV-Spitzenverband bittet das IQTIG hierzu um Stellungnahme.

#### Abschließende Bewertung

Problematisch erscheint dem GKV-Spitzenverband die Darstellung des IQTIG zur „abschließenden Bewertung“ durch Institutionen wie beispielsweise das IQTIG (Seite 150).

Auf Seite 150 stellt das IQTIG dar, dass eine „abschließende Bewertung“ nicht durch die Fachkommission erfolgen soll:

*„Die zuständige Stelle trifft die abschließende Entscheidung in Bezug auf die Einordnung des Indikatorergebnisses in das Bewertungsschema.“*

Auf Seite 175 wird konkretisiert, was das IQTIG unter einer zuständigen Stelle versteht: „zuständige Stelle (LQS, LAG, IQTIG)“. Vor diesem Hintergrund sollten nach Ansicht des GKV-Spitzenverbandes unbedingt die Legitimation und der Prozess der „abschließenden Bewertung“ noch klarer erläutert und begründet werden. Gerade vor dem Hintergrund der „wissenschaftlichen Unabhängigkeit“ des IQTIG ist nachvollziehbar, dass das IQTIG in dieser Bewertung eine wichtige Rolle zukommt. Aus der Formulierung im Abschlussbericht geht jedoch nicht hervor, welche Personen (mit welcher Qualifikation) nach welcher Systematik (und welchen Kriterien) die systematisch ermittelte Bewertung der Fachkommissionen „überstimmen“ können. Müssen die Mitarbeiter des

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 13 von 28

IQTIG, die diese „abschließende Bewertung“ vornehmen, nicht ebenfalls die Anforderungen an die fachliche Qualifikation erfüllen wie die Kommissionsmitglieder?

Auch im Methodenpapier des IQTIG findet sich hierzu bislang keine entsprechende Beschreibung des Vorgehens zur Bewertung von Indikatorenergebnissen.

#### Umgang mit Dokumentationsfehlern

Ebenfalls problematisch erscheint die Darstellung zum Umgang mit Dokumentationsfehlern (Seite 150 ff.). Es wird dargestellt, dass Datenfehler berücksichtigt werden sollen, es wird jedoch nicht konkret beschrieben, wie diese identifiziert werden. Sollen von Krankenhäusern „behauptete“ (nicht extern validierte) Dokumentationsfehler akzeptiert werden? Wie lässt sich dieser Rückschritt gegenüber dem Verfahren der planungsrelevanten Indikatoren begründen, bei dem das Stellungnahmeverfahren auf der Grundlage einer verlässlichen, validierten Datengrundlage geführt wird?

## **II. Stellungnahme zu einzelnen Aspekten des Vorberichts**

### **Kapitel 1: Einleitung**

Das IQTIG formuliert in diesem Abschnitt des Vorberichtes sein Auftragsverständnis. Das Kapitel beinhaltet konkrete Empfehlungen für die Auswertung und Bewertung von leistungserbringerbezogenen Qualitätsindikatorergebnissen auf Basis von Dokumentationsdaten der Leistungserbringer, auf Basis von Sozialdaten sowie auf Basis durch Patientenbefragungen erhobener Daten.

Das Institut setzt die Weiterentwicklung in entscheidenden Punkten vom bisherigen Konzept zum Strukturierten Dialog im Rahmen der QSKH-Richtlinie ab. Dieser versteht sich als Bewertung auffälliger Ergebnisse und Einleitung von qualitätsfördernden Maßnahmen unter Einbezug von Experten auf Bundes und Landesebene. Bisher ist es möglich, mit rechnerisch auffälligen Ereignissen „individuell“, d. h. nicht standardisiert und daher auch mit großer Heterogenität umzugehen: Der Versand von Hinweisen, Stellungnahmen, Begehungen oder Zielvereinbarung sind in diesem Prozess möglich. Das IQTIG ersetzt den Begriff des „Strukturierten Dialogs“ durch „Stellungnahmeverfahren“, damit wird der gesamte Prozess eingegrenzt auf die Bewertung von Ergebnissen.

### **Kapitel 1.3 „Auftragsverständnis“**

Nicht nachvollziehbar erscheint, dass das IQTIG ausdrücklich die Bewertung der Daten- und Dokumentationsqualität aus der Konzeptentwicklung ausschließt (Seite 17).

In der QSKH-RL ist eine solche Bewertung durch den Verweis in §9 (7) selbstverständlicher Bestandteil des Strukturierten Dialogs und sollte somit auch Bestandteil einer Weiterentwicklung des Strukturierten Dialogs sein.

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 14 von 28

#### **Kapitel 1.4 „Einbeziehung der auf Landesebene beauftragten Stellen“**

Das IQTIG stellt auf Seite 18 ff. dar, dass dieser Auftragsbestandteil durch einen Workshop umgesetzt wurde. Damit erscheint dieser Auftragsbestandteil formal erfüllt. Allerdings hätte eine intensivere und ggf. kontinuierliche Nutzung der praktischen Erfahrung dieser Institutionen möglicherweise vorteilhaft sein können.

#### **Kapitel 2: Analyse der Ausgangssituation**

Das Kapitel stellt die Funktionen des Strukturierten Dialogs und seine Durchführung übersichtlich und sachgerecht dar. Es liefert eine biometrische Analyse des Ist-Zustandes der Heterogenität und zu Aufwand und Effizienz des Verfahrens. Diese Ergebnisse sind bereits bekannt, aber bisher noch nicht in dieser Tiefe und Ausführlichkeit dargestellt worden. Sie liefern die solide Grundlage für die Ansätze im Weiterentwicklungskonzept.

#### **Kapitel 2.1 „Funktionen des Strukturierten Dialogs“**

Problematisch erscheint eine Darstellung auf Seite 21 zur Sensitivität und Spezifität. Dargestellt wird in Tabelle 1 lediglich der „zweite Prüfschritt“ (qualitative Bewertung von Indikatorenergebnissen durch Experten), nicht aber der „erste Prüfschritt“ (entspricht der Feststellung einer Abweichung vom Referenzbereich). Die Sensitivität des Bewertungsprozesses wird aber weitestgehend durch den „ersten Prüfschritt“ bestimmt. Nachvollziehbar ist die Darstellung in Bezug auf die Spezifität. Deutlicher herausgestellt werden sollte allerdings nach Auffassung des GKV-Spitzenverbandes, dass die Sensitivität methodenimmanent nicht bewertet werden kann. Dies wird zwar auf Seite 21 angesprochen (*„Ein auffälliges Indikatorergebnis (Ergebnis erster Prüfschritt) kann im Rahmen des Strukturierten Dialogs (zweiter Prüfschritt) zwar widerlegt werden, bei einem unauffälligen Indikatorergebnis kann dieses jedoch nicht widerlegt werden, da nur auffällige Ergebnisse einem zweiten Prüfschritt unterzogen werden“*), es wird jedoch nicht ausreichend klar gemacht, dass die Sensitivität maximal vom „ersten Prüfschritt“ abhängt und daher bei Verwendung beispielsweise von 5%-Perzentilen als Referenzbereichen eine äußerst niedrige Sensitivität anzunehmen ist.

#### **Kapitel 2.4 „Durchführung des Strukturierten Dialogs“**

Es erfolgt eine sehr differenzierte und aussagekräftige Beschreibung des Ist-Zustands der Durchführung des Strukturierten Dialogs und die sachgerechte Ableitung von wesentlichen Schlussfolgerungen und Handlungsbedarfen.

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 15 von 28

### **Kapitel 2.5 Biometrische Analyse der Heterogenität**

Das IQTIG stellt hier eine eindrucksvolle Analyse zur Validität der Bewertungen im Strukturierten Dialog vor. Diese Analyse zeigt an beispielhaft ausgewählten Indikatoren einerseits die recht erhebliche Heterogenität zwischen den Bundesländern in Bezug auf die Einholung von Stellungnahmen und die qualitative Bewertung. Abbildung 2 (Seite 35) zeigt das recht bemerkenswerte Ergebnis, dass bei Einrichtungen mit starken Abweichungen vom Referenzbereich (4. Quartil) von fast allen Ländern eine Stellungnahme angefordert wird, bei geringeren Abweichungen allerdings die Anforderung durchgehend heterogen erfolgt. Die Abbildung zeigt weiterhin, dass die Wahrscheinlichkeit einer Einstufung als „qualitativ auffällig“ scheinbar praktisch linear mit dem Ausmaß der Abweichung vom Referenzbereich steigt. Insgesamt zeigt die Analyse eine erhebliche Heterogenität zwischen den Ländern, dies betrifft vor allem den Umgang mit Einrichtungen, die nur relativ moderate Abweichungen vom Referenzbereich zeigen. Mit dieser Analyse begründet das IQTIG wissenschaftlich fundiert die hohe Bedeutung eines möglichst einheitlichen Vorgehens in Bezug auf die Anforderung von Stellungnahmen und die folgende qualitative Bewertung.

### **Kapitel 2.6 Biometrische Analyse zu Aufwand und Effizienz des Strukturierten Dialogs**

Die Analyse liefert Hinweise auf die Effizienz des Strukturierten Dialogs durch eine Bewertung der Spezifität des „Prüfschritts 2“ (qualitative Bewertung). Diese Analyse kann aus mehreren Gründen allerdings nur als grobe Annäherung angesehen werden. Zum einen zeigt gerade die Darstellung der Heterogenität in Kapitel 2.5, dass der „Endpunkt“ einer qualitativ auffälligen Bewertung nicht als valide angesehen werden kann. Darüber hinaus kann die Sensitivität der Feststellung eines Qualitätsmangels aufgrund der sehr niedrigen Sensitivität des „Prüfschritts 1“ (Referenzbereich) praktisch nicht bewertet werden. Aus dieser Analyse leitet das IQTIG vor allem drei Handlungsfelder ab (Seite 52):

1. Es soll geprüft werden, bei welchen Indikatoren Weiterentwicklungsbedarf besteht.
2. Es soll durch Berücksichtigung statistischer Signifikanz der Einfluss zufälliger Faktoren auf das Indikatorergebnis berücksichtigt werden.
3. Es wird thematisiert (Analyse hierzu auf Seite 40), dass die aktuelle Berechnung der perzentilbasierten Referenzbereiche (Ermittlung des Perzentils auf Grundlage der Einrichtungen mit mehr als 20 Fällen) dazu führt, dass teilweise deutlich mehr Einrichtungen auffällig werden, als das Perzentil suggeriert.

Diese Handlungsfelder sind nicht überraschend und waren im Wesentlichen bereits im Auftrag an das IQTIG konkret benannt. Dennoch ist als positiv zu bewerten, dass die empirische Analyse des

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 16 von 28

IQTIG nun nochmals bestätigt, dass diese Handlungsfelder tatsächlich von praktischer Relevanz sind.

### **Kapitel 3: Methodischer Hintergrund**

#### **Kapitel 3.2 „Reichweite der Qualitätsaussage von Stellungnahmen“**

Das IQTIG formuliert auf Seite 56:

*„Die fachliche Bewertung im Stellungnahmeverfahren soll sich explizit auf das vom Indikator adressierte Qualitätsmerkmal beziehen und nicht eine über dieses Qualitätsmerkmal hinausgehende Aussage anstreben“.*

Dies ist einerseits nachvollziehbar. Es entspricht dem im Bericht angesprochenen Verständnis, Indikatoren nicht als „Aufgreifkriterien“ zu verstehen, sondern als quantitative Maße. Es soll qualitativ nur bewertet werden, ob die zur Auffälligkeit führenden Fälle als Mangel angesehen werden können. Dies führt allerdings zu einer relevanten Limitation: es kann nicht bewertet werden, ob Strukturen und Prozesse einer Einrichtung (die beispielsweise für die Vermeidung von Komplikationen, die in einem Indikator erfasst werden relevant sind – aber eben auch nicht nur hierfür) Qualitätsmängel aufweisen. Dies hat die Konsequenz, dass für das Modul „Qualitätsförderung“ praktisch keine Informationsgrundlage geschaffen werden kann. Dies führt auch dazu, dass über Ursachen von Qualitätsdefiziten praktisch kaum Aussagen gemacht werden können – und damit ein „Lernen von den Fehlern Anderer“ nur sehr limitiert möglich ist.

#### **Kapitel 3.3 „Gütekriterien für Bewertungsprozesse“**

Im Kapitel 3.3 erfolgt eine sehr gute Darstellung zum wissenschaftlichen Hintergrund des Konzepts der qualitativen Bewertung von Indikatorenergebnissen durch Experten. Es wird eine Analogie zum diagnostischen Prozess der Urteilsbildung hergestellt und nachvollziehbar begründet. Diese Darstellung baut auf Kapitel 2.4.1 im Bericht des IQTIG zu Stufe 1 vom 30.11.2018 auf. Diese Darstellung kann letztlich als theoretischer Unterbau für die seit Jahren praktizierten Verfahren im Strukturierten Dialog angesehen werden und begründet darüber hinaus die Empfehlungen zur Weiterentwicklung. Insbesondere wird wissenschaftlich begründet, warum die Verwendung expliziter Bewertungskriterien einem auf impliziten Regeln fundierenden Urteilsprozess unterlegen ist.

#### **Kapitel 3.5 „Vor- und Nachteile explizierter Beurteilungsregeln im Stellungnahmeverfahren“**

Die differenzierte Abwägung von Vor- und Nachteilen einer impliziten und einer expliziten Bewertung in Kapitel 3.5 begründet die vom IQTIG letztlich abgeleiteten Schlussfolgerungen und Empfehlungen. Die Abwägung macht auch die Limitationen jeglichen Verfahrens von qualitativen

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 17 von 28

Urteilen durch Experten deutlich und stellt heraus, dass zwar eine möglichst hohe Objektivität angestrebt wird, jedoch aufgrund der Komplexität der zu bewertenden Prozesse letztlich immer implizite Elemente in der Urteilsfindung in Kauf genommen werden müssen. Unklar erscheint die Formulierung auf Seite 60:

*„Durch den Vergleich eines bereinigten Indikatorwerts mit einem (angepassten) Referenzbereich würde berücksichtigt, dass der Referenzbereich eines Indikators in der externen Qualitätssicherung durch Beschluss des G-BA normative Wirkung hat“.*

Was ist in diesem Zusammenhang mit einem „angepassten Referenzbereich“ in Bezug auf die „normative Wirkung“ gemeint?

Dies sollte im Abschlussbericht erläutert werden.

### **Kapitel 3.6 „Methodische Einordnung der Qualitätsförderung“**

Kapitel 3.6 stellt eine wissenschaftliche und theoretische Einordnung des Elements der Qualitätsförderung im Strukturierten Dialog dar.

Es wird klar dargestellt, dass eine direkte wissenschaftliche Bewertung aufgrund fehlender Datenbasis nicht möglich ist. Es wird daher eine Analogie zu Peer-Review-Verfahren hergestellt, die allerdings als nur eingeschränkt angemessen anzusehen ist, da die Freiwilligkeit eines Peer-Review in starkem Kontrast zur gesetzlich vorgegebenen Verpflichtung zur Teilnahme im Rahmen des Strukturierten Dialogs steht.

### **Kapitel 4: Eckpunkte des Rahmenkonzeptes für die Qualitätsbewertung und -förderung“**

Die Darstellung in Kapitel 4 macht noch einmal deutlich, dass das gesamte Verfahren ausschließlich darauf ausgelegt ist, Qualitätsdefizite zu identifizieren und Qualitätsbewertung von Förderung zu trennen. Es legt ausführlich die Bewertung der Auswertungsergebnisse in zwei Schritten dar. Die Darstellung erscheint weniger als Rahmenkonzept für ein richtlinienübergreifendes Vorgehen zur Bewertung verschiedener Ausprägungen von „Qualität“, sondern als Beschreibung eines Algorithmus zur Feststellung eines „Qualitätsdefizits“.

Im Auftrag vom 18.1.2018 war die Entwicklung eines Rahmenkonzeptes gefordert:

*„Vor dem Hintergrund der zunehmenden Anforderungen an das Verfahren der externen stationären Qualitätssicherung und damit des Strukturierten Dialogs (z.B. unter planungsrelevanten oder vergütungsabhängigen Aspekten) wird nunmehr das IQTIG mit der Weiterentwicklung beauftragt. Dabei müssen bei allen Schritten der Beauftragung ein zu entwickelndes Rahmenkonzept und eine Abstimmung der verschiedenen Richtlinien mitberücksichtigt werden“.*

Nach Ansicht des GKV-Spitzenverbandes liefert das Verständnis des IQTIG in Bezug auf ein Rahmenkonzept nicht die im Auftrag geforderte „Abstimmung“ mit anderen Richtlinien, in denen qualitative Bewertungen von Auswertungsergebnissen vorgenommen werden. Insbesondere bleibt

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 18 von 28

unklar, in welchem Verhältnis die in den vorliegenden Empfehlungen des IQTIG resultierenden Bewertungskategorien „Qualitätsdefizit“ und „kein Hinweis auf Qualitätsdefizit“ zu den Bewertungen beispielsweise in der „Plan-QI“-RL und zu Anforderungen, abgestufte Bewertungen von Qualität vorzunehmen, stehen.

Unter dem Aspekt der Effizienz des Verfahrens erscheint der Vorschlag des IQTIG nachvollziehbar, dass eine Stellungnahme des Leistungserbringers freiwillig ist, die Nicht-Abgabe einer Stellungnahme dann jedoch als Bestätigung eines Qualitätsdefizits zu werten ist (Seite 69).

### **Kapitel 5: Statistische Auswertungsmethodik**

Dieses Kapitel erscheint auch auf Grund seiner umfangreichen und umfassenden Darstellung als Kernstück des vorliegenden Konzeptes: In Fortführung der Überlegungen des Berichts der Stufe 1 werden Modelle der statistischen Auswertungsmethodik aufgeführt und in Tiefe diskutiert.

Zu einzelnen Themen:

#### „Analytische“ oder „enumerative“ Herangehensweise

Auf Seite 137 weist das IQTIG auf folgenden Sachverhalt hin:

*„Ein wichtiger Bestandteil des Rahmenkonzepts ist, dass für jeden Indikator bei der Entwicklung entschieden wird, ob der Indikator nach einer analytischen oder enumerativen Herangehensweise ausgewertet werden soll. Der Großteil der Indikatoren in der DeQS-RL ist aus methodischer Sicht des IQTIG nach der analytischen Herangehensweise auszuwerten.“*

In der Konsequenz könnte es daher für eine Festlegung in der DeQS-RL erforderlich sein, dies für jeden Indikator festzulegen. Das IQTIG müsste daher zukünftig für jeden Indikator die entsprechende Einordnung transparent machen. Dies müsste erstmalig (z.B. mit Veröffentlichung der prospektiven Rechenregeln) für das Erfassungsjahr erfolgen, in dem die entsprechenden Regelungen in der Richtlinie in Kraft treten. Der GKV-Spitzenverband bittet das IQTIG um eine Stellungnahme, zu welchem Erfassungsjahr dies erstmalig möglich sein wird und in welchem Rahmen (z.B. prospektive Rechenregeln) dies aus Sicht des IQTIG erfolgen sollte.

#### Signifikanzniveau

Auf Seite 141 stellt das IQTIG dar *„Somit ist  $\alpha$  sorgfältig und behutsam zu wählen: Möchte man eine einheitliche Grundlage für die Operationalisierung des Begriffs „hinreichender Hinweis auf ein Qualitätsdefizit“ über alle QS-Verfahren haben, dann ist ein Wert von  $\alpha$  für alle QS-Verfahren festzulegen. Möchte man dagegen berücksichtigen, dass unterschiedliche Konsequenzen, Ressourcen und Verfahrenszwecke eine Rolle spielen, kann es ggf. eine unterschiedliche Wahl von  $\alpha$  für die quantitative Auffälligkeitseinstufungsmethode geben.“*

*„Das IQTIG empfiehlt für die Wahl des Signifikanzniveaus für die statistische signifikante Auffälligkeitseinstufung einen der folgenden Werte, ...“*

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 19 von 28

Für eine Umsetzung der Empfehlungen des IQTIG in der DeQS-RL sind an dieser Stelle Konkretisierungen erforderlich.

## **Kapitel 6: Fachliche Bewertung**

Die Darstellung auf Seite 143 und das „Bewertungsschema“ auf Seite 145 machen deutlich, dass die qualitative Bewertung durch Experten im Wesentlichen nur zwei Ergebnisse produzieren kann: „Qualitätsdefizit“ oder „kein Hinweis auf Qualitätsdefizit“. Darüber hinaus wird nur noch eine Kategorie „Sonstiges“ vorgeschlagen, die (nachvollziehbarerweise) nicht im Einzelnen operationalisiert ist. Das heißt, das Konzept richtet sich ausschließlich auf Identifikation von Qualitätsdefiziten. Dies erscheint angesichts des beauftragten Rahmenkonzepts stark reduziert und nicht ausreichend.

Es wird nicht deutlich, in welchem Verhältnis diese Bewertung zur Bewertung im Verfahren der planungsrelevanten Indikatoren („unzureichende Qualität“) steht. Es berücksichtigt ebenfalls nicht, wie andere „Kategorisierungen“ von Qualität (z.B. „Exzellenzqualität“ im Sinne der Regelungen zu qualitätsbasierten Zu- und Abschlägen) identifiziert werden könnten. Vor diesem Hintergrund weist das IQTIG auch darauf hin, *dass die Feststellung „kein Hinweis auf Qualitätsdefizit“ nicht mit „zureichender Qualität“ gleichzusetzen ist.* Dieser Hinweis stellt noch einmal sehr deutlich heraus, dass mit Hilfe der Empfehlungen des IQTIG „zureichende“ Qualität nicht festgestellt werden kann. Das IQTIG erläutert: *„Daher wird empfohlen, die Methodik dahingehend weiterzuentwickeln, dass hinreichende Sicherheit über das Erreichen des Solls in einem Indikator (z. B. des Referenzbereichs für erwartbare Qualität) mit einer entsprechenden Bewertungskategorie (z. B. „zureichende Qualität“) bezeichnet werden kann.“*

Eine solche Entwicklung sollte nach Auffassung des GKV-Spitzenverbandes schnellstmöglich erfolgen.

### „Partielle Nachberechnung“

Auf Seite 152 stellt das IQTIG dar, wie die „partielle Nachberechnung“ erfolgen soll:

*„Dazu wird ein Indikatorwert berechnet, der um diejenigen nicht vom Leistungserbringer zu verantwortenden Datenfehler und fachlich-inhaltlichen Einflussfaktoren bereinigt ist, deren Auswirkung auf das Indikatorergebnis eindeutig benannt werden kann. Diese Bereinigung erfolgt z. B. dadurch, dass Behandlungsfälle, die vom Indikator nicht hätten adressiert werden sollen, aus der Grundgesamtheit des Indikators ausgeschlossen werden (Details siehe Abschnitt 6.4.1). Der so bereinigte Wert wird im Folgenden als informativer Indikatorwert bezeichnet und dient ausschließlich als Information im Rahmen der fachlichen Bewertung; er hat nicht die Funktion, den originalen Indikatorwert der Einrichtung zu ersetzen.“*

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 20 von 28

Des Weiteren wird aus dem originalen Referenzbereich des Indikators ein modifizierter Referenzbereich abgeleitet. Dieser ist – wie der informative Indikatorwert – um den Anteil bereinigt, der schon im originalen Referenzbereich für die gerichteten, nicht vom Leistungserbringer zu verantwortenden Einflussfaktoren einkalkuliert ist (siehe Abschnitt 6.4.2). Dieser Referenzbereich kann als der Referenzbereich verstanden werden, der festgelegt worden wäre, wenn die Berechnungsvorschrift des bereinigten Indikators von vornherein die gerichteten Einflüsse berücksichtigt hätte, und wird im Folgenden als bereinigter Referenzbereich bezeichnet.“

Das Vorgehen, sowohl die Ergebnisse der Einrichtung als auch der Referenzbereiche zu bereinigen, erscheint prinzipiell nachvollziehbar. Die Praktikabilität dieses Vorgehens sollte im Abschlussbericht weitergehend erläutert werden. Klargestellt werden sollte auch nach Ansicht des GKV-Spitzenverbandes, ob folgende Formulierung auf Seite 60 im Anhang so zu verstehen ist, dass die zur nachträglichen Bereinigung der Referenzbereiche beschriebene Methodik noch nicht eingesetzt werden kann, sondern dass hier noch Entwicklungsbedarf besteht:

*„Zum Zeitpunkt der Berichtslegung entwickelt das IQTIG eine Methodik, wie bei Festlegung von Referenzbereichen die Größe der beiden Anteile des Referenzbereichs angemessen und nachvollziehbar festgelegt werden kann. In diesem Rahmen soll auch untersucht werden, ob eine Schätzung der Anteile bei bestehenden Referenzbereichen mittels komplexerer statistischer Verfahren Vorteile hat.“*

Falls die Formulierung so zu verstehen ist, sollte klargestellt werden, wann eine nutzbare Methodik zur Verfügung steht.

Weiter fragt sich der GKV-Spitzenverband ob dieses Konzept so zu verstehen, dass Referenzbereiche nachträglich – während des Stellungnahmeverfahrens – für einzelne Einrichtungen neu berechnet werden? Falls ja: müssten diese Referenzbereiche nicht konsequenterweise dann für alle Einrichtungen eingesetzt werden?

Erläuterungs- und diskussionsbedürftig erscheint auch die Darstellung auf Seite 58 im Anhang zur Nachberechnung der Referenzbereiche:

*„Es wird empfohlen, die Differenz zwischen dem Idealwert des Indikators (0 % bzw. 100 %) und dem empirischen bundesweiten Wert (der Wert, der bei Berechnung des Indikators auf Grundlage aller Behandlungsfälle bundesweit resultiert) als Näherung für den Anteil gerichteter Einflussfaktoren zu verwenden (höchstens aber die Breite des Referenzbereichs). Dieses Vorgehen folgt der Überlegung, dass der Bundeswert einerseits per Definition auf dem durchschnittlichen Risikoprofil der Patientinnen und Patienten basiert, dass also alle gerichteten Einflussfaktoren mit ihrem mittleren Ausmaß in diesen Wert eingehen. Andererseits kann aus der Tatsache, dass der Indikator*

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 21 von 28

*Qualitätsverbesserungen bewirken soll, abgeleitet werden, dass die bundesweite Versorgungsqualität nicht optimal ist. Für den Abstand des empirischen Bundeswertes vom Idealwert kann also angenommen werden, dass er mindestens so groß ist wie der mittlere Einfluss der gerichteten Faktoren auf den Indikatorwert<sup>10</sup>, und dass er zusätzlich einen Anteil enthält, der einer nicht optimalen bundesweiten Versorgungsqualität geschuldet ist“.*

Ist diese Darstellung so zu verstehen, dass der Unterschied zwischen dem maximal erreichbaren Wert eines Indikators und dem ermittelten Bundesdurchschnitt als Schätzung der (nicht von der Einrichtung beeinflussbaren) „gerichteten Einflussfaktoren“ verwendet werden soll? Wie soll dabei der Anteil des von den Einrichtungen zu verantwortenden „Verbesserungspotentials“ ermittelt bzw. geschätzt werden? Dies sollte im Abschlussbericht weitergehend erläutert werden. Erläuterungsbedürftig erscheint dabei vor allem, was mit „höchstens aber die Breite des Referenzbereichs“ gemeint ist. Ist diese Formulierung so zu verstehen, dass es in keinem Fall zu einer „Aufweichung“ des Referenzbereichs kommen kann (sondern nur zu einer „Verschärfung“)?

#### Berücksichtigung von „Einflussfaktoren“

Die Beispiele legen nahe, dass es sich wie im Verfahren der planungsrelevanten Indikatoren um „Ausnahmetatbestände“ handelt, d.h. um Konstellationen, die bei der Berechnung des Indikators nicht ausreichend berücksichtigt sind. Um eine konkrete Umsetzung für ein bundesweites Verfahren zu ermöglichen, sollten die in den Beispielen genannten Konstellationen in abstrakt beschriebene Kriterien / Verfahrensregeln gefasst werden. Insgesamt erscheint dem GKV-Spitzenverband das vom IQTIG vorgelegte Konzept als bestmögliche Annäherung an ein möglichst objektives Verfahren. Das IQTIG weist auf diese Limitationen des empfohlenen Vorgehens offen und transparent hin (z.B. Seite 163).

#### Umgang mit Dokumentationsfehlern

Diskussionsbedürftig erscheint der Umgang mit Dokumentationsfehlern. Auf Seite 150ff. wird dargestellt, dass Datenfehler berücksichtigt werden sollen. Im entsprechenden Abschnitt (Seite 153) wird jedoch nicht konkret beschrieben, wie diese identifiziert werden. In der Anlage zum Bericht (Seiten 54 und 55) wird an einem Beispiel erläutert, dass im Stellungnahmeverfahren vorgebrachte („behauptete“) Dokumentationsfehler nicht berücksichtigt werden. Praktisch führt diese Konstellation dann zu einer Bewertung als „Qualitätsdefizit“. Dieses Vorgehen unterscheidet sich von dem im Verfahren der planungsrelevanten Indikatoren, in dem das Vorliegen von Dokumentationsfehlern vor dem Stellungnahmeverfahren durch eine externe, unabhängige Validierung bestätigt wird. Praktisch erscheint das vom IQTIG empfohlene Vorgehen zumindest konfliktträchtig. Der GKV-Spitzenverband bittet das Institut um Klarstellung.

*Im Bericht auf Seite 155 wird formuliert:*

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 22 von 28

*„Je nach betroffenem Datenfeld können auch Datenfehler unterschiedlich gut durch eine Nachberechnung adressiert werden. Wurden durch Datenfehler Behandlungsfälle fälschlich in die Indikatorberechnung einbezogen, so lässt sich deren Auswirkung durch Ausschluss der Fälle aus der Grundgesamtheit des Indikators ermitteln. Hat ein Datenfehler dazu geführt, dass das interessierende Ereignis (z. B. Komplikationen oder die Durchführung eines Prozesses) fälschlich als gegeben bzw. als nicht gegeben gewertet wurde, so kann das Ausmaß des Datenfehlers auf den Indikatorwert ebenfalls einfach angegeben werden. Die Nachberechnung besteht dann darin, den Behandlungsfall als Behandlungsfall ohne interessierendes Ereignis zu werten, also ihn z. B. aus dem Zähler des Indikators auszuschließen (siehe Anhang, Kapitel 6). Geht ein fehlerhafter Wert dagegen eher indirekt in das Indikatorergebnis ein, beispielsweise, wenn Werte für Risikofaktoren in einer falschen Einheit dokumentiert wurden, so könnte die Auswirkung des Datenfehlers auf den Indikatorwert nur durch mit hohem Aufwand ermittelt werden.“*

Auf Seite 55 im Anhang wird – scheinbar im Widerspruch hierzu – formuliert:

*„Bei der Ermittlung des informativen Indikatorwerts im Rahmen der partiellen Nachberechnung werden die Angaben zu den Komplikationen so berücksichtigt, wie sie ursprünglich dokumentiert waren. Die Dokumentationsfehler können nicht als besondere Konstellation, die den Hinweis auf ein Qualitätsdefizit entkräftet, geltend gemacht werden.“*

Wie kann die im Bericht auf Seite 155 beschriebene Konstellation zustande kommen, so dass sie nicht von der Einrichtung zu verantworten ist? Der GKV-Spitzenverband bittet das Institut um Begründung, warum dies abweichend von bisherigem Vorgehen empfohlen wird. Das Grundkonzept der planungsrelevanten Indikatoren – das Stellungnahmeverfahren findet auf der Grundlage einer validierten, verlässlichen Datengrundlage statt – erscheint für eine belastbare Bewertung sehr vorteilhaft.

#### Umgang mit perzentilbasierten Referenzbereichen

Auf Seite 158 wird klargestellt, dass perzentilbasierte Referenzbereiche, keine „erwartbare“ Qualität definieren:

*„Verteilungsbezogene Referenzbereiche erlauben keine Aussagen über die Erfüllung definierter Standards, sondern lediglich vergleichende Aussagen relativ zu den Ergebnissen aller Leistungserbringer (IQTIG 2019a: 161 f.). Ob beispielsweise ein Leistungserbringer, der im 5. Perzentil der Verteilung aller Leistungserbringer in diesem Qualitätsindikator liegt, erwartbare Standards in der Versorgung erfüllt, kann nicht aus dem Indikatorergebnis erschlossen werden.“*

Als Konsequenz dieser Feststellung wird formuliert:

*„Dennoch könnte das Ergebnis des Leistungserbringers absolut gesehen keinen Qualitätsmangel darstellen, sondern lediglich schlechter sein als das Ergebnis von z. B. 90 % der Leistungserbringer (bei einem Referenzbereich auf Basis des 10. Perzentils).“*

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 23 von 28

und *„Strenggenommen müsste daher aus methodischer Sicht für Indikatoren mit einem verteilungsbezogenen Referenzbereich das Einstufungsergebnis „keine Qualitätsbewertung“ lauten, wenn nicht bereits die partielle Nachberechnung zu einer Entkräftung des Hinweises auf ein Qualitätsdefizit geführt hat.“*

Ist diese Darstellung nicht so zu verstehen, dass Indikatoren mit perzentilbasiertem Referenzbereich praktisch nicht als quantitatives Maß zur Feststellung der Erfüllung von Qualitätsanforderungen anzusehen sind?

*Die vorgeschlagene qualitative Bewertung hat ja letztlich nicht das Ziel, klar benennbare Qualitätsmängel zu identifizieren, sondern verfolgt ja letztlich ein „Ausschlussprinzip“ zur Validierung der quantitativen Ergebnisse (Widerlegung „des hinreichenden Verdachts auf ein Qualitätsdefizit“, der ausschließlich quantitativ generiert wird).*

Wäre damit in der Konsequenz für perzentilbasierte Indikatoren nicht das gesamte, hier vorgeschlagene Konzept zur qualitativen Bewertung ungeeignet? Der GKV-Spitzenverband bittet um Klarstellung.

#### Subkategorien zur Einstufung

Die in Kapitel 6.4.4 beschriebene Einstufung von Bewertungsergebnissen in Subkategorien ist nachvollziehbar und erscheint sachgerecht.

#### **Kapitel 6.5 „Mindestanforderungen an Stellungnahmen“**

Die empfohlenen Anforderungen an Stellungnahmen erscheinen sachgerecht und geeignet, die Ziele einer möglichst weitgehenden Vereinheitlichung des Vorgehens und hoher formaler Korrektheit zu erreichen. Erläuterungsbedürftig erscheint die Formulierung auf Seite 165:

*„In den Stellungnahmen können für das Verfehlen des Referenzbereichs zum einen auf einzelne Behandlungsfälle bezogene Gründe angegeben werden (z. B. dass ein Notfall ein Abweichen vom üblichen Vorgehen erforderlich machte) und zum anderen übergeordnete Gründe (z. B. dass wegen einer Spezialisierung der Einrichtung auf bestimmte Behandlungen die Patientinnen und Patienten ein besonderes Risikoprofil haben). Vor dem Hintergrund der folgenden Argumente sollte die Stellungnahme möglichst auf eine Analyse von übergeordneten Gründen fokussieren ...“*  
und

*„Allerdings ist es sowohl bei Einzelfallbezogenen als auch bei übergeordneten Gründen wichtig, dass der Leistungserbringer in seiner Stellungnahme für jeden Grund angibt, auf welche seiner Behandlungsfälle dieser Grund zutrifft bzw. ob der Grund für alle Behandlungsfälle zutrifft. Dazu sind die Vorgangsnummern der entsprechenden Behandlungsfälle mit der Stellungnahme zu übermitteln. Diese Angabe ist auch für diejenigen Behandlungsfälle zu machen, bei denen kein*

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 24 von 28

*interessierendes Ereignis aufgetreten ist, da dies für die Berechnung des informativen Indikatorergebnisses erforderlich ist (siehe Abschnitte 6.4.1 und Anhang, Kapitel 6). Gründe, bei denen aus der Stellungnahme nicht hervorgeht, ob sie nur bei den Behandlungsfällen mit einem interessierenden Ereignis oder auch bei Fällen ohne interessierendes Ereignis vorgelegen haben, können in der fachlichen Bewertung nicht berücksichtigt werden“*

Die Ausführung erscheint dem GKV-Spitzenverband widersprüchlich, da die Anforderung der Übermittlung der betroffenen Vorgangsnummern praktisch immer „übergeordnete Gründe“ zu einer einzelfallbezogenen Betrachtung macht? Es sollte im Abschlussbericht darauf eingegangen werden, wie der Aufwand (insbesondere bei den Einrichtungen) für dieses Vorgehen zu bewerten ist.

#### ***Kapitel 6.6 „Zusammensetzung von Fachkommissionen“***

*Es wird eine Empfehlung zur Berufung der Fachkommissionen gemacht.*

Auf Seite 167 empfiehlt das IQTIG eine einheitliche Aufwandsentschädigung:

*„Zur Steigerung der Akzeptanz des Verfahrens gegenüber den Expertinnen und Experten der Fachkommission wird eine bundesweit einheitliche Regelung zur Aufwandsentschädigung, die gleichermaßen für alle Expertinnen und Experten gilt, empfohlen“*

Dieser Vorschlag würde entweder eine Änderung der DeQS-RL (einheitliche Regelung) oder eine Abstimmung aller LAGen erfordern. Dieser Vorschlag wird vom GKV-Spitzenverband abgelehnt, die bisherigen Regelungen in der aktuellen DeQS-RL richten sich nach den Grundsätzen wie der G-BA sie anwendet und werden als sachgerecht betrachtet.

Auch die weiteren detaillierten Vorgaben greifen nach Auffassung des GKV-Spitzenverbandes zu weit in die Regelungshoheit auf Landesebene ein und bauen einen enormen Bürokratieaufwand auf.

Das IQTIG macht insgesamt nur sehr knapp Vorschläge zum Umgang mit Interessenkonflikten (Seite 168):

*„Damit eine unabhängige Bewertung der Versorgungsqualität gewährleistet werden kann, müssen Konflikte mit sekundären Interessenslagen weitestgehend vermieden werden. Solche Interessenkonflikte können beispielsweise auffällige Indikatorergebnisse der eigenen Einrichtung sein oder eine Konkurrenzsituation geografisch benachbarter Einrichtungen.“*

Diese Darstellung sollte im Abschlussbericht deutlich erweitert und konkretisiert werden.

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 25 von 28

### **Kapitel 7: Modul Qualitätsförderung**

Ab Kapitel 7 und damit leider auch in Bezug auf die konkreten Umsetzungsempfehlungen wird der Vorbericht inhaltlich deutlich knapper. Das IQTIG macht aber nach Ansicht des GKV-Spitzenverbandes einen klar strukturierten und formalisierten Vorschlag zum Vorgehen, der als erste Grundlage geeignet erscheint, ein möglichst homogenes Vorgehen bei länderbezogenen Verfahren zu ermöglichen. An einigen Stellen besteht Erläuterungs- und Ergänzungsbedarf. Auf Seite 172 wird dargestellt:

*„Nach Abschluss der Qualitätsbewertung wird dem internen Qualitätsmanagement bei erstmaliger Feststellung eines Qualitätsproblems zunächst die Möglichkeit eingeräumt, intern und damit im sanktionsfreien Raum entsprechende Maßnahmen einzuleiten. Nur bei wiederholten Qualitätsmängeln soll die zuständige Stelle aktiv werden, da wiederholte Qualitätsdefizite bei einem oder inhaltlich ähnlichen Indikatoren darauf hindeuten, dass die leistungserbringerinternen Anstrengungen, das Qualitätsproblem zu beheben, nicht ausreichend wirksam waren.“*

Unklar erscheint, was an dieser Stelle mit „sanktionsfreiem Raum“ gemeint ist.

Dies sollte im Abschlussbericht erläutert werden.

Nicht ganz klar erscheint das Konzept, wie geeignete qualitätsverbessernde Maßnahmen identifiziert werden sollen. Hierzu wird auf Seite 174 formuliert:

*„Qualitätsfördernde Maßnahmen sollten möglichst individuell auf den einzelnen Leistungserbringer angepasst sein, um das spezifische Qualitätsproblem optimal zu adressieren“.*

Diesem Statement kann zweifellos zugestimmt werden. Offen bleibt allerdings, wie die Ursachen der Qualitätsdefizite identifiziert werden sollen. Nur so kann ja eine gezielt wirksame Maßnahme festgelegt werden. Zudem erscheint es erforderlich (und ist auch im Auftrag unter den Punkten 2d und 2e angesprochen), Ursachen von Qualitätsdefiziten für eine strukturierte Berichterstattung systematisch zu kategorisieren. Nur so kann transparent werden, ob und ggf. in welcher Form die datengestützte Qualitätssicherung wirksam ist. Eine solche Berichterstattung ermöglicht aber vor allem auch, dass auch Einrichtungen, die nicht auffällig waren, systematisch von der Qualitätssicherung profitieren können. Eine solche Kategorisierung könnte auch eine gute Grundlage liefern, um „besonderen Handlungsbedarf“ auf übergeordneter Ebene feststellen zu können.

Vor diesem Hintergrund ist der Feststellung auf Seite 174

*„So lange das Verfahren zur Qualitätsmessung jedoch nur das Nicht-Erreichen von Qualitätsanforderungen feststellen kann, ist ein Lernen von den Besseren nur schwierig zu realisieren.“*

zwar zuzustimmen. Dennoch kann das Verfahren bei systematischer Klassifikation der Ursachen von Qualitätsdefiziten ein „Lernen von den Fehlern Anderer“ ermöglichen.

Auf Seite 175 wird dargestellt:

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 26 von 28

*„Als erster Schritt zum Abschluss von Zielvereinbarungen erfragt die zuständige Stelle, welche Maßnahmen die Einrichtung vorschlägt, um bestehende Qualitätsdefizite zu beheben. Auch die prüfende Fachkommission kann Rückmeldungen geben, wo sie Möglichkeiten für Verbesserungen sieht.“*

Es erscheint unklar, wie die Fachkommission diese Informationen erhält.

Die Ursachenanalyse erscheint weitestgehend auf die Einrichtung verlagert (Seite 175 „... der von der Einrichtung weiter durchzuführenden Ursachenanalysen“). Dies erscheint für ein Konzept der Qualitätsförderung unter Einbeziehung von Experten als „Peers“ sehr problematisch.

Die „Verzahnung“ der beiden Module „Bewertung“ und „Förderung“ ist nach Ansicht des GKV-Spitzenverbandes noch ergänzungsbedürftig.

#### **Kapitel 8: Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens**

Auf Seite 181 weist das IQTIG auf folgenden wichtigen Aspekt hin:

*„Anders als in den themenspezifischen Bestimmungen gemäß § 12 DeQS-RL festgelegt, würde damit die Einholung der Stellungnahmen nicht durch die Entscheidung der Fachkommission, sondern unmittelbar durch die zuständige Stelle erfolgen, woraus sich eine deutliche Zeitersparnis ergeben dürfte“*

Dieser Empfehlung des IQTIG sollte gefolgt werden, da sich praktisch ohne „Qualitätsverlust“ des Prozesses eine deutliche Effizienzsteigerung erreichen ließe.

Auf Seite 183 wird Folgendes als Möglichkeit zur Verkürzung von Fristen:

*„Folglich könnte das Stellungnahmeverfahren am 15. September des EJ + 1 abgeschlossen werden ... . Wie oben beschrieben steht diese Schätzung jedoch unter Vorbehalt, da bisher keine Erfahrungen mit der empfohlenen Methodik zur Bewertung von Stellungnahmen vorliegen.“*

Falls der Prozess in dieser Form durchführbar sein sollte, wäre dies als wichtiger Fortschritt gegenüber dem Status quo anzusehen, da die Bewertungen für den Qualitätsbericht der Krankenhäuser damit bereits am 15.9. eines Jahres (anstatt bisher am 31.10. eines Jahres) zur Verfügung stehen würden. Ggf. ließe sich damit erreichen, dass die Veröffentlichung der einrichtungsbezogenen Ergebnisse im auf das Erfassungsjahr folgende Jahr erfolgen könnte.

Auf Seite 185 wird auf eine sehr bedeutsame Möglichkeit der Verkürzung der Fristen im Verfahren insgesamt hingewiesen:

*„Eine weitere Möglichkeit für einen früheren Start wäre, die 1-monatige Korrekturfrist (siehe Abbildung 39) für die Daten zu streichen. Angesichts der quartalsweisen Auswertungen und Rückmeldeberichte und der damit verbundenen Möglichkeit, nötige Korrekturen rechtzeitig und bereits während des EJ durchzuführen, erscheint eine zusätzliche Korrekturfrist nicht notwendig.“*

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 27 von 28

*Darüber hinaus bestünde eine weitere Möglichkeit das Stellungnahmeverfahren vorzuverlegen darin, dass man die Datenlieferfrist vorzieht. Gemäß der DeQS-RL besteht der Datensatz des EJ nur aus den Dokumentationsbögen für die Patientinnen und Patienten, die noch im gleichen EJ entlassen werden, sodass Überlieger dem EJ nicht mehr zugerechnet werden. Dadurch ist die Dokumentation der letzten Fälle bereits am 31. Dezember des EJ möglich, sodass prinzipiell eine frühere Datenannahmefrist beispielsweise bis zum 31. Januar des EJ + 1 festgelegt werden könnte.“*

Diese Optionen sollten aus Sicht des GKV Spitzenverbandes – zunächst auch unabhängig von der Umsetzung der Empfehlungen des IQTIG zum Stellungnahmeverfahren – in der AG DeQS geprüft und ggf. realisiert werden. Eine entsprechende Diskussion sollte gemeinsam mit der vom IQTIG vorgeschlagenen Veränderung der „Zeitschiene“ geführt werden.

#### **Kapitel 9: Berichterstattung**

Das IQTIG schlägt vor, die Berichterstattung aus dem Stellungnahmeverfahren in die beiden Elemente „Berichterstattung zur Versorgungsqualität: Qualitätsergebnisse am Ende des Moduls „Qualitätsbewertung“ und „Berichterstattung über die Maßnahmen der Qualitätsförderung“ zu trennen (Seite 188). Die Ergebnisse zum Modul „Qualitätsbewertung“ könnten dadurch bereits im dem Erfassungsjahr folgenden Jahr zur Verfügung stehen (und damit fast 6 Monate früher als bislang). Der Vorschlag des IQTIG erscheint dem GKV-Spitzenverband als sehr begrüßenswerte Möglichkeit, Teile der Berichterstattung deutlich früher als bislang veröffentlichen zu können.

#### **Kapitel 10: Empfehlungen zur Evaluation**

Die „Empfehlungen“ des IQTIG erscheinen sehr allgemein und generisch. Im Wesentlichen wird empfohlen, sowohl eine formative wie auch eine summative Evaluation durchzuführen und die Erarbeitung eines Evaluationsplans zu beauftragen. Der Empfehlung kann grundsätzlich gefolgt werden.

#### **Kapitel 11: Zusammenfassung der Empfehlungen und Konsequenzen**

In Kapitel 11.3 formuliert das IQTIG in Bezug auf die Nutzbarkeit von Indices *„Daher sieht das hier vorgeschlagene Konzept vor, dass das Ergebnis jedes einzelnen Qualitätsindikators abschließend bewertet werden soll.“*

Vor diesem Hintergrund sollte das IQTIG im Abschlussbericht dazu Stellung nehmen, wie zukünftig mit Indikatoren umgegangen werden soll, die als Indices bezeichnet sind oder als solche aufgefasst werden können.

Stellungnahme des GKV-Spitzenverbandes vom 21.11.2019  
zum Vorbericht zu Stufe 1 und 2 Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern des IQTIG vom 10. Oktober 2019  
Seite 28 von 28

- Als Index bezeichnet:  
„QI 51901: Qualitätsindex der Frühgeborenenversorgung“ im Verfahren Neonatologie
- Ggf. als Index anzusehen:  
„QI 54018: Spezifische Komplikationen bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ im Verfahren HEP aggregiert unterschiedliche Ereignisse wie Gefäßläsionen, motorische Nervenschäden, Frakturen und Wundinfektionen und könnte daher als Index aufgefasst werden.

Auf Seite 222 formuliert das IQTIG

*„Aufgrund der Vielzahl der Änderungen, die die Umsetzung des empfohlenen Vorgehens mit sich bringen würde und aufgrund der Vielzahl der beteiligten Akteure, deren Handeln aufeinander abgestimmt werden müsste, empfiehlt das IQTIG die Entwicklung eines Umsetzungskonzepts. Im Rahmen eines solchen Umsetzungskonzepts sollte dann auch die Entscheidung über die konkrete Ausgestaltung des Stellungnahmeverfahrens unter Abwägung der Möglichkeiten zur Standardisierung des Verfahrens getroffen werden“*

Der Vorschlag des IQTIG erscheint überraschend, da er darauf hinweist, dass das IQTIG seine vorgelegten Empfehlungen trotz ca. zweijähriger Bearbeitungszeit als nicht umsetzungsreif einschätzt. Der GKV-Spitzenverband vertritt hier die Ansicht, dass die Ausgestaltung, Standardisierung und Klärung der Praktikabilität ausdrücklich Inhalte der hier zu bearbeitenden Beauftragung zur Weiterentwicklung des Strukturierten Dialogs sind.



KASSENÄRZTLICHE  
BUNDESVEREINIGUNG

---

# STELLUNGNAHME DER KASSENÄRZTLICHEN BUNDESVEREINIGUNG

ZUR WEITERENTWICKLUNG DES STRUKTURIERTEN DIALOGS MIT  
KRANKENHÄUSERN – VORBERICHT ZU STUFE 1 UND STUFE 2  
(STAND 21. NOVEMBER 2019)

DEZERNAT VERSORGUNGSQUALITÄT

21. NOVEMBER 2019

VERSION 1.0

# INHALT

---

<b>1.</b>	<b>PRÄAMBEL</b>	<b>3</b>
<hr/>		
<b>2.</b>	<b>ZUSAMMENFASSUNG UND FAZIT</b>	<b>3</b>
<hr/>		
<b>3.</b>	<b>KAPITELBEZOGENE ANMERKUNGEN</b>	<b>5</b>
3.1	Einleitung	5
3.2	Analyse der Ausgangssituation	5
3.3	Methodischer Hintergrund	6
3.4	Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und -förderung	7
3.5	Statistische Auswertungsmethodik	7
3.6	Fachliche Bewertung	8
3.7	Modul Qualitätsförderung	10
3.8	Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens	10
3.9	Berichterstattung	12
3.10	Empfehlungen zur Evaluation des neuen Konzepts	12
3.11	Zusammenfassung der Empfehlungen und Konsequenzen	12

---

## 1. PRÄAMBEL

Gemäß Plenumsbeschluss des Gemeinsame Bundesausschusses (G-BA) wurde das wissenschaftliche Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) am 18. Januar 2018 mit einer „Weiterentwicklung des Strukturierten Dialoges (SD) mit Krankenhäusern“ beauftragt. Als SD wird eine Abfolge von Prozessen innerhalb der „Richtlinie über Maßnahmen der Qualitätssicherung für nach § 108 SGB V zugelassene Krankenhäuser“ (QSHK-RL) beschrieben, die zum Ziel haben, eine durch Indikatoren abgebildete rechnerische Auffälligkeit der Qualität einer Leistungserbringung in Krankenhäusern zu überprüfen, abzugleichen sowie ein möglicherweise bestehendes klinisches Qualitätsdefizit zu bestätigen. Analysen des Strukturierten Dialoges mit Krankenhäusern hatten in den Vorjahren ein heterogenes Vorgehen der jeweils auf Landesebene mit der Durchführung des SD beauftragten Stellen ergeben. Die o. g. Beauftragung des G-BA vom 18. Januar 2018 hatte im Wesentlichen zum Ziel, ein einheitliches Vorgehen für den Prozess des SD vorzuschlagen, sowie Wege der Aufwandsreduktion für den Verfahrensablauf aufzuzeigen.

Weiterhin erhielt das IQTIG am 18. Mai 2018 die Beauftragung, „Kriterien für die Qualitätsbewertung sowie die Einleitung und Durchführung von QS-Maßnahmen“ im Rahmen der sektorenübergreifenden Qualitätssicherung (sQS) zu entwickeln. Wenige Tage vor geplanter Abgabe des Abschlussberichtes zu dieser Beauftragung erreichte den G-BA die Nachricht, das Institut habe sich dazu entschieden, Inhalte der Beauftragung vom 18. Mai 2018 in seinen Arbeiten Weiterentwicklung des SD mit abzubilden. Obwohl thematische Überschneidungen zwischen diesen Beauftragungen erkennbar sind, ergab sich nach Einschätzung der KBV eine klare Abgrenzung der Inhalte, so dass die Entscheidung des IQTIGs unverständlich erschien. Auf Aufforderung des Unterausschusses Qualitätssicherung erklärte sich das IQTIG im Nachgang bereit, spezifische Aspekte der Beauftragung vom 18. Mai 2018 in einem gesonderten Dokument darzulegen. In einer Stellungnahme an die betreuende Fach-AG im Mai und August 2019 sowie im Unterausschuss Qualitätssicherung im September 2019 hat die KBV dargelegt, dass sie die Beauftragung vom 18. Mai 2018 lediglich als teilweise erfüllt sieht. Das IQTIG sicherte zu, fehlende Anteile im Rahmen des hier vorliegenden Berichtes nachzuliefern.

Dies ist nach Einschätzung der KBV **nicht im ausreichenden Maße** geschehen. Es lassen sich lediglich generische Zusicherungen finden, dass alle vorliegenden Ergebnisse auf die sQS zu übertragen sind. Der Nachweis hierzu bleibt größtenteils aus. Nach Ansicht der KBV kann der Auftrag des G-BA vom 18. Mai 2018 weiterhin **lediglich als teilweise erfüllt** bezeichnet werden.

Aufgrund der Rolle des Instituts als Akteur im System der Gesetzlichen Krankenversicherung und des Stellenwerts der Qualitätssicherung für alle GKV-Versicherten, stellen sich an die vom Institut angewandten Methoden in besonderer Weise legitimatorische Anforderungen. Die „fachlich-wissenschaftliche Unabhängigkeit“ wird sowohl im Gesetz als auch in der Satzung des Instituts betont. Gleichwohl kann eine Weiterentwicklung der Qualitätssicherung nur in enger Zusammenarbeit mit dem G-BA als primärer Auftraggeber des IQTIG funktionieren. Vor diesem Hintergrund dankt die KBV für die Einbeziehung der Träger in den Diskurs und versteht nachfolgende Anmerkungen als einen Beitrag zum konstruktiven Dialog. Im Sinne einer vertrauensvollen Zusammenarbeit sollte zukünftig **frühzeitig der Dialog** mit Beteiligten und Auftraggebern im G-BA gesucht werden, falls seitens des IQTIGs Abweichungen von Beauftragungsinhalten gewünscht sind.

## 2. ZUSAMMENFASSUNG UND FAZIT

Die KBV sieht den vertragsärztlichen Versorgungsbereich im vorliegenden Bericht **nicht** als gleichberechtigten und gleichwertigen Partner der sektorenübergreifenden Qualitätssicherung abgebildet. Weitestgehend werden **Konzepte, die das IQTIG für Krankenhäuser auf Basis von Daten der stationären**

**QS entwickelt hat, auf die vertragsärztliche Versorgung übertragen.** Dabei bleiben wesentliche strukturelle Unterschiede zwischen Praxis und Krankenhaus unerwähnt. Die vertragsärztliche Praxis ist kein „kleines Krankenhaus“ und sollte nicht als solches abgehandelt werden; die eigenständige Rolle im Gesamtkontext des gesundheitlichen Versorgungsspektrums ist zu berücksichtigen. **Insbesondere der bewusste Ausschluss vertragsärztlicher Expertise beim fachlichen Austausch mit den auf Landesebene beauftragten Stellen, ist kritisch zu bewerten** und durch entsprechende Nacharbeiten des IQTIGs zu ergänzen.

Kernforderungen zur Weiterentwicklung des SD sind in der DeQS-RL bereits implementiert und umgesetzt worden. Somit ist eine Vielzahl der im Bericht aufgeführten Empfehlungen zur Neugestaltung der Prozesse obsolet. **Die normativen Vorgaben der DeQS-RL sind im Abschlussbericht, insbesondere vor dem Hintergrund der vollständigen Überführung aller QSKH-Verfahren in die DeQS-RL im Jahre 2020, unbedingt zu berücksichtigen.**

Ebenfalls bleibt **das Konzept einer „Mandantenfähigen Datenbank“**, wie sie in der DeQS-RL dargelegt wird, bedauerlicherweise vollständig unbeachtet. Es finden sich **deutliche Widersprüche** zwischen Empfehlungen des IQTIGs zu Nachberechnung von Qualitätsindikatoren (QI) auf Landesebene in diesem Bericht sowie Vorschläge zur Umsetzung der Mandantenfähigen Datenbank in anderen Berichten des IQTIGs. Hier ist eine Prüfung der Inhalte zu empfehlen.

**Die im vorliegenden Bericht vorgestellten statistischen Methoden sind der Problemstellung angemessen und werden fachlich korrekt und umfangreich erläutert.** Allerdings erschließt sich der Bericht in dieser Form nur, wenn die methodische Kompetenz eines Statistikstudiums mitgebracht wird. Im Sinne einer größeren Verständlichkeit sollten im Bericht die wesentlichen Punkte und Handlungsempfehlungen klar dargestellt und erläutert werden. Dies würde die Akzeptanz des vorgeschlagenen Ansatzes erheblich fördern.

**Folgende Vorschläge des IQTIGs zur Weiterentwicklung des SD/Stellungnahmeverfahrens sind im vollen Umfang zu unterstützen und Entwicklungen auf diesem Gebiet verstärkt zu fordern:**

- die Abkehr von einer „rechnerischen Auffälligkeit“, hin zu einer „statistischen Signifikanz“ bei der Identifikation quantitativer Auffälligkeiten;
- die konsequente Verwendung von Indikatoren mit festen Referenzbereichen anstelle von perzentilbasierten Indikatoren oder Indikatoren ohne Referenzbereiche.

Insbesondere sind **Erprobungen** der vorgeschlagenen Modelle mit einer limitierten Anzahl von Verfahrensteilnehmern anzustreben, bevor diese bundesweit eingeführt werden. Dabei sollten **Kosten-/Nutzen-** sowie **Aufwand-/Nutzen-Darstellungen** im Rahmen dieser Erprobungen erfolgen.

Ebenfalls werden alle Bestrebungen des IQTIGs hin zur **Entwicklung eines transparenten, effizienten und für alle Verfahrensteilnehmer aufwandsarmen Vorgehen** begrüßt. Die KBV möchte allerdings betonen, dass die übergreifende Bewertung von Ergebnisauswertungen und individuelle Auswahl qualitätsfördernder Maßnahmen durch eine qualifizierte Fachkommission essentieller Bestandteil einer effizienten Qualitätssicherung ist. Der Vorschlag des IQTIGs, dass **aus jeder quantitativen Auffälligkeit eines Indikators zwingend jeweils ein Stellungnahmeverfahren zu erfolgen habe, ist abzulehnen**, da dies bei der hohen Anzahl von Indikatoren pro Verfahren gerade nicht zu einer Aufwandsreduktion führen wird.

**Ebenso ist es abzulehnen, dass formale Fehler bei der Beantwortung von Stellungnahmen als Qualitätsdefizite der medizinischen Versorgung gewertet werden.** Ziel der QS ist eine qualitativ hochwertige Patientenversorgung, nicht das möglichst korrekte Ausfüllen von Formblättern

## 3. KAPITELBEZOGENE ANMERKUNGEN

### 3.1 Einleitung

Das IQTIG betont in den einleitenden Kapiteln, alle Ergebnisse und Empfehlungen des vorliegenden Berichtes seien auf die sQS, wie sie über die „Richtlinie zur datengestützten, einrichtungsübergreifenden Qualitätssicherung (DeQS-RL)“ implementiert wird, anwendbar. Allerdings wird insbesondere durch die Erläuterungen zum Auftragsverständnis in Kapitel 1.3. deutlich, dass dies bei der Entwicklung des Konzeptes nicht vorgesehen war und lediglich der Bezug zur externen, stationären QS gesucht wurde.

Seitens der KBV wird um eine Korrektur der Darstellung der Abläufe zur Beauftragung des G-BA vom 18. Mai 2018 auf S. 15 in Kapitel 1.2 erbeten. Der an dieser Stelle erwähnte Bericht des IQTIGs zum 15. März 2019 wurde nach Aufforderung des Unterausschusses QS erstellt; eine gesonderte Bearbeitung der Beauftragung vom 18. Mai 2018 war vom IQTIG nicht vorgesehen. Ebenso ist der Umgang der Beauftragung vom 18. Mai 2018 in keinem formalen Zusammenhang mit der Außerkraftsetzung der Qesü-RL zum Ende des Jahres 2018 zu setzen. Die DeQS-RL als Nachfolge-Richtlinie der Qesü-RL hat den nahtlosen Fortgang der sQS sowie aller laufenden Beauftragungen aus den Vorjahren normativ gewährleistet.

Weiterhin ist in Frage zu stellen, ob im Jahre 2013 ein Workshop mit den Landesarbeitsgemeinschaften (LAG) stattgefunden hat; zu diesem Zeitpunkt war in keinem Bundesland eine solche gegründet. Im Anhang zum Bericht werden die Protokolle zweier weiterer Treffen mit Vertretern der Landesebene beigefügt, allerdings fehlt bei dem Protokoll des als „Workshop zum SD“ benannten Treffens die Angabe eines Datums. Vermutlich handelt es sich hierbei um das im Kapitel 1.4 (S. 18) erwähnte Treffen vom 27. Mai 2019 mit „Vertretern der LQS und LAG“. Ein weiteres Treffen fand am 18. Juni 2019 statt – hier werden als Teilnehmer die Vertreter der Landesqualitätsstellen (LQS) benannt, ohne Erwähnung der LAG. Die KBV geht davon aus, dass unter den in Kapitel 1.4 (S. 18) genannten „Vertretern der Landesebene und Experten“ bei allen aufgeführten Treffen keine Personen mit nachgewiesener Fachexpertise der vertragsärztlichen Versorgung beteiligt waren. Durch Verwendung des Begriffes „LAG“ an dieser Stelle wird fälschlicherweise ein sektorenübergreifendes Vorgehen suggeriert. Die KBV bittet, diese Darstellung zu korrigieren und erwartet bei zukünftigen Entwicklungsschritten die aktive Integration und Berücksichtigung der vertragsärztlichen Perspektive.

### 3.2 Analyse der Ausgangssituation

Es wird in Kapitel 2.1 bis 2.4 dargelegt, dass die QSKH-RL keine Definition des Begriffs „Strukturierter Dialog“ beinhaltet. Im Rahmen des SD würden unter Einbeziehung von Experten Ergebnisse von Einrichtungen bewertet, Maßnahmen zur Qualitätsverbesserung und –förderung eingeleitet und die Validität der Dokumentation überprüft. Dabei fehle eine klare Struktur. So sei zum Beispiel nicht definiert, welche Form oder welchen Inhalt eine Stellungnahme haben solle. Experten seien Ehrenamtliche, von denen ausschließlich Kenntnisse aus dem Bereich Qualitätsmanagement und Qualitätssicherung gefordert seien. Auf welchem Gebiet die Mitglieder darüber hinaus fachkundig sein sollen, ist nicht näher beschrieben. Auch variere die Anzahl der Mitglieder pro Fachgruppe stark, sowohl zwischen den Bundesländern als auch zwischen den Fachgruppen.

Gleichzeitig wird festgestellt, dass die Betrachtung des „Strukturierten Dialogs“ der QSKH-RL mit Blick auf die Implementation eines weiterentwickelten Konzepts in der DeQS-RL wichtig sei. Dabei wird im Wesentlichen außer Acht gelassen, dass die DeQS-RL an dieser Stelle bereits über abweichende, weiterentwickelte Regelungen verfügt, die die identifizierten Defizite des SD aufgreifen und korrigieren. Eine Gegenüberstellung von QSKH-RL und DeQS-RL sowie eine Bewertung im Detail bleibt aus und sollte im Abschlussbericht entsprechend ergänzt werden. Durch die alleinige Darstellung der Abläufe der stationären QS wird an dieser Stelle deutlich, dass eine Betrachtung des Ist-Zustandes in der sQS mit vertragsärztlicher Beteiligung nicht stattgefunden hat.

In Kapitel 2.5 wird gezeigt, dass sich Bewertungsstellen systematisch unterscheiden:

- 1) Die Aufnahmebereitschaft eines Stellungnahmeverfahren nach quantitativer Auffälligkeit;
- 2) die Bewertung als qualitativ auffällig, nach einem Stellungnahmeverfahren.

Die Unterschiede zwischen den Bewertungsstellen lassen sich nicht allein durch die Unterschiede zwischen den Leistungserbringern erklären. Es wird abgeleitet, dass eine einheitliche (methodisch-statistische) Bewertung durch die Bewertungsstellen für die Aufnahme eines Stellungnahmeverfahrens notwendig ist (siehe Kapitel 5).

Die Unterschiede zwischen den Bewertungsstellen wurden methodisch mit einem gemischten logistischen Regressionsmodell am Beispiel von drei verschiedenen Qualitätsindikatoren bestimmt. Das Modell ist gut gewählt, da es sowohl die Wahrscheinlichkeit für quantitative/qualitative Auffälligkeit als auch die Hierarchisierung von Leistungserbringern und Bewertungsstellen beschreiben kann. Das Modell wird im Anhang Kapitel 4 schlüssig und nachvollziehbar erklärt, setzt aber erweiterte statistische Kenntnisse voraus.

Ergänzend sollte im Abschlussbericht dargestellt werden, ob die drei gewählten Beispiele, die für die exemplarische Analyse der Heterogenität verwendeter Indikatoren repräsentativ für die Gesamtmenge der Qualitätsindikatoren der QSKH-RL sind. Ebenfalls fehlt hier die Erwähnung der sQS-Indikatoren aus den DeQS-Verfahren, insbesondere eine Erläuterung, ob das angewendete Modell auf Sozialdatenindikatoren, Follow-Up-Indikatoren oder das Index-Konzept der Einrichtungsbefragung QS WI theoretisch übertragbar sei.

Im Kapitel 2.6 wird zu der Einschätzung gelangt, dass für Leistungserbringer mit kleiner Fallzahl die derzeitige Methodik der „rechnerische Auffälligkeit“ zu restriktiv sei: Es komme häufiger zu quantitativer Auffälligkeit, die sich nicht in einer qualitativen Auffälligkeit bestätige. Es wird vorgeschlagen, deshalb die Fallzahl beziehungsweise statistische Unsicherheit für die Bestimmung der quantitativen Auffälligkeit zu berücksichtigen (siehe auch Kapitel 5), und diese damit weniger oft quantitativ auffällig werden zu lassen. Damit könne der Aufwand im Strukturierten Dialog verringert werden, da weniger Stellungnahmen angefordert würden.

Diese Überlegungen werden seitens der KBV begrüßt. Bei einer weitaus höheren Anzahl vertragsärztlicher Einrichtungen im Bundesgebiet als es Krankenhäuser gibt – die durchschnittlich über weniger Fallzahlen verfügen – ist die Verringerung der Anzahl an Stellungnahmeverfahren durch Präzisierung der Berechnung einer „Auffälligkeit“ in einem Indikator unbedingt anzustreben.

### 3.3 Methodischer Hintergrund

Ab S. 58 in Kap 3.3 werden methodische Überlegungen beschrieben, den Strukturierten Dialog/das Stellungnahmeverfahren zu vereinheitlichen und Transparenz in Bewertung und Entscheidung herbeizuführen. Dabei darf nach Einschätzung der KBV nicht übersehen werden, dass die Fachkommissionen mit ihrer Bewertung letztendlich Einzelfallentscheidungen auf Basis langjähriger, medizinischer Expertise treffen sollen. Dies bedeutet, dass per definitionem spezielle Gegebenheiten und Sonderheiten sowie besondere Umstände des Leistungserbringers und seines Patientenkollektivs berücksichtigt werden. Einzelfallentscheidungen haben eine eingeschränkte Vergleichbarkeit, dennoch können sie selbstverständlich transparent und nachvollziehbar sein. Dieser Tatsache muss das vom IQTIG angestrebte Konzept der „einheitlichen Entscheidungsregeln“ (vgl. S. 58) unbedingt Rechnung tragen.

In Kapitel 3.6 ab S. 61 vergleicht das IQTIG die Abläufe des SD mit dem eines „Peer-Review-Verfahrens“ und zeigt auf, dass „eine Evaluation des Strukturierten Dialogs bislang nicht durchgeführt wurde, sodass derzeit

keine wissenschaftlichen Daten zu sogenannten „Erfolgsfaktoren“ vorliegen“. Es erscheint jedoch problematisch, Empfehlungen zur Weiterentwicklung von qualitätsfördernden Maßnahmen auf Basis von Literatur über Peer-Review-Verfahren abzuleiten. Vielmehr erscheinen Evaluation und Erprobung „vor Ort“ an dieser Stelle angezeigt, wenn man tatsächliche Aussagen zum Weiterentwicklungspotential von qualitätsfördernden Maßnahmen treffen will.

### 3.4 Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und -förderung

Es wird in diesem Kapitel ein zukünftiges Rahmenkonzept für den SD beschrieben, das separate Module innerhalb des Prozesses abgrenzt:

- 1) Qualitätsbewertung (Stellungnahmeverfahren),
- 2) Qualitätsförderung (QS-Maßnahmen),
- 3) Überprüfung der Daten- und Dokumentationsqualität (Validierung).

Diese Herangehensweise wird seitens der KBV zwar grundsätzlich unterstützt, jedoch fehlt an dieser Stelle der Bezug zur DeQS-RL, in der diese Trennung bereits normativ festgelegt ist. Die DeQS-RL ist in diesem Sinne als Weiterentwicklung der QSKH-RL zu verstehen. Der methodische Zugewinn der Erläuterungen in Kapitel 4 erschließt sich der KBV daher an dieser Stelle nicht und könnte vollständig entfallen, wenn der Bezug zur DeQS-RL im Abschlussbericht entsprechend ergänzt wird.

Auf S. 68 wird der angestrebte Ablauf innerhalb des Moduls für die Qualitätsbewertung beschrieben. Qualitätsindikatoren, die statistisch auffällig sind, sollen als „hinreichender Hinweis für ein Qualitätsdefizit“ verstanden werden. „Die zuständige Stelle“ - vermutlich die LAG - soll daraufhin den Leistungserbringer für jeden auffälligen QI getrennt zur Stellungnahme auffordern. Eine Gesamtbetrachtung aller QI und eine Abwägung soll es zukünftig vor einer Aufforderung zur Stellungnahme nicht mehr geben.

An dieser Stelle bleiben seitens der KBV in Anbetracht der Vielzahl von Qualitätsindikatoren pro Verfahren innerhalb der DeQS-RL (bei letzter Zählung alleine 39 QIs für die QS PCI) berechnete Zweifel, dass dieses empfohlene Vorgehen zur Effizienz im Sinne einer „Verschlankung“ des Verfahrens führen wird. Die Gesamtbetrachtung aller Qualitätsindikatoren eines Verfahrens sowie die anschließende Möglichkeit zur Einzelfallabwägung bei der Entscheidung zur Stellungnahme und Auswahl der qualitätsfördernden Maßnahmen ist eine Kernaufgabe der Fachkommissionen. Zentraler Faktor für eine zielgerichtete, effiziente Qualitätsförderung ist die Vermeidung von Redundanzen - die sich aus multiplen Stellungnahmeverfahren zu einem QS-Verfahren beim Leistungserbringer sicherlich ergeben würden.

### 3.5 Statistische Auswertungsmethodik

Es werden in diesen Kapiteln verschiedene Möglichkeiten beschrieben, die quantitative Auffälligkeit zu bestimmen.

- 1) Rechnerische Auffälligkeit: die bisher verwendete Methode, im Bericht stellenweise gleichgesetzt mit quantitativer Auffälligkeit, basiert auf „harten“ Referenzwerten für die Qualitätsindikatoren, benachteiligt aber Leistungserbringer mit kleinen Fallzahlen.
- 2) Statistisch signifikante Auffälligkeit: berücksichtigt statistische Unsicherheit, die sich bei Leistungserbringern mit kleiner Fallzahl stärker auswirkt und dadurch weniger häufig Auffälligkeit auslöst.
- 3) Statistisch relevante Auffälligkeit: berücksichtigt darüber hinaus die Abweichung vom Referenzwert und die zukünftige Fallzahl.

Für alle drei Fälle wird im Rahmen eines als Bayesschem Netzwerk modellierten Entscheidungsprozesses mit Verlustfunktion eine optimale Entscheidungsstrategie hergeleitet.

Weiterhin wird der Entscheidungsprozess um eine zweistufige Variante erweitert, die auf den Daten von zwei aufeinander folgenden Jahren basiert. In jedem Fall muss zwischen dem Aufwand für ein Stellungnahmeverfahren und den negativen Auswirkungen eines nicht entdeckten Qualitätsmangels abgewogen werden.

Insgesamt empfiehlt der Bericht, die statistische Signifikanz für die quantitative Auffälligkeit zu benutzen und bei kleiner Fallzahl die Daten von zwei aufeinander folgenden Jahren zusätzlich in einer zweistufigen Entscheidung zu verwenden. Bei der Verwendung von statistischer Signifikanz stellt sich die Frage nach dem Signifikanzniveau  $\alpha$ . Der Bericht macht verschiedene Vorschläge, er legt sich in dieser entscheidenden Frage aber nicht fest.  $\alpha = 0,05$  entspricht der wissenschaftlichen Konvention,  $\alpha = 0,465$  würde ähnliche Ergebnisse wie die bisher verwendete Methodik liefern. Ein dynamisches  $\alpha$  könnte je nach Aufwand-Nutzen-Abwägung pro Indikator, sowie  $\alpha = 0,25$  für ein einheitliches Reporting der Qualitätsbericht-Ergebnisse gewählt werden. Hier sollte unbedingt eine eindeutige Empfehlung ausgesprochen werden.

Kapitel 5 ist zwar insgesamt schlüssig und umfangreich erläuternd geschrieben, mit seinen vielen mathematischen Details zur Modellierung (u. a. Bayessche Netzwerke, Konzept der statistischen Relevanz, A-priori-Wahrscheinlichkeiten), sicherlich in einem Anhang besser aufgehoben. Eher unüblich in der Statistik und für das Verständnis nicht notwendig ist die aufgeführte Unterscheidung zwischen „enumerativer“ und „analytischer“ Statistik. Inhaltlich problematisch ist die Gleichsetzung des Verlusts durch Falsch-Positive und Richtig-Positive bzw. die Annahme, dass der Aufwand eines Stellungnahmeverfahrens nicht vom Ausgang desselben abhängt.

Die KBV begrüßt ausdrücklich das Bestreben des IQTIGs, die systematische Benachteiligung von Leistungserbringern mit kleineren Fallzahlen durch Verlassen des Konzeptes der „rechnerischen Auffälligkeit“ zugunsten der Verwendung der „statistischen Signifikanz“ erheblich abzumildern. Zudem ist die ausführliche Analyse des SD der vergangenen Jahre wissenschaftlich interessant, wenn auch die Erkenntnis, dass ein Signifikanzniveau von  $\alpha = 0,465$  der jetzigen Methodik entspricht, überraschend ist und somit jegliche Anpassung eine Verbesserung darstellt.

### 3.6 Fachliche Bewertung

Das in Kapitel 6.1 (Tabelle 13, S. 145) dargelegte Bewertungsschema, die dahinterliegende Methodik (Kapitel 6.2) sowie der Ablauf (Kapitel 6.3) der fachlichen Bewertung einschließlich des Bewertungsalgorithmus (Kapitel 6.4) sind aus Sicht der KBV insgesamt sehr komplex und werden zum einen sehr detailliert und zum anderen aber auch nur in Ansätzen beschrieben. Abschließend wird lediglich darauf verwiesen, dass die Methodik noch weiterentwickelt werden muss. Dies betrifft sowohl die Weiterentwicklung der biometrischen Methodik zur Festlegung einer „zureichenden“ Qualität bezogen auf das Qualitätsziel eines Indikators (S. 146/S. 174), wie auch die Bestimmung der Anteile eines Referenzbereiches (Toleranzanteil und gerichtete Einflussfaktoren, S. 157).

Weiterhin äußert sich das IQTIG, dass aus methodischer Sicht für Indikatoren mit einem verteilungsbezogenen Referenzbereich immer das Einstufungsergebnis „keine Qualitätsbewertung“ (S) lauten müsse (S. 158), da die beschriebene Berechnung des Referenzbereichs nur bei fest definierten Referenzwerten anwendbar sei. Gleichzeitig sollen jedoch in einem direkten Widerspruch hierzu perzentilbasierte Referenzbereiche im Stellungnahmeverfahren vorerst wie feste Referenzbereiche behandelt werden.

Die KBV bittet um Erläuterung, in welchem Rahmen und bis wann diese Entwicklungsarbeiten erfolgen sollen. Zum jetzigen Zeitpunkt ist der Bewertungsalgorithmus unverständlich. Auch ist nicht nachvollziehbar, wieso nicht erst feste Referenzbereiche definiert werden und anschließend die Bewertung nach neuer Methodik vorgenommen wird.

Hinsichtlich des Bewertungsschemas scheint die Einstufung in die Unterkategorien – insbesondere nach dem in den Abbildungen 34 bzw. 35 dargestellten Algorithmus – für die Fachkommissionen eine Herausforderung darzustellen und voraussichtlich eine ausführliche Schulung und Einarbeitung zu benötigen. Insbesondere lehnt die KBV die automatische Einstufung in die Kategorie „Qualitätsdefizit“ A0 ab, sollte eine Stellungnahme nicht den formalen Anforderungen entsprechen. In diesem Falle sollte dem Leistungserbringer nach Information durch die zuständige Stelle die Möglichkeit gegeben werden, die Stellungnahme in einer bestimmten Frist zu überarbeiten. Alternativ kann die Einstufung „keine Qualitätsbewertung“ (S) vorgenommen werden.

Das IQTIG spricht in diesem Zusammenhang im Fazit (Kapitel 6.4.6, S. 163) davon, dass für die Umsetzung der Empfehlungen der fachlichen Bewertung eine Kosten-Nutzen-Abwägung durchgeführt werden sollte. Leider wird dies nicht weiter ausgeführt oder beispielhaft für ein QS-Verfahren mit allen Auswirkungen dargestellt. Die KBV empfiehlt daher vor der bundesweiten Einführung und normativen Umsetzung dieser Vorschläge in der DeQS-RL eine Erprobungsphase und anschließende Evaluation auf Basis einer oder mehrerer QS-Verfahren in einem oder mehreren Bundesländern.

Die KBV bittet um Erläuterung, wer das „einheitliche Web-basierte Portal“ (S. 164) für alle Bewertungsstellen zur Durchführung der fachlichen Bewertung erstellen und betreiben soll. Auch soll dieses Portal Funktionalitäten für die Berechnung bereinigter Ergebnisse vorhalten – wie ist dies im Kontext der gemäß DeQS-RL vom IQTIG vorzuhaltenden mandantenfähigen Datenbank einzuordnen? Das IQTIG führt hierzu ergänzend an, dass die partielle Nachberechnung von Ergebnissen zwangsläufig im Austausch mit der Fachkommission erfolgen soll (S. 150). Aus Sicht der KBV können die LAGen ihre Aufgaben gemäß den Empfehlungen des IQTIG nur umsetzen, wenn ein entsprechend umfänglicher Zugriff auf alle QS-Daten möglich ist. Die KBV bittet darum – im Hinblick auf die aktuellen Beratungen im G-BA zur mandantenfähigen Datenbank – in diesem Kontext genauer zu beschreiben, welche Daten in welcher Form zur Verfügung stehen müssen und wie eine angemessene technische Unterstützung der Landesebene aussehen soll.

Im Rahmen des Stellungnahmeverfahrens soll laut IQTIG bei nicht zu verantwortenden, auffälligen Indikatorergebnissen nachgewiesen werden, dass mit der gebotenen Sorgfaltspflicht alles Zumutbare getan wurde (S. 166). Die KBV bittet um Erklärung, in welcher Form dieser Nachweis erfolgen soll. Außerdem ist kritisch anzumerken, inwieweit Vertragsärzte und -ärztinnen hier gegenüber welcher Organisation in welchem Umfang überhaupt nachweispflichtig sind. Bei dem Stellungnahmeverfahren handelt es sich aus Sicht der KBV allerdings um ein Verfahren zur Klärung möglicher Qualitätsdefizite mit dem Ziel, Versorgungsqualität zu verbessern und nicht um ein Gerichtsverfahren. Insgesamt scheint dieser Vorschlag das Aufgabenfeld des IQTIGs als wissenschaftliches Institut zu verlassen und einer rechtlichen Prüfung zu bedürfen.

In Kapitel 6.6 fehlt aus Sicht der KBV leider der Bezug zu den aktuellen Regelungen zur Zusammensetzung von Fachkommissionen in der DeQS-RL und QSKH-RL. So geht das IQTIG von einem Bewerbungsverfahren aus und definiert formale Anforderungskriterien dafür. Die Richtlinien legen jedoch fest, dass die Experten jeweils von den im QS-Verfahren betroffenen Organisationen entsendet werden. Auch wenn der Wunsch nach einer bundesweit einheitlichen Regelung zur Aufwandsentschädigung nachvollziehbar erscheint, hinterfragt die KBV, wie dies im Kontext umzusetzen ist, dass die entsendenden Organisationen die Kosten für die Mitglieder übernehmen. Des Weiteren ist die Empfehlung nicht nachvollziehbar, dass grundsätzlich für jedes QS-Verfahren folgende ständige Experten für die Fachkommissionen zu benennen sind: Fachärzte, Pflegefachkräfte, klinische Qualitätsmanager und Patientenvertreter. Dies sollte unbedingt verfahrensabhängig festgelegt werden. Beispielsweise ist das eigenständige Berufsbild des klinischen Qualitätsmanagers in der vertragsärztlichen Versorgung nicht bekannt und wäre in einem Verfahren zur ambulanten Psychotherapie fehlbesetzt.

### 3.7 Modul Qualitätsförderung

Aus Sicht der KBV spricht grundsätzlich nichts dagegen, das für bundesbezogene Verfahren entwickelte Zielvereinbarungskonzept auch bei den landesbezogenen Verfahren zu verwenden. Wie vom IQTIG beschrieben, existiert die Zielvereinbarung als solche bisher nicht in der DeQS-RL, wird aber in der QFD-RL als eigene, fördernde Maßnahme aufgeführt. In den Beratungen dazu bei der Erstellung der DeQS-RL wurde explizit zwischen einer Vereinbarung und einer Zielvereinbarung unterschieden, und bei der Überführung der Qesü-RL auf einer Anpassung der Vereinbarung in Zielvereinbarung verzichtet. Es bleibt in diesem Zusammenhang offen, was genau das IQTIG unter einer Zielvereinbarung vs. Vereinbarung versteht bzw. wie sie aus Sicht des IQTIGs einzuordnen ist.

Die Empfehlung des IQTIG, Gespräche und Begehungen zukünftig nicht mehr im Rahmen des Stellungnahmeverfahrens vorzunehmen, sondern als qualitätsfördernde Maßnahmen der Maßnahmenstufe 1 zuzuordnen (S. 173), kann von der KBV nicht mitgetragen werden. Gerade im vertragsärztlichen Bereich werden zur Verminderung des Aufwandes für Ärzte und für die zuständige Stelle bisher auch persönliche oder telefonische Gespräche im Rahmen des Stellungnahmeverfahrens durchgeführt, die häufig deutlich schneller und effizienter zur Klärung des Sachverhaltes führen, als langwierige schriftliche Korrespondenzen. Natürlich ist es problemlos möglich, auch Gespräche transparent und nach einem möglichst einheitlichen Schema in ihren Ergebnissen zu protokollieren.

Der Vorschlag, qualitätsfördernde Maßnahmen auch ohne identifiziertes Defizit einzuleiten (S. 174), erscheint aus Sicht der KBV sehr irritierend. Wozu wird der große Aufwand einer QS-Dokumentation und daraus folgender Qualitätsmessung mit Indikatoren betrieben? Diese könnte dann vollständig entfallen und auf Basis von Stichproben, Zufall oder nach eigenem Ermessen Maßnahmen durchgeführt werden. Diese Vorgehensweise erscheint bei der nachvollziehbaren Forderung nach mehr Transparenz und Standardisierung in der Qualitätsbewertung verwunderlich.

Die KBV kann die Überlegungen des IQTIG, bei fördernden Maßnahmen den Schwerpunkt auf „Lernen von den Besseren“ zu legen zwar gut nachvollziehen, gibt jedoch zu bedenken, dass dies in der Umsetzung mit einigen Herausforderungen verbunden sein wird. Zum einen liegt bisher keine Methodik vor, Leistungserbringer mit „zureichender“ Qualität zu identifizieren zum anderen wird die Gewinnung und Zusammensetzung der Peers nicht ganz unproblematisch sein, je nachdem, wie die regionale Verteilung und bisherige Zusammenarbeit der entsprechenden Leistungserbringer ist. Die KBV würde sich hier noch konkretere Überlegungen zur Umsetzung dieses Vorschlages wünschen.

### 3.8 Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens

Das IQTIG führt in diesem Kapitel verschiedene Optionen auf, die das Stellungnahmeverfahren verkürzen können. Hierbei schränkt das IQTIG die zu verwendenden Datenquellen auf ausschließlich einrichtungs- und fallbezogene QS-Dokumentation ein (Kapitel 8.1 S. 177). Der zukünftige Einbezug von Patientebefragungen als auch die schon erhobenen Sozialdaten werden in diesem Konzept der Neuausrichtung von Zeitfristen in der DeQS-RL nicht berücksichtigt. Die KBV fordert den Einbezug aller Datenquellen der DeQS-RL für das theoretische Konstrukt einer Neuausrichtung aller Zeitfristen.

Bei der Erläuterung der bisherigen Fristen gibt das IQTIG fälschlicherweise eine allgemeine Frist für die Beendigung des Stellungnahmeverfahrens laut DeQS-RL an, welche nicht korrekt ist. Der auf S. 178 zitierte § 12 Absatz 6 der Themenspezifischen Bestimmungen bezieht sich ausschließlich auf die Beendigung des Stellungnahmeverfahrens in den Krankenhäusern zum 31. Oktober des Erfassungsjahres (EJ) +1 und nicht auf die vertragsärztlichen Leistungserbringer in der DeQS-RL.

Weiterhin erörtert das IQTIG drei Optionen zur Verkürzung des Stellungnahmeverfahrens:

- 1) **Aufwandsreduktion durch Volumenreduktion:** Hier wird vor allem die neue Methodik zur Berechnung statistisch signifikanter Auffälligkeiten thematisiert, in der das IQTIG eine Zeitersparnis sieht, da die Einholung einer Stellungnahme nicht mehr durch die Fachkommission entschieden wird, sondern direkt durch die LAG (bzw. durch den Algorithmus des IQTIG) vorgenommen werden soll. Dieses Vorgehen kann seitens der KBV nicht unterstützt werden (siehe Kapitel 3), da eine indikatorübergreifende Ergebnisbewertung durch die Fachkommission als zentrales Element der Qualitätsbeurteilung und -förderung gesehen wird. Zudem wird unter der Thematik Zeitersparnis auch der Wegfall von Hinweisen verortet, was in der Umsetzung sicher nicht zu einer Gesamtaufwandsverringerung führen wird, da voraussichtlich anstelle von Hinweisen die wesentlich umfangreicheren Stellungnahmeverfahren vermehrt zum Einsatz kommen werden. Dagegen begrüßt die KBV ausdrücklich den Vorschlag des IQTIG, eine Aufwandsreduktion durch eine Streichung von Qualitätsindikatoren zu erreichen.
  
- 2) **Effizienzsteigerung durch Standardisierung:** Zwar wird durch die Einschränkung auf ausschließlich schriftliche Stellungnahmen der Prozess standardisiert, jedoch gibt die DeQS-Richtlinie wesentlich mehr Möglichkeiten vor, die in der Anwendung aufwandsärmer sein können. Dies betrifft insbesondere mündliche/telefonische Stellungnahmen, welche derzeit in den laufenden Verfahren auch häufig angewandt werden (siehe Kapitel 7). Weiterhin führen die rigiden Vorgaben von Formulierungen für den schriftlichen Stellungnameprozess zu Irritationen bei den LAGen, die teilweise bereits (Software-)Lösungen mit eigenen Formulierungen entwickelt und implementiert haben. Hierzu folgender Protokollauszug aus dem Workshop zur Weiterentwicklung des Strukturierten Dialogs: „Einige Teilnehmerinnen und Teilnehmer berichten, dass es Fachgruppen gebe, die einen strukturierten Fragebogen (teilweise zwischen Geschäftsstellen abgestimmt) entwickelt und Anfragetexte zusammengestellt haben, welche dann scheinbar nicht mehr genutzt werden können, wenn die Texte zur Einholung von Stellungnahmen vorgeschrieben seien. Dies könne für die Akzeptanz des Verfahrens abträglich seien.“
  
- 3) **Frist für die Abgabe der Stellungnahme innerhalb von vier Wochen:** Diese Vierwochenfrist liegt in der Sommer- bzw. Urlaubszeit – dem gesamten Monat Juli eines Jahres. Anders als bei Krankenhäusern kann eine Stellungnahme im vertragsärztlichen Bereich nur von einem Individuum, dem angesprochenen Vertragsarzt/der Vertragsärztin selbst durchgeführt werden. Zuarbeiten durch Mitarbeiter einer QM-Abteilung, Medizincontrolling oder eines mehrköpfigen ärztlichen Teams mit Vertretungsregelungen für Chefärzte, wie sie im Krankenhaus üblich sind, finden sich nicht in der vertragsärztlichen Versorgung. Die KBV spricht sich gegen eine Verkürzung der Stellungnahmefrist aus, da dies voraussichtlich zu einer vermehrten Anzahl von nicht-fristgerecht erbrachten Stellungnahmen bei Vertragsärzten führen wird, und als eine systematische Benachteiligung dieser Leistungserbringergruppe zu bezeichnen ist. Dem oben genannten Protokoll ist zu entnehmen, dass die Verkürzung der Stellungnahmefrist auch für die Krankenhäuser problematisch gesehen wird. Auch hier wird in der Haupturlaubszeit mit einer deutlich reduzierten Personaldecke die Patientenversorgung aufrechterhalten und deshalb administrative Aufgaben nicht erste Priorität erhalten.

Weitere Vorschläge des IQTIG, wie die Streichung der einmonatigen Korrekturfrist als auch die Vorverlegung der Datenlieferfrist auf den 31. Januar des EJ +1 ergibt für die fallbezogenen Daten mit quartalsweiser Lieferfrist sicherlich Sinn, sind aber für die einrichtungsbezogene Dokumentation abzulehnen, da es zu einer enormen Einschränkung der Datenlieferfrist von neun auf vier Wochen für die Vertragsärzte führen würde.

### 3.9 Berichterstattung

Im Kapitel „Berichterstattung“ erläutert das IQTIG Empfehlungen für eine optimierte Berichterstattung zu den Ergebnissen des Strukturierten Dialogs bzw. dem Stellungnahmeverfahren mit Blick auf die DeQS-Richtlinie. Die Empfehlungen aus Kapitel 4, Trennung von Qualitätsbewertung und Qualitätsförderung, wie sie bereits in der DeQS-Richtlinie geregelt sind, sowie die Empfehlungen aus Kapitel 8 zur Berichterstattung, ermöglichen eine frühzeitige Zurverfügungstellung von Qualitätsergebnissen, um der Qualitätsförderung mehr Zeit einzuräumen.

Die KBV begrüßt es, dass zukünftig über die Ergebnisse der Qualitätsbewertung (Bundesauswertung sowie die Ergebnisse der Qualitätsbewertung) ein Jahr früher im Bundesqualitätsbericht berichtet werden soll. Ergebnisse zu den Maßnahmen der Qualitätsförderung und zur Datenvalidierung erfolgen aus Zeitgründen weiterhin im Berichtsjahr EJ +2.

### 3.10 Empfehlungen zur Evaluation des neuen Konzepts

In diesem Kapitel werden die in Kapitel 5 vorgeschlagen Methoden der Operationalisierung von „Hinreichenden Hinweisen“ aufgegriffen. Die mögliche Aufwandsreduktion für den Strukturierten Dialog bildet sich über die Reduktion von quantitativen Auffälligkeiten ab und soll für die statistisch signifikante und statistisch relevante Auffälligkeit jeweils abgeschätzt werden.

Als „Tuning-Parameter“ werden unterschiedliche Werte für Schwellenwert  $\zeta$  und Signifikanzniveau  $\alpha$  eingesetzt und anhand perzentil-basierter Indikatoren und Indikatoren mit festen Referenzbereichen die unterschiedlichen Wirkungsgrade auf die Anzahl von quantitativen Auffälligkeiten aufgezeigt. Hierbei zeigt sich deutlich, dass eine Aufwandseinsparung – unabhängig vom eingesetzten „Tuning-Wert“ – nur für Indikatoren mit festem Referenzbereich zu erreichen ist. Zwar wird anschließend schlüssig und nachvollziehbar dargelegt, dass die Anzahl an quantitativen Auffälligkeiten Stellungnahmen erst ab einem  $\zeta = 2$  und  $\alpha = 0,2$  verringert werden kann, allerdings fehlt auch an dieser Stelle weiterhin die klare Empfehlung zur Umsetzung und das Fazit bleibt vage und ziellos. Ebenso werden Überlegungen zu der „Mehrjahreseinstufung“ aus Kapitel 5 aufgegriffen. Es fehlt hier die konkrete Empfehlung für den Einsatz spezifischer „Tuning-Werte“ und welche Grundbedingungen erfüllt sein müssen, um eine solche mehrjährige Betrachtung durchführen zu können (Fallzahlen? Verfahrensspezifisch? Indikatorspezifisch?). Da sich in Kapitel 5 bereits für die Verwendung der „statistischen Signifikanz“ ausgesprochen wurde stellt sich die Frage, weshalb hier erneut die „statistische Relevanz“ als gleichwertige Option dargestellt wird.

Die KBV begrüßt grundsätzlich die Empfehlung des IQTIG zu einer formativen (prozessbegleitend mit Implementation der Empfehlungen) und summativen (retrospektiv nach Abschluss der Implementation) Evaluation des vorliegenden Konzeptes.

### 3.11 Zusammenfassung der Empfehlungen und Konsequenzen

Die Überlegungen des IQTIG, grundsätzlich Indikatoren mit festem Referenzbereich in der sQS einzusetzen, wird von der KBV ausdrücklich begrüßt. Allerdings sollten diese Empfehlung vom IQTIG auch konsequent umgesetzt werden und das bisher übliche Vorgehen, Verfahren mit Indikatoren völlig ohne Referenzbereiche oder mehrheitlich perzentil-basierte Indikatoren (QS PCI, QS NET) in die bundesweite Versorgung einzubringen, kritisch überdacht werden. Die Weiterentwicklung bestehender Indikatoren (S. 210/211) sieht die KBV als Kernaufgabe des IQTIG im Rahmen der Verfahrenspflege – gesonderte Beauftragungen durch den G-BA sind hierfür nicht notwendig.

Die auf S. 213 in Kapitel 11.3 dargelegte grundsätzliche Ablehnung von Indizes als Aufgreifkriterium kann nur mit Verwunderung zur Kenntnis genommen werden; insbesondere die Annahme, dass „die Bewertung einzelner Indikatoren durch Verwendung eines Index verzichtbar“ sei. Es stellt sich die Frage, wie diese

Aussage in Bezug auf die Einrichtungsbefragung QS WI zu verstehen ist. Auch bittet die KBV um Ergänzung, zu welchen nicht näher aufgeführten „anderen Zwecken“ die Ergebnisse von QIs weiterverwendet werden. Gerade mit dem Ziel der Effizienzsteigerung und Verschlankung des SD könnten Indizes einen wertvollen Beitrag leisten – es sei an dieser Stelle erneut der Hinweis erlaubt, dass in der QS PCI zwei kardiologische Interventionen mit 39 Indikatoren abgebildet werden.

Bedauerlich ist, dass ein aus Sicht der KBV ebenfalls vielversprechendes Konzept zur Aufwandsreduktion – die Entwicklung Leistungsbereiche übergreifender Indikatoren – nur sehr kurz angerissen wird. An dieser Stelle sei darauf verwiesen, dass die KBV insbesondere bei der Entwicklung von Patientenbefragungen wiederholt ein Modulsystem für Indikatoren anregt und gerade für dieses Instrument der QS ein Leistungsbereiche übergreifendes Vorgehen als sehr sinnvoll erachten würde

»  
**Stellungnahme zum Vorbericht „Weiterentwicklung des  
Strukturierten Dialogs mit Krankenhäusern – Stufe 1 und 2“  
(IQTIG Stand 10. Oktober 2019)**

Stand 20.11.2019

»

Kassenärztliche Bundesvereinigung  
Behrenstraße 42  
10117 Berlin  
[www.kzbv.de](http://www.kzbv.de)  
Tel. 030 280179-0

## Inhalt

I. Vorbemerkung.....	3
II. Zusammenfassung.....	5
III. Grundlegende Anmerkungen.....	8

## I. Vorbemerkung

Der Vorbericht ist mit der Überschrift „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“ betitelt. Der Entwicklungsbericht geht auf die Beauftragung vom 18. Januar 2018 des Gemeinsamen Bundesausschusses (G-BA) zurück und adressiert die QS-Verfahren nach der QSKH-Richtlinie. Auf den ersten Blick sind die Vertragszahnärzte bei diesem Entwicklungsauftrag nicht betroffen.

Nahezu parallel erging am 17. Mai 2018 ein ähnlicher Entwicklungsauftrag auf Basis der Qesü-Richtlinie an das IQTIG. Das IQTIG wurde beauftragt, bundesweit einheitliche Bewertungskriterien für die qualitative Bewertung der Indikatorergebnisse der sektorenübergreifenden QS-Verfahren PCI und WI zu entwickeln. Hierzu liegt der AG DeQS ein vom IQTIG am 15. März 2019 nicht fristgerecht bereitgestellter Abschlussbericht vor. Die AG DeQS vom 28. Mai 2019 stellte einvernehmlich fest, dass die Beauftragung vom 17. Mai 2018 vom IQTIG nicht vollständig bearbeitet wurde und bat das Institut, die Anmerkungen der AG DeQS bei der Erstellung des Berichts zur Stufe 2 gemäß Beauftragung vom 18. Januar 2018 entsprechend zu berücksichtigen. Das IQTIG hatte die Situation geprüft und war zu der Einschätzung gekommen (E-Mail vom 20. Juni 2019), dass eine weitergehende Bearbeitung des Auftrags hinsichtlich des Vorgehens im Stellungnahmeverfahren für Follow-up-Indikatoren im Rahmen der Beauftragung vom 17. Mai 2018 nicht möglich war. Bezüglich der Entwicklung von Kriterien zur Beurteilung von Sozialdaten empfiehlt das IQTIG eine Folgebeauftragung für das Jahr 2020 zu formulieren und bietet hierfür seine Unterstützung an.

Die KZBV legte – wie andere Bänke des G-BA auch - großen Wert darauf, dass beide Beauftragungen getrennt voneinander bearbeitet werden. Um das Verfahren zeitlich abzukürzen, sprach sich die Mehrheit im G-BA jedoch dafür aus, dass das IQTIG die noch fehlenden Punkte aus der Beauftragung zur Qesü-RL in diesem Auftrag ergänzt. Vor dem angesprochenen Hintergrund hat sich die KZBV entschieden, sich am Stellungnahmeverfahren gemäß § 137a Abs. 7 SGB V zu beteiligen.

In den Kapitel 1.2 „Schnittstellen zu anderen Beauftragungen und Richtlinien“ und Kapitel 1.3 „Auftragsverständnis“ des Vorberichts zur Beauftragung vom 18. Januar 2018 führt das IQTIG aus, dass es „sinnvoll erscheint, ein einheitliches Rahmenkonzept für ein Stellungnahmeverfahren zu den quantitativen Ergebnissen der datengestützten Qualitätssicherung zu

entwickeln, das prinzipiell in allen indikatorbasierten QS-Verfahren zur Anwendung kommen kann.“ Mit diesem Ziel legt das IQTIG in seinem Abschlussbericht seine Empfehlungen für ein Stellungnahmeverfahren im Rahmen der jetzt übergreifenden DeQS-Richtlinie vor. Der Bericht beinhaltet vor allem konkrete Empfehlungen für die Auswertung und Bewertung von leistungserbringerbezogenen Qualitätsindikatoren-Ergebnissen auf Basis der QS-Dokumentation durch die Leistungserbringer.

Die KZBV wird in ihrer schriftlichen Stellungnahme nur grundlegende Aspekte ansprechen. Aufgrund der willkürlichen Vermengung beider Aufträge hat der Bericht einen hohen Grad an Unüberschaubarkeit erreicht. Die KZBV behält es sich vor, etwaige weitere Kommentare oder Kritikpunkte in späteren Anhörungen einzubringen.

## II. Zusammenfassung

Mit dem vorgelegten Bericht zur „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“ wurde notdürftig versucht, nicht nur dem Auftrag vom 18. Januar 2018, sondern auch der Beauftragung „Kriterien für die Qualitätsbewertung sowie für die Einleitung und Durchführung von Qualitätssicherungsmaßnahmen im Rahmen der Qesü-RL“ vom 17. Mai 2018 gerecht zu werden. So gehen an vielen Stellen des Vorberichtes Ebenen aufgrund der unterschiedlichen Rechtsgrundlagen (QSKH-RL; Qesü-RL/DeQS-RL; Plan-QI) durcheinander. Die Entwicklung des Berichtes ist vor allem durch die Qualitätssicherung im stationären Sektor geprägt. Auch das begleitende Panel zur Weiterentwicklung des Strukturierten Dialogs fand nur mit Vertreterinnen und Vertreter aus dem stationären Sektor statt.

Das IQTIG schlägt ein rechnerisches Modell vor, das zu einer höheren Spezifität<sup>1</sup> der Qualitätsindikatoren und zu einer eindeutigeren Zuschreibbarkeit der Ergebnisse führen soll. Die KZBV wird auf das statistische Modell im Rahmen ihrer Stellungnahme nicht weiter eingehen.

Es kann davon ausgegangen werden, dass eine sehr weitgehende Standardisierung im Prozess der Bewertung und im Stellungnahmeprozess große Anteile an Ressourcen bei den LAGen und Fachkommissionen einsparen kann; dies geht jedoch zu Lasten der Bürokratie bei den Leistungserbringern. Für die Abgabe von Stellungnahmen hat das IQTIG hohe Anforderungen formuliert, technische und formale Vorgaben gemacht, die den Aufwand bei den Leistungserbringern aller Voraussicht nach steigen lassen. Hinweise, kurze telefonische Absprachen oder individuelle Vereinbarungen im gegenseitigen Einvernehmen sind im vorliegenden Modell nicht mehr vorgesehen. Die Berechtigung des IQTIG, solche Vorgaben zu machen, ist nicht ersichtlich.

Feedbackberichte an die Leistungserbringer sollen künftig alle statistischen Berechnungen der Qualitätsindikatoren inklusive der Angaben für statistische Auffälligkeiten (z. B. Nichterreichen des Referenzwertes oder des Signifikanzniveaus) enthalten. Darüber hinaus sollen die Aufforderungen zum Stellungnahmeverfahren beigefügt sein. Dabei erzeugt jeder Qualitätsindikator eine eigene Aufforderung zur Stellungnahme. Die KZBV begrüßt eine erhöhte Zuverlässigkeit bzw. Spezifität der Qualitätsindikatoren, plädiert aber für mehr Spielraum im

---

<sup>1</sup> Die Spezifität gibt die Wahrscheinlichkeit an, mit der unauffällige Leistungserbringer auch tatsächlich kein Qualitätsdefizit aufweisen.

Stellungnahmeprozess ansich. Die statistischen Berechnungen pauschal und alleine mit der Aufforderung zu einer Stellungnahme an die Leistungserbringer zu versenden, ohne dass eine Bewertung oder Entscheidung seitens der LAG und/oder der Fachkommission vorgeschaltet ist, ist aus Sicht der KZBV kontraproduktiv. Die Sinnhaftigkeit eines solchen Vorgehens ist verfahrensspezifisch zu betrachten. Im Ergebnis dürfte dies ansonsten zu einem hohen Vertrauensverlust seitens der Leistungserbringer führen. Ein erster Bewertungsprozess vor Versendung der Aufforderungen zu Stellungnahmen durch die LAG als Lenkungsgremium muss erhalten bleiben.

Zudem sind bestehende Richtlinienvorgaben zu beachten. An dieser Stelle wird u. a. die willkürliche Vermischung zwischen dem Strukturierten Dialog nach der QSKH-RL und des Stellungnahmeverfahrens gemäß DeQS-RL deutlich. Nach § 17 Abs. 1 DeQS-RL sind die Fachkommissionen mit ihrem Sachverstand bei der Entscheidung zur Einholung von Stellungnahmen mit zu beteiligen.

Obgleich das IQTIG vorgibt, mit dem vorliegenden Bericht auch die Beauftragung vom 17. Mai 2018 (Qesü-RL) erfüllt zu haben, fehlt die Darstellung der Aufgaben und der Rolle der KVen/KZVen im Stellungnahmeprozess und bei der Umsetzung von Qualitätsförderungsmaßnahmen. Sie unterstützen ganz wesentlich die LAGen und die Fachkommissionen bei der Umsetzung der Prozesse. Sie haben die unmittelbare Verbindung zu den Vertrags(zahn)ärztinnen und Vertrags(zahn)ärzten und können den Prozess aktiv mitgestalten. Die aktuellen Regelungen gemäß § 17 DeQS-RL sind bei der Neugestaltung des Strukturierten Dialogs zwingend zu beachten und können nicht vom IQTIG ins Belieben gestellt werden.

Zentrale Beauftragungsgegenstände aus der Beauftragung vom 17. Mai 2018 (zur Qesü-RL) blieben bislang nahezu unbearbeitet:

- Es sollte geprüft werden, ob die Einbeziehung mehrerer Leistungserbringer in das Stellungnahmeverfahren zu ein- und demselben Indikator fachlich geboten erscheint (vor allem bei den Langzeit-Follow-up-Indikatoren) und welche Modifikationen des Stellungnahmeverfahrens dann ggf. erforderlich sind.

- Die Weiterentwicklung für die Auswertung und Bewertung von leistungserbringerbezogenen Qualitätsindikatoren-Ergebnissen auf Basis von Sozialdaten oder Daten aus Patientenbefragungen fehlt im Bericht.

Die KZBV bittet das IQTIG, entsprechende Textpassagen kritisch zu prüfen und richtlinienkonform anzupassen. Noch fehlende Beauftragungsbausteine sollten nachgeliefert werden.

### III. Grundlegende Anmerkungen

#### 1. Auftragsverständnis (Kap. 1.3)

Im Kapitel 1.3 nimmt das IQTIG eine Konkretisierung des Auftrags vor. Notwendig sei eine Definition von auffälligen Indikatorenergebnissen, die Konkretisierungen für das Stellungnahmeverfahren sowie die Entwicklung eines Einstufungsschemas zur Bewertung von Indikatorenergebnissen.

Dafür hält es das IQTIG für sinnvoll, eine richtlinienunabhängige Definition von auffälligen Indikatorenergebnissen, Konkretisierungen für das Stellungnahmeverfahren und ein Einstufungsschema zur Bewertung von Indikatorenergebnissen zu entwickeln. Begründet wird dies damit, dass die DeQS-RL künftig das neue Rahmenkonzept für alle QS-Verfahren darstellt. Die QS-Verfahren der Qesü-RL wurden bereits in die DeQS-Richtlinie integriert, die QS-Verfahren der QSKH-RL werden stufenweise überführt. Somit solle die Bewertung übergreifend sein und nicht davon abhängen, in welcher Richtlinie die Qualitätsindikatoren verortet sind (IQTIG 2019, S. 15).

Die KZBV sieht es kritisch, dass mit diesem Entwicklungsbericht beide Beauftragungen gemeinsam abgedeckt werden sollen. Eine richtlinienunabhängige Bearbeitung ist abzulehnen, da es grundlegende Unterschiede zwischen der Qesü-RL bzw. der DeQS-RL, der QSKH-RL und der im Bericht oft zitierten PlanQI-RL gibt:

- Die Inhalte, der Ablauf sowie der sprachliche Duktus unterscheiden sich beim Stellungnahmeverfahren und dem Strukturierten Dialog.
- Ergebnisse zur Struktur-, Prozess- und Ergebnisqualität der stationären Gesundheitsversorgung sollen Verwendung bei den Landesplanungsbehörden für Zwecke der qualitätsorientierten Krankenhausplanung finden (Planungs- und Vergütungsentscheidungen). Dies erfordere einen hohen Grad an Justiziabilität; planungsrelevante Qualitätsindikatoren entbehren im ambulanten Sektor jedoch jeder Rechtsgrundlage.
- Die Zuständigkeiten unterscheiden sich. Nach der Qesü-RL sind die LAGen unter Beteiligung der Fachkommissionen für die Bewertung der Auffälligkeiten (Stufe 1) zuständig – für Stufe 2 u. a. auch die KVen/KZVen; nach der QSKH-RL binden die auf der Landesebene beauftragten Stellen (LQS) Fachgruppen in die Bewertung von

Qualitätsindikatoren ein. Diese Gremien setzen sich jeweils unterschiedlich zusammen und erfordern laut Richtlinien unterschiedliche Beteiligungsanforderungen. Für die Mitarbeit in Fachgruppen werden ausschließlich Kenntnisse aus dem Bereich Qualitätsmanagement und der Qualitätssicherung gefordert. Auf welchem Gebiet die Mitglieder darüber hinaus fachkundig sein sollten, ist nicht näher in der QSKH-RL beschrieben. Laut DeQS-RL (Teil 1, § 5 Abs. 5), die die Qesü-RL zum 1. Januar 2019 abgelöst hat, müssen in den Fachkommissionen mindestens 2/3 Ärztinnen und Ärzte der entsprechenden Fachrichtungen bzw. Zahnärztinnen und Zahnärzte bzw. Psychotherapeutinnen und Psychotherapeuten unter den stimmberechtigten Mitgliedern sein. Dass das IQTIG in seinem Vorbericht ab Kapitel 3 für den Begriff Fachgruppe den Begriff Fachkommission gemäß DeQS-RL verwendet, löst das Problem nicht (vgl. IQTIG 2019, S. 24, Fußnote 11).

## 2. Analyse der Ausgangssituation (Kap. 2)

Obwohl sich die Empfehlungen im Bericht auf ein richtlinienunabhängiges Stellungnahmeverfahren beziehen, liegt der Fokus bei den Ausführungen zur Ausgangssituation ausschließlich auf einer Analyse des Strukturierten Dialogs gemäß QSKH-RL (vgl. IQTIG 2019, S. 20). Auch an dieser Stelle wird wieder deutlich, dass der Bericht zum Strukturierten Dialog nachträglich um entsprechende Passagen in Bezug zum Bericht ergänzt wurde.

Dargestellt werden die strukturellen Voraussetzungen des Strukturierten Dialogs hinsichtlich der beauftragten Stellen auf Bundes- und Landesebene und der Fachgruppen im stationären Sektor. Des Weiteren erfolgt eine Übersicht über die derzeitige Durchführung des Strukturierten Dialogs<sup>2</sup> (rechnerische Auffälligkeit, Bewertung der Versorgungsqualität, Hinweise, schriftliche Stellungnahmen, Besprechungen und Begehungen, Entscheidungsfindung, Bewertung der Dokumentationsqualität, Qualitätsförderung und Zielvereinbarungen).

---

<sup>2</sup> Zur Orientierung: Im Erfassungsjahr 2015 wurden gemäß QSKH-RL bei 15.856 rechnerischen Auffälligkeiten 9.797 schriftliche Stellungnahmen seitens der LQS und des IQTIG angefordert, es gab 5.939 Hinweise, 1.121 Ziele wurden vereinbart und es fanden 282 Besprechungen und 19 Begehungen statt.

Im Rahmen biometrischer Analysen (statistische Berechnungen wie gemischte logistische Regressionsmodelle<sup>3</sup>) stellte das IQTIG eine große Heterogenität in der Vorgehensweise und bei der Bewertung im Strukturierten Dialog fest (IQTIG 2019, S. 32 ff). Die Analysedaten stammen wiederum ausschließlich aus dem stationären Sektor, aus dem QS-Verfahren „Hüftendoprothesenversorgung“, was hier nochmals kritisch angemerkt wird.

Das IQTIG kam in einem Zwischenfazit zu dem Schluss, dass einerseits zu prüfen sei, bei welchen Qualitätsindikatoren es indikatorspezifischen Weiterentwicklungsbedarf gibt (Evidenzgrundlage u. a.). Andererseits sollte eine methodische Weiterentwicklung in der rechnerischen Auffälligkeitseinstufung erfolgen. Tendenziell werden zu viele Standorte mit kleiner Fallzahl als auffällig eingestuft, weil die statistische Unsicherheit in der Bestimmung der Auffälligkeitseinstufung nicht berücksichtigt werde (IQTIG 2019, S. 52-53).

### 3. Methodischer Hintergrund (Kap. 3)

Die bisherige Umsetzung des Strukturierten Dialogs zeige eine starke Heterogenität. So seien nicht nur die Ergebnisse zur tatsächlichen Versorgungsqualität eines Leistungserbringers maßgebend, es spiele auch eine Rolle, wer die Stellungnahme beim Leistungserbringer anfordert, wie die Stellungnahme formuliert ist und welche Fachkommission diese bewertet (vgl. IQTIG 2019, S. 59). Methodisch werden explizite Beurteilungsregeln im Stellungsverfahren (d. h. Rechenvorschriften und Gewichte für Einflussfaktoren für die fachliche Bewertung und für Nachberechnungen) angestrebt, die einem implizitem Vorgehen (Bewertung durch Expertinnen und Experten) vorgeschaltet werden. Grundsätzlich – vor allem aus ökonomischen und ressourcentechnischen Gründen – kann dem Vorgehen gefolgt werden.

---

<sup>3</sup> Die Bewertungsergebnisse des Strukturierten Dialogs wurden mit den rechnerischen Ergebnissen der Leistungserbringer verknüpft und ausgewertet. Es wurden zwei Fragestellungen aufgeworfen: Wie groß ist die landesstellenabhängige Heterogenität bezüglich der Aufnahme eines Stellungsverfahren bei rechnerischer Auffälligkeit eines Leistungserbringers? Und: Wie groß ist die landesstellenabhängige Heterogenität bezüglich der Bewertung als qualitativ auffällig nach Aufnahme eines Stellungsverfahren? Die Methodik wurde vom IQTIG im Anhang Kapitel 4 beschrieben.

#### 4. Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und –förderung (Kap. 4)

Das IQTIG schlägt eine klare Trennung zwischen der „Qualitätsbewertung“ (Indikatorenbeurteilung, Einholung und Bewertung der Stellungnahmen) und „Qualitätsförderung“ (Steigerung der Versorgungsqualität bei einem Leistungserbringer nach abgeschlossener Qualitätsbewertung) vor. Diese Trennung ist nachvollziehbar und sinnvoll.

Der Prozess zur Bewertung von Qualitätsindikatorenergebnissen ist wiederum als zweistufiges Verfahren vorgesehen: 1. Formale Prüfung durch die beauftragte Stelle und 2. Fachliche Bewertung durch die beauftragte Stelle auf Empfehlung der Fachkommission. Inwieweit sich die einzelnen Bewertungsinhalte wirklich klar voneinander hinsichtlich ihrer Adressaten trennen lassen (Aufgabe LAG oder Aufgabe Fachkommission), sieht die KZBV an dieser Stelle eher skeptisch. Hier bedarf es sicherlich den Erfahrungen aus der späteren Praxis für eine abschließende Beurteilung.

Das neue Konzept sieht die verpflichtende Einholung von Stellungnahmen vor, wenn ein Indikatorergebnis eines Leistungserbringers einen „hinreichenden Hinweis auf ein Qualitätsdefizit“ nahelegt. Mit dem Versand der Auswertungen wird der Leistungserbringer auf rechnerische Auffälligkeiten hingewiesen und anhand der Vorjahresergebnisse soll für den Leistungserbringer zu erkennen sein, ob es sich um wiederholte oder um eine erstmalige Auffälligkeit handelt. Die Stellungnahmen sollen nach dem neuen Konzept durch konkrete Mindestanforderungen formal wie inhaltlich so weit wie möglich standardisiert werden. Dieses Vorgehen ist nachvollziehbar und im Sinne einer Ökonomisierung und einem einheitlichen Bewertungsprozess prinzipiell sinnvoll. Allerdings wird vor zu vielen Standard-Vorgaben und vor Dogmatismus gewarnt. Wie bereits unter Punkt II dieser Stellungnahme angesprochen, obliegt das Einholen von Stellungnahmen den LAGen als Lenkungsgremium (vgl. § 17 DeQS-RL). Die vom IQTIG vorgeschlagenen Anforderungen an die Stellungnahmen können hier bestenfalls als Anregungen verstanden werden, da es dem IQTIG an jeglicher Kompetenz für formelle Vorgaben fehlt.

Aus Sicht der KZBV werden die Qualitätsindikatoren einseitig im Kontext von „accountability“ (für PlanQI; Zu- und Abschläge etc.) verfolgt. Indikatoren sollen einen hohen Grad an Spezifität mitbringen, d. h. sie sollen vor allem justiziabel sein. Dadurch sind die Anforderungen an die Methodik sehr hoch (IQTIG 2019, S. 70). Dies trifft nach Auffassung der KZBV nur auf

einen Teil der QS-Verfahren bzw. nur auf bestimmte Qualitätsindikatoren zu. Planungsrelevante Qualitätsindikatoren finden sich ausschließlich im stationären Sektor. Auch das Thema Zu- und Abschläge ist nur für die stationären Einrichtungen von Relevanz. An dieser Stelle zeigt sich nun mal mehr, dass eine klare Trennung beider Beauftragungen sinnvoll gewesen wäre.

Nach Auffassung der KZBV soll durch die Qualitätssicherung vor allem die intrinsische Motivation zur Qualitätsverbesserung gestärkt werden. Sanktionierende Maßnahmen werden nach der DeQS-Richtlinie erst in zweiter Stufe verfolgt (vgl. § 17 Abs. 3 DeQS-RL).

#### 5. Statistische Auswertungsmethodik (Kap. 5)

In einem sehr ausführlichen und umfangreichen Kapitel (IQTIG 2019, S. 71 – 142) stellt das IQTIG seine komplexe Auswertungsmethodik dar, mit der künftig belastbare quantitative Aussagen zur Versorgungsqualität der Leistungserbringer geliefert werden sollen. Ferner sollen belastbare Informationen zur Auslösung des Stellungnahmeverfahrens für die Leistungserbringer gewonnen werden (u. a. Indikatorwert; zugehöriges Unsicherheitsintervall; quantitative Auffälligkeitseinstufung). Die statistische Auswertungsmethodik erfolgt in Anlehnung an das Verfahren für Planungsrelevante Qualitätsindikatoren (bayesianische Auswertungsmethodik). Stellschrauben für eine „hinreichende statistische Evidenz für ein mögliches Qualitätsdefizit“ sind demnach das Signifikanzniveau und der Referenzwert eines Qualitätsindikators (vgl. Kap. 5.5, S. 137). Vermisst werden Modellrechnungen hinsichtlich von Einrichtungen unterschiedlicher Größe. Insbesondere für kleine Einrichtungen bzw. Einrichtungen mit kleinen Fallzahlen – wie das auf Vertrags(zahn)arztpraxen zutreffen kann - sollte eher ein höheres Signifikanzniveau<sup>4</sup> z. B. von  $\alpha = 0,01$  gewählt werden.

Die KZBV wird zu diesem Kapitel keine weitere Stellung nehmen. Folgendes sei jedoch noch angemerkt: Dass die Ergebnisse der statistischen Auswertung als Information für Patientinnen und Patienten im Sinne einer erhöhten Transparenz zur Verfügung gestellt werden sollen, (IQTIG 2019, S. 71), lehnt die KZBV ab. Zum einen handelt es sich bei den statistischen Berechnungen zunächst um rechnerische und noch nicht abschließend um qualitative Auffälligkeiten. Erst, wenn der Bewertungsprozess durch das Lenkungsgremium (LAG) unter

---

<sup>4</sup> Die maximale Irrtumswahrscheinlichkeit beträgt 1%.

Einbezug der Fachkommissionen erfolgt ist, kann von qualitativen Auffälligkeiten gesprochen werden. Zum anderen erschließen sich die hoch komplexen statistischen Berechnungen nur, wenn genügend statistische Kenntnisse mitgebracht werden.

- Die KZBV fordert das IQTIG auf, die Patientinnen und Patienten als Adressatengruppe für die Ergebnisse der statistischen Auswertung zu streichen (vgl. S. 71).

## 6. Fachliche Bewertung (Kap. 6)

Die Standardisierung der fachlichen Bewertung erfolgt anhand der Definition eines Ausnahmetatbestandes, der die übergeordnete Zuordnung in die Kategorien „kein Hinweis auf ein Qualitätsdefizit“ (Kategorie U = unauffällig) und „Vorliegen eines Qualitätsdefizits“ (Kategorie A = auffällig) ermöglicht. Die präzise Zuordnung ergibt sich durch das Bewertungsschema, das in Tabelle 13 (IQTIG 2019, S. 145) abgebildet wird. Dieses soll dazu dienen, eine abschließende Einstufung der Ergebnisse nach dem Qualitätsbewertungsprozess auf Basis einer Stellungnahme durch den Leistungserbringer einzuordnen.

Die Einstufungen sind mit Ziffern weiter unterteilt (0 bis 3). Wird ein Leistungserbringer statistisch auffällig, beispielsweise über Qualitätsindikatoren mit Perzentilbereichen, so werden seine Ergebnisse, wenn es keinen Hinweis auf ein Qualitätsdefizit gibt, immer noch unter Ziffer 0 „kein hinreichender Hinweis auf ein Qualitätsdefizit“ eingestuft. Das weckt den Eindruck eines Generalverdachtes; ein „Quäntchen Qualitätsdefizit“ bleibt immer bestehen.

- Die KZBV plädiert für eine Umbenennung der Ziffer 0 in „kein Hinweis auf ein Qualitätsdefizit“.

Nach dem neuen Bewertungsschema (Tabelle 13) beabsichtigt das IQTIG, formale Qualitätsmängel wie „Verzicht auf Einreichung einer Stellungnahme“, „Stellungnahme nicht fristgerecht eingereicht“ oder „Stellungnahme entsprach nicht den Anforderungen“ den fachlich-inhaltlichen Qualitätsmängeln gleich zu setzen. So werden die formalen Qualitätsmängel grundsätzlich einem „Qualitätsdefizit“ zugeordnet (Kategorie A, Stufe 0). Dies ist aus Sicht der KZBV nicht sachgerecht und nicht hinnehmbar.

- Die KZBV fordert für „Nicht-Einreichen einer Stellungnahme“, „das nicht fristgerechte Einreichen einer Stellungnahme“ oder „Stellungnahme entsprach nicht den Anforderungen“ eine eigene Bewertungskategorie. Leistungserbringer sollten wenigstens an die Stellungnahme erinnert und aufgefordert werden, die Stellungnahme nachzureichen. An dieser Stelle ist das abgestufte Vorgehen gemäß QFD-Richtlinie zu beachten. Der Grundsatz der Verhältnismäßigkeit muss gewahrt bleiben. Die Vermischung von formalen und inhaltlichen Qualitätsdefiziten führt nur zu mangelnder Akzeptanz bei den Leistungserbringern und entbehrt jeder Grundlage.
- Nicht abgebildet wird in Tabelle 13, wie mit Stellungnahmen umgegangen wird, die sowohl Dokumentationsfehler als auch Ausnahmetatbestände angeben. Wie erfolgt hier die Endbewertung? Die KZBV bittet um entsprechende Vorschläge.

Nach Ansicht des IQTIG fehlt es noch an einer Methodik, um zwischen „Abwesenheit eines hinreichenden Hinweises auf ein Qualitätsdefizit“ und „Hinweis auf zureichende Qualität“ unterscheiden zu können. Vor dem Hintergrund der unterschiedlichen Verwendungszwecke (Versorgungssteuerung; Qualitätsförderung; Qualitätszu- und abschlüsse etc.) wird die Weiterentwicklung der biometrischen Methodik um diese Komponente empfohlen (IQTIG 2019, S. 146). So soll auch zwischen „zureichende Qualität“ und „besonders gute Qualität“ (IQTIG 2019, S. 146) unterschieden werden können.

- Die Entwicklung einer Methodik zur Feststellung von „Exzellenzqualität“ ist nicht Gegenstand der Beauftragung.

Für die fachliche Bewertung der Stellungnahmen schlägt das IQTIG ein kombiniertes Vorgehen vor, das sowohl auf impliziten und expliziten Regeln basiert:

- a) Eine Bewertung, die auf impliziten Regeln basiert, d. h. individuelle Einschätzungen von Bewertenden gehen in die Gesamtbeurteilung mit ein.
- b) Bewertungen auf Basis von expliziten Regeln, sofern für Indikatorberechnungen eindeutige Einflussfaktoren vorab angegeben werden können. Indikatorwerte können dann partiell nachberechnet werden.

Aus Abbildung 31 (IQTIG 2019, S. 149) geht hervor, dass die zuständige Stelle (LQS, LAG) partielle Nachberechnungen durchführt und auch die heuristische Beurteilung für sonstige

Einflussfaktoren unter Hinzuziehung von Expertinnen und Experten durch die Fachkommission durchführt. Enthält eine Stellungnahme des Leistungserbringers beispielsweise über die in der Nachberechnung berücksichtigten Faktoren hinaus noch zusätzliche Einflüsse auf den Indikatorwert, obliegt es den Expertinnen und Experten der Fachkommission, diese qualitativ zu beurteilen.

Unterschieden wird zwischen Datenfehlern und fachlich-inhaltlichen Einflussfaktoren. Die Beurteilung, welche der in der Stellungnahme eines Leistungserbringers angeführten Datenfehler vom Leistungserbringer zu verantworten sind, soll die zuständige Stelle (LQS, LAG) treffen. Die Beurteilung, ob ein fachlich-inhaltlicher Einflussfaktor vorliegt, den der Leistungserbringer zu verantworten hat, soll grundsätzlich unter Beratung der zuständigen Stelle durch die Experten der Fachkommission erfolgen. Am Ende entsteht ein sehr detaillierter Bewertungsalgorithmus mit Binnendifferenzierungen, ob vom Leistungserbringer zu verantwortende Datenfehler eine Fehlmessung verursacht haben; ob ein Qualitätsdefizit vorliegt, der teilweise durch vom LE zu verantwortende Datenfehler bedingt ist oder ob der Hinweis auf ein Qualitätsdefizit zurückzuführen ist (IQTIG 2019, Abbildung 34, S. 159).

- Die KZBV hält den Prozess der Bewertung für sehr aufwändig und sehr komplex. Die etwaigen partiellen Nachberechnungen stellen sich äußerst anspruchsvoll dar. Die Aufgaben, die das IQTIG den LQS/LAGen und den Fachkommissionen zuweist, erscheinen jeweils schwierig voneinander abgrenzbar zu sein. Die Aufgabenzuweisung erfordert eine klare Moderation und Abgrenzung.

Im Kapitel 6.6 (IQTIG 2019, 166) hat das IQTIG Empfehlungen in Bezug auf die Anforderungen an die Zusammensetzung der Fachkommissionen und an die Expertinnen und Experten (z. B. Qualifikation, Erfassung und Umgang mit Interessenskonflikten und Befangenheiten) ausgesprochen. So plädiert das IQTIG für eine standardisierte Zusammensetzung von Fachkommissionen. Angeführt werden Kriterien für die Auswahl der Expertinnen und Experten (fachlich qualifiziert, unabhängig neutral und interdisziplinär zusammengestellt). Dies ist grundsätzlich sinnvoll und nachvollziehbar. Entscheidend für eine kompetente und auch akzeptable Beurteilung kann nur allein eine für das jeweilige Thema bestehende fachliche Kompetenz sein, die die Gegebenheiten vor Ort ebenfalls kennt. Auch müssen dabei stets die regionalen Gegebenheiten einbezogen werden, z. B. die Voraussetzungen bei kleinen Bundesländern.

Nicht nachvollziehbar ist, dass über die Beschreibung der erforderlichen Qualifikationen hinaus, die beteiligten Berufsgruppen vorgegeben werden: Fachärzte, Pflegekräfte, Patientenvertreter, weitere Angehörige der Gesundheitsfachberufe, Krankenhaushygieniker, assoziierte Mitglieder (z. B. medizinische Dokumentare; Expertinnen und Experten, die sich im Ruhestand befinden).

- Im Rahmen § 5 Abs. 5 der DeQS-RL ist die Zusammensetzung der Expertengremien bzw. der Fachkommissionen klar geregelt. Details werden in den Themenspezifischen Bestimmungen konkretisiert. Ausführungen zu den beteiligten Berufsgruppen zu machen (vgl. Kap. 6.6.2, IQTIG 2019, S. 169) ist nicht Beauftragungsgegenstand.

## 7. Modul Qualitätsförderung (Kap. 7)

Unter der Überschrift „Empfehlungen zur Einleitung qualitätsfördernder Maßnahmen“ (Kap. 7.1) empfiehlt das IQTIG, über die Vorgaben der DeQS-RL und der QFD-RL hinaus die Möglichkeit einzuräumen, „qualitätsfördernde Maßnahmen auch ohne identifiziertes Defizit einzuleiten, wenn diese vom Leistungserbringer oder von der zuständigen Stelle für notwendig und sinnvoll erachtet werden, um eine weitere Verbesserung der Qualität zu erreichen.“ (IQTIG 2019, S. 174).

- Qualitätsfördernde Maßnahmen können jederzeit vom Leistungserbringer selbst aus intrinsischer Motivation heraus durchgeführt oder in Anspruch genommen werden. Ohne Qualitätsdefizit besteht seitens der zuständigen Stelle keine Rechtsgrundlage, etwaige Qualitätsförderungsmaßnahmen einzuleiten. Die KZBV bittet das IQTIG, diesen Passus entsprechend zu korrigieren.

Weiter heißt es in diesem Kapitel: „Der zuständigen Stelle und der Fachkommission kommt damit hinsichtlich der Qualitätsförderung eine entscheidende Rolle zu. (...) Eine Voraussetzung dafür ist der direkte Kontakt zwischen der zuständigen Stelle, der Fachkommission und dem Leistungserbringer.“ (IQTIG 2019, S. 174).

- An keiner Stelle in diesem Kapitel geht das IQTIG darauf ein, dass die zuständige Stelle von der zuständigen KV/KZV oder durch die Krankenkassen unterstützt werden

kann (vgl. § 17 Abs. 1 Satz 4; Abs. 4 und Abs. 7 DeQS-RL). Die KZBV fordert das IQTIG auf, in diesem Kapitel eine entsprechende Korrektur dahingehend vorzunehmen, dass die KVen/KZVen bei der Umsetzung von Qualitätsförderungsmaßnahmen eine aktive, originäre Aufgabe übernehmen. Die KVen/KZVen haben per se einen unmittelbaren, direkten Kontakt zum Leistungserbringer und sind in der Lage, die Vertrags(zahn)ärzte bei der Qualitätsverbesserung adäquat zu unterstützen.

Im Kapitel 7.2 greift das IQTIG als mögliche Maßnahme zur Qualitätsförderung (aus der QFD-RL) die „Zielvereinbarungen“ heraus und beschreibt diese detailliert hinsichtlich der Umsetzung. Bezug genommen wird hierbei auf ein Zielvereinbarungskonzept aus dem Jahr 2018, das für bundesbezogene Verfahren dem G-BA vorgelegt worden sei. Nicht alle Stellungnahmeberechtigten kennen dieses Konzept, da es sich ausschließlich um einen Konzeptvorschlag im stationären Sektor handelt.

- Die KZBV bittet das IQTIG um eine Herausnahme dieses Kapitels aus dem Entwicklungsbericht. Die Formulierung von Zielvereinbarungen und Empfehlungen zum Controlling gehören nicht zur Beauftragung (weder zum Auftrag gemäß QSKH-RL vom 18.01.2018 noch zum Auftrag gemäß Qesü-RL vom 17.05.2018).

## 8. Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens (Kap. 8)

Gemäß Auftrag vom 18.01.2018 (Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern) sollte das IQTIG prüfen, ob es Optionen zur Verkürzung des Verfahrens gibt (IQTIG 2019, S. 177 ff.). Das IQTIG stellt in diesem Kapitel verschiedene Optionen vor.

Der Vorschlag, sukzessive die Anzahl an Qualitätsindikatoren anhand von Eignungskriterien zu prüfen und ggf. die Aussetzung von Indikatoren zu empfehlen, wird seitens der KZBV hinsichtlich der Verfahrensökonomie sehr begrüßt.

- Die KZBV plädiert für überschaubare QS-Verfahren mit kleinen und ggf. wechselnden QI-Sets.

Der Vorschlag, dass die zuständigen Stellen bei Nicht-Einhaltung der formalen Anforderungen der Stellungnahmen direkt eine abschließende Bewertung ohne weitere Einholung von

Präzisierungen vornehmen sollen (IQTIG 2019, S. 180), kann von der KZBV nicht geteilt werden.

- Die KZBV fordert, dass für die Einholung der Stellungnahmen genügend Zeit eingeplant wird, so dass nicht erfüllte Anforderungen korrigiert werden können. Wenn ein Leistungserbringer zum Beispiel aus Krankheits- oder Urlaubsgründen im genannten Zeitraum nicht in der Lage ist, eine Stellungnahme entsprechend den Anforderungen abzugeben, muss er eine Erinnerung mit einer Nachfrist erhalten. Gleiches gilt, wenn die Anforderungen an die Stellungnahmen nicht ad hoc erfüllt werden.

Das IQTIG schlägt vor, dass eine IT-Unterstützung den Stellungnahmeprozess verkürzen kann (eine dem Bewertungsprozess angepasste Softwarelösung, welche einfache und komplexe Schritte - wie z. B. den Zugriff auf die Stellungnahmen und die Nachberechnung von Ergebnissen - für die Anwender erleichtert). Vor dem Hintergrund der Einrichtung Mandantenfähiger Datenbanken, ist dies ein wichtiges Thema. Der Vorschlag sollte aus KZBV-Sicht weiterverfolgt werden.

- Das IQTIG wird gebeten, für eine IT-Unterstützung entsprechende Vorschläge zu unterbreiten.

Die Empfehlung des IQTIG, den Prozess dadurch zu verkürzen, dass die Einholung der Stellungnahmen ohne Entscheidung der Fachkommissionen durch die zuständigen Stellen erfolgt (IQTIG 2019, S. 181), kann von der KZBV nicht mitgetragen werden.

- Die KZBV verweist auf bestehende Richtlinienregelungen. Die Aufgaben und die Kompetenz der Fachkommissionen müssen erhalten bleiben (§ 17 Abs. 1 DeQS-RL).

Eine Reduktion des Zeitraumes im Modul der Qualitätsbewertung (Stellungnahme-Verfahren) von ursprünglich 4 auf knapp 3 Monate erscheint der KZBV eher unrealistisch und ist abzulehnen.

- Es ist zu berücksichtigen, dass das Stellungnahmeverfahren für alle Beteiligten in die Haupturlaubszeit im Sommer hineinfällt (Juli bis September). Somit ist sowohl auf

Seiten der LQS/LAG, der Fachkommissionen und auch auf Seiten der Leistungserbringer mit Zeitverzögerungen im Verfahren zu rechnen. Soweit Aufforderungen zur Stellungnahme mit der Post versendet werden, müssen auch hierfür Bearbeitungstage einkalkuliert werden.

Abschließend schlägt das IQTIG unter Einbeziehung der Rückmeldungen der LQS/LAG-Geschäftsstellen ein Konzept zur Vereinfachung der Qualitätsbewertung vor. Momentan sieht die DeQS-RL vor, dass die LAGen in einem Prozess wechselnder Zuständigkeiten mit der Fachkommission die Bewertung der auffälligen Indikatorergebnisse vornehmen und Empfehlungen für qualitätsfördernde Maßnahmen abgeben (IQTIG 2019, S. 183 ff). Der Prozess könnte abgekürzt werden, wenn das IQTIG die Bewertung der Ergebnisse (Fertigstellung Auswertung) übernimmt und den LAGen zur Verfügung stellt. „Die Aufforderungsschreiben für die auffälligen Indikatorergebnisse könnten quasi zeitgleich mit Veröffentlichung der finalen Rechenregeln und der Übermittlung der Auswertungen an die Leistungserbringer versendet werden, falls entsprechende IT-technische Strukturen aufgebaut werden würden.“, so das IQTIG (2019, S. 185).

- Die KZBV sieht alleine schon aus datenschutzrechtlichen Gründen keine andere Möglichkeit, als dass die Bewertung der Ergebnisse und die Aufforderung zur Stellungnahme seitens der LQS/LAG erbracht werden. Das IQTIG hat keine Befugnis, leistungserbringeridentifizierende Daten einzusehen. Daher kommt dieser Vorschlag eindeutig nicht in Betracht.

In **Kapitel 7 des Anhangs** zum Vorbericht macht das IQTIG detaillierte Textvorschläge zur Einholung von Stellungnahmen (Formulierung von Anforderungen).

Im ersten Absatz findet sich sogleich der Hinweis: „Erreicht uns innerhalb der Frist keine den untenstehenden Kriterien entsprechende Stellungnahme, wird das Ergebnis Ihrer Einrichtung als ‚Qualitätsdefizit‘ bewertet“.

- Dieser Satz ist zwingend zu streichen. Die KZBV trägt nicht mit, dass eine nicht fristgerecht abgegebene Stellungnahme sofort zu einem „Qualitätsdefizit“ führt. Hier ist mindestens ein Erinnerungsschreiben mit einer Fristverlängerung vorzusehen.

Die Stellungnahme soll von den Leistungserbringern in digitaler Form eingereicht werden.

- Dies ist ggf. eine Hürde, die einer Grundlage entbehrt. Die Aufforderung müsste lauten: „möglichst in digitaler Form“. Im Fall der digitalen Übersendung müssten zudem entsprechende Verschlüsselungen aus Gründen des Datenschutzes angewandt werden. Die Voraussetzungen zur Verschlüsselung müssten zur Verfügung gestellt werden.

Für jedes auffällige Indikatorergebnis soll eine separate Stellungnahme erfolgen.

- Was versteht das IQTIG unter einer „separaten Stellungnahme“? Die KZBV geht davon aus, dass zu jedem einzelnen Indikator einzeln Stellung genommen wird, dass aber nur eine Stellungnahme abzugeben ist. Unnötige Bürokratie ist zu vermeiden.

Der Spiegelstrich „Die fachlich verantwortliche Instanz Ihrer Einrichtung ist über den indikatorbasierten Hinweis auf ein Qualitätsdefizit in Ihrer Einrichtung zu informieren und hat die angefertigte Stellungnahme zu autorisieren. Bitte bestätigen Sie in Ihrer Stellungnahme, dass dies geschehen ist.“ scheint sich auf den stationären Sektor zu beziehen und ist als solcher zu kennzeichnen.

- Diese Aufforderung sollte daher zwingend umformuliert werden. Unklar ist zum einen, wer ist die „fachlich verantwortliche Instanz“? Im ambulanten Bereich gibt die ärztliche Leitung der Einrichtung die Stellungnahme ab und trägt die Verantwortung. Daher ist es ausreichend, wenn im Anschreiben steht, „Die Stellungnahme ist durch die ärztliche Leitung der Einrichtung zu unterschreiben.“

Bei den weiteren Spiegelstrichen „Bei einem auffälligen Ergebnis in einem risikoadjustierten Prozessindikator“ und „Bei einem auffälligen Ergebnis in einem risikoadjustierten Ergebnisindikator“ werden die Leistungserbringer aufgefordert, Gründe anzuführen, warum trotz Risikoadjustierung das Ergebnis außerhalb des Referenzbereiches liegt und welche Einflussfaktoren bei den Patientinnen und Patienten so häufig auftraten, dass diese nicht mehr adäquat durch die Risikoadjustierung berücksichtigt werden konnten.

- Im Stellungnahmeverfahren können Risikofaktoren als Ausnahmetatbestand geltend gemacht werden, die zuvor bei der Risikoadjustierung nicht berücksichtigt wurden. Dies ist grundsätzlich zu begrüßen, setzt allerdings voraus, dass der Datenschutz durch zusätzliche Angaben gewahrt bleibt. Die Anonymität der Patientinnen und Patienten ist auch im Stellungnahmeprozess sicherzustellen.

#### 9. Berichterstattung (Kap. 9)

In den §§ 10, 19 und 20 DeQS-RL werden die Berichtsformate klar vorgegeben. Das IQTIG spricht sich in dem Bericht für veränderte Fristen aus. Das Modul „Qualitätsförderung“ müsse für die Berichterstattung noch im Detail entwickelt werden. Bis Rückmeldungen der Einrichtungen zu den mehrstufigen Maßnahmen und Ergebnissen der Qualitätsförderung vorliegen, würden mehrere Jahre vergehen. Daher sollen Details erst zu einem späteren Zeitpunkt entwickelt werden. Voraussetzung sei auch, dass das hier vorgestellte neue Rahmenkonzept zur Qualitätsbewertung und –förderung durch den G-BA beschlossen wurde. Insofern ist eine Stellungnahme zu diesem Punkt entbehrlich.

#### 10. Empfehlungen zur Evaluation des neuen Konzepts (Kap. 10)

In Kapitel 10 werden verschiedene Optionen hinsichtlich ihrer Eignung für eine mögliche Evaluation des neuen Bewertungskonzepts auf abstrakter Ebene erörtert. Zu diesem Kapitel erfolgt keine schriftliche Stellungnahme durch die KZBV.



**Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV**

## **Stellungnahme der maßgeblichen Organisationen nach § 140f SGB V zum Vorbericht des IQTIG:**

### **Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern**

Vorbericht zu Stufe 1 und Stufe 2

(Stand: 10. Oktober 2019)

**21.11.2019**

#### **Ansprechpartner/in:**

Cordula Mühr, DBR / Sozialverband Deutschland e.V. (SOVD)

[cordulamuehr@yahoo.de](mailto:cordulamuehr@yahoo.de)

Herbert Weisbrod-Frey, DBR / Sozialverband VdK Deutschland e.V. (VdK)

[herbert.weisbrod-frey@web.de](mailto:herbert.weisbrod-frey@web.de)



## Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV

### Inhaltsverzeichnis

1. Generelle Kommentare .....	3
2. Spezifische Kommentare.....	4
Zu Sensitivität und Spezifität des SD (S.21).....	4
Zu den veränderten Anforderungen an die Qualitätsmessung und -bewertung (S.22) .....	5
Zu den strukturellen Voraussetzungen des Strukturierten Dialogs (S. 24 ff).....	6
Zum Einstieg in den Strukturierten Dialog (S. 26 und Anhang 3 + 4).....	6
Zur Bewertung der Versorgungsqualität im Strukturierten Dialog (S. 27 ff) .....	7
Qualitative Auffälligkeiten je Qualitätsindikator (u.a. S. 41 ff) .....	11
Zu Berichtserstattung (u.a. Kapitel 9).....	12
3. Zusammenfassung.....	12

Hinweis: Die im Text aufgeführten Seitenzahlen, Tabellen-Angaben und wörtliche Zitate beziehen sich, so weit nicht anders gekennzeichnet, auf den Vorbericht des IQTIG Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern Vorbericht zu Stufe 1 und Stufe 2 (Stand: 10.10.2019).

## Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV

### 1. Generelle Kommentare

Vor dem Hintergrund der derzeitigen Heterogenität in der Durchführung des Strukturierten Dialogs (SD) in den einzelnen Bundesländern unterstützt die Patientenvertretung (PatV) ausdrücklich das Ziel eines einheitlichen, standardisierten Vorgehens aller beauftragten Stellen bei der Auswertung und Bewertung von Indikatorergebnissen. Im vorgelegten Bericht stellt das IQTiG anhand der aktuellen Vorgehensweise gemäß QSKH-RL die Ausgangssituation und den daraus abgeleiteten Weiterentwicklungsbedarf umfassend dar. Obwohl sich die Empfehlungen laut Beauftragung auf ein richtlinienunabhängiges Stellungnahmeverfahren beziehen sollen ist es akzeptabel, als Ausgangssituation die Regelungen und Vorgehensweisen der QSKH-RL zu wählen, da hier die meisten Erfahrungen mit dem Strukturierten Dialog vorliegen und dessen Funktionen in ähnlicher Form in dem Stellungnahmeverfahren der DeQS-RL vorgesehen sind. Aus Sicht der Patientenvertretung bestehen auch im Stellungnahmeverfahren der DeQS-RL vergleichbare „Gefahrenquellen“ für Heterogenität.

Wie vom IQTiG vorgeschlagen muss auch aus Sicht der PatV in den der QSKH nachfolgende Richtlinien ein einheitliches, richtlinienübergreifendes Vorgehen implementiert werden.

Die PatV unterstützt auch ausdrücklich das vom IQTiG als zentral beschriebene Ziel, durch eine Weiterentwicklung des SD und damit aller QS-Verfahren eine Steigerung der realen Versorgungsqualität zu erreichen.

Hinsichtlich des Zieles der „Steigerung der Effizienz des Strukturierten Dialogs“, finden sich im Anhang des Berichtes (Protokollierung des Workshops mit LQS/LAG) Aussagen des IQTiGs, dass zur Steigerung der Effizienz die Treffergenauigkeit der Indikatoren entscheidend sei und dass diese derzeit noch nicht umfänglich gegeben sei. Angesichts

## Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV

der teilweise massiven inhaltlichen Kritik der Workshop-Teilnehmer an den aktuellen Indikatoren (der das IQTiG teilweise zustimmt<sup>1</sup>) ist die Frage berechtigt, ob nicht als erster Schritt die angekündigte Verbesserung der Indikatoren hätte erfolgen müssen, - sowohl im Vorfeld der im Bericht dargestellten empirischen Überprüfung ihrer Bewertung als auch im Vorfeld der im Bericht vorgelegten Vorschläge zur Weiterentwicklung des SD.

Der vorgelegte Bericht „(...) beinhaltet konkrete Empfehlungen für die Auswertung und Bewertung von leistungserbringerbezogenen Qualitätsindikatorergebnissen auf Basis von Dokumentationsdaten der Leistungserbringer, auf Basis von Sozialdaten sowie auf Basis durch Patientenbefragungen erhobener Daten.“(S.15). Diese Herangehensweise des IQTiG wird von der PatV ausdrücklich unterstützt, da sie von der gleichberechtigten Wertigkeit aller normativ geregelten Datenquellen in den Verfahren der datengestützten Qualitätssicherung ausgeht.

## 2. Spezifische Kommentare

### Zu Sensitivität und Spezifität des SD (S.21)

*„Ein auffälliges Indikatorergebnis (Ergebnis erster Prüfschritt) kann im Rahmen des Strukturierten Dialogs (zweiter Prüfschritt) zwar widerlegt werden, bei einem unauffälligen Indikatorergebnis kann dieses jedoch nicht widerlegt werden, da nur auffällige Ergebnisse einem zweiten Prüfschritt unterzogen werden.“*

Aus dieser sehr zutreffenden Aussage können bzw. müssen die Empfehlungen des IQTiG in Kapitel 6 für ein neues Bewertungsschema abgeleitet werden (statt „qualitativ auffällig“ und „qualitativ unauffällig“ neu: „Qualitätsdefizit“ / „kein Hinweis auf Qualitätsdefizit“). Die tabellarische Darstellung von Sensitivität und Spezifität des Gesamtverfahrens zur Qualitätsmessung, bezogen auf einen Qualitätsindikator, veranschaulicht bildlich, warum allein

---

<sup>1</sup> siehe dazu S. 10 in Anlage 1: „Es wird zudem erläutert, dass es notwendig ist, ein Konzept zur Weiterentwicklung des Strukturierten Dialogs auf der Basis geeigneter Qualitätsindikatoren (siehe Eignungskriterien von Qualitätsindikatoren, Methodische Grundlagen des IQTiG 2019) zu erstellen und nicht auf der Basis ungeeigneter Indikatoren.“)

## Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV

das Fehlen von Hinweisen auf ein Qualitätsdefizit nicht erlaubt, ein Ergebnis als gute Qualität zu bezeichnen.

### Zu den veränderten Anforderungen an die Qualitätsmessung und -bewertung (S.22)

Hier werden zutreffend die veränderten Anforderungen an die Qualitätsmessung und -bewertung beschrieben und eingeordnet. Anstelle des ehemaligen "Lernen von den Besseren in einem geschützten Raum" stehen jetzt Transparenz, öffentliche Berichterstattung und Steuerungsfunktion von Versorgungsqualität inkl. Vergütungs- und Planungsrelevanz im Vordergrund (weniger „improvement“ - z. B. Peer-Reviews - stattdessen mehr „accountability“). Mit diesen neuen Zweckbindungen steigen die methodischen Anforderungen an die Aussagekraft von Indikatorergebnissen (weg von reinen Aufgreifkriterien hin zu quantitativen Größen für Qualität).

Interessant ist an dieser Stelle der Ausblick des IQTiG auf die mögliche methodische Weiterentwicklung im Zusammenhang mit der Beauftragung zum Qualitätsportal: können metrische Ergebnisse mehrerer Indikatoren verdichtet werden? Das IQTiG weist selbst darauf hin, dass die Aggregation von Indikatorergebnissen voraussetzt, dass die einzelnen Indikatorergebnisse mehr als bloße Aufgreifkriterien darstellen und ihre numerischen Ergebnisse selbst eine Aussagekraft haben müssen. Eine Aggregation würde somit voraussetzen, dass die vom IQTiG angekündigte „*umfassende Prüfung aller Qualitätsindikatoren anhand von Eignungskriterien*“ gemäß Methodischer Grundlagen des IQTiG 2019 bereits erfolgt ist. Solange der Nachweis durch das IQTiG jedoch nicht erbracht wurde, dass numerische Ergebnisse einzelner Indikatoren selbst eine Aussagekraft haben, ihre „Trefferquote“ also deutlich erhöht wurde, ist eine solche Aggregation von Indikatorergebnissen auf Basis der gegebenen Indikatoren nicht angemessen. Ebenfalls obsolet ist es nach dem Verständnis der Patientenvertretung, unter diesen Voraussetzungen den Prozess der Qualitätsmessung mittels automatisierbarer statistischer Methoden zu standardisieren und auf diesem Weg die derzeit tatsächlich z.T. aufwendigen dialogischen Verfahrensanteile verzichtbar machen zu wollen.

## **Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV**

### **Zu den strukturellen Voraussetzungen des Strukturierten Dialogs (S. 24 ff)**

Die Heterogenität der strukturellen Voraussetzungen des SD gemäß QSKH-RL und die fehlenden bzw. unkonkreten Regelungen, - z.B. hinsichtlich der Qualifikation von Fachgruppenmitgliedern - werden im Bericht zutreffend dargestellt.

Leider nicht eingegangen wird in diesem Zusammenhang jedoch auf die organisatorische Anbindung der LQSen an die Landes-Krankenhausgesellschaften bzw. Landes-Ärztekammern und die damit möglicherweise verbundenen Interessenkonflikte sowie - in die Zukunft gerichtet - auf die notwendigen strukturellen (z.B. vertraglichen) Voraussetzungen zur geforderten Neutralität der LAG-Geschäftsstellen.

### **Zum Einstieg in den Strukturierten Dialog (S. 26 und Anhang 3 + 4)**

Bereits bezüglich des relevanten ersten Schrittes, also dem Einstieg in den SD, ist gemäß QSKH-RL große Heterogenität gegeben. So sind bei rechnerisch auffälligen Ergebnissen grundsätzlich zwei nicht operationalisierte Handlungsoptionen erlaubt: die Versendung von Hinweisen oder die Aufforderung zur Stellungnahme. Regularien dafür, anhand welcher Kriterien diese erste wichtige Entscheidung zu treffen ist, sind in der Richtlinie nicht normiert. Das IQTiG vermutet, dass mögliche, derzeit von den LQSen angewendete Kriterien die Relevanz des Indikators, die statistischen Signifikanz von Abweichungen und das Vorliegen einer Risikoadjustierung sind.

Als wesentlicher weiterer Hintergrund von Heterogenität kommt hinzu, dass den LQSen ermöglicht wird, landesindividuelle Qualitätsindikatorensätze einzusetzen oder sogar landesindividuelle Referenz- und Vertrauensbereichen anzuwenden. Vor diesem Hintergrund verwundern die Ergebnisse der Analysen des IQTiG in Anhang 3 und 4 nicht, dass die Unterschiedlichkeit der Vorgehensweise der LQSen bereits bei diesem ersten Schritt, der Auslösung des SD, signifikant ist und maßgeblich von den Bewertenden

Stellen abhängt. Die Patientenvertretung bedankt sich an dieser Stelle für diesen letzten Beweis dafür, wie dringlich die Vereinheitlichung der Vorgehensweisen ist, wenn Transparenz und faire Leistungserbringervergeiche das Ziel von QS-Verfahren sein sollen.

### Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV

Gemäß QSKH-RL können sich rechnerisch auffällige Ergebnisse entweder auf einen Qualitätsindikator oder auf ein Auffälligkeitskriterium beziehen, weswegen auch die Bewertung der Ergebnisse jeweils entweder den Fokus „Versorgungsqualität“ oder den Fokus „Dokumentationsqualität“ hat. Das IQTiG will im vorliegenden Bericht bzgl. seiner Empfehlungen leider nicht auf den Problemkreis „Dokumentationsqualität“ eingehen. Dies wird von der PatV kritisch gesehen, zumal dem Bericht gleichzeitig zu entnehmen ist, dass z.B. für das EJ 2015 bundesweit durchschnittlich 10,4 % der rechnerisch auffälligen Indikatorergebnisse aufgrund nachträglicher Angaben von Dokumentationsfehlern nicht bewertet werden konnten und insgesamt 17,9 % aller rechnerisch auffälligen Ergebnisse aufgrund von nachträglich angegebenen Dokumentationsproblemen entweder indirekt oder direkt zu einer Entlastung des Leistungserbringers führten. Vor diesem Hintergrund hat die PatV erwartet, dass das IQTiG in seinem vorgelegten Bericht auch Vorschläge zum Umgang mit mangelnder Dokumentations- und Datenfeldqualität macht. Dies umso mehr, als das IQTiG mangelnde Dokumentations- und Datenfeldqualität als wesentliche Ursache für eine schlechte „Trefferquote“ von Indikatoren und somit als eine relevante Ursache der Ineffektivität des SD identifiziert hat (siehe dazu ebenfalls 2.6.2 und 2.6.3).

### Zur Bewertung der Versorgungsqualität im Strukturierten Dialog (S. 27 ff)

Im Zusammenhang mit den Ergebnissen der rein stationären QS gemäß QSKH-RL wird gemäß vorliegendem Bericht bundesweit im Durchschnitt für mehr als ein Drittel aller rechnerisch auffälligen Indikatorergebnisse lediglich ein Hinweis an den Leistungserbringer versendet. Weil dies bedeutet, dass in vielen Bundesländern für eine große Anzahl an rechnerisch auffälligen Ergebnissen keine Qualitätsbewertung vorgenommen wird, ist es aus Sicht der PatV überfällig, dass das Versenden von Hinweisen als „Maßnahme“ in den nachfolgenden Richtlinien nicht mehr vorzusehen ist. Vor diesem Hintergrund begrüßt die Patientenvertretung grundsätzlich den Vorschlag des IQTiG, durch die statistische Operationalisierung eines „hinreichenden Hinweises auf ein Qualitätsdefizit“ das Versenden von Hinweisen überflüssig zu machen, obschon die PatV sich nicht in der Lage sah, die vorgeschlagene statistische Methodik nachzuvollziehen. In jedem Fall bedarf die tatsächliche Treffsicherheit der vorgeschlagenen statistischen Operationalisierung eines

### Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV

solchen „hinreichenden Hinweises“ aus Sicht der PatV einer nachfolgenden empirischen Überprüfung.

Weil in der QSKH-RL weder Form noch Inhalt von Stellungnahmen definiert werden begrüßt die PatV ausdrücklich die Textvorschläge des IQTiGs zur standardisierten Einholung von Stellungnahmen (Anhang 7) als auch den Vorschlag, die Methodik des Stellungnahmeverfahrens im Internet zu veröffentlichen und Leistungserbringern im Anschreiben zur Aufforderung einer Stellungnahme den Link mitzuteilen. Dies schafft größtmögliche Transparenz und ggf. Akzeptanz des Verfahrens bei allen Beteiligten und entspricht damit dem Ziel der Beauftragung zur Erhöhung der Transparenz und Steigerung der Effizienz.

Im Kapitel 3 (S. 65) schlägt das IQTiG vor, dass die qualitative Beurteilung im Stellungnahmeverfahren „...nicht als ein eigenständiges Bewertungsinstrument zu verstehen (ist), sondern in der Prüfung der Validität der Qualitätsaussage eines Qualitätsindikators in dem speziellen Fall eines Leistungserbringers (besteht), der auffällig wurde (...)" " „Die Vorteile einer standardisierten Erhebung und Bewertung durch Qualitätsindikatoren gingen verloren bzw. wären nicht erforderlich, wenn die Qualitätsindikatoren lediglich als Aufgreifkriterien zur Ermittlung auffälliger Leistungserbringer genutzt würden, das Stellungnahmeverfahren aber eine eigene, vom Qualitätsziel des Indikators unabhängige Bewertung vornähme.“ Diese Bewertung des IQTiG ist nachvollziehbar. In der Konsequenz bedeutet der auf den Seiten 149 und Seite 64 im Anhang vorgeschlagene Ablauf der fachlichen Bewertung von Stellungnahmen eine Annäherung an das Vorgehen beim PlanQI-Verfahren. Dies wird von der Patientenvertretung ausdrücklich unterstützt, weil es zu einer Steigerung der Objektivität im Verfahren beiträgt. Dementsprechend ist es folgerichtig, dass - wie beim QS-Verfahren PlanQI - Nachberechnungen des Indikatorwertes pro Leistungserbringer durchgeführt werden, falls geeignete Einflussfaktoren im Stellungnahmeverfahren geltend gemacht werden können, und dass erst im Anschluss daran die abschließende Bewertung der Qualität vorgenommen wird. Im Anhang 5 zum Bericht werden nachvollziehbar Beispielfälle für die Qualitätsbewertung im Stellungnahmeverfahren beschrieben, offenbar in Analogie zur Definition von „Ausnahmetatbeständen“ (ATB) im PlanQI-Verfahren. Solche iterativ weiterentwickelten und dadurch zunehmend standardi-

### Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV

sierten Definitionen von ATB sind aus Sicht der PatV auch bei den indirekten QS-Verfahren prinzipiell ein richtiger Weg zur Vereinheitlichung der Bewertungen im SD. Es stellt sich aber die Frage, wie dies angesichts von dezentralen Landesgremien organisiert umgesetzt werden kann.

Auf S.148 analysiert das IQTiG nachvollziehbar, warum eine vollständige, explizite und quantitative Bewertung der in den Stellungnahmen geltend gemachten Einflussfaktoren nicht möglich ist und schlägt daher ein kombiniertes Vorgehen bei der Qualitätsbewertung vor: 1. Heuristische Beurteilung anhand impliziter Regeln durch Experten und 2. parallele partielle Nachberechnung. Erst danach soll sich der Bewertungsalgorithmus - dargestellt auf S.151 - anschließen, der für die PatV verständlich abgebildet ist.

Gemäß Empfehlung des IQTiG soll weiterhin eine Bewertungskategorie „Sonstiges“ beibehalten werden z.B. wenn vom Leistungserbringer unverschuldet kein Stellungnahmeverfahren durchgeführt werden konnte. Die Hintergründe dieser Kategorisierung müssen bei den kommenden Berichten geprüft werden – auch das Gremium zur Systempflege sollte eine Auflistung der Gründe für diese Einstufung erhalten.

Auf Seite 158 empfiehlt das IQTiG als Interimslösung, perzentilbasierte Referenzbereiche im Stellungnahmeverfahren wie feste Referenzbereiche zu behandeln, fordert jedoch perspektivisch die Festlegung fester, kriteriumsbezogener Referenzbereiche für alle Indikatoren, weil nur so Klarheit darüber geschaffen wird, welche Qualitätsstandards von den Leistungserbringern gefordert werden können. Eine solche Weiterentwicklung unterstützt die PatV nachdrücklich, weil perzentilbasierte Referenzbereiche immer allein abhängig von den Ergebnissen der Grundgesamtheit sind und damit nicht geeignet, um zuverlässigen Aussage zur Qualität machen zu können.

Auf Seite 72 schlägt das IQTiG vor, dass die zuständige Stelle „*ab dem zweiten Qualitätsdefizit bei einem Indikatorergebnis oder bei inhaltlich ähnlichen Indikatoren innerhalb von drei Jahren immer qualitätsfördernde Maßnahmen einleitet*“ (siehe auch Anhang 1). Aus Sicht der PatV ist die Konkretisierung des IQTiG unbedingt zu unterstützen, dass die zuständige Stelle bereits bei einem erstmalig auftretenden Qualitätsdefizit unter Angabe

### Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV

von Gründen jederzeit in einem Qualitätsindikator qualitätsfördernde Maßnahmen bei einem Leistungserbringer einleiten“ kann. Die Formulierung „*innerhalb von 3 Jahren immer.*“ ist hingegen zu unkonkret und lässt den Akteuren auf Landesebene zu großen Handlungsspielraum.

Mehr am Rande macht das IQTiG in seinem Bericht darauf aufmerksam, dass auch Unterschiede bzgl. des Umfangs der Einbeziehung ihrer Fachgruppen in die Bewertung der schriftlichen Stellungnahme durch die Landesgeschäftsstellen bekannt seien. Diesen Hinweis hält die Patientenvertretung für sehr relevant, lässt sich doch ggf. ein Teil der Heterogenität von Bewertungen auch hierdurch erklären. Leider fehlen jedoch im Bericht Empfehlungen des IQTiG dazu, auf welchem Weg auch hier - in die Zukunft gerichtet - eine Vereinheitlichung erreicht werden könnte.

Obwohl im Bericht dargestellt wird, dass die Entscheidungsfindung in den Fachkommissionen respektive in den LQSen (das wird leider nicht differenziert untersucht, auch nicht, ob mit oder ohne Patientenbeteiligung) auf unterschiedlichem Wege zustande kommen, nämlich mal im Konsens, mal durch formale Abstimmungen, mal durch eine Mischung aus Beidem, macht das IQTiG in seinem Bericht leider keine Vorschläge dazu, wie - in die Zukunft gerichtet - hier eine sinnvolle Vereinheitlichung erfolgen sollte.

Die PatV unterstützt ausdrücklich den Vorschlag des IQTiG, dass die Aufgaben der Fachkommissionen darin bestehen sollen, Ergebnisse aus den Stellungnahmen fachlich zu beurteilen und - als Folge des Bewertungsprozesses - dem IQTiG gegenüber ihre Empfehlungen zu qualitätsfördernden Maßnahmen zu geben. Aus Sicht der PatV sollte das IQTiG - unter Berücksichtigung dieser - ggf. auch divergierenderen - Bewertungen der Fachkommissionsmitglieder auf Basis seiner fachlich / wissenschaftlichen Expertise und gebotenen Neutralität die Empfehlungen der Fachkommissionen prüfen, die abschließende Bewertung vornehmen, daraus Maßnahme-Empfehlungen ableiten und den maßgeblichen Stellen zur Umsetzung empfehlen. Leider fehlen im vorliegenden Bericht die Empfehlungen des IQTiG zur - in die Zukunft gerichteten - Funktionszuweisung für Geschäftsstelle und Lenkungsgremium der LAGen im Zusammenhang mit diesen Bewertungs- und Entscheidungsprozessen. Aus Sicht der PatV kann unter

### **Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV**

ordnungspolitischen Gesichtspunkten das Lenkungsgremium der LAG - als originäres Gremium der Trägerorganisationen - nicht über Maßnahmen gegenüber den von ihnen zu vertretenden Mitgliedern entscheiden. Diese Funktion sollte allein den neutralen Geschäftsstellen überlassen bleiben. In Abgrenzung dazu sollten die Lenkungsgremien der LAGen dem G-BA gegenüber lediglich die Verantwortung über die richtliniengemäße Einhaltung des Gesamtprozesses tragen, nicht die für inhaltliche Bewertungen und Maßnahme-Entscheidungen. Andernfalls müsste normiert werden, dass Entscheidungen der Lenkungsgremien über Maßnahmen, die von den diesbezüglichen Empfehlungen des IQTiG abweichen, dem G-BA gegenüber berichtet und begründet werden müssen. Nur so wäre aus Sicht der PatV eine einheitliche und interessenunabhängige Umsetzung des Verfahrens denkbar.

Ausdrücklich begrüßt wird jedoch schon einmal die Empfehlung des IQTiG, dass für eine größere Akzeptanz des Verfahrens die interdisziplinäre Perspektive durch die Perspektive der Patientinnen und Patienten, die die eigentlichen Adressaten der externen Qualitätssicherung sind, ergänzt werden soll, was nur bedeuten kann, dass Vertreterinnen und Vertreter der Patientinnen und Patienten gleichwertig und auf Augenhöhe in alle Expertengremien aufgenommen werden müssen. Die Patientenvertretung bittet das IQTiG, bei seiner Überarbeitung des Berichts diesbezüglich seine in die Zukunft gerichteten, richtlinienübergreifende Empfehlungen auszuformulieren.

### **Qualitative Auffälligkeiten je Qualitätsindikator (u.a. S. 41 ff)**

Das IQTiG weist in seinem Bericht nach, dass der Anteil der qualitativen Auffälligkeiten an den rechnerischen Auffälligkeiten je nach Qualitätsindikator sehr unterschiedlich ist. Unter Kapitel 11.2 (Seite 212) wird dieser Sachstand auch noch mal im Zusammenhang mit möglichen Interessenkonflikten diskutiert. So könnte eine höhere „Spezifität“ vermeiden, dass einem Leistungserbringer irrtümlich unzureichende Qualität attestiert wird. Bei höher gewichteter Sensitivität des Verfahrens könnten Leistungserbringer mit „tatsächlich“ unzureichender Qualität besser identifiziert werden. Das IQTiG möchte aber erst im Rahmen einer neuen Beauftragung die Frage beantworten, welchen Gewinn das Stellungnahmeverfahren als nachgeschalteter Prüfschritt zur Sicherstellung der Aussagekraft

## Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV

eines Indikatorergebnisses darstellt. Es ist zu klären, ob und warum das nicht noch nachträglich im Rahmen dieses Auftrags erfolgen kann.

### Zu Berichtserstattung (u.a. Kapitel 9)

Der vom IQTiG vorgelegte Vorschlag zur Umbenennung der Begrifflichkeiten im Bewertungsschema (statt „*qualitativ auffällig*“ und „*qualitativ unauffällig*“ neu: „*Qualitätsdefizit*“ / „*kein Hinweis auf Qualitätsdefizit*“) wird von der PatV ausdrücklich begrüßt, weil so Klarheit darüber verschafft wird, welche Form der „Auffälligkeit“ vorliegt und auch der Öffentlichkeit gegenüber künftig nicht automatisch „gute Qualität“ beim Fehlen von Hinweisen auf Defizite bescheinigt wird.

Die Empfehlung des IQTiG, künftig eine Trennung der Qualitätsbewertung und der Qualitätsförderung vorzusehen, wird von der Patientenvertretung ausdrücklich begrüßt, weil so eine frühzeitigere Berichterstattung zu den Ergebnissen der Qualitätsbewertung (im EJ+1) ermöglicht wird. Die Berichterstattung zur Qualitätsförderung kann ohne Schaden erst nachgelagert erfolgen (im EJ+2.). Diese frühere Berichterstattung ist insbesondere für die Information und damit zur Unterstützung von Auswahlentscheidungen für Patient\*innen von großer Bedeutung. Darüber hinaus kann aber auch das interne Qualitätsmanagement auf Seiten der Leistungserbringer frühzeitiger zielgerichtete, qualitätsverbessernde Maßnahmen initiieren.

## 3. Zusammenfassung

Die Empfehlung des IQTiG, explizite Regeln für die Prozessschritte der Auslösung des Strukturierten Dialogs und für die nachfolgenden Qualitätsbewertungen zu normieren werden von der Patientenvertretung unterstützt, weil eine Standardisierung zur Verringerung der derzeitigen v.a. bewertungsstellenabhängigen Heterogenität führen wird.

Eine Steigerung der Komplexität des Verfahrens (3.5) ist vor diesem Hintergrund aus Sicht der Patientenvertretung in Kauf zu nehmen, wenn damit die Transparenz und Nachvollziehbarkeit des Verfahrens insgesamt erhöht und faire Leistungserbringervergleiche ermöglicht werden.

### **Stellungnahme der maßgeblichen Patientenorganisationen nach § 140 f SGBV**

Offen geblieben ist eine Konkretisierung des Begriffs „Zuständige Stelle“ vor dem Hintergrund der unterschiedlichen Richtlinien, sowie eine Konkretisierung der künftigen Funktionen von Lenkungs-gremien und neutralen Geschäftsstellen der LAGen.

Sowohl die im vorliegenden Bericht vorgeschlagene Trennung von Qualitätsbewertung und Qualitätsförderung, als auch der modulhafte Aufbau des Rahmenkonzeptes werden von der PatV vollumfänglich unterstützt. Besonders begrüßenswert erscheint dabei die frühzeitige Berichterstattung (EJ+1) und ein somit ermöglichtes schnelleres Aufgreifen des internen Qualitätsmanagements zur Einleitung qualitätsverbessernder Maßnahmen bei auffälligen Qualitätsergebnissen.

Die vom IQTiG mit seinen Vorschlägen angestrebte verbesserte Akzeptanz des Verfahrens kann aber letztlich nur dann maßgeblich erhöht werden, wenn die Treffsicherheit der einzelnen Indikatoren weiter erhöht wird. Dies könnte aus Sicht der PatV bereits zu einem erheblichen Teil ohne eine separate Beauftragung des IQTiG im Rahmen der regelmäßigen Systempflege der Verfahren erfolgen.

Davon unbenommen ist eine generelle Überprüfung aller Indikatoren und Verfahrensbereiche auf ihre Aktualität und Repräsentativität notwendig, verbunden mit einer Methodik zur Erarbeitung und Implementierung neuer Indikatoren und Leistungsbereiche in einem akzeptablen Zeitraum.

# ÄRZTEKAMMER SACHSEN-ANHALT

## Körperschaft des öffentlichen Rechts



Ärztekammer · Landesgeschäftsstelle · PSF 1561 · 39005 Magdeburg

### PROJEKTGESCHÄFTSSTELLE QUALITÄTSSICHERUNG

Bearbeiter : Frau Dr. Wolf  
Tel. : (03 91) 60 54-79 50  
FAX : (03 91) 60 54-79 51  
E-Mail : pgs-quali@aeksa.de

Magdeburg, 19.11.2019

Sehr geehrte Damen und Herren,

für die Möglichkeit einer Stellungnahme zum Vorbericht „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“ danken wir Ihnen.

Unsere Anmerkungen dazu sind wie folgt:

Für die Durchführung des Stellungnahmeverfahrens ist gem. Abschnitt 8.1 ein verkürzter Zeitraum von 2,5 Monaten (derzeit 4 Monate) zu Lasten der umsetzenden Landesebene und der Leistungserbringer vorgesehen. Dies wird in der ohnehin in diese Zeit fallenden Sommerferien als kontraproduktiv und fraglich umsetzbar empfunden. Es sollte an dieser Stelle eher über eine Straffung der Prozesse beim auswertungserstellenden IQTIG nachgedacht werden, zumal der Datenannahmeschluss um 4 Wochen auf den 31.01. des Folgejahres vorverlegt werden soll. Hierdurch sollte es möglich sein, die Auswertungen der Landesebene auch 4 Wochen früher zur Verfügung zu stellen, welche dann für das Stellungnahmeverfahren genutzt werden können (wie im Plan-QI-Verfahren mit im April vorliegenden Auswertungen).

Zu berücksichtigen ist weiterhin, dass in einem derart verkürzten Zeitraum jede Fachkommission mindestens zweimal beraten muss und auch die LAG zweimal (über Auslösung sowie Bewertung) entscheiden muss. Gerade hinsichtlich der möglichen Bedeutung der Bewertung, z.B. im Rahmen von Veröffentlichungen bzw. ggf. weitreichender Konsequenzen (z.B. P4P), wäre die empfohlene Zeiteinschränkung auch nicht zu verantworten gegenüber den beteiligten Leistungserbringern und umsetzenden Geschäftsstellen auf Landesebene.

Die vorgesehene komplexe, nicht nachvollziehbare und nicht praxistaugliche Nachberechnung von Indikatoren führt zu einer weiteren Verzögerung und Aufwandssteigerung des Stellungnahmeverfahrens, da letztlich auch die Fachkommissionen wiederholt beraten und bewerten müssen.

Eine umfassende Information und Einbeziehung sowohl der Leistungserbringer, der Öffentlichkeit als auch der umsetzenden Stellen und deren Fachkommissionen erscheinen aufgrund der komplexen Abläufe zwingend, denn Nachvollziehbarkeit bildet die Grundlage allgemeiner Akzeptanz.

Es ist nicht nachvollziehbar, weshalb das Nicht-Einhalten von Verfahrensregeln den Bewertungsprozess unterbrechen soll. Es wird in diesen Fällen die Einstufung

„auffällig“ (A) mit der erklärenden Ziffer „0“ (=Verzicht auf Einreichung einer Stellungnahme oder Stellungnahme nicht fristgerecht eingereicht oder Stellungnahme entsprach nicht den Anforderungen) ergänzt, jedoch wird mit „A“ die Klassifikation „Qualitätsdefizit“ vergeben.

Das Akzeptieren der Feststellung des Defizites von Seiten des Leistungserbringers durch Nicht-Abgabe einer erläuternden Stellungnahme sollte nicht gleichgesetzt werden mit dem Nicht-Einhalten von gesetzten formalen Anforderungen an eine Stellungnahme. Es handelt sich hier zwar auch um ein Qualitätsdefizit, jedoch bezüglich des Qualitäts-/Kommunikationsmanagements des Leistungserbringers und nicht der medizinischen Leistungserbringung. Letztere ist jedoch für Patienten/ die Öffentlichkeit im Rahmen der Transparenz der Qualität der Leistungserbringer von Interesse.

Mit freundlichen Grüßen



Dr. med. M. Wolf



## **Stellungnahme der Bayerischen Arbeitsgemeinschaft für Qualitätssicherung in stationären Versorgung (BAQ) zum Vorbericht des IQTIGs zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern:**

Vielen Dank für die Möglichkeit einer Stellungnahme zum Vorbericht des IQTIGs zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern. Prinzipiell ist eine Weiterentwicklung von Prozessen der Qualitätsbewertung und Qualitätsförderung wünschenswert und auch sinnvoll.

Vorbemerkend irritiert der Titel des Vorberichts im Hinblick auf die in 2021 geplante Überführung aller indirekten stationären QS-Verfahren von der QSKH-RL in die DeQS-RL. Dieser erweckt den Anschein, dass die Weiterentwicklung des Strukturierten Dialogs ausschließlich auf stationäre Einrichtungen reduziert sei. Ein differenzierter Umgang mit den unterschiedlichen Leistungserbringern ist sicherlich nicht zielfördernd und ebenso wenig vermittelbar und stellt das gesamte vorliegende Konzept bzw. die Beauftragung des G-BA insofern in Frage, dass nicht von vorneherein ein Bericht im Hinblick auf die DeQS-RL für alle Leistungserbringer beauftragt wurde. Sowohl Wording als auch Inhalt dieses Berichtes stehen zum Teil im Widerspruch zu der DeQS-RL.

Konkret zum Inhalt wollen wir folgende Anmerkungen geben:

1. Ziel der Weiterentwicklung soll eine Reduktion der Heterogenität der Bundesländer im Rahmen des Strukturierten Dialogs sein. Die Ursachenanalyse hierzu ist unseres Erachtens nicht ausreichend – Strukturen der Bundesländer (Anzahl der zu betreuenden Krankenhäuser, Flächenland/Stadtstaaten, Personalausstattung der Geschäftsstellen etc.) finden keine Berücksichtigung.
2. Der vorgegebene Zeitrahmen zwischen Beginn und Abschluss des Bewertungsprozesses wird als realitätsfern und kaum umsetzbar eingeschätzt:  
Der Dialog findet in den Sommermonaten (Schulferien) statt. Das ist nicht nur für Leistungserbringer bei ausgedünnter Personaldecke schwierig, sondern auch bei Terminabsprachen mit Fachkommission und Lenkungsgrremium. Bitte beachten Sie, dass gemäß DeQS-RL die Fachkommissionen sowohl die Einleitung als auch den Abschluss „nur empfehlen“ und erst das Lenkungsgrremium der LAG Beschlüsse fasst, woraufhin eine Geschäftsstelle aktiv z.B. ein Stellungnahmeverfahren einleiten oder abschließen kann. Die entsprechenden Sitzungen oder seien es nur Beschlüsse im schriftlichen Verfahren machen eine Zeitschiene gemäß Abb. 39 unmöglich. Aus unseren langjährigen Erfahrungen einer bisher selbstausswertenden Stelle auf Landesebene empfehlen wir zur Verkürzung des Zeitrahmens eine frühere Veröffentlichung der Jahresberichte/Ergebnisse. Gerade im Hinblick auf die nun prospektiven Rechenregeln ist es für uns nicht nachvollziehbar, dass für die Fertigstellung der Auswertungen ein ähnlicher Zeitrahmen vorgesehen ist, wie für den aufwändigen Prozess der Qualitätsbewertung. Die verwendeten Rechenregeln müssen dem Gebot der Transparenz und Nachvollziehbarkeit folgend zum Zeitpunkt der Veröffentlichung der Jahresauswertungen ebenfalls öffentlich vorliegen.
3. Qualität (der Bewertung) vor Quantität und Schnelligkeit: wir wollen, dass sich Leistungserbringer bei rechnerischen Auffälligkeiten mit den Fällen auseinandersetzen und diese analysieren. Nur so kann bei Bedarf eine gezielte Qualitätsförderung erarbeitet werden. Dies steht im Widerspruch dazu, dass eine Stellungnahme freiwillig sein soll oder vorschnell „Qualitätsdefizite“ wegen unzureichenden Stellungnahmen oder wegen Form-/Fristfehler bescheinigt werden. Dies mag womöglich im Sinne der Arbeitsreduktion für Geschäftsstellen/Fachkommissionen gut gemeint sein, verschleiert aber womöglich

tatsächlich vorliegende medizinisch-relevante Qualitätsprobleme, für die dann erstmal keine Qualitätsförderung erfolgt.

4. Nachberechnung eines informativen Indikatorwerts im Rahmen des Bewertungsprozesses führt vermutlich zu Verzögerungen, erscheint sehr komplex (erheblicher Ressourcenverbrauch) und erschwert die Transparenz und Nachvollziehbarkeit des Verfahrens. Vermutlich wird hierdurch erneut Heterogenität erzeugt.
5. Bewertungsprozess: Wir möchten für eine klare Trennung zwischen „fehlender Stellungnahme“, „Datenfehler“ und medizinischen Problemen plädieren. Diese Trennung sollte nicht erst durch erklärende Ziffern erfolgen, sondern bereits durch eine eindeutige Klassifikation. Die Konsequenzen einer fehlenden Trennung waren in den vergangenen Jahren bei der Veröffentlichung der Plan-QI zu sehen (siehe z.B. Geodarstellung einer Zeitung). Für Patienten und deren Angehörigen wird es somit unmöglich, sich adäquat über die Versorgungsqualität von Leistungserbringern zu informieren.
6. Bewertung von Qualitätsindikatoren sollten nicht ohne Hinzuziehen von Vollzähligkeits- bzw. Auffälligkeitskriterien erfolgen; ohne diese sind sie nur wenig aussagekräftig.
7. Die derzeit verfügbaren Qualitätsindikatoren lassen aufgrund Ihrer Güte in ihrem rechnerischen Ergebnis keinen Rückschluss auf die Qualität einer Behandlung zu. Sie dienen lediglich als Aufgreifkriterium für die Einleitung eines Strukturierten Dialogs. Die verwendete Begrifflichkeit „hinreichender Hinweis auf ein Qualitätsdefizit“ ist bereits wertend und ohne Stellungnahme des Leistungserbringers nicht zulässig.
8. Zusammensetzung Fachkommission: Die Festlegung der Zusammensetzung der Fachkommission gemäß Bericht ist widersprüchlich zu den Vorgaben der DeQS-RL (Themenspezifische Bestimmungen). Die derzeitigen Vorgaben in den themenspezifischen Bestimmungen sind ausreichend um vergleichbare Fachkommissionen aufzustellen. Ein aufwändiges Bewerbungsverfahren wird voraussichtlich als abschreckend empfunden und könnte zu Problemen bei der Besetzung in den Gremien führen.

Fazit: leider wurde die Landesebene, die über einen langen Zeitraum Erfahrungen mit dem Strukturierten Dialog gesammelt hat, erst zu spät und zu wenig in die Konzeptgestaltung des IQTIG zur Weiterentwicklung eingebunden. Viele angesprochenen Punkte wurden während der Workshops bereits angemerkt, fanden aber wohl aufgrund der fortgeschrittenen Zeit keine Berücksichtigung mehr in den nun vorliegenden Vorbericht. Dem Verfahren der Plan-QI-RL folgend findet mit dem angestrebten Vorgehen eine Umkehr der Beweislast statt: Der Leistungserbringer muss nun sein Unverschulden nachweisen. Bevor Qualitätsindikatoren nicht in Ihrer Güte überarbeitet werden und eine bestimmte Sensitivität bzw. ein hoher Vorhersagewert hinsichtlich der Identifizierung von Qualitätsdefiziten erreicht wird, sollte ein solches Konzept mit diesem Grundverständnis nicht in Kraft treten.

Stellungnahme der Geschäftsstelle Qualitätssicherung Hessen zum Vorbericht des IQITG zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern

#### A. Vorbemerkungen:

Im Hinblick auf die in 2021 mutmaßliche Überführung aller indirekten stationären QS-Verfahren von der QSKH-RL in die DeQS-RL wäre eine Weiterentwicklung des Strukturierten Dialogs reduziert auf Krankenhäuser problematisch. Im Rahmen der DeQS-RL beziehen sich Qualitätssicherungsmaßnahmen (gem. §17) immer auf Leistungserbringer. Ein differenzierter Umgang (z.B. mit Praxen gegenüber Krankenhäusern) ist hier nicht vorgesehen und wäre auch nicht vermittelbar. Gleichzeitig sind auch einige Punkte in dem Bericht (Qualitätsbewertung versus Qualitätsförderung) nicht kompatibel mit der aktuellen DeQS-RL und müssten zu entsprechenden Anpassung führen. **Somit stellt sich die Frage warum der Auftrag des G-BA nicht von vorneherein im Kontext zur DeQS-RL gestellt wurde und weiterhin wie mit den im Bericht formulierten Vorschlägen des IQITG im Rahmen der DeQS-RL ab 2021 umgegangen werden soll.**

Bei der ausführlichen Analyse zur Heterogenität der Bundesländer im Rahmen des strukturierten Dialogs werden im Bericht die Gründe nicht ausreichend hinterfragt. So spielt mutmaßlich die **Güte der Qualitätsindikatoren hinsichtlich Evidenz zum einen und dem positiven prädiktiven Wert (im Rahmen des strukturierten Dialogs) zum anderen eine wichtige Rolle**. Eine differenzierte Analyse der Berichte zum strukturierten Dialog der letzten Jahre wäre hier sinnvoll und sicher auch erkenntnisreich gewesen. Bei der Analyse zur Heterogenität wird zwar auf den Einfluss der Anzahl der zu prüfenden Indikatorergebnisse auf die Art des Strukturierten Dialoges eingegangen, eine **differenzierte Korrelation bzgl. Anzahl KH pro Bundesland zu Art und Ergebnis des Strukturierten Dialoges** bleibt jedoch aus. Zur geplanten Überprüfung der Qualitätsindikatoren hinsichtlich ihrer Güte wird mehrfach im Bericht eingegangen, auf die Bedeutung der qualitativen sowie quantitativen Ausstattung einer Geschäftsstelle wird jedoch nur kurz und oberflächlich eingegangen.

Auch wenn es nicht zu erwarten ist, dass ein Werk zu einem bedeutenden Bestandteil der externen Qualitätssicherung aus einem „Guss“ geschrieben ist, fällt bei dem Bericht zur „Weiterentwicklung des Strukturierten Dialog“ die doch zum Teil **sehr unterschiedliche Vertiefung** der Themen auf. Im Besonderen sticht hier das Kapitel 5 (Seite 71-142) heraus, welches ohne akademische Vorkenntnisse in den Methoden der Statistik nicht zu verstehen ist. Hier hätte ein großzügiges Verlagern in den Anhang sowie eine kurze prägnante und nachvollziehbare Zusammenfassung nicht nur den Bericht an sich verkürzt, sondern auch dem Leser bzw. der Leserin eine klare Nachricht zu dem Thema verschaffen können.

#### B. Ungenauigkeiten im Bericht:

Seite 24, Kapitel 2.3: Grundsätzlich sind die Fachexperten zwar ehrenamtlich in den Arbeitsgruppen tätig, zu erwähnen wäre jedoch, dass in einigen Bundesländern den Mitgliedern Aufwandsentschädigungen gezahlt werden. Inwieweit dies die Zusammensetzung bzw. die Arbeit beeinflusst ist unklar.

Seite 26, Kapitel 2.4: Das Einstufungsschema ist keine Erfindung des AQUA-Instituts. Es wurde noch von der Vorgängereinrichtung der BQS in Zusammenarbeit mit den Landesgeschäftsstellen entwickelt. Hintergrund der Entwicklung war die Notwendigkeit einer Darstellung im strukturierten Qualitätsbericht.

Seite 27, Kapitel 2.4.2: Auch bei der Versendung eines Hinweises kann die Beurteilung einer rechnerischen Auffälligkeit erfolgen. So werden Kliniken, bei denen z.B. bei einem kollegialen Gespräch klare Hinweise auf Qualitätsdefizite zu Tage kommen, welche mutmaßlich in dem Jahr nach der rechnerischen Auffälligkeit noch bestehen, weiterhin qualitativ auffällig bewertet ohne erneut eine Stellungnahme anzufordern.

Seite 28, Kapitel 2.4.2 (Hinweise): Bei dem Thema Hinweise wird in dem Bericht zu wenig auf die Hintergründe eingegangen. Diese wurden ausführlich im Bericht der Projektgruppe erläutert. Ein häufiger Grund für die Versendung von Hinweisen ist die rechnerische Auffälligkeit durch einen Einzelfall. Hierzu gab es eine klare Empfehlung der LQS-Arbeitsgruppe „Strukturierter Dialog“, statt, wie in §10 QSKH-RL angeboten, auf einen SD zu verzichten, zumindest einen Hinweis zu versenden. Weiterhin wurden Hinweise u.a. bei Neueinführung von Qualitätsindikatoren oder Modifikation bestehender Qualitätsindikatoren oder im Rahmen einer Verlaufsbeurteilung (s.o.) eingesetzt.

### C. Einschätzung zu den Vorschlägen des IQITG

1. Grundsätzliche Voraussetzung vor Umsetzung des neuen Konzepts ist die Überprüfung hinsichtlich der **Güte der Indikatoren**. Diese sollte zu einer **Priorisierung bzw. einer Reduktion** der aktuell gültigen Qualitätsindikatoren führen
2. Die streng **modulare Betrachtungsweise (Qualitätsbewertung, Qualitätsförderung und Daten- und Dokumentationsqualität)** ist kritisch zu hinterfragen. Zum einen sollte nicht ausgeschlossen werden, dass qualitätsfördernde Maßnahmen (wie kollegiale Gespräche) auch zu Bewertungszwecken genutzt werden. Noch wichtiger ist es aber **Daten- und Dokumentationsqualität bei der Bewertung miteinzubeziehen**. (Einem KH mit einer Unterdokumentation oder einer in der Stichprobenprüfung aufgefallenen relevanten Fehldokumentation die Bewertung „kein hinreichender Hinweis für ein Qualitätsdefizit“ zu attestieren ist nicht akzeptabel.)
3. Der Ansatz **statistische Methoden** bei der Auslösung eines Stellungnahmeverfahrens zu berücksichtigen ist grundsätzlich nachvollziehbar. Wünschenswert wäre hierbei ein differenziertes Vorgehen. Zum einen sollte unterschieden werden zwischen **Prozess- und Ergebnisindikatoren**. Bei Prozessindikatoren ist die Relevanz bei betroffenen Einzelfällen anders zu werten als bei häufig betroffenen Fällen in einer Klinik mit hoher Fallzahl (z.B. im Rahmen einer konsequenten Missachtung von Leitlinien). Bei relevanten Ergebnisindikatoren, die häufiger auch mit einem Defizit in der Ergebnisqualität in Verbindung zu bringen sind (z.B. intraoperative Komplikationen), wäre für Kliniken mit geringer Fallzahl eine automatische Bewertung mit „kein hinreichender Hinweis für ein Qualitätsdefizit“ kritisch zu sehen. Hier könnte von Anfang an eine jahresübergreifende Betrachtung (besser 3 als nur 2 Jahre) zur Auslösung eines SD erfolgen.

4. Der Begriff „**Qualitätsdefizit**“ im Rahmen der Qualitätsbewertung ist hier kritisch zu hinterfragen. Zum ersten kann die Aussage, dass ein (statistisch) abweichendes Indikatorergebnis als „Hinreichender Hinweis für ein Qualitätsdefizit“ gesehen wird, in Frage gestellt werden. Hierfür müssten die Qualitätsindikatoren zumindest eine entsprechende Güte aufweisen und auch einen hohen positiven Vorhersagewert hinsichtlich der Identifizierung von Qualitätsdefiziten darstellen. Die aktuell eingesetzten Qualitätsindikatoren erreichen, bis auf wenige Ausnahmen, diese Güte nicht. Angebracht wäre daher nur die Aussage „statistisch signifikante Abweichung vom Referenzbereich“. Erst nach dem strukturierten Dialog kann eine Aussage zu möglichen Qualitätsdefiziten gemacht werden. Auf Grund der nur limitierten Möglichkeiten im Rahmen der Bewertung sollte hier grundsätzlich von „vorliegenden Hinweisen“ gesprochen werden, und das sowohl bei „keinen Hinweisen auf ein Qualitätsdefizit“ als auch bei „Hinweisen auf ein Qualitätsdefizit“. Dass schon allein die Nichtabgabe einer Stellungnahme bzw. die Nichteinhaltung der formalen Kriterien als „Qualitätsdefizit“ abgestempelt wird, muss hier ebenfalls in Frage gestellt werden.
5. **Nachberechnungen** im Rahmen der Bewertung sind grundsätzlich kritisch zu werten. Hier muss aber vorab die Bedeutung eines Referenzwertes diskutiert werden. Für einen ratenbasierten Qualitätsindikator müsste im Idealfall 0%, 100% oder 1 (O/E) als Referenzbereich stehen. In solch einem Fall würden im Rahmen von Stratifizierung oder Risikoadjustierung alle patientenbezogenen Faktoren berücksichtigt werden. Hierfür müssten die Instrumente jedoch stark ausgeweitet werden, was mit einem entsprechend hohen Dokumentationsaufwand verbunden wäre. Aus diesem Grund wird zumeist ein perzentilbasierter Referenzbereich definiert. Dieser stellt an sich noch kein Indiz für ein „Hinweis auf ein Qualitätsdefizit“ dar, sondern könnte vielmehr als Schwellenwert zur Auslösung eines Strukturierten Dialoges verstanden werden. Die Bewertung der rechnerisch auffälligen Klinik sollte daher völlig unabhängig von Nachberechnungen auf Grund von Einzelfällen erfolgen. Das wird indirekt im Bericht auf Seite 165 (Inhaltlicher Fokus) vom IQTIG selbst auch so formuliert und muss daher als Widerspruch im Bericht gesehen werden. Falls es auf Grund einer eingegangenen Stellungnahme Hinweise auf ein Qualitätsdefizit gibt, sind diese entsprechend zu bewerten (auch wenn es sich nur um Einzelfälle handelt). Die hier dargestellt Nachberechnung erscheint kompliziert, wenig nachvollziehbar und nicht praxistauglich.
6. Bei dem vorgeschlagenen **Bewertungsschema** gibt es in der ersten Ebene (Einstufung) quasi nur U für Unauffällig und A für Auffällig, gleichbedeutend mit kein Hinweis auf Qualitätsdefizit und Qualitätsdefizit. Bei der Bewertung von Dokumentationsfehlern wird nur noch zwischen zu verantwortenden und nicht zu verantwortenden Datenfehlern unterschieden. Welche Kriterien für die Verantwortung herangezogen werden können bleibt unklar. In der Annahme dass die Datenfehler fast ausschließlich von den Leistungserbringern zu verantworten sind, führt damit unabhängig von der Relevanz und dem Ausmaß der Fehldokumentation eine Einstufung in A (Qualitätsdefizit). Damit führt die Bewertung letztendlich zu einem undifferenzierten Bild hinsichtlich tatsächlicher medizinischer Qualitätsdefizite und beispielsweise vereinzelter Dokumentationsunschärfen, die in keinem Bezug zur medizinischen Versorgung stehen. In diesem Zusammenhang steht auch die

Bewertung A0, welche bereits aus formalen Gründen ein Qualitätsdefizit bescheinigt. Damit wird ein zum Teil komplexer Bewertungsvorgang auf letztendlich zwei Schlüssel beschränkt (U/A), welche nach den Erfahrungen der letzten Jahre bei Veröffentlichung maßgeblich z.B. für Einfärbung oder Geodarstellung stehen, und ggf. auch bei zukünftigen Konsequenzen (P4P) undifferenziert Anwendung finden.

7. Gemäß dem Vorschlag vom IQTIG empfiehlt die Fachkommission eine **abschließende Bewertung**. Die **abschließende Entscheidung** zur Einstufung gemäß Bewertungsschema trifft jedoch die zuständige Stelle. Abweichungen zur Empfehlung sollen begründet und dokumentiert werden. Unklar bleibt hierbei wann und mit welcher Legitimation z.B. die Landesarbeitsgemeinschaft der Empfehlung eines mutmaßlich neutralen Fachgremiums widersprechen kann. Letztendlich sind hierbei nur formale Gründe vorstellbar. Eine Abweichung der Empfehlung der Fachkommission durch die LAG birgt die Gefahr einer interessensgeleiteten Beeinflussung der Bewertung und sollte daher vermieden werden.
8. Die Vorschläge zur Zusammensetzung von **Fachkommissionen** sind grundsätzlich positiv zu bewerten, vor allem die Auswahl der Personen durch die zuständige Stelle, und nicht, wie gem. DeQS-RL praktiziert, die Berufung primär durch die entsprechenden Interessensvertretungen. Dies würde zu einer höheren Wahrscheinlichkeit führen, dass die Bewertungen auch neutral und nicht interessensgeleitet abgegeben werden. Auch eine einheitliche Aufwandsentschädigung wäre zu begrüßen. Die aktuelle Praxis zeigt sowohl Unterschiede zwischen den Bundesländern als auch Unterschiede innerhalb einer Fachkommission (z.B. KV-Vertreter versus KH-Vertreter). Ob das **aufwändige Bewerbungsverfahren** letztendlich nicht eher eine abschreckende Wirkung zeigt und zu Problemen bei der Besetzung der Gremien führt, bleibt abzuwarten.
9. Auch wenn im Bericht an diversen Stellen Vorschläge gemacht werden, welche nicht mit der aktuellen DeQS-Richtlinie vereinbar sind (siehe auch Vorbemerkung), werden gerade bei dem Kapitel zur **Verkürzung des Stellungnahmeverfahrens** streng die Fristen aus der DeQS-RL übernommen. Bei der Empfehlung des IQTIG wird der Zeitraum von der Einholung der Stellungnahmen bis zur Bewertung auf 2,5 Monate reduziert (bisher 4 Monate bei berichtspflichtigen Indikatoren). Hierbei wird weder berücksichtigt, dass der entsprechende Zeitraum in die Ferienzeit fällt, noch dass in diesem Zeitraum die Fachkommission mindestens zweimal tagen muss und auch die LAG zweimal (über Auslösung sowie Bewertung) entscheiden muss. Gerade hinsichtlich der möglichen Bedeutung der Bewertung, z.B. im Rahmen von Veröffentlichungen bzw. ggf. weitreichender Konsequenzen (z.B. P4P), wäre die empfohlene Zeiteinschränkung auch nicht zu verantworten gegenüber den beteiligten Leistungserbringern und Geschäftsstellen auf Landesebene. Jedoch ergibt sich durch die Umstellung des Auswertungszeitraums auf das Entlassungsdatum und die prospektiven Rechenregeln die Möglichkeit einer deutlich früheren Bereitstellung der Rückmeldeberichte (zumindest für fall- und einrichtungsbezogene QS-Dokumentation). Beispielgebend dafür ist das Plan-QI-Verfahren, in welchem die Berichte bereits Mitte April zur Verfügung gestellt werden und es damit auch möglich wäre, frühzeitig und mit ausreichend Zeit den strukturierten Dialog durchzuführen. Hierauf geht das IQTIG nicht ein. Auch geht es in dem Bericht nicht auf die Möglichkeit ein, die Stellungnahmeverfahren bereits unterjährig durchzuführen. Damit könnte auch einem wesentlichen Kritikpunkt

externer Qualitätssicherungsmaßnahmen, dem langen Zeitverlauf (bis zu 1 ½ Jahre) zwischen einem Qualitätsdefizit und dem Hinweis darauf bzw. dem Bewerten als solches, entgegengetreten werden. Da im Rahmen der DeQS-RL Quartalslieferungen der Daten sowie Quartalsberichte vorgesehen sind, erfolgt (neben internen Möglichkeiten wie z.B. dem QS-Monitor) ggf. der Hinweis in Form eines rechnerisch auffälligen Indikatorergebnisses bereits früher und sogar schon unterjährig. Damit stellt sich die Frage, warum der Strukturierte Dialog nicht auch unterjährig durchgeführt wird. Gerade zu gezielten Prozessindikatoren könnte hier frühzeitig einer möglichen Fehlversorgung entgegengewirkt werden. In Hessen wird dies bereits im Landesverfahren QS-MRE mit Erfolg praktiziert.

#### **D. Fazit**

**Der erste Schritt zur Harmonisierung des Strukturierten Dialoges ist die Überprüfung der bestehenden Qualitätsindikatoren auf Ihre Güte tatsächlich Qualitätsdefizite aufzudecken und die damit verbundene Reduktion der Indikatoren. Dies wird im Bericht des IQTIG auch mehrfach erwähnt. Ein weiterer Punkt, auf welchen im Bericht nur rudimentär eingegangen wird, ist die Unabhängigkeit der Landesgeschäftsstellen bzw. Landesarbeitsgemeinschaften sowie die personelle Ausstattung hinsichtlich Qualifikation und Quantität. Gerade bei den vorgeschlagenen qualitätsfördernden Maßnahmen ist dies maßgeblich zur positiven Beeinflussung der Versorgungsqualität. Ob der formale und komplexe Bewertungsvorgang und das enge Zeitfenster zu einem weniger heterogenen Ergebnis zwischen den Bundesländern führt, muss bezweifelt werden. Die vereinfachte und undifferenzierte Bewertung in ausschließlich zwei Ergebnisse (Kein Hinweis auf Qualitätsdefizit und Qualitätsdefizit), welche gerade im Hinblick auf die zukünftig vermehrte Nutzung (z.B. Veröffentlichung, Planung, P4P) vorteilhaft scheint, wird dem komplexen Verfahren in keiner Weise gerecht und führt daher vermutlich eher zu adversen Effekten bei den betroffenen Leistungserbringern. Einige Ansätze erscheinen durchaus sinnvoll, so ist ein statistischer Ansatz bei der Berechnung zu begrüßen, solange hier Leistungserbringer mit wenigen Fällen und schlechter Qualität nicht aus dem Radar fallen. Auch die Überlegungen zur Besetzung der Fachkommissionen sind grundsätzlich nachvollziehbar, wenngleich hierdurch auch Hürden aufgestellt werden, die möglicherweise die Besetzung des Gremiums erschweren. Die Möglichkeit der zuständigen Stelle eine Empfehlung der Fachkommission revidieren zu können sollte jedoch klar geregelt werden um hier interessensgesteuerte Bewertungen zu verhindern.**



**Stellungnahme  
zum Vorbericht  
„Weiterentwicklung des Strukturierten Dialogs  
mit Krankenhäusern“**

**Geschäftsstelle Qualitätssicherung Nordrhein-Westfalen (QS-NRW)**

Gartenstraße 210-214  
48147 Münster  
Telefon 0251 929-2601  
Fax 0251 929-2649

Tersteegenstraße 9  
40474 Düsseldorf  
Telefon 0211 4302-2700  
Fax 0211 4302-2709

Die Geschäftsstellen QS-NRW bedankt sich für die Möglichkeit der Teilnahme am schriftlichen Beteiligungsverfahren zum Vorbericht Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern.

Zu Beginn möchten wir die umfassende Aufarbeitung des Themas Weiterentwicklung strukturierter Dialog würdigen.

Ausführlich wird sich in dem Bericht der Qualitätsindikatoren angenommen. Dargelegt werden die Aspekte der Auswertung, der Festlegung von Referenzbereichen, Bedeutung bzw. Einfluss von Spezifität und Sensitivität im Stellungnahmeverfahren, dass letztlich der Verbesserung der Validität der Indikatorergebnisse dient. Eingegangen wird auf die erforderliche Differenzierung zwischen Qualitätsbewertung und Qualitätsförderung. Die Beauftragung zielte auch auf die Steigerung der Effizienz und den damit verbundenen verbesserten Ressourcen-Einsatz. Thematisiert werden zudem Anforderungen an und Bewertung von Stellungnahmen. Auch werden Vorgaben zur Zusammensetzung medizinischer Arbeitsgruppen diskutiert.

Nachfolgend finden Sie zu einigen Aspekten unsere Anregungen:

### **Qualitätsindikatoren (QI)**

Bei der Vielzahl der zurzeit existierenden Indikatoren ist das im Vorbericht angesprochene Verfahren, über die Bundesfachgruppen eine Neubewertung der Indikatoren vorzunehmen, um so die Anzahl der QI zu verringern, zu begrüßen. Anstelle der zweischrittigen Vorgehensweise schlagen wir vor, dass die Landesebene frühzeitig einbezogen wird, z.B. im Rahmen von Bund-Land-Kommissionen zu den einzelnen Leistungsbereichen. Durch diesen direkten Austausch wären die Erkenntnisse, die mit den unterschiedlichen Perspektiven und verschiedenen Erfahrungen einhergehen, gewahrt und könnten frühzeitig berücksichtigt werden.

Des Weiteren ist die systematische mathematische und statistische Aufarbeitung ein wichtiger Punkt wie auch die Schaffung von mehr Transparenz in den Rechenregeln. Der im Vorbericht beschriebene Ansatz, die Qualitätsindikatoren so zu gestalten, dass eine rechnerische Auffälligkeit gleich einer qualitativen Auffälligkeit ist, würde Ressourcen sparen, ist aber im Kontext der medizinischen Versorgung von Menschen unrealistisch und mit keiner derzeit bekannten mathematisch/statistischen Methode verlässlich umsetzbar.

⇒ Begrüßenswert ist das Ziel der Verbesserung der Qualitätsindikatoren bei gleichzeitiger Verringerung ihrer Anzahl.

### **Neuberechnung**

Das Thema Neuberechnung erfordert eine differenzierte Betrachtung. Eine Neuberechnung nach erfolgter Bewertung durch ein Stellungnahmeverfahren bietet die Möglichkeit der Abbildung eines bewerteten Ergebnisses z.B. in der Außendarstellung. Ohne Frage muss diese Neuberechnung aus unserer Sicht alle Leistungserbringer in diesem Indikator umfassen. Die Gründe hierfür sind ausführlich in den Kapitel 6.3 und 6.4 (Seite 150 und folgende) des Berichts zur Weiterentwicklung des SD dargelegt. Ansonsten wäre die Konsequenz, dass eine bundesweite Vergleichbarkeit der Ergebnisse nicht gewährleistet wäre.

Der unter Punkt 2.6 (Seite 40) dargestellte Anpassungsbedarf bezüglich der Perzentilenberechnung wird von uns geteilt. Differenziert ist die Darstellung des bestehenden Problems, dass Sensitivität und Spezifität unzureichend gemessen werden, erfolgt. Hier muss ein Kompromiss eingegangen werden, da wie beschrieben eine hohe Spezifität den Leistungserbringer interessiert und eine hohe Sensitivität mehr den potentiellen Patienten (öffentliche Berichterstattung).

- ⇒ Heterogenität ist zu vermeiden. Anzustreben sind Ergebnisse mit hoher Aussagekraft für alle im System Beteiligten.
- ⇒ Zusammenfassend scheint das Verfahren der Nachberechnung in der dargestellten Form nicht mit angemessenem Aufwand umsetzbar und könnte durch unterschiedliche Umsetzungsstrategien zu Unterschiedlichkeiten führen.

### **Strukturierter Dialog / Stellungnahmeverfahren (SD/STNV)**

Sinnvoll für das dargestellte Prozedere (1. Sichtung der Ergebnisse, 2. Versenden der STN-Anforderungen, 3. Erstellung STN durch LE, 4. Formale Prüfung der STN, 5. ggf. Rückmeldung an LE und erneute formale Prüfung) ist, dass den LE die formalen Anforderungen für die STN-Erstellung bekannt und vertraut sind. Ein Effizienzgewinn ist im Laufe der Zeit zu erwarten. Gleichzeitig gehen bei einer stärkeren Formalisierung des STNV für die Klärung möglicherweise relevante Informationen verloren. Eine Beurteilung allein wegen Nichteinhaltung der Formalien als Qualitätsdefizit A0 ist, wie unten dargelegt, aus unserer Sicht problematisch.

Die Güte der Qualitätsindikatoren und ihrer Berechnung ist nach aktuellen Erfahrungen nicht in einem so hohen Maße fehlerfrei, valide und unstrittig, dass ein außerhalb des Referenzwerts gelegenes Rechenergebnis Ausnahmetatbestände und entlastende Gründe erfordert, um die Annahme eines Qualitätsdefizits aufzugeben. Dies trifft insbesondere auch bei der Anwendung von perzentilgebundenen Referenzwerten im Kontext seltener Ereignisse zu. Die sollte beim Wording im Verfahren beachtet werden.

Zusammenfassend sind aus unserer Sicht

- ⇒ Regeln für das Erstellen einer Stellungnahme sinnvoll. Ein solches Vorgehen benötigt Zeit bis es sich bei allen Beteiligten eingespielt hat. Eine hohe Personalfuktuation steht dem entgegen.
- ⇒ bei der Erstellung von Stellungnahmen auch gewisse Freiräume erforderlich.

### **Bewertungsschema**

Die Eingrenzung auf letztlich drei Bewertungskategorien ist aus unserer Sicht zu begrüßen. Hinsichtlich der Unterkategorien sehen wir jedoch Änderungsbedarf. Beispielsweise umfasst A0 drei sehr unterschiedliche Aspekte, die unseres Erachtens eine Differenzierung erfordern:

1. „Verzicht auf Einreichung einer Stellungnahme“ sollte eine eigene Unterrubrik darstellen entsprechend Plan QI-Prozedere „Zusicherung“.
  2. „Stellungnahme entsprach nicht den Anforderungen“ ist zu allgemein gefasst und wird den verschiedenen möglichen Konstellationen nicht gerecht.
- ⇒ Die Eingrenzung der Bewertungskategorien wird befürwortet, ohne dass dabei, ein relevanter Aussageverlust in Kauf genommen werden darf (ein Qualitätsdefizit aufgrund formaler Fehler der Stellungnahme ist nicht als medizinisches Defizit zu werten, dies muss klar erkennbar bleiben).

### **Zeitlicher Ablauf**

Die Vorstellungen zum zeitlichen Ablauf sind aus unserer Sicht unrealistisch und unbefriedigend. Der demnach den Ländern zur Verfügung stehende Zeitrahmen ist zu kurz. Um mehr Zeit für den/das anspruchsvollen SD/STNV zu gewinnen, ist es erforderlich, dass das IQTIG die Auswertungserstellung spätestens bis zum 15. Mai jedes Jahres zur Verfügung stellt. Ein Zeitgewinn wäre damit erfüllt und zudem gewährleistet, dass die Befassung nicht nur in die Sommerferienphase fällt. Infolgedessen würde sich auch der zeitliche Rahmen für die Bewertung des Stellungnahmeverfahrens verlängern, was der Komplexität der zu erfüllenden Aufgaben und der Vielzahl an einzubeziehenden Personen aus den med. AG/Fachkommissionen Rechnung trägt. Straffungen sind im Bereich der Auswertungserstellung bei etablierten statistischen Verfahren und Formaten sicher eher möglich als im Bereich der Stellungnahmeverfahren und Bewertungen mit einer Vielzahl von beteiligten Institutionen und Gremien. Hier sollte die Verteilung der zeitlichen Vorgaben kritisch überarbeitet werden.

Eine Verlängerung des den Ländern zur Verfügung stehende Zeitrahmens um mindestens vier Wochen wird als erforderlich angesehen.

### **Besetzung der Fachkommissionen**

Zu dem unter 6.6 beschriebenen Besetzungsverfahren sehen wir Bedarf für Anmerkungen. Wir stimmen überein, dass der Gewinn von fachkompetenten Personen für die Wahrnehmung eines Ehrenamts unter Einhaltung gewisser Kriterien erfolgen sollte. Die Handhabung sollte jedoch realitätsangepasst sein und eine gewisse Flexibilität ermöglichen. So sollte die Aufgabe, in einer Fachkommission mitzuwirken keinesfalls durch Formalismus unattraktiv gestaltet werden, die Fachkommissionen sind darauf angewiesen, kompetente Mitglieder zur aktiven Mitarbeit gewinnen zu können. Grundsätzlich ist das Besetzungsverfahren in den fachspezifischen Regelungen der DeQs-RL hinterlegt und bedarf somit keiner Neuregelung, ebenso in der plan. QI-RL.

- ⇒ Ergänzend zu dem bereits in verschiedenen Richtlinien geregelten Verfahren (DeQs-RL, plan. QI-RL) sollten Aufgaben und Aufwand im Vorfeld dargestellt werden, um die Erwartungen an die Ehrenamtsträger transparent zu machen.

### **Qualitätsförderung**

Eine Ressourceneinsparung ist möglicherweise dann zu erwarten, wenn die Maßnahmen der Stufe 1 (DeQS) erst nach der Bewertung des Stellungnahmeverfahrens erarbeitet werden. Dies wird jedoch vielfach bereits so gehandhabt, so z.B. in NRW. Klinikgespräche wie auch die Vereinbarung von Zielvereinbarungen werden nach Abschluss des „Strukturierten Dialogs“ durchgeführt.

- ⇒ Die Thematik der Qualitätsförderung sollte mehr Gewicht bekommen, als dies in der aktuellen Aufarbeitung erfolgt ist. Neben der Ermittlung von Defiziten ist die Schaffung von Angeboten an Leistungserbringer im Sinne einer Hilfe zur Selbsthilfe bedeutsam, um die Defizite zu beseitigen und die Qualität zu fördern. Patientensicherheit kann man nicht herbeibewerten, sie entwickelt sich durch Herausstellung der Qualitätsförderung im Rahmen der Qualitätssicherung.

Abschließend möchten wir noch auf einen Aspekt eingehen. Der vorliegende Bericht zielt auf die DeQs-RL ab. Er bezieht sich aber in vielen Punkten auf die plan. QI.-RL. Der beschriebene Ablauf des „Strukturierten Dialogs“ bzw. „Stellungnahmeverfahrens“ ähnelt sehr dem der plan.-QI.-RL, auch in der Verwendung der Begrifflichkeiten. Im Prinzip ist es als positiv zu werten, wenn alle gültigen Verfahren demselben Schema entsprechen.

Allein die zurzeit gültigen Richtlinien erfordern einen hohen Aufwand für die Leistungserbringer. Es gibt z.T. unterschiedliche Datenannahmestellen bei unterschiedlichen Richtlinien. Leistungserbringer werden von unterschiedlichen Stellen (IQTIG und Landesstellen) zu gleichen Indikatoren angeschrieben (plan.-QI – statistisch oder rechnerisch auffällig).

Weniger wäre hier wahrscheinlich mehr, weniger Qualitätsindikatoren mit höherer Aussagekraft und gleichzeitig ein intensiver Umgang mit diesen. Das auch im Vorbericht angesprochene PEER-Review Verfahren wäre hier zu nennen. Es hat sich gezeigt, dass ein direkter Kontakt mit den Leistungserbringern zu besseren Ergebnissen führt. Qualitätsförderung hat somit einen positiven Effekt auf Qualitätsergebnisse.

Wichtig ist in unseren Augen auch das Darstellen von Zwischenberichten. Wenn diese zeitnah zur quartalsweisen Datenlieferung erfolgen, haben die Leistungserbringer die Möglichkeit, im laufenden Erfassungsjahr gegenzusteuern und damit ein Stellungnahmeverfahren zu vermeiden. Damit würden erhebliche Ressourcen gespart und die Effektivität gesteigert, da mit der gesparten Zeit mehr Qualitätsförderung betrieben werden kann. Sinnvolle Zwischenauswertungen erfordern allerdings eine auswertbare Datenmenge und stoßen bei kleinen Fallzahlen an die Grenzen der Sinnhaftigkeit.

Als letzten Punkt sollte noch die Evaluation angesprochen werden. Nur durch die Rückmeldung aller Beteiligten kann ein Verfahren verbessert werden. Die Befragung muss sich an die beauftragten Stellen (Geschäftsstellen auf Landesebene, Fachkommissionen, aber auch an die Leistungserbringer) richten. Nur aus der Zusammenschau aller kann es zu einer wirklichen Verbesserung kommen.

## Landesgeschäftsstelle einrichtungs- und sektorenübergreifende Qualitätssicherung



Landesgeschäftsstelle einrichtungs- und  
sektorenübergreifende Qualitätssicherung  
c./o. Sächsisches Landesärztekammer  
PF10 04 65 · 01074 Dresden

**Institut für Qualitätssicherung und  
Transparenz im Gesundheitswesen (IQTIG)  
Katharina-Heinroth-Ufer 1  
10787 Berlin**

Dresden: 21. November 2019  
 Bearbeiterin: Dr. Beate Trausch  
 Telefon: 03 51 - 82 67-271  
 Telefax: 03 51 - 82 67-382  
 E-Mail: [mail@gesue-sachsen.de](mailto:mail@gesue-sachsen.de)

### **Stellungnahme der Landesgeschäftsstellen zur externen Qualitätssicherung in Sachsen – schriftliches Beteiligungsverfahren zum Vorbericht „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“**

Sehr geehrte Damen und Herren,

als Landesgeschäftsstelle einrichtungs- und sektorenübergreifende Qualitätssicherung der LAG Sachsen bzw. Projektgeschäftsstelle Qualitätssicherung Sachsen nehmen wir Stellung zum Vorbericht „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“ vom 10. Oktober 2019. Dazu möchten wir in diesem Brief eine allgemeine Stellungnahme darlegen. Die beiden Geschäftsstellen waren beteiligt am Workshop zu diesem Thema am 27. Mai 2019 sowie an der LAG- bzw. LQS-IQTIG-Sitzung am 17. und 18. Juni 2019 und haben sich zu der Thematik schriftlich am 13. Juni 2019 geäußert.

Im Kapitel 2 „Analyse der Ausgangssituation“ (S. 20 – 53) wurde vom IQTIG sehr detailliert und nachvollziehbar eine IST-Analyse durchgeführt, die sich allerdings teilweise auf eine Auswertung aus dem Jahre 2012 bezieht und deren Aktualität damit zu hinterfragen ist. Gründe für die Heterogenität des Strukturierten Dialogs in den verschiedenen Bundesländern wurden aber auf dieser Datenbasis dargelegt. Sehr aufschlussreich sind die Betrachtungen zu qualitativen Auffälligkeiten in Relation zu rechnerischen Auffälligkeiten und angeforderten Stellungnahmen bezogen auf Daten des Erfassungsjahres 2017 (S. 42 ff). Heterogenität im Strukturierten Dialog wurde unter anderem bezüglich der Entscheidung: „Versendung eines Hinweises“ vs. „Anforderung einer schriftlichen Stellungnahme“ diskutiert. Es wurde des Weiteren als eine Grundaussage hervorgehoben, dass sich die gesetzlich verpflichtenden Qualitätsmessungen weg vom „improvement“ hin zu „accountability (z. B. öffentliche Berichterstattung, Vergütungsentscheidungen“ (siehe S. 23) entwickelt, was allgemein Grundlage für kontroverse Diskussionen sein dürfte.

Im Kapitel 3 „Methodischer Hintergrund“ (S. 54 – 65) ging das IQTIG sehr differenziert auf die methodische Funktion von Stellungnahmen, die Reichweite der Qualitätsaussage von Stellungnahmen, Gütekriterien für Bewertungsprozesse, die Ableitung von Anforderungen an das Stellungnahmeverfahren, Vor- und Nachteile expliziter Beurteilungsregeln im Stellungnahmeverfahren sowie die methodische Einordnung der Qualitätsförderung ein, die für den Leser nachvollziehbar sind.

Dabei können allerdings die Aussagen im ersten Absatz (S. 56) nicht unkommentiert bleiben. Es wurde dort ausgeführt, dass es nicht Ziel der fachlichen Bewertung sei, in den Behandlungsfällen, die zu einem auffälligen Indikatorergebnis geführt haben, grobe Behandlungsfehler zu finden, wenn diese keinen Bezug zum dargestellten Qualitätsmerkmal haben. – Woher stammt diese Feststellung? Sie kann nicht aus den erklärten Zielen zu den einzelnen Qualitätssicherungsverfahren geschöpft werden. Es sollte der Fachkommission aus Gründen der Umsetzung der Qualitätssicherungsverfahren und der ärztlichen Ethik unbenommen bleiben, bei offensichtlich werdenden groben Behandlungsfehlern die Leistungserbringer darauf aufmerksam zu machen und einen Hinweis zu erteilen.

**Landesgeschäftsstelle einrichtungs- und  
sektorenübergreifende Qualitätssicherung**  
c./o. Sächsische Landesärztekammer  
Schützenhöhe 16 · 01099 Dresden  
Postanschrift: PF10 04 65 · 01074 Dresden

T +49 (0) 351 82 67-271

F + 49 (0) 351 82 67-382  
[mail@gesue-sachsen.de](mailto:mail@gesue-sachsen.de)

**Kapitel 4 „Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und –förderung“ (S. 66 – 70):**

In diesem Kapitel wurde auf die modulare Betrachtungsweise - bestehend aus dem Modul „Qualitätsbewertung“, dem Modul „Qualitätsverbesserung“ und dem Modul „Daten- und Dokumentationsqualität“ - eingegangen, wobei das letzte Modul nicht Gegenstand des vorgestellten Konzeptes sei.

Dem in Abbildung 9 (S. 68) dargelegten Ablauf der Qualitätsbewertung und –förderung kann von Seiten der Geschäftsstellen zugestimmt werden.

Nachfolgend (S. 69) wurde hervorgehoben, dass die Abgabe einer Stellungnahme freiwillig zu erfolgen habe und es dem Leistungserbringer überlassen sei, von dieser Möglichkeit Gebrauch zu machen. Bedeutet dies nicht einen Verzicht darauf, sich kritisch mit dem rechnerisch auffälligem Ergebnis auseinanderzusetzen? Unter Berücksichtigung der weiteren Ausführungen in diesem Vorbericht stellt sich die Frage, ob nicht diese vermeintliche Großzügigkeit sich am Ende in einen Nachteil für den Leistungserbringer umkehrt und dieser qualitativ auffällig bewertet wird, da er seine Möglichkeiten zur Problemlösung unterschätzt/verkennt?

**Kapitel 5 „Statistische Auswertungsmethodik“ (S 71 - 142):**

Hier wurden vor allen Dingen eine entscheidungstheoretische Modellierung des Bewertungsprozesses mit großer Detailtiefe, die für Nicht-Mathematiker nicht oder nur partiell nachvollziehbar sind, dargestellt. Dabei wurden in einem ersten Teil zur Optimierung der quantitativen Auffälligkeitseinstufung anhand der existierenden Datengrundlage (d. h. jährliche Erfassung) und in einem zweiten Teil die Erweiterung auf einen Zweijahreszeitraum dargelegt und kritisch bewertet.

Auf Seite 116 wurde dazu eine Strategie für eine zweijährige Auswertung entworfen, die auf den Folgeseiten noch detailreich dargelegt wurde und letztendlich Eingang in das vorgeschlagene Konzept des IQTIG findet, wie sie bereits bei den Plan-QI realisiert wurden. An dieser Stelle sei jedoch bereits bemerkt, dass es aus Sicht der Landesgeschäftsstellen Sachsen dadurch zu einer deutlichen Verzögerung in der Bewertung der Ergebnisse eines Leistungserbringers zu einem Erfassungsjahr kommen dürfte bei bisher bereits einer Betrachtung im Berichtsjahr von Daten aus dem EJ -1, bei Follow-Up-Kriterien und Sterblichkeit EJ -2, so dass dann qualitätsfördernde Maßnahmen möglicherweise nicht mehr zielführend oder viel zu spät übermittelt werden. Eine kritische Stellungnahme dazu war bereits in dem Impulsvortrag „Erste Erfahrungen planungsrelevante QI – Perspektive LQS“ von Frau Dr. Susanne Macher-Heidrich, Geschäftsstelle Qualitätssicherung NRW, anlässlich der 11. Qualitätssicherungskonferenz des GBA 2019 dargelegt worden (<https://www.g-ba.de/service/veranstaltungen/qs-konferenz2019/>).

Des Weiteren wurden im **Abschnitt 5.5 „Zusammenfassung und Empfehlungen“** die statistische Auswertungsmethodik präsentiert und darauf aufbauend neue Begrifflichkeiten definiert. Auf die Problematik von Auswertung bei Leistungserbringern mit kleinen Fallzahlen wurde eingegangen. Die Wahl des Signifikanzniveaus für die statistische signifikante Auffälligkeitseinstufung wurde vom IQTIG empfohlen.

**Kapitel 6 „Fachliche Bewertung“ (S. 143 – 171):**

Dort wurden die bisher verwendeten Begrifflichkeiten „qualitativ auffällig/unauffällig“ durch die neuen Formulierungen „hinreichender/kein hinreichender Hinweis auf ein Qualitätsdefizit“ ersetzt.

Mit dem in Tab. 13 aufgeführten Bewertungsschema (S. 145) wurden die im Workshop diskutierten Aspekte berücksichtigt.

Unter **Punkt 6.2 „Methodik der fachlichen Bewertung“** wurde auf die Beurteilung anhand impliziter (= individuelle Einschätzung der beratenden Experten) bzw. expliziter (= auf Basis standardisierter Informationen) Regeln eingegangen, deren mathematischer Hintergrund in den vorangegangenen Kapiteln hergeleitet wurde. Dabei wurden die Grenzen einer vollständigen expliziten und quantitativen Bewertung dargelegt (S. 148). Ausgehend davon empfiehlt das IQTIG auf Seite 149 „... ein kombiniertes Vorgehen, bei dem Einflussfaktoren, deren Beitrag zum Indikatorergebnis eindeutig angegeben werden kann, in Form einer partiellen Nachberechnung des Indikatorwertes und die übrigen angeführten Einflussfaktoren durch heuristische Beurteilung mit Unterstützung der Expertinnen und Experten berücksichtigt werden.“, welches zusätzlich in Abbildung 31 (S. 149) dargestellt wird.

Unter **Punkt 6.3 „Ablauf der fachlichen Bewertung“** wurde nachfolgend dargelegt, dass durch eine „zuständige Stelle“, die im Text nicht genauer bezeichnet ist (annehmbare Landesgeschäftsstellen bei Verfahren die auf Landesebene ausgewertet werden, IQTIG bei auf Bundesebene auszuwertenden Verfahren) verschiedene Aufgaben zu erledigen sind.

Im weiteren Text des Kapitels (S. 150 ff.) sowie in Abb.32 (S. 151) wurde dann auf die partielle Nachberechnung eingegangen. An dieser Stelle stellt sich die Frage, durch wen und in welchen Zeitraum bzw. mit welchen Ressourcen diese Neuberechnungen erfolgen sollen?

Die Einführung eines sogenannten „informativen“ Indikatorwertes neben dem „originalen“ Indikatorwert führt möglicherweise zu zusätzlichen Verwirrungen bei allen am Verfahren Beteiligten, auch wenn die Begriffe in diesem Vorbericht definiert und dargelegt wurden. Es sei an dieser Stelle betont, dass die Grundvoraussetzung für eine optimale und zielführende Durchführung des Strukturierten Dialogs bzw. des Stellungnahmeverfahrens die Auswahl geeigneter Qualitätsindikatoren (QI) darstellt – Qualitätsindikatoren mit hoher Sensitivität und Spezifität, in bestimmten Fällen risikoadjustiert sowie mit einer eindeutigen und klaren Formulierung der QI in den Dokumentationsbögen und Ausfüllhinweisen für die Leistungserbringer, die Nachberechnungen gar nicht erst erforderlich machen.

Zudem ist die Befürchtung zu äußern, dass sich durch partielle Nachberechnungen, so, wie sie aus dem Berichtstext herauszulesen sind, das Stellungnahmeverfahren erheblich verzögert.

Im letzten Abschnitt zu diesem Unterpunkt (S. 150) wurde formuliert, dass durch „... die zuständige Stelle ... die abschließende Entscheidung in Bezug auf die Einordnung des Indikatorergebnisses in das Bewertungsschema ...“ erfolgt. Aus Sicht der Geschäftsstellen Sachsen wird damit von einem bewährten Vorgehen, das die Einordnung einvernehmlich durch die Arbeitsgruppe/Fachkommission erfolgt, abgewichen.

Hinsichtlich der „Beurteilung der Verantwortlichkeit für einen Einflussfaktor“ (S. 153 – 154) wurde der „zuständigen Stelle“ die Aufgabe zugewiesen, darüber zu befinden, „... ob vom Leistungserbringer mit der gebotenen Sorgfaltspflicht alles Zumutbare getan wurde, um den aus seiner Sicht nicht von ihm zu verantwortenden Einflussfaktor aufzufangen (z. B. durch Einplanung von genügend Ressourcen und genügend Zeit)“. Anhand welcher Kriterien soll dies erfolgen? Das angeführte Beispiel reicht dazu nicht aus Klarheit zu vermitteln. Bereits bei Plan-QI zeigen sich deutliche Schwierigkeiten in der Form, dass „Ausnahmetatbestände“ nicht für alle Konstellationen zu definieren sind.

Insgesamt stellt sich die Frage, in welcher Form tatsächlich die Neuberechnungen erfolgen sollen. Nach dem Verständnis der Geschäftsstellen Sachsen würde bei einer Vielfältigkeit von angeführten Gründen eher noch eine größere Heterogenität der Beurteilungen im Strukturierten Dialog/Stellungnahmeverfahren zu erwarten sein.

Im Unterpunkt 6.4.6 dargelegten „Fazit“ (S. 163 ff.) wurde durch das IQTIG eine kritische Betrachtung der eigenen Vorschläge hinsichtlich deren Machbarkeit bereits vorgenommen, die jedoch den absehbaren Aufwand in einer zu abgemilderten Form aus Sicht der Geschäftsstellen Sachsen abbilden. Es wurden Vorschläge das IQTIG unterbreitet, wie man dieses Procedere technisch unterstützen könnte, zum Beispiel in Form eines webbasierten Portals bzw. einer einheitlichen Softwarelösung für alle Bewertungsstellen. Unberücksichtigt bleibt jedoch die personelle Ausstattung der zuständigen Stelle bzw. Landesstellen.

Unter Punkt 6.5 „Mindestanforderungen an Stellungnahmen“ wurde darauf hingewiesen, dass es zwischen den Bundesländern Unterschiede zwischen den abgeforderten Stellungnahmen von Leistungserbringern hinsichtlich Inhalt und Form gebe. Die vom IQTIG dargelegten formalen Kriterien und der inhaltliche Fokus (S. 165) können Zustimmung durch die Geschäftsstellen Sachsen erhalten bis auf eine Ausnahme: bezüglich Diskussion von Einzelfällen, dem „u. U. kein zusätzlicher Erkenntnisgewinn für die Qualitätsbeurteilung gegenübersteht...“ (S. 166, 2. Abschnitt), die in dieser Verallgemeinerung nicht haltbar ist.

Im Unterpunkt 6.6 „Zusammensetzung von Fachkommissionen“ wurden die anlässlich des Workshops und der LAG- bzw. LQS-IQTIG-Sitzung geäußerten Bemerkungen der Geschäftsstellen berücksichtigt/gewürdigt und Anregungen aufgenommen. Zu begrüßen ist diesbezüglich, dass eine bundesweit einheitliche Regelung für eine angemessene Aufwandsentschädigung empfohlen wird, die eine Wertschätzung dieser ehrenamtlichen Tätigkeit zum Ausdruck bringt.

### **Kapitel 7 „Modul Qualitätsförderung“ (172 – 176):**

In diesem Kapitel wurden die verschiedenen möglichen qualitätsfördernden Maßnahmen dargelegt mit Verweis auf Kompetenzen, die der „zuständigen Stelle“ zugewiesen werden, wie dies bereits in vorausgegangenen Abschnitten und Kapiteln erfolgte. Aus Sicht der Landesgeschäftsstellen ist es zwingend erforderlich, diese „zuständige Stelle“ klar zu definieren (z. B. gem. DeQS-RL § 17 abs. 1 und 2). In der Darlegung ergibt sich teilweise aber auch der Eindruck (S. 174, Abs. 4, S. 175, letzter Absatz), dass damit auf Landesebene z. B. die LAG-Geschäftsstelle nach DeQS-RL oder die LQS nach QS KH-RL gemeint sind

- dann sollten diese auch so benannt oder an irgendeiner relevanten Stelle im Text definiert werden (z. B. im Abkürzungsverzeichnis).

### **Kapitel 8 „Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens“ (S 177 – 186):**

In **Unterpunkt 8.1 „Das Verfahren gemäß DeQS-RL“** wurden aus Sicht der Geschäftsstellen einige Sachverhalte nicht korrekt im Text und in der Abbildung 38 unter Berücksichtigung des Kapitels 7 wiedergegeben bzw. sind die Kapitel in diesem Vorbericht in sich nicht konsistent: Dies betrifft zum einen die Korrekturfrist für die Datenlieferung, die gemäß DeQS-Richtlinie bis zum 15. März des dem Erfassungsjahr folgenden Jahr (= EJ + 1) und nicht, wie angegeben, zum 23. März zu erfolgen hat. Zum anderen soll das Stellungnahmeverfahren gemäß Richtlinie bis zum 31. Oktober abgeschlossen sein mit Empfehlungen für eine Maßnahmenstufe 1, die dem Verständnis der Richtlinie nach und den in diesem Bericht enthaltenen Ausführungen in Kapitel 7 den qualitätsfördernden Maßnahmen entsprechen, die über eine zu benennende Zahl von Monaten durch den Leistungserbringer zu realisieren sind.

Im **Unterpunkt 8.2 „Empfehlungen zur Verkürzung des Stellungnahmeverfahrens“** wurden die Überlegungen des IQTIG dargelegt. Diese beinhalten

- a. Verminderung der Anzahl zu prüfender Qualitätsindikatoren
- b. effizientere Gestaltung des Stellungnahmeverfahrens
- c. mehr Bereitstellung von Personal in den LQS oder LAG, um die Bearbeitung zu beschleunigen.

Bemerkung zu a): Korrekterweise sollte man von einer „Qualitätsverbesserung der anzuwendenden Qualitätsindikatoren“ sprechen.

Bemerkung zu c): Bereitstellung von mehr Personal in den Geschäftsstellen, welches zusätzlich nur kurzfristig mit einer höheren Arbeitsintensität zu betrauen ist, führt nicht automatisch zu einer Prozessbeschleunigung, denn es ist zu berücksichtigen, dass Fristen für Leistungserbringer und für die Vorbereitung der Sitzungen durch die Fachkommission einzuhalten sind. Viel problematischer ist die Tatsache, dass das gesamte Stellungnahmeverfahren über die Sommermonate, d. h. in die Urlaubszeit Juli und August aller Beteiligten am Verfahren (resultierende kritische Besetzungssituation bei den Leistungserbringern) sowie teilweise in die kongressintensiven Zeiten der Fachkommissionsmitglieder fallen. Dazu erfolgen zudem weitere Ausführungen in den entsprechenden Kapiteln.

Die unter **Punkt 8.2.2 „Effizienzsteigerung durch Standardisierung im Vorgehen“** dargelegten Vorschläge für Softwarelösungen bedürfen einer differenzierten Betrachtung. Die „explizite Vorgehensweise“, die mit Nachberechnungen von Qualitätsindikatoren einhergeht, wird keineswegs zu einer Verschlinkung des Verfahrens. Softwarelösungen hingegen, die bei den einrichtungs- und sektorenübergreifenden Verfahren eine bessere Implementierung bestehender IT-Strukturen aus den bisherigen QS KH-Verfahren mit den zur Verfügung stehenden IT-Strukturen der Kassenärztlichen Vereinigung ermöglicht, ist jedoch sehr wünschenswert.

Zu **Punkt 8.2.3 „Auswirkungen der Empfehlungen auf den Beginn der Bewertung von Stellungnahmen“**: Die zeitlichen Betrachtungen in diesem Abschnitt wurden vom IQTIG sehr optimistisch gefasst und entsprechen leider nicht den Tatsachen, die wir unter anderem 2019 feststellen mussten (Datenlieferungen für die DeQS-Verfahren QS PCI: 1. und 4. Jul 2019, wegen Fehlern Nachberechnung 9. Juli 2019), QS WI: 2. und 9. bis 11. Juli 2019 je nach Adressat). Grundvoraussetzung ist, dass eine korrekte Datenlieferung termingerecht erfolgt. Der Aussage, dass eine Aufwandsreduktion erzeugt werden kann, indem „... Aufforderungen zur Stellungnahme nicht mehr individuell angepasst werden müssen (u. a. generisch und nicht QI-spezifisch sind) ... und somit von den zuständigen Stellen ausschließlich die Auswertungen für einen Leistungserbringer zusammen mit einem Standard-Begleitschreiben weitergeleitet werden müssten ...“ (S. 181, Ende erster Absatz) ist vehement zu widersprechen, war auch nicht Tenor im Workshop und entspricht nicht dem bewährten Vorgehen – zumindest in Sachsen. Vielmehr müssen die Aufforderungen zur Stellungnahme zu den einzelnen Qualitätsindikatoren mit möglichst differenzierten und präzisen Fragen erfolgen, was durch vorgefertigte Textbausteine und datenbankgestützte Routinen möglich wird. Derartige differenzierte Texte können selbstverständlich vor Eintreffen der Ergebnisse abgestimmt mit den Fachkommissionen erarbeitet werden. Dadurch kann die Effizienz der Fachkommission erhöht werden. Der vom IQTIG vorgeschlagene dreiseitige Text im Anhang zum Vorbericht unter Kapitel 7 (Seite 61 – 63) enthält relativ allgemeine Formulierungen und verwendet eine Nomenklatur, die für die Leistungserbringer wenig zielführend ist. Dieser Text eignet sich allenfalls für eine allgemeine Erläuterung bzw. zur Einführung.

Der unter **Punkt 8.2.4 „Auswirkungen der Empfehlungen auf den Abschluss des Stellungnahmeverfahrens“** in Abb. 39 aufgezeigte zeitliche Rahmen für das Stellungnahmeverfahren bis zum Abschluss am 15. September ist als unrealistisch bzw. sachlich nicht nachvollziehbar zu betrachten. Das IQTIG berücksichtigte nicht, dass die Fachkommission selbst für die Vorbereitung auf die Sitzung Zeit benötigt. Nach der Sitzung müssen die Bewertungen bearbeitet und, wie bisher lt. DeQS-Richtlinie festgeschrieben, durch das Lenkungsgremium bestätigt werden. Des Weiteren ist zu bedenken, dass eine Vielzahl von Qualitätssicherungs-Verfahren in kürzester Zeit parallel zu bearbeiten sind. Die Fachkommissionssitzungen nehmen Zeit in Anspruch und in der Kardiologie und Orthopädie/Unfallchirurgie reicht meist eine Sitzung der Fachkommission nicht aus. All dies findet sich in den zeitlichen Überlegungen in diesem Vorbericht nicht wieder. Eine Personalaufstockung ist nicht sinnvoll, um die Zeit zu verkürzen, da aus Personalmanagementgründen eine kontinuierliche Tätigkeit mit langfristig zu bindendem, gut qualifiziertem Personal zu bevorzugen ist. Es wurde neben den bereits weiter oben angegebenen Gründen vom IQTIG nicht berücksichtigt, dass auch noch das Lenkungsgremium der Einleitung der Maßnahmenstufe 1 zur Qualitätsförderung gemäß DeQS-Richtlinie zuzustimmen hat – es sei denn, man beabsichtigt, davon Abstand zu nehmen. Dies setzt voraus, dass die Ergebnisse des Stellungnahmeverfahrens vorliegen und dem Lenkungsgremium vorgestellt werden (im Allgemeinen innerhalb einer Sitzung). Da der Prozess der Erstellung des Qualitätssicherungsergebnisberichtes durch die LAG über einen automatisierten Prozess erfolgt, ist nicht nachvollziehbar, weshalb diesen wiederum mehr als fünf Monate zugedacht werden.

Im **Abschnitt 8.3 „Vorschläge der zuständigen Stellen auf Landesebene“** wurde in Abb. 41 (S. 184) ein „schematischer Ablauf der Bewertung von Auffälligkeiten gemäß Empfehlungen“ dargelegt, der in sich nicht logisch ist, denn der Leistungserbringer ist auch über die einzuleitende Maßnahmenstufe 1 oder 2 zu informieren, was logischerweise erst dann erfolgen kann, wenn das Lenkungsgremium zugestimmt hat. Ein Aspekt sei hier noch erwähnt: Die KV Sachsen hebt hervor, dass diese Entscheidung zum Abschluss des Stellungnahmeverfahrens einen Bescheidcharakter trage, und deshalb an die ambulanten Leistungserbringer durch die KV Sachsen selbst übermittelt werden müsse.

Im Workshop wurde von den Landesgeschäftsstellen ausschließlich die Frage gestellt, ob die Lieferung der Rückmeldeberichte früher als am 30.06. erfolgen könnte, nicht aber, dass man die Fristen für die Leistungserbringer und das gesamte Stellungnahmeverfahren verkürzt. Dazu wurden vom IQTIG auf Seite 185 des Vorberichtes Überlegungen dargelegt, wie dies erfolgen könnte. Die Korrekturfrist für die Leistungserbringer besteht bis zum 15.03. nach DeQS-Richtlinie, die nach der QSKH-Richtlinie nicht besteht, dort ist der 28.02. als Frist fixiert. Man könnte über diese 15 Tage diskutieren, d. h. auch unter DeQS-Richtlinie Lieferfrist ohne Korrekturfrist 28.02. eines Jahres, sollte aber nicht beim Leistungserbringer weiter Fristen kürzen, da dieser ohnehin unter einem erheblichen Dokumentations- bzw. allgemeinen Leistungsdruck steht und die Hintergründe für eine derartige Änderung schwierig mit der nötigen Akzeptanz zu vermitteln wären. Bitte führen Sie sich vor Augen, dass der Quartalsbericht für das I. Quartal erst am 1. Oktober geliefert wird, für das II. Quartal am 02.01. des Folgejahres. Die Leistungserbringer haben damit nur wenige Gelegenheiten, eine unterjährige Korrektur anhand der Auswertungen vorzunehmen. Diese vom IQTIG vorgeschlagenen zeitlichen Änderungen bewirken letztlich aber keine Prozessbeschleunigung der Auswertung beim IQTIG. Man sollte aus Sicht der Geschäftsstellen Sachsen den zeitlichen Ablauf derzeit so belassen, wie er in der DeQS-Richtlinie steht, da dieser nach den bisherigen Erfahrungen eher eine Ausgewogenheit zeigt im Vergleich zu den gemachten Vorschlägen.

In **Kapitel 9 „Berichterstattung“** (S. 187 – 191) kann die auf Seite 188, im 1. Abschnitt gemachte Aussage nicht unwidersprochen bleiben. Dieser Abschnitt ist nicht nachvollziehbar, denn die Bewertung der Qualität setzt die Analyse der Ursachen voraus. Mit der getroffenen Entscheidung über die Qualität kann auch sofort entschieden werden, welche Maßnahmen eingeleitet werden sollen. Die Umsetzung dieser qualitätsfördernden Maßnahmen benötigt dann angemessene Zeit beim Leistungserbringer.

#### **Kapitel 10 „Empfehlungen zur Evaluierung des neuen Konzepts“** (S. 192 – 209):

Darin wurde „eine prozessbegleitende, formative Evaluation des Entwicklungskonzepts ...“ empfohlen, „... um bereits vorgeschlagene Änderungen des Strukturierten Dialogs während dessen Implementierung zu begleiten und gegebenenfalls zu optimieren ...“. Dass ein derartiges Vorgehen nicht die Akzeptanz erhöht, zeigt sich derzeit bei QS WI. Man kann eigentlich nur ein Pilotprojekt starten, wenn man Teilaspekte des vorgeschlagenen, aber kritikwürdigen Vorgehens überprüfen möchte.

**Kapitel 11 „Zusammenfassung der Empfehlungen und Konsequenzen“ (S. 210 – 220):**

In dieser Zusammenfassung wurde erstmals im Rahmen dieses Berichtes ausführlicher auf die Anforderungen an und die Weiterentwicklung von Qualitätsindikatoren eingegangen, einer wesentlichen Voraussetzung für die Optimierung des Strukturierten Dialogs bzw. Stellungnahmeverfahrens.

Die im 2. Abschnitt auf S. 216 gemachten Aussagen sind für den Leser nicht schlüssig. Die DeQS-Richtlinie sieht bereits keine „Hinweise“ im Stellungnahmeverfahren vor, wodurch ein Kritikpunkt aus Kapitel 2 bezüglich Heterogenität bei Auslösung des Stellungnahmeverfahrens beseitigt wurde. Es stellt sich aber die Frage, weshalb weniger Stellungnahmen bei Leistungserbringern mit kleinen Fallzahlen ausgelöst werden sollen, da diese, wie die Erfahrungen zeigen, häufig tatsächlich qualitativ auffällig sind und Struktur- und Prozessmängel aufweisen, die zügig beseitigt werden sollten.

Da unter **Punkt 11.5.1 „Überblicksliste der Empfehlungen“** eine Zusammenfassung der bis dahin umfangreichen Betrachtungen erfolgte, sei an dieser Stelle nochmals der Verweis auf die Mehrjahresbetrachtung der Qualität der Leistungserbringer (S. 218) gestattet. Eine Längsschnittbetrachtung der Qualität eines Leistungserbringers ist sicher wünschenswert, sollte aber nicht dazu führen, dass die jährlichen Qualitätsbewertungen mit Empfehlungen zur Qualitätsförderung verlassen werden.

Auf Seite 220 wurde dargelegt, dass unter bestimmten Umständen eine Nachberechnung des Indikatorergebnisses durch die Fachkommission „... mittels technischer Unterstützung“ durchgeführt werden und nachfolgend eine erneute Bewertung erfolgen sollen. Genau dieser Punkt dürfte jedoch zu einer erheblichen Verzögerung des Stellungnahmeverfahrens führen.

Die Zuordnung von Begehungen, Besprechungen bzw. kollegialen Gesprächen zur Maßnahmenstufe 1 ist zu begrüßen.

**Zusammenfassend** ist einzuschätzen, dass in diesem Vorbericht die entscheidungstheoretischen, mathematischen und statischen Modellierung sowie Betrachtungsweisen des Bewertungsprozesses einen großen Anteil einnehmen. Daraus abgeleitete Vorschläge wurden bezüglich Umsetzbarkeit vom IQTIG selbst kritisch betrachtet. Die trotzdem vorgeschlagenen Änderungen können aber von den Landesgeschäftsstellen Sachsen nicht unwidersprochen bleiben. Auf Details wurde im vorangegangenen Text eingegangen.

Die Erfahrungen von 2 Jahren Stellungnahmeverfahren im Rahmen der DeQS-Richtlinie wurden in diesem Vorbericht leider nicht berücksichtigt und sollten in diesen einfließen. Die Erfahrungen von Sachsen z. B. zeigen, dass der in der DeQS-Richtlinie derzeit fixierte zeitliche Rahmen ausgewogen und angemessen erscheint und die Bewertungsvorgaben im Stellungnahmeverfahren nutzbar sind, sinnvollerweise aber einer Präzisierung bedürfen. Die Fortschreibung der Regelungen zum Stellungnahmeverfahren nach DeQS-Richtlinie in der aktuellen Form (ggf. mit geringen Präzisierungen im Bewertungsschema und Verlagerung von „Begehungen“ und „Gesprächen“ unter die qualitätsfördernden Maßnahmen sollte folglich erwogen werden.

Eine Kürzung von Zeiten/Fristen zu Lasten der Leistungserbringer und im Stellungnahmeverfahren sind aus Sicht der Geschäftsstellen Sachsen nicht zielführend.

Nach DeQS-Richtlinie sind „Hinweise“ im Stellungnahmeverfahren, die als ein Grund für die Heterogenität des Strukturierten Dialogs“ herausgearbeitet worden waren, bereits nicht vorgesehen.

Aus Sicht der Geschäftsstellen Sachsen führen die vom IQTIG vorgeschlagenen Änderungen im Strukturierten Dialog/Stellungnahmeverfahren nicht zu der gewünschten Verschlinkung des Verfahrens, sondern vielmehr zu einer im Routinebetrieb wenig strukturierbaren Vorgehensweise und zu einer weiteren Aufblähung, auch wenn die mathematischen Modellierungen eine Diskussionsgrundlage bilden können. Formale Kriterien sollten nicht über medizinische im Stellungnahmeverfahren und in der Qualitätsbeurteilung gestellt werden.

Verschlinkung des Verfahrens bedeutet nach Auffassung der Geschäftsstellen Sachsen, wie bereits im Schreiben vom 13. Juni 2019 zum Rahmenkonzept für die Qualitätsbewertung und –förderung des IQTIG vom 14. Mai 2019 dargelegt, unter anderem:

- Verbesserung der Qualität der entwickelten Qualitätsindikatoren (u. a. mit festen Referenzbereichen, valider Risikoadjustierung), mit den dazugehörigen optimierten Datensätzen und Ausfüllhinweisen für Leistungserbringer sowie der Reduktion von Qualitätsindikatoren bei mangelnder Aussagekraft. Leitlinienänderungen müssen schneller Berücksichtigung finden. Anregungen zur Entwicklung von Qualitätsindikatoren werden regelmäßig von den

Fachkommissionen und in Qualitäts-Konferenzen (siehe unter anderem 11. Qualitätssicherungskonferenz des G-BA vom 26./27. September 2019; Bund-Land-Konferenz Kardiologie vom 5. September 2019) dargelegt.

- Pünktliche, zuverlässige und fehlerfreie Bereitstellung der Spezifikation und einer QIDB (unbedingt auch in maschinenlesbarer Form).
- Beschlussfassung des G-BA zu den endgültigen Rechenregeln mit Bereitstellung der Jahresauswertungen, da bei relevanten Änderungen in den Rechenregeln, diese den Fachkommissionen und Geschäftsstellen („zuständigen Stellen“) zum Stellungnahmeverfahren zur Verfügung stehen müssen (siehe QS WI).
- Pünktliche und fehlerfreie Bereitstellung der Jahresauswertung durch das IQTIG – nach Möglichkeit vor dem 30.06. – ohne Kürzung von Zeiten für Leistungserbringer oder Geschäftsstellen – bereits 15.06. wäre hilfreich.
- Die in der DeQS-Richtlinie derzeitig geregelte jeweilige Zustimmung durch das Lenkungsgremium, sollte zumindest bei Aufforderung des Leistungserbringers zu einer schriftlichen Stellungnahme überdacht/verschlinkt werden.
- Bereitstellung der mandantenfähigen Datenbank mit praktikablem, rasch realisierbarem Zugriff zur Bereitstellung wichtiger Informationen für die Fachkommissionen im Stellungnahmeverfahren als Ergänzung zu den Stellungnahmen der Leistungserbringer.

Bei grundsätzlichen Änderungen, z. B. in Form eines vielschichtig gegliederten Bewertungsvorgehens, wie im Bericht des IQTIG dargelegt, ist es aus Sicht der Geschäftsstellen Sachsen unbedingt notwendig, zunächst eine Machbarkeitstestung im Rahmen eines Pilotprojektes durchzuführen, bevor ein Übergang in den Routinebetrieb vorgenommen werden kann.

Mit freundlichen Grüßen



Dr. med. Beate Trausch  
Leiterin  
Landesgeschäftsstelle einrichtungs- und  
sektorenübergreifende Qualitätssicherung



Dipl.-Med. Annette Kaiser  
Leiterin  
Projektgeschäftsstelle Qualitätssicherung

## **Stellungnahme**

### **zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern**

**Vorbericht zu Stufe 1 und Stufe 2 vom 10.10.2019**

### **des IQTIG**

der Landesgeschäftsstellen für Qualitätssicherung

Berlin

Mecklenburg-Vorpommern

Rheinland-Pfalz

Saarland

Thüringen

Stand: 18.11.2019

**Inhalt**

Allgemein .....	3
Beweislastumkehr von der zuständigen Stelle zum Leistungserbringer .....	3
Trennung von Qualitätsbewertung und Qualitätsförderung .....	4
Auslösung des Stellungnahmeverfahrens.....	5
Fachliche Bewertung der Stellungnahmen .....	6
Mindestanforderungen an Stellungnahmen .....	12
Zusammensetzung der Fachkommissionen .....	12
Qualitätsförderung .....	13
Zielvereinbarung.....	14
Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens.....	14
Anhang Textvorschläge Einholung Stellungnahmen.....	16
Zusammenfassung .....	18

## Allgemein

Der Bericht wurde aufgrund des Auslaufens der QSKH-RL vor allem mit Blick auf das in § 17 DeQS-RL beschriebene Stellungnahmeverfahren entwickelt. Gleichzeitig erfolgt bei vielen Empfehlungen eine Adaption des Prüfverfahrens bei den planungsrelevanten Qualitätsindikatoren. Dies mag im Sinne einer Vereinheitlichung des Verfahrens sinnvoll sein, negiert aber die unterschiedlichen Zielsetzungen beider Richtlinien. Das IQTIG versucht in dem vorliegenden Bericht die verschiedenen Ziele, wie Versorgungsplanung, Vergütung und Berichterstattung in einem einheitlichen Stellungnahmeverfahren zusammenzufassen. Dies muss jedoch grundsätzlich hinterfragt werden, denn gerade die verschiedenen Ziele drücken sich in unterschiedlichen Richtlinien aus und müssen nicht zwingend ein einheitliches Vorgehen erfordern. Aufgrund der aktuell gleichermaßen verwendeten Qualitätsindikatoren könnte dies implizit gewünscht sein, kann sich aber ändern, wenn zukünftig andere Arten von Qualitätsindikatoren verwendet werden. Besonders deutlich wird dies beim Thema: Indizes.

*Das vorgeschlagene Konzept vom IQTIG sieht vor, dass das Ergebnis jedes einzelnen Qualitätsindikators abschließend bewertet werden soll. Dieses Vorgehen ist damit auch im Einklang mit dem Vorgehen gemäß plan. QI-RL. Die diesem Punkt der Beauftragung scheinbar zugrunde liegende Annahme, dass eine Bewertung einzelner Indikatorergebnisse verzichtbar sei, wird daher nicht geteilt. Aus diesen Gründen wird eine Verwendung von aggregierten Qualitätsindikatorergebnissen als Aufgreifkriterien nicht empfohlen. (S.213)*

Dieser Argumentation muss in Teilen widersprochen werden, denn Indizes können sehr wohl als Aufgreifkriterien für die Auslösung des Stellungnahmeverfahrens verwendet werden. Als Beispiel sei hier der Qualitätsindex zum kritischen Outcome bei Reifgeborenen in der Geburtshilfe genannt. Dieser basiert auf Ergebnissen einzelner Qualitätsindikatoren, so dass das Ergebnis jedes einzelnen Qualitätsindikators abschließend bewertet werden kann. Werden also Qualitätsindikatoren einzeln berechnet und dann zu einem Index zusammengefasst, könnte sowohl die Auslösung des Stellungnahmeverfahrens spezifischer erfolgen und gleichzeitig die Indikatoren bewertet werden. Diese Methodik wird in dem Bericht des IQTIG jedoch nicht untersucht und ist daher zu kritisieren, gerade weil dies erwartet werden durfte.

## Beweislastumkehr von der zuständigen Stelle zum Leistungserbringer

*In der aktuellen Form des Strukturierten Dialogs liegt die Beweislast bei den Bewertungsstellen, ein bestehendes Qualitätsdefizit anhand der Stellungnahme des Leistungserbringers genau benennen zu können (S.43) Das IQTIG empfiehlt nunmehr eine Beweislastumkehr*

*von der zuständigen Stelle zum Leistungserbringer. Zukünftig soll die Abgabe einer Stellungnahme durch den Leistungserbringer freiwillig sein. Wird seitens des Leistungserbringers auf eine Stellungnahme verzichtet, gilt gemäß des Verständnisses von Qualitätsindikatoren (siehe Abschnitt 3.1), dass ein hinreichender Hinweis für ein Qualitätsdefizit vorliegt, der Indikator in seiner Validität nicht angezweifelt wird und somit ein Qualitätsdefizit vorliegt. (S.69)* Dies erschließt sich aus den vorliegenden Richtlinien nicht und sollte daher einer rechtlichen Prüfung unterzogen werden. Das IQTIG selbst begründet die Beweislastumkehr in seinem Bericht nicht.

### **Trennung von Qualitätsbewertung und Qualitätsförderung**

*Es wird empfohlen, den Begriff des Stellungnahmeverfahrens nur für den Teil des Vorgehens nach § 17 DeQS-RL zu verwenden, der die Einholung und abschließende Bewertung der Stellungnahmen umfasst (Qualitätsbewertung). Die darauf aufbauenden Maßnahmen der Qualitätsförderung sollen im Sinne der Trennung von Qualitätsbewertung und Qualitätsförderung nicht als Teil des Stellungnahmeverfahrens verstanden werden. Diese Trennung kann auch dazu beitragen, dass die Qualitätsergebnisse früher verfügbar sind als bisher. (S.215)* Dies kann grundsätzlich befürwortet werden, wirft aber die Frage auf, woran sich Sanktionen der Maßnahmenstufe 2 nach § 17 DeQS-RL anschließen. So sind Fördermaßnahmen wenig akzeptiert, wenn bereits aufgrund der Qualitätsbewertung ein Vergütungsabschlag erfolgt bzw. eine Abrechnungsgenehmigung entzogen wurde. Folgt man dem Vorschlag des IQTIG müssen Bewertung, Förderung und Sanktion zwingend in der Richtlinie aufeinander aufbauend beschrieben werden.

*Das IQTIG vertritt die Auffassung, dass Empfehlungen im Rahmen eines Peer Reviews weniger gut angenommen werden können, wenn gleichzeitig die Qualitätsbewertung noch aussteht. (S.66)*

Hier zeigt die Praxis das genaue Gegenteil. Die Möglichkeit des Austausches auf Augenhöhe schafft Akzeptanz bei den Leistungsringern und ermöglicht eine bessere Bewertung der Versorgungssituation. Gerade bei wiederholt rechnerischen Auffälligkeiten kann eine Begehung oder ein Kollegiales Gespräch effektiver sein als eine erneute Stellungnahme. Daher sollte nicht wie vom IQTIG empfohlen, auf Kollegiale Gespräche und Begehungen verzichtet werden, sondern diese Instrumente weiterentwickelt werden, um eine Standardisierung und Objektivität der Bewertung zu ermöglichen. Beispielweise ist nicht festgelegt, welche formalen und inhaltlichen Kriterien die Gespräche beinhalten.

*Das IQTIG geht davon aus, dass auch für so genannte Follow-up-Indikatoren, deren Auswertung sich auf Ereignisse (z. B. Revisionseingriffe oder der Tod einer Patientin oder eines Patienten) innerhalb eines bestimmten Beobachtungszeitraumes im Anschluss an ein interessierendes Ereignis (z. B. eine Organtransplantation) bezieht, das vorgeschlagene Verfahren der Qualitätsbewertung angewendet werden kann. (S. 67)*

Dies scheint angesichts der wenigen praktischen Erfahrungen im Strukturierten Dialog und den Schwierigkeiten bei der qualitativen Bewertung der Follow-up-Indikatoren nicht sinnvoll zu sein. Das IQTIG selbst räumt Probleme der Zuschreibbarkeit der Verantwortung und der Datenqualität ein.

### **Auslösung des Stellungnahmeverfahrens**

*Feste Referenzwerte führen bei kleinen Fallzahlen eher zur rechnerischen Auffälligkeit. (S.49) Problematisch ist auch die Berechnung von perzentilbasierten Referenzwerten und die damit verbundene Auffälligkeitseinstufung, da die gegenwärtig verwendete Methodik dazu führt, dass insbesondere bei Qualitätsindikatoren mit einem hohen Anteil von Standorten mit einer Fallzahl kleiner 20 weit mehr Standorte auffällig werden, als durch das nominell angestrebte Niveau des Perzentilreferenzbereiches. (S.53)*

Daher wird die Einführung der statistischen Auffälligkeit als Hinweis auf ein Qualitätsdefizit befürwortet, aber ein Signifikanzniveau wird nicht festgelegt. *Das IQTIG empfiehlt eine sehr pragmatische Vorgehensweise, wie z. B. übergangsweise das Signifikanzniveau so zu wählen, dass in etwa der gleiche Aufwand durch Stellungnahmen resultiert. (S.216)*

Ziel war jedoch eine Reduktion der Stellungnahmen.

*Für die Auslösung des Stellungnahmeverfahrens empfiehlt das IQTIG die bayesianische Version der statistisch signifikanten Auffälligkeitseinstufung basierend auf den Daten zweier Erfassungsjahre. (S.141)*

Unklar geblieben ist, wie in den Folgejahren die Auslösung und Bewertung erfolgen soll. In der Vergangenheit haben die auf Landesebene zuständigen Stellen stets betont, dass eine erneute Ergebnisbewertung erst im Erfassungsjahr +2 sinnvoll ist, da die Qualitätsbewertung im Erfassungsjahr +1 zu einem Zeitpunkt erfolgt, wo Verbesserungsmaßnahmen erst spät greifen können. Bei einer Berechnung zweier aufeinander folgenden Jahre ist zu erwarten, dass die Ergebnisse im aktuellen Erfassungsjahr zum gleichen Ergebnis wie im Vorjahr führen. Interessant wäre daher, eine Modellrechnung vom IQTIG, wie die Ergebnisse der statistisch signifikanten Auffälligkeitseinstufung basierend auf den Daten von drei Erfassungsjahren aussehen. Da diese Ergebnisse nicht untersucht wurden, stellt sich die Frage, ob die

statistisch signifikante Auffälligkeitseinstufung eventuell nur alle 2 Jahre berechnet wird. Dies widerspricht jedoch dem Ansatz des IQTIG, für jedes Indikatorenergebnis eine jährliche Qualitätsbewertung vorzunehmen. Dem Bericht war nicht eindeutig zu entnehmen, ob die Fälle zweier Erfassungsjahre oder nur die Fälle des letzten Erfassungsjahres in das Stellungnahmeverfahren eingehen.

*Es wird zwar ausgeführt, dass wenn im Vorjahr bei einem Leistungserbringer ein Stellungnahmeverfahren ausgelöst wurde, bei dem der Leistungserbringer bei der Bewertung dann abschließend als „unauffällig“ eingestuft wurde, so könnte man dies als Anlass sehen, zur Bewertung des aktuellen Jahres keine Vorjahresdaten hinzuzunehmen. (S.165)*

Dies bezieht sich nach unserem Verständnis jedoch auf die Indikatorenberechnung, aber nicht zwingend auf die „Fälle mit einem möglichen Qualitätsdefizit“. Dies sollte eindeutig klargestellt werden, da diese Entscheidung die sich anschließenden Prozessschritte beeinflusst. Die 2jährige Datenbasis wirft zudem weitere Fragen auf. So stellt sich die Frage, wie bei kleinen Fallzahlen besser stochastischen Einflüssen begegnet werden kann, wenn die Vorjahresdaten aufgrund eines Stellungnahmeverfahrens im Vorjahr nicht bei der quantitativen Qualitätsbewertung berücksichtigt werden. Hier gebe es dann ausschließlich die statistische Auffälligkeit im aktuellen Erfassungsjahr. Gerade vor dem Hintergrund der gewünschten Reduktion des Aufwandes und der zu erreichenden Effizienz des Stellungnahmeverfahrens würden mit einer nur teilweisen Anwendung der 2 Jahres-Laufregel nicht die gewünschten Effekte erzielt.

### **Fachliche Bewertung der Stellungnahmen**

*Im Bewertungsschema gemäß Tabelle 13 wird zwischen „Qualitätsdefizit“ auf der einen Seite und „kein Hinweis auf Qualitätsdefizit“ auf der anderen Seite unterschieden. (S.145)*

Die Kategorie „Qualitätsdefizit“ beinhaltet eine Bewertung „A0“, die beim Verzicht auf Einreichung einer Stellungnahme oder wenn die Stellungnahme nicht fristgerecht eingereicht oder wenn die Stellungnahme nicht den Anforderungen entsprach, anzuwenden ist. Diese Bewertungskategorie impliziert eine fehlende Mitwirkung des Leistungserbringers, was jedoch nicht zwingend gegeben sein muss. Es sollte schon unterschieden werden, ob der Leistungserbringer bewusst verzichtet, z.B. weil das Qualitätsdefizit aufgrund der quantitativen Auffälligkeitsprüfung anerkannt wird, oder die Stellungnahmefrist nicht eingehalten wurde.

*Die fachliche Bewertung hat die Funktion, die in der Stellungnahme angeführten Gründe für das Verfehlen des Referenzbereichs zu prüfen und zu beurteilen, ob eine besondere Konstellation vorliegt, die den Hinweis auf ein Qualitätsdefizit aus dem Indikatorergebnis entkräft-*

*tet. Unter einer besonderen Konstellation wird in diesem Bericht das Vorliegen eines oder mehrerer Einflussfaktoren auf das Indikatorergebnis bei einem Leistungserbringer verstanden, die nicht von diesem zu verantworten sind und die in ihrer Gesamtheit das Verfehlen des Referenzbereichs bei diesem Qualitätsindikator erklären. (S.147)*

Offen hingegen bleibt, wann dem Leistungserbringer eine Verantwortung zugeordnet werden kann. Da dies im Bericht nicht definiert wird, ist zu erwarten, dass die Verantwortlichkeit bei den zuständigen Stellen sehr unterschiedlich interpretiert wird. Folgt man der Argumentation des IQTIG, so führt genau diese fehlende Definition zu der angesprochenen Heterogenität der Bewertungen. Das Ziel einer Standardisierung und Objektivität der Bewertung würde verfehlt. Das IQTIG selbst räumt ein, dass es bei manchen Einflussfaktoren erforderlich sein kann, die Verantwortlichkeit des Leistungserbringers je nach Konstellation des Behandlungsfalls unterschiedlich einzuschätzen.

*Für solche Einflussfaktoren muss entschieden werden, ob die Verantwortlichkeit für jeden Behandlungsfall eindeutig als gegeben oder nicht gegeben klassifiziert werden kann und sie sich für eine partielle Nachberechnung auf Grundlage der anerkannten Behandlungsfälle eignen, oder ob die Auswirkung auf das Indikatorergebnis für diesen Einflussfaktor nicht zuverlässig angegeben werden kann. (S.154)*

Die Einzelbeurteilung je Behandlungsfall macht das Verfahren sehr aufwendig und ist unter dem Aspekt der geforderten Effizienz kritisch zu hinterfragen.

*Die in einer Stellungnahme genannten Gründe müssen das Vorliegen solcher Einflussfaktoren nachvollziehbar darstellen. Dazu würden für die Einflussfaktoren vorab ein Algorithmus und quantitative Gewichte festgelegt werden, anhand derer eine Neubewertung unter Berücksichtigung dieser Faktoren durchgeführt würde (z. B. in Form einer Nachberechnung des Indikatorwerts für diesen Leistungserbringer und eines erneuten Vergleichs mit einem ggf. angepassten Referenzbereich). Dazu würden für die Einflussfaktoren vorab ein Algorithmus und quantitative Gewichte festgelegt werden, anhand derer eine Neubewertung unter Berücksichtigung dieser Faktoren durchgeführt würde (z. B. in Form einer Nachberechnung des Indikatorwerts für diesen Leistungserbringer und eines erneuten Vergleichs mit einem ggf. angepassten Referenzbereich). Es lässt sich jedoch für manche Faktoren nicht zuverlässig angeben, wie sie bei einer Nachberechnung berücksichtigt werden sollten. (S.147/148) Das IQTIG empfiehlt daher ein kombiniertes Vorgehen, bei dem Einflussfaktoren, deren Beitrag zum Indikatorergebnis eindeutig angegeben werden kann, in Form einer partiellen Nachberechnung des Indikatorwertes und die übrigen angeführten Einflussfaktoren durch heuristische Beurteilung mit Unterstützung der Expertinnen und Experten berücksichtigt werden. (S.149)*

Der Prozess der Nachberechnung wird vom IQTIG hinreichend beschrieben und soll im Nachgang weiter diskutiert werden. Unklar hingegen bleibt, wie konkret die Gewichtung zwischen partieller Nachberechnung und heuristischer Beurteilung erfolgen. Die Ausführungen des IQTIG hierzu werden so verstanden, dass es zwar einen Bewertungsalgorithmus gibt, dieser wird aber nicht näher beschrieben.

*Hierzu führt das IQTIG aus, dass die Bewertung sich daran orientieren soll, in welchem Ausmaß die Datenfehler einerseits und die inhaltlichen Einflussfaktoren andererseits ausschlaggebend für die Bewertung als „Qualitätsdefizit“ waren. Eine exakte Quantifizierung des jeweiligen Anteils ist jedoch nicht möglich. (S. 160)*

Es ist also nicht erkennbar, wie Datenfehler und z.B. nicht leitlinienkonforme Behandlung zueinander gewertet werden. Aus dem jetzigen Bericht und den Erfahrungen bei den planungsrelevanten Indikatoren entsteht der Eindruck, dass im ersten Schritt unter Berücksichtigung der anerkannten Einflussfaktoren eine Nachberechnung erfolgt, die zu einer Änderung des Nenners, Zählers und/oder Referenzbereiches führen kann, und dass das bereinigte Ergebnis (informative Indikatorwert) nur dann einer heuristischen Beurteilung unterzogen wird, wenn der informative Indikatorwert vom Referenzbereich abweicht. Dies würde jedoch in unzulässiger Weise die fachliche Expertise der Fachkommission vernachlässigen.

*Nach Abschluss der formalen Prüfung werden anhand der in der jeweiligen Stellungnahme angeführten Gründe die adressierten Einflussfaktoren von der zuständigen Stelle extrahiert und ihre Art entweder als Datenfehler oder als fachlich-inhaltlicher Einflussfaktor klassifiziert. (S. 149)*

Üblicherweise werden in der Praxis jedoch Datenfehler im Rahmen der Datenvalidierung, also im Rahmen der Überprüfung der Dokumentationsqualität, identifiziert. Anhand der Patientenakte wird je Datenfeld die Übereinstimmung mit den QS-Daten ermittelt. Im Rahmen des Stellungnahmeverfahrens kann eine solche Prüfung in der Regel nur grob ohne Bezug zum Datenfeld und Patientenpseudonym erfolgen. Die Praxis zeigt bereits heute, wie schwer es ist, die fehlerhaften Datenfelder, die zur rechnerischen Auffälligkeit geführt haben, dem Leistungserbringer aufzuzeigen. Je komplexer die Berechnung des Indikators, desto schwieriger ist die Fehlerermittlung. Beispielsweise kann der Nachweis beim Indikator zu leitlinienkonformen Indikation bei HSM/ICD aufgrund der Komplexität der Berechnung der verschiedenartigen Indikationen kaum geführt werden. Insofern stellt sich die Frage, wie verlässlich sind vom Leistungserbringer angegebene Datenfehler und wie können diese validiert werden.

*Die partielle Nachberechnung erfolgt also zwangsläufig im Austausch mit der Fachkommission, was eine angemessene technische Unterstützung erfordert. (S. 150)*

Neben der technischen Unterstützung wird dieser Part der fachlichen Bewertung jedoch zusätzliche Ressourcen bei den zuständigen Stellen erfordern, da dieser Verfahrensschritt neu in den Bewertungsablauf zu integrieren ist und bislang kein obligates Handeln darstellt. Der Umfang wird erheblich sein, da ggf. jede Stellungnahme die Berechnung eines informativen Indikatorwertes erfordert. Dies sollte mit Bezug auf die Fristen im Stellungnahmeverfahren, die nachfolgend thematisiert werden, zwingend berücksichtigt werden.

*Die Beurteilung, welche der in der Stellungnahme eines Leistungserbringers angeführten Datenfehler vom Leistungserbringer zu verantworten sind, trifft die zuständige Stelle. Die Beurteilung, ob ein fachlich-inhaltlicher Einflussfaktor vom Leistungserbringer zu verantworten ist, erfolgt grundsätzlich unter Beratung der zuständigen Stelle durch die Expertinnen und Experten der Fachkommission. (S.154)*

Diese Regelung entspricht nicht den derzeit gültigen Richtlinienbestimmungen (QSKH-, DeQS-RL). Da wie bereits beschrieben, Datenfehler nicht einfach eindeutig zu identifizieren sind, sollten alle Einflussfaktoren – auch die Datenfehler – durch die Expertinnen und Experten der Fachkommission beraten werden.

*Das IQTIG schreibt, dass aus methodischer Sicht für Indikatoren mit einem verteilungsbezogenen Referenzbereich das Einstufungsergebnis „keine Qualitätsbewertung“ (S) lauten sollte, wenn nicht bereits die partielle Nachberechnung zu einer Entkräftung des Hinweises auf ein Qualitätsdefizit geführt hat. Da derzeit für viele Qualitätsindikatoren der externen Qualitätssicherung noch keine festen Referenzbereiche definiert sind, hätte ein solches Vorgehen den Nachteil, dass trotz aufwändiger Datenerhebung keine Qualitätsaussage getroffen werden würde. Als Interimslösung empfiehlt das IQTIG daher, perzentilbasierte Referenzbereiche im Stellungnahmeverfahren wie feste Referenzbereiche zu behandeln. (S.158)*

Diese Aussage ist nicht nachvollziehbar, da auch bei den planungsrelevanten Qualitätsindikatoren perzentilbasierte Referenzbereiche verwendet werden (z.B. beim Qualitätsindex zum kritischen Outcome bei Reifgeborenen in der Geburtshilfe). Diese werden auch im Rahmen des Stellungnahmeverfahrens mit zureichender bzw. unzureichender Qualität bewertet. Da bei den Indikatoren der datengestützten einrichtungsübergreifenden Qualitätssicherung zu klären ist, ob sich ein hinreichender Verdacht für ein Qualitätsdefizit bestätigt oder nicht – also nicht geklärt wird, ob ein zureichende Qualität vorliegt – besteht keine generelle Notwendigkeit, die perzentilbasierten Referenzbereiche in feste Referenzbereiche zu überführen.

*Das IQTIG vermutet, dass einige der im dargestellten Bewertungsalgorithmus explizit formulierten Schritte schon so oder so ähnlich implizit angewendet werden. Es hält es beispiels-*

*weise für plausibel, dass Ad-hoc-Nachberechnungen von Indikatorwerten nach Ausschluss bestimmter Fälle durchgeführt werden, um zu prüfen, ob der Ausschluss der Fälle dazu führt, dass der Referenzwert erreicht wird. (Seite 163)*

Dieser Vermutung muss explizit widersprochen werden. Der Ausschluss bestimmter (begründeter) Fälle und ein Heranziehen des „korrigierten“ Ergebnisses zur Bewertung würden den in der DeQS-RL unter § 1 Abs. 2 formulierten Zielen deutlich entgegenstehen, weil bei einem solchen Prozedere unter Umständen anhand der übrigen Fälle festgestellte Struktur- und Prozessmängel unberücksichtigt blieben.

*Für die Umsetzung des vorgeschlagenen Bewertungsprozesses empfiehlt das IQTIG, geeignete technische Unterstützung für die Durchführung der fachlichen Bewertung zur Verfügung zu stellen. Dies könnte beispielsweise in Form eines für alle Bewertungsstellen einheitlichen Web-basierten Portals erfolgen, welches Funktionalitäten für die Berechnung bereinigter Ergebnisse bietet und die einzelnen Schritte des Bewertungsprozesses sequentiell abfragt. Eine einheitliche Softwarelösung für alle Bewertungsstellen würde zudem Heterogenität in der Bewertung aufgrund unterschiedlicher Softwarelösungen bei den Landesstellen vermeiden. (S.164)*

Diese Empfehlung beinhaltet zwei Probleme. Zum einen wird deutlich, dass eine höhere Standardisierung auch mit teilweise komplexeren Prozessen einhergeht. Demzufolge darf erwartet werden, dass der Prozess der Nachberechnung mit einem hohen Aufwand für die zuständigen Stellen verbunden ist. Daher muss die Nachberechnung zwingend als vorgelagerter Schritt vor den Sitzungen der Expertengremien implementiert werden.

*Allerdings beschreibt das IQTIG auch, dass nicht jeder Einflussfaktor kategorisiert werden kann (S.155), und damit frühestens in den Sitzungen der Expertengremien über die Berücksichtigung bzw. Nichtberücksichtigung bei der Nachberechnung entschieden werden kann. Praktisch ist aber nur sehr schwer vorstellbar, dass Nachberechnungen mittels eines Webportals in der Sitzung erfolgen können. Folglich wird es auch Nachberechnungen geben, die erst in einer zweiten Sitzung mit den Experten abschließend beraten werden können. Dies widerspricht aber dem Ziel, dass Verfahren zu verkürzen und die Ergebnisse schneller verfügbar zu machen.*

Zum zweiten lässt sich aufgrund der beschriebenen Prozessparameter, die bei einer Nachberechnung zu berücksichtigen sind (bereinigter Referenzbereich, Toleranzanteil), nur schwer vorstellen, wie diese mittels eines Web-basierten Portals ohne Berücksichtigung der besonderen Versorgungskonstellation zum jeweiligen Indikatorergebnis abgebildet werden können. Soll der Prozess der Nachberechnung wirklich umgesetzt werden, kann dies nur mit dem Software-Know-how bei den Landesstellen gelingen. Bis heute gibt es keine prakti-

schen Erfahrungen in dem Verfahren, die derart komplexe Datenauswertungen webbasiert ermöglichen.

*Das IQTIG beschreibt formale Kriterien, die eine Stellungnahme erfüllen soll. Entspricht eine Stellungnahme nicht diesen Kriterien, soll die Stellungnahme als nicht abgegeben gelten. Ist in diesem Fall die empfohlene 4-Wochen-Frist für die Einreichung einer Stellungnahme noch nicht verstrichen, soll dem Leistungserbringer die Möglichkeit gegeben werden, eine den Kriterien entsprechende Stellungnahme einzureichen. (S.165)*

Hier stellt sich die Frage, ob ein solches Vorgehen einer rechtlichen Bewertung standhält. Ist dem Leistungserbringer mit der 4-Wochen-Frist ausreichend Möglichkeit zur Stellungnahme gegeben oder wird es nicht bestimmte Konstellationen geben, wo die Nichteinhaltung der Frist nicht durch den Leistungserbringer zu verantworten ist. Denkbar wären technische Probleme bei der Abgabe der Stellungnahme. Wäre es daher nicht ratsam, eine weitere Eskalationsstufe einzubauen (Erinnerung, Mahnung, Fristverlängerung)? Die meisten zuständigen Stellen auf Landesebene haben solche Eskalationsstufen implementiert, die sich auch in der Praxis bewährt haben. Darüber hinaus stellt das IQTIG zur Erreichung der Bewertung formale Fehler sehr deutlich über fachliche Inhalte der Stellungnahmen. Ein solches Vorgehen muss im Sinne einer am Qualitätsziel orientierten Bewertungen kritisch hinterfragt werden.

*In den Stellungnahmen können für das Verfehlen des Referenzbereichs zum einen auf einzelne Behandlungsfälle bezogene Gründe angegeben werden (z.B. dass ein Notfall ein Abweichen vom üblichen Vorgehen erforderlich machte) und zum anderen übergeordnete Gründe (z.B. dass wegen einer Spezialisierung der Einrichtung auf bestimmte Behandlungen die Patientinnen und Patienten ein besonderes Risikoprofil haben). Das IQTIG empfiehlt jedoch die Stellungnahme möglichst auf eine Analyse von übergeordneten Gründen zu fokussieren. (S.165)*

Dies hängt letztendlich aber von der Art des Indikators ab. Komplikationsraten werden immer auf den einzelnen Behandlungsfall fokussieren, während bei der Bestimmung der Gehfähigkeit in der Regel übergeordnete Gründe eine Rolle spielen. Unabhängig davon sieht das Konzept vom IQTIG jedoch auch bei übergeordneten Gründen die Möglichkeit zur Nachberechnung vor. Diese kann aber nur über eine Zuordnung der Behandlungsfälle erfolgen, d.h. auch hier wird künftig eine Einzelfallbewertung erfolgen müssen. In Kombination mit der Dokumentation des Bewertungsprozesses ergibt sich ein hoher bürokratischer Aufwand, der zwar Transparenz schafft, aber ein höherer Nutzen nicht mit Sicherheit vorherbestimmt werden kann.

*Sowohl bei Einzelfall-bezogenen als auch bei übergeordneten Gründen ist es wichtig, dass der Leistungserbringer in seiner Stellungnahme für jeden Grund angibt, auf welche seiner Behandlungsfälle dieser Grund zutrifft bzw. ob der Grund für alle Behandlungsfälle zutrifft. Dazu sind die Vorgangsnummern der entsprechenden Behandlungsfälle mit der Stellungnahme zu übermitteln. (S.166)*

An dieser Stelle wird darauf hingewiesen, dass nicht alle Indikatoren so konzipiert sind, dass Vorgangsnummern zur Verfügung stehen (z.B. AK 851805: Relative Differenz zwischen den Angaben in der QS-Dokumentation und der Risikostatistik: mehr Dekubitus in der QS-Dokumentation als in der Risikostatistik). Hier wird expliziert ein Auffälligkeitskriterium genannt, in der Annahme, dass die statistische Basisprüfung ebenfalls von der Weiterentwicklung des Strukturierten Dialoges umfasst ist.

### **Mindestanforderungen an Stellungnahmen**

*Die Stellungnahme ist in Schriftform zu erstellen und einschließlich eventuell notwendiger Anhänge in einem Dokument in digitaler Form einzureichen. (S.165)*

Aus welchem Grund wird den Geschäftsstellen vorgeschrieben, ein digitales Dokument entgegennehmen zu müssen? Dies wird im Konzept nicht begründet. Bei vielen medizinischen Dokumenten (z.B. Kurven A3 Format) hat sich die Papierform bewährt, da diese auch wieder den Fachkommissionen in Papierform zugänglich gemacht wird. Auch hier sollte jede LAG gemäß ihren gewachsenen Strukturen vorgehen dürfen.

### **Zusammensetzung der Fachkommissionen**

*Das IQTIG empfiehlt, eine fachlich qualifizierte, unabhängige und interdisziplinäre Fachkommission zusammenzusetzen, um die Akzeptanz des Gesamtverfahrens sicherzustellen und Qualitätsdefizite erfolgversprechend abzubauen. Die Verantwortung für die Zusammensetzung einer fachlich qualifizierten, unabhängigen und interdisziplinären Fachkommission soll zukünftig bei der zuständigen Stelle liegen. Zur Steigerung der Akzeptanz des Verfahrens gegenüber den Expertinnen und Experten der Fachkommission wird eine bundesweit einheitliche Regelung zur Aufwandsentschädigung, die gleichermaßen für alle Expertinnen und Experten gilt, empfohlen. Auch im IQTIG-LQS-Treffen am 18. Juni 2019 wurde eine einheitliche Aufwandsentschädigung seitens der Teilnehmerinnen und Teilnehmer befürwortet. (S.167)*

Dieser Aussage muss entschieden widersprochen werden. Dem Protokoll zum IQTIG-LQS-Treffen am 18. Juni 2019 ist zu entnehmen, dass seitens einiger Teilnehmerinnen und Teilnehmer angemerkt wird, dass eine einheitliche Aufwandsentschädigung aller Mitglieder der Expertengremien geleistet werden sollte. Ein einheitliches Votum wurde in der Sitzung und auch danach nicht abgefragt.

Bezogen auf die Empfehlung nach einer fachlich qualifizierten, unabhängigen und interdisziplinären Fachkommission stellt sich die Frage, ob die Experten zukünftig nicht mehr durch die Vertragspartner entsendet werden sollen. Wenn dem so ist, müsste die Besetzung der Expertengremien zukünftig ausgeschrieben werden. Dies scheint jedoch angesichts des Personalmangels sowohl bei den Ärzten als auch in der Pflege ein wenig erfolversprechendes Modell zu sein. Hier sollte der bisherige Status quo beibehalten werden, wonach die Vertragspartner die Besetzung der Expertengremien sicherstellen.

*Zur Sicherstellung der fachlichen Qualifikation empfiehlt das IQTIG, dass die zu berufenden Ärzte nach der Facharztausbildung die letzten drei Jahre vor ihrem Mitwirken in der Fachkommission im entsprechenden Versorgungsbereich praktisch tätig gewesen sein sollen. Dies soll analog auch für die Pflege und Therapeuten gelten. (S.168)*

Damit wäre unter Umständen eine Mitarbeit von Mitarbeitern des Medizinischen Dienstes der Krankenkassen (MDK) zukünftig ausgeschlossen. Diese sind aber bereits seit Jahren in den Expertengremien etabliert und anerkannt. Von daher ist nicht ersichtlich, warum diese zukünftig nicht mehr mitarbeiten sollen.

*Das IQTIG formuliert formale Anforderungskriterien für eine Bewerbung als Experte. (S.168)*

Die hier genannten Unterlagen lassen einen zusätzlichen bürokratischen Aufwand vermuten, der nicht erforderlich ist, wenn die Besetzung über die Vertragspartner – wie in den themenspezifischen Bestimmungen der DeQS-RL vorgegeben – erfolgt.

## **Qualitätsförderung**

*Bei einem wiederholten Qualitätsdefizit empfiehlt das IQTIG, dass ab dem zweiten Qualitätsdefizit bei einem Indikatorergebnis oder bei inhaltlich ähnlichen Indikatoren innerhalb von drei Jahren die zuständige Stelle immer qualitätsfördernde Maßnahmen einleitet. (S.173)*

Eine Begründung, warum hier ein 3-Jahres-Zeitraum und nicht wie bei der Auslösung des Stellungnahmeverfahrens ein 2-Jahres-Zeitraum gewählt wird, erfolgt nicht. Insofern ist anzunehmen, dass das IQTIG bei der Qualitätsförderung anerkennt, dass Qualitätsverbesserungen erst nach drei Jahren sinnvoll gemessen werden können. Daher stellt sich die Frage,

warum diese Erkenntnis nicht auch bei der Auslösung des Stellungnahmeverfahrens, also bei der statistisch signifikanten Auffälligkeitseinstufung Berücksichtigung findet. Sollen die Ergebnisse sinnvoll gemessen werden, muss auch die Auslösung des Stellungnahmeverfahrens über einen 3-Jahres-Zeitraum erfolgen. Nur so kann in der Prozesskette die notwendige Stringenz hergestellt werden.

### **Zielvereinbarung**

*Im Rahmen von Zielvereinbarungen einigen sich ein Leistungserbringer und die zuständige Stelle (LQS, LAG, IQTIG) auf Empfehlungen zur Behebung von Qualitätsdefiziten (fachliches Ziel) in einem festgelegten Zeitraum (Zeitziel) sowie ggf. auf begleitende Maßnahmen, die dieser Zielerreichung dienlich sind (flankierende Umsetzungsziele). (S.175)*

Nicht beschrieben wird, ob im Rahmen der Qualitätsförderung die Zielvereinbarungen pseudonymisiert – also ohne Identifikation des Leistungserbringers – abgeschlossen werden.

*Im Rahmen der fachlichen Bewertung sollen die Leistungserbringer gegenüber den Mitgliedern der Fachkommission pseudonymisiert werden, um Verzerrungen der fachlichen Beurteilung zu vermeiden. (S.149)*

Wird die aktuelle Umsetzung zugrunde gelegt, erfolgt eine Depseudonymisierung innerhalb der Fachkommissionen und bleibt damit in einem geschützten Raum. Nur in Ausnahmefällen werden die Lenkungsorgane in den Prozess der Qualitätsförderung eingebunden. Es empfiehlt sich, dieses Procedere beizubehalten, dies bedarf aber einer Konkretisierung im Bericht des IQTIG.

### **Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens**

*Das vorliegende Konzept sieht im Sinne einer Standardisierung und Ressourcenschonung ein schriftliches Stellungnahmeverfahren zur Qualitätsbewertung vor. Das bedeutet, dass die zuständigen Stellen bei Nicht-Einhaltung der Anforderungen direkt eine abschließende Bewertung ohne weitere Einholung von Präzisierungen vornehmen sollen. Diese Empfehlung soll die zuständigen Stellen einerseits dadurch entlasten. (S.180)*

Die in dem Bericht beschriebene Qualitätsbewertung, bestehend aus formaler Prüfung, Nachberechnung und fachlicher Bewertung, schafft grundsätzlich mehr Objektivität und Transparenz, der Aufwand wird jedoch sowohl für die Leistungserbringer als auch die zuständigen Stellen deutlich umfangreicher als bisher.

*Das IQTIG selbst räumt ein, dass auch die formale Prüfung einen gewissen Aufwand darstellt und für die neuen Empfehlungen zur fachlichen Bewertung keine Erfahrungswerte vor-*

*liegen und somit noch unklar ist, wie groß die Aufwandsreduktion im Vergleich zum bisherigen Vorgehen gemäß QSKH-RL tatsächlich ist. (S.180/182)*

Insofern basieren die Vorschläge des IQTIG zur Verkürzung des Stellungnahmeverfahrens auf falschen und nicht belegbaren Annahmen.

*Eine Standardisierung von Prozessen steigert z.B. dadurch die Effizienz, dass Entscheidungen (Versand von Hinweisen), die zuvor eingehend von den zuständigen Stellen oder der Fachkommission abgewogen wurden, zukünftig klar geregelt wären. (S.180)*

Der Versand von Hinweisen wird bislang auch dazu genutzt, Indikatoren zu erläutern, auf Kriterien des QS-Filters und auf folgerichtige Dokumentation hinzuweisen. Da es sich bei den inzwischen sehr komplexen Verfahren keineswegs um selbsterklärende Formalien handelt, entsteht trotz des Wegfalls der Hinweise auf rechnerische Abweichungen ein nicht zu unterschätzender weiterer Aufwand durch Erläuterungen zum Verfahren.

*Unter Berücksichtigung des Stellungnahmezeitraums von vier Wochen und einem einwöchigen Zeitraum bis zum Eingang der Aufforderung zur Stellungnahme bei den Leistungserbringern könnte somit die Abgabefrist für Stellungnahmen bzw. der Beginn der Bewertung der Stellungnahmen auf den 04. August des EJ+1 festgelegt werden. (S.181)*

Diese Empfehlung berücksichtigt weder die bundesweit unterschiedlichen Ferienzeiten, die zu unterschiedlichen Bearbeitungsständen in den Bundesländern während dieser Zeit führen, noch die bislang sehr fehleranfälligen Auswertungen des IQTIG, die eine Korrektur der Auswertungen erfordern und damit zur Verschiebung des Stellungnahmeverfahrens führen. Aufgrund der fehlenden Erfahrungswerte mit dem vorgeschlagenen Stellungnahmeverfahren des IQTIG ist dringend davon abzuraten, parallel die Fristen zu verkürzen.

*Die Abbildung 39 stellt mögliche Änderungen der Fristen dar. (S182)*

Wie folgerichtig geschlussfolgert wird (Seite 150), können geeignete Einflussfaktoren ausschließlich mit fachlicher Expertise der Fachkommissionen von „sonstigen“, heuristisch zu beurteilenden Einflussfaktoren separiert werden. Praktisch nicht vorstellbar erscheint aktuell, dass partielle Neuberechnungen während derselben Sitzung erfolgen können. Die eventuelle Notwendigkeit einer weiteren Sitzung findet im straffen Zeitplan des IQTIG keine Berücksichtigung.

Bezüglich der Frist zum Erhalt der Rückmeldeberichte der Länder am 30.06. muss angemerkt werden, dass den Ländern die Auswertungen im eigenen Zuständigkeitsbereich bislang sehr viel früher (ab Anfang Mai) zur Verfügung stehen. Ziel der Änderung der Fristen soll eine frühzeitige Berichterstattung der Versorgungsqualität sein. Aus welchem Grund die Fristen der Länder derart strikt gekürzt werden, der vorgelagerte Prozess der Erstellung der

Rückmeldeberichte auf Bundesebene jedoch nicht effizienter gestaltet wird, ist nicht nachvollziehbar.

*Durch die Aufteilung der Prozesse in zwei Module stünde den zuständigen Stellen im Jahr nach der Erfassung somit mehr Zeit für die Einleitung und Umsetzung von Maßnahmen zur Qualitätsförderung zur Verfügung, sodass der Erfolg der Maßnahmen ggf. bereits im EJ + 1 gemessen werden könnte. (S. 183)*

Die Annahme, dass durch eine Vorverlagerung der Bewertung um 6 Wochen und eine Ausdehnung der Qualitätsförderung auf Ende Dezember eine Messung des Erfolgs bereits im Erfassungsjahr + 1 gemessen werden könnte, ist nicht belegt. Die Rückmeldeberichte und somit die finalen Jahresergebnisse stehen den Einrichtungen in den DeQS-Verfahren erst im 2. Halbjahr zur Verfügung.

*Die Aufforderungsschreiben für die auffälligen Indikatorergebnisse könnten quasi zeitgleich mit Veröffentlichung der finalen Rechenregeln und der Übermittlung der Auswertungen an die Leistungserbringer versendet werden, falls entsprechende IT-technische Strukturen aufgebaut werden würden. (S.185)*

Die IT-Strukturen sind weniger das Problem. Problematisch sind vielmehr wie bereits oben beschrieben, die fehleranfälligen Auswertungen des IQTIG, die einer sofortigen Auslösung des Stellungnahmeverfahrens entgegenstehen. Dazu kommen Schnittstellenprobleme beim Datenaustausch zwischen Bundes- und Landesstellen, die aufgrund von einseitigen Änderungen des IQTIG an den Spezifikationen der Austauschformate entstehen und daher ständige Anpassungen der IT-Strukturen auf Landesebene erfordern. Dem könnte man begegnen, indem die Auswertungen zumindest übergangsweise auf Landesebene belassen würden.

*Das IQTIG beschreibt weitere Möglichkeiten für einen früheren Start des Stellungnahmeverfahrens durch Streichung der 1-monatigen Korrekturfrist der QS-Daten und durch Festlegung einer früheren Datenannahmefrist. Es spricht darüber hinaus jedoch von weitreichenden Konsequenzen, falls die Frist für die Übermittlung der Rückmeldeberichte nach vorn verlagert werden sollte. (S. 185, 186)*

Die Änderungen der Datenannahmefristen werden aus den vom IQTIG benannten Gründen befürwortet. Sie ändern jedoch für sich allein nichts am Zeitpunkt des Starts des Stellungnahmeverfahrens, sofern die Fristen für die Rückmeldeberichte beibehalten werden.

## **Anhang Textvorschläge Einholung Stellungnahmen**

Die Geschäftsstellen verfügen über eine zum Teil mehr als 15 jährige Expertise im Einholen von Stellungnahmen im Rahmen des Strukturierten Dialogs gemäß QSKH-RL. Die beinhaltet auch, dass jahrelang an den Anfragetexten für die einzelnen Qualitätsindikatoren gearbeitet wurde. Auch wenn das Stellungnahmeverfahren im Rahmen der DeQS-RL anders strukturiert ist als der Strukturierte Dialog, so ist es dennoch erforderlich, dass Leistungserbringer um eine Stellungnahme gebeten werden (ob verpflichtend oder freiwillig). Dadurch dass die QS Verfahren gemäß QSKH-RL voraussichtlich ab 2021 in die DeQS-RL überführt werden, sollte den einzelnen LAGen aufgrund ihrer Expertise eine gewisse Freiheit in der Art der Anfragen an die Leistungserbringer gewährt werden, um an das Ziel einer qualifizierten Rückmeldung durch die Leistungserbringer zu gelangen. Es wird daher empfohlen, dass neben den recht formalistischen Textvorschlägen des IQTIGs die in den Bundesländern bewährten Formulierungen ergänzend verwendet werden können.

*Bei einem auffälligen Ergebnis in einem Indikationsindikator: Bitte schildern Sie in Ihrer Stellungnahme, warum die Prozedur häufiger durchgeführt wurde, als dies nach der Rechenvorschrift des Qualitätsindikators indiziert war, respektive warum die Prozedur trotz bestehender Indikation laut Qualitätsindikator nicht häufig genug durchgeführt wurde, und warum dies dennoch nicht als Qualitätsdefizit einzustufen ist. (S.62)*

Die Arbeit aller im Medizinbetrieb tätigen Menschen sollte sich nach Leitlinien und Studienergebnissen orientieren. Bei einem auffälligen Ergebnis in einem Indikationsindikator sehen die Formulierungsvorschläge des IQTIG vor, dass die Indikationsstellung von „den Rechenvorschriften des Qualitätsindikators“ abhängen soll. Hier sollte entweder die Formulierung geändert werden („..., als die nach den von den Fachgesellschaften erarbeiteten Leitlinien indiziert war...“) oder das IQTIG sollte die Indikatoren so valide gestalten, dass die getroffene Aussage zutrifft.

Die Formulierung bei einem auffälligen Ergebnis in einem risikoadjustierten Ergebnisindikator ist sehr zu begrüßen.

## Zusammenfassung

Abschließend werden hier die aus unserer Sicht kritischen Punkte im Vorbericht des IQTIG nochmals zusammenfassend dargestellt:

1. Die Übertragbarkeit der Aufgabenstellung zur Weiterentwicklung des Strukturierten Dialoges auf die DeQS-RL wird zwar im Vorbericht beschrieben, aber fraglich bleibt in diesem Zusammenhang, inwieweit das Verfahren auch auf den ambulanten Bereich übertragbar ist.
2. Die Beweislastumkehr wird rechtlich kritisch gesehen und lässt keine Ressourceneinsparung auf Landesebene zu.
3. Fachliche Aspekte im Rahmen der Bewertung im Stellungnahmeverfahren sind gegenüber formalen Anforderungen zu priorisieren.
4. Die fachlichen Bewertungen der Stellungnahmen müssen weiterhin durch die Fachkommissionen erfolgen, nicht durch die Geschäftsstellen.
5. Das Procedere der Bewertung ist so anzupassen, dass im Stellungnahmeverfahren festgestellte Qualitätsdefizite – auch im Einzelfall – als solche zu bewerten sind, auch wenn nach Neuberechnung ein rechnerisch unauffälliges Ergebnis besteht.
6. Die im Konzept genannten Fristen sind in der Praxis nicht zu halten und daher abzulehnen, insbesondere da mögliche Ressourcen in vorgelagerten Prozessen unbeachtet bleiben.
7. Die Nutzung individuell landespezifischer und mit den Einrichtungen erprobter Anfragetexte oder durch die Fachkommissionen entwickelter Fragebögen müssen weiter möglich sein. Nur so kann sichergestellt werden, dass die Einrichtungen die mit dem Qualitätsindikator verbundene Fragestellung in gebotener Art und Weise verstehen und beantworten können.
8. Die Benennung von Mitgliedern der Fachkommissionen obliegt entsprechend der Richtlinie der LAG und hat sich in Praxis bewährt, sodass es keiner Änderung bedarf.

# Niedersächsische Krankenhausgesellschaft e.V.

## – Projektgeschäftsstelle Qualitätssicherung –

Nieders. Krankenhausgesellschaft · Postfach 44 49 · 30044 Hannover

IQTIG  
Institut für Qualitätssicherung und Transparenz im  
Gesundheitswesen  
Katharina-Heinroth-Ufer 1

D-10787 Berlin

Hausadresse:  
Thielenplatz 3  
30159 Hannover

Postfachadresse:  
Postfach 44 49  
30044 Hannover

Telefon (05 11) 3 07 63-31  
Telefax (05 11) 3 07 63-33  
E-Mail pgs-qs@nkgev.de  
Internet <http://www.nkgev.de>

Ihr Zeichen

Ihre Nachricht vom

Unser Zeichen

Datum:

Lü  
33.04-15

21. November 2019

### Qualitätssicherung in der stationären Versorgung

**hier: Vorbericht des IQTIG zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern  
Stellungnahme**

Sehr geehrte Damen und Herren,

für die Übersendung des vertraulichen Vorberichtes des IQTIG zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern bedanken wir uns recht herzlich.

Gern machen wir als Projektgeschäftsstelle Qualitätssicherung in Niedersachsen Gebrauch von der Möglichkeit, eine Stellungnahme zu diesem Berichtsentwurf des IQTIG abzugeben.

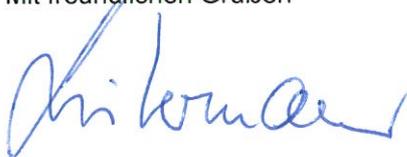
Aus Sicht der Projektgeschäftsstelle Qualitätssicherung sollten bei dem Konzept zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern folgende Punkte hinterfragt und überdacht werden:

- Es sollte in dem Konzept ein einheitlicher Sprachgebrauch nach DeQS-RL gewählt werden. Vermischungen der Begrifflichkeiten verschiedener Richtlinien führen nur zur Verwirrung.
- Außerdem sollte eine Einbindung auch der niedergelassenen Leistungserbringer in das Konzept erfolgen. Bisher sind im Titel nur die Krankenhäuser genannt.
- Weiterhin sollte für eine größere Akzeptanz der Qualitätsindikatoren eine kurzfristige Überprüfung der bisher im Rahmen des strukturierten Dialogs verwendeten Qualitätsindikatoren auf Eignung für die richtige Messung bezüglich des geforderten Qualitätszieles erfolgen. Können die verwendeten Qualitätsindikatoren wirklich zwischen „Qualität“ und „unzureichender Qualität“ unterscheiden? Dies sollte zu einer Reduktion bzw. Priorisierung der Qualitätsindikatoren führen.
- Um eine bessere Aussage bzgl. der Qualität treffen zu können, sollte auch über Indikatoren-Sets, die die Leistung einer Abteilung abbilden, nachgedacht werden. Diese sollten auch im Strukturierten Dialog genutzt werden, um die Treffsicherheit der Ergebnisse zu erhöhen.
- Darüber hinaus sollte eine Vorverlegung der Datenannahmefrist auf den 31.01. d. J. erfolgen, um nachgelagerte Prozesse im Jahresverlauf früher beginnen zu können. Die Rückmeldeberichte/Auswertungen an Leistungserbringer könnten somit bis zum 15.05.d.J. versandt werden. Auch der Strukturierte Dialog könnte somit anstatt am 01.07. d.J. bereits am 15.05. d.J. und somit 6 Wochen früher als bisher beginnen.
- Eine Durchführung des strukturierten Dialogs von Anfang Juli d. J. bis Mitte September d. J. inklusive Gremiensitzungen halten wir aufgrund der Ferienzeit für nicht durchführbar. Der Zeitrahmen, den das IQTIG vorsieht stellt einen kritischen Pfad dar. Jegliche Probleme im Ablauf führen automatisch zu Verlängerungen im Verfahren. Außerdem wird dabei nicht die EDV-technische und personelle Ausstattung der einzelnen Geschäftsstellen berücksichtigt. Auch die entstehenden Kosten werden nicht thematisiert.

- Das Vorgehen bei der Nachberechnung von Qualitätsindikatoren, Änderungen von Referenzbereichen ist sehr komplex und wenig transparent, so dass dies weder Gremienmitglieder, Krankenhausmitarbeiter noch Patienten verstehen. Diese Methodik trägt weder zur Verschlankung noch zur Nachvollziehbarkeit von Ergebnissen bei. Außerdem hat man leider, wenn man das theoretische Konzept gedanklich in die Praxis überführt den Eindruck, dass solange an den Indikatorergebnissen herumgerechnet wird bis eine der beiden Kategorien (U oder A) in dem Bewertungsschema erreicht wird.
- In dem Konzept fehlen Informationen, wie sich die EDV-technische Unterstützung der mandantenfähigen Datenbank / des Webportals zeitlich/inhaltlich auf den Prozess des Stellungnahme-/Bewertungsverfahrens auswirkt. Welche Konsequenzen hat ihr Einsatz auf die Zusammenarbeit in den Gremien auf Landesebene? Welche Kosten entstehen?
- Die Bewertungskategorien U0 = „kein hinreichender Verdacht auf ein Qualitätsdefizit“ lässt diesen Schluss ohne Prüfung durch eine Stellungnahme nicht zu. Der Text impliziert aus unserer Sicht, dass eine Ergebnisprüfung erfolgt ist. Mit dieser Kategorie kann aber nur gesagt werden, dass der Qualitätsindikator nach Berechnung im Referenzbereich liegt.
- Die konzeptionelle Trennung in Qualitätsbewertung und Qualitätsförderung ist zu hinterfragen, da in Niedersachsen kollegiale Gespräche als qualitätsfördernde Maßnahme auch für Bewertungszwecke genutzt werden. Damit haben wir gute Erfahrungen gesammelt.
- Weiterhin sollte die Kategorie A0 = „Verzicht auf Einreichung einer Stellungnahme oder Stellungnahme nicht fristgerecht eingereicht oder Stellungnahme entsprach nicht den Anforderungen“ überdacht werden. Man sollte generell zwischen medizinischen Qualitätsdefiziten und eventuellen Dokumentationsproblemen unterscheiden können. Dies ist in der Kategorie nicht möglich. Dort gibt es nur eine Unterscheidung in U (Kein Hinweis auf ein Qualitätsdefizit) und A (Qualitätsdefizit), egal um welche Art von Problemen es sich handelt.
- Der Begriff Datenfehler ist nicht trennscharf. Der bisherige Begriff Dokumentationsfehler umschreibt das Problem besser oder sind andere Fehler gemeint?
- Zusätzlich zum Konzept zur Weiterentwicklung des Strukturierten Dialogs in Krankenhäusern sollte vom IQTIG ein Konzept zur Schulung der am Strukturierten Dialog beteiligten Personen (Gremienmitglieder/Mitarbeiter der LAGs) erfolgen, da nur im gleichen Umgang mit den Hilfsmitteln und dem gleichen Verständnis im Zustandekommen der Bewertungen die Heterogenität bundesweit zumindest vermindert werden kann.
- Vor einer Einführung des Weiterentwicklungskonzeptes sollte außerdem die Öffentlichkeit ausführlich über die Änderungen informiert werden, damit sich Fehler, wie sie im Zusammenhang mit dem PlanQI-Verfahren aufgetreten sind, nicht wiederholen.

Für Rückfragen stehen wir unter Tel. (05 11) 3 07 63-31 gern zu Ihrer Verfügung.

Mit freundlichen Grüßen



Uwe Lükermann



Vorbericht des IQITGs zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern:

Stellungnahme des Qualitätsbüros Bremen

Bremen, am 21. November 2019

Das Ziel, verschiedene Prozesse im Gesamtkontext von strukturiertem Dialog (sD) bzw. Stellungnahmeverfahren zu verkürzen, zu vereinheitlichen, auf aktuelle Normgrundlagen anzupassen und transparenter darzustellen, wird begrüßt.

Da der Vorbericht den strukturierten Dialog mit Krankenhäusern beschreibt, beziehen sich auch die folgenden Ausführungen ausschließlich auf stationäre Einrichtungen.

**Stellungnahmeverfahren:** Die Möglichkeit für Kliniken in Stellungnahmen auf Berechnungsgrundlagen Bezug nehmen zu können, erfordert die Veröffentlichung der Rechenregeln zum Zeitpunkt des Auswertungsversandes.

**Nachberechnung:** Das dargestellte Verfahren zur Nachberechnung, Änderung von Referenzbereichen etc. ist in der Umsetzung sehr komplex und für Gremien und Patienten wenig transparent. Zur Verschlankung des Verfahrens und Nachvollziehbarkeit der Ergebnisse trägt diese Methodik daher weder für die Kliniken, noch die Experten- und Lenkungs-gremien oder die Öffentlichkeit bei. Es ist davon auszugehen, dass die Vergleichbarkeit von Auswertungsergebnissen zwischen den Kliniken im Land oder Bund abnehmen und die Verschiedenheit der Indikatorergebnisse methodenbedingt zunehmen wird. Eine Hilfe für Patienten ist die fehlende Vergleichbarkeit sicher nicht.

Wie genau die Prozessschritte ab Zeitpunkt Abgabe der Stellungnahme über Nachberechnung und endgültiger Bewertung auf einem Zeitstrahl und unter Einbindung der Fachkommission und der LAG angedacht sind, wird nicht deutlich.

Nicht thematisiert ist, in welcher Höhe Kosten für technischen und personellen Aufwand in den LAGen durch Nachberechnungen zu erwarten sind.

Im Konzept ist nicht vorgesehen, wie sich Abfragen über die **mandantenfähige Datenbank (mDB)** in den Prozess Stellungnahmeverfahren/Bewertung zeitlich einfügen werden. Für den Fall, dass eine Abfrage über die mDB erfolgen soll, um weitere Informationen zu QI und Stellungnahme zu erhalten, die für eine Bewertung relevant sind, muss das Abfrageergebnis der Fachkommission rechtzeitig vor der Fachkommissionssitzung vorliegen. Von Interesse ist daher, welcher Zeitrahmen von Anmeldung einer Anfrage über die mDB bis zur tatsächlichen Durchführung der Abfrage vorgesehen ist. Dies ist auch vor dem Hintergrund bedeutsam, dass das bundesweite Anfragevolumen zur mDB in einer Zeit sehr knapper Ressourcen (Ferienzeit) sehr hoch sein wird.

**Kleine Fallzahlen:** Die Ergebnisberechnung kann bei wenigen Fallzahlen im Indikator aggregiert über zwei Jahre erfolgen. Der Reduktion des Aufwandes für ein Stellungnahmeverfahren/strukturierten Dialog steht die mögliche Aufdeckung von Qualitätsproblemen mit einem Jahr Verzögerung entgegen.

Änderungen von Rechenregeln oder Referenzbereichen zwischen den Erfassungsjahren können die Vergleichbarkeit der Indikatoren im Longitudinalvergleich einschränken, was bei einer Aggregation zu beachten ist.

Die Möglichkeit der jahresübergreifenden Indikatorberechnung bei kleiner Fallzahl bedeutet auch, dass Kliniken mit höherem Fallvolumen im Indikator jährlich Daten erhalten und zu veröffentlichen haben, Häuser mit geringem Volumen dagegen nicht.

**Bewertung** von Stellungnahmen, Ergebnissen, Einschätzung von Einflussfaktoren u. Ä. soll ausschließlich durch Fachkommissionen und abschließend durch die LAG vorgenommen werden, hingegen nicht durch die Geschäftsstelle der LAG. Zur fachlichen Bewertung der Indikatorergebnisse hat die LAG hierzu eigens Fachkommissionen mit Experten eingerichtet.

Die vorgeschlagene Bewertungskategorie „Qualitätsdefizit“ nicht mindestens nach formal und medizinisch zu unterscheiden, greift zu kurz. Für Patienten und Angehörige wird es schwierig sein, diese nicht differenzierende Information veröffentlichter Ergebnisse und Bewertungen einzuschätzen und für sich zu bewerten.

Inwiefern eine Reihe der aktuellen Indikatoren tatsächlich geeignet ist, „Qualitätsdefizite“ zu detektieren, ist fraglich. Vor dem Hintergrund der Güte verschiedener Qualitätsindikatoren sollte die

Begrifflichkeit dieser Bewertungskategorie grundsätzlich überdacht werden.

Die Bewertungskategorie „Uo“ – kein hinreichender Hinweis auf ein Qualitätsdefizit – ist zu verwenden „bei Ergebnissen innerhalb des Referenzbereiches“ (s. S. 144). Da in der Regel hier jedoch keine qualitative Ergebnisprüfung z. B. über ein Stellungnahmeverfahren und somit keine Bewertung durch ein Expertengremium vorliegt, sondern lediglich ein statistisches Ergebnis, sollte die Kategorie neutral formuliert sein (z. B. Ergebnis innerhalb des Referenzbereiches).

Wird eine Stellungnahme nicht abgegeben, so sollte auch sprachlich dargestellt werden, dass das Ergebnis als auffällig eingestuft wird, weil der Verdacht auf einen Qualitätsmangel nicht ausgeräumt wurde. Die Empfehlung, ohne fehlende Kenntnis das Vorliegen eines Qualitätsdefizits zu attestieren (Ao), sollte überdacht werden.

Abbildung 9 auf Seite 68 im Vorericht sowie der Algorithmus im Anhang (S. 64) verdeutlichen die Komplexität des Bewertungsverfahrens, das in dieser Form und in dem vorgesehenen Zeitrahmen kaum umsetzbar sein wird.

**Fristen:** Sowohl Einleitung der Stellungnahmen, deren Beantwortung durch Kliniken (und Autorisierung) in einer Vierwochenfrist als auch Beratung und Bewertung durch zuständige Expertengremien und der LAG bis Mitte September ist aufgrund der ausgedünnten Personaldecke in allen beteiligten Institutionen während der Ferienzeit sicher nicht umsetzbar.

Der Zeitstrahl auf S. 182 (Abbildung 39) zeigt, dass zur Verkürzung des Verfahrens Fristen erst ab dem 30.6. enger gesetzt werden und somit ausschließlich die Prozesse der Stellungnahmeverfahren berühren.

Die auf S. 185 eröffnete Möglichkeit, die Datenannahmefrist um einen Monat auf dem 31.1. vorzuziehen, wird daher ausdrücklich begrüßt, sofern alle anschließenden Prozesse entsprechend früher starten und Aufgaben in den in allen beteiligten Institutionen ohnehin stark verdichteten Sommermonaten besser verteilt werden.

**Fachkommissionen:** Es ist zu befürchten, dass die Besetzungen der Fachkommissionen im besten Fall nur sehr langwierig sein werden und im schlechtesten Fall das Interesse an einer ehrenamtlichen Mitarbeit aufgrund des aufwendigen Prozederes (Bewerbungsverfahren) und detaillierter Formalia äußerst gering sein wird.

Die bisherigen Regelungen der themenspezifischen Bestimmungen nach DeQS-RL zur Zusammensetzung der Fachkommissionen haben sich bewährt und sollten so fortgeführt werden. Die Vorgaben zu Qualifikation und (Mindest-) Anzahl der Experten sind hier eindeutig geregelt und gelten für alle Bundesländer gleichermaßen.

Der Vorschlag im Vorbericht nach Veröffentlichung von Kriterien zur Auswahl von Expertinnen und Experten und zur Zusammensetzung der Fachkommissionen gefährdet das angestrebte Ziel der Vereinheitlichung.

Aufwandspauschalen auf Kosten der LAGen für Mitglieder der Fachkommissionen stehen der DeQS-RL in der gültigen Fassung entgegen, die eine Übernahme der Kosten der Fachexperten durch die LAG explizit ausschließt (§ 5 Abs. 5; Ausnahme: Patientenvertretung).

**Begrifflichkeiten:** Es erschließt sich nicht, weshalb die Bewertungskategorie Uo in der Erläuterung des Bewertungsschemas das Attribut „hinreichend“ trägt, die Einstufung aber, wie auch bei U1-U3, als „kein Hinweis auf Qualitätsdefizit“ definiert ist.

Anhang, Kapitel 7: Das Schreiben „Anforderung Stellungnahmeverfahren“ sollte deutlich gekürzt werden.

Anhang, Kapitel 9: Das abgebildete Formular zum Umsetzungsstand von Zielvereinbarungen verliert sich in Detailliertheit, so dass die wesentlichen Informationen, die für Lenkungsgruppen von Interesse sind, zu verschwimmen drohen.

QIG BW GmbH – Birkenwaldstraße 145 – 70191 Stuttgart

IGTIG  
Institut für Qualitätssicherung und Transparenz  
im Gesundheitswesen  
Katharina-Heinroth-Ufer 1  
10787 Berlin

Dr. Ingo Bruder  
Geschäftsführer

Telefon 0711 184278-01  
Telefax 0711 184278-29  
bruder@qigbw.de

Unser Zeichen  
Ib/Ci

191118\_IQTIG\_Weiterentw-  
SD\_SN-LAG-BW.docx

18.11.2019

## **Stellungnahme der LAG gemäß DeQS-RL Baden-Württemberg (Fachbeirat DeQS) der QIG BW GmbH**

Sehr geehrte Damen und Herren,

Der Fachbeirat DeQS bedankt sich für die Möglichkeit zur Stellungnahme zum Vorbericht des IQTIG zur „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“. Die Re-Strukturierung der zum Teil in die Jahre gekommenen Prozesse der Bewertung von Indikatorergebnissen und die Straffung des Verfahrens durch die Trennung von Qualitätsbewertung, Dokumentationsqualität und Qualitätsförderung ist begrüßenswert, damit der Prozess stärker vereinheitlicht werden kann. Eine Übergangsregelung sollte gefunden werden, um speziell bei der Einrichtung von Fachkommissionen nach DeQS-RL die etablierte Beratung im Strukturierten Dialog durch die fachärztlichen Arbeitsgruppen nach QSKH-RL für einen definierten Übergangszeitraum (z. B. 2 Jahre) auf das Stimmnahmeverfahren umzustellen.

Folgende weitere Punkte werden von Seiten der LAG kritisch gesehen:

1. Eine klare Abgrenzung zwischen „beauftragter Stelle“ und „zuständiger Stelle“ sollte erfolgen. Die Bezeichnung „beauftragte Stelle“ wird in der Regel verwendet, wenn Aufgaben der aktuellen LQS gemäß QSKH-RL, „zuständige Stelle“ wenn die Funktionen der LAG gemäß DeQS-RL beschrieben werden. Geht man von einer realistischen Zeitschiene zur Umsetzung der Weiterentwicklung aus (Beschluss des Konzeptes => Richtlinienanpassung etc.), so muss klar sein, dass zum Zeitpunkt der Implementierung die QSKH-RL außer Kraft gesetzt sein wird und alle Leistungsbereiche nach den Regeln der DeQS-RL umgesetzt werden. Aus diesem Grund sollte sich das gesamte Konzept an den Vorgaben der DeQS-RL orientieren. Wieso unter 7.2 Zielvereinbarungen empfohlen wird, dass die LAG (die die Zielvereinbarungen schließt und überwacht) an das Lenkungsgremium (LKG) gemäß § 5 DeQS-RL berichten soll, erschließt sich nicht, da das LKG das Beschlussorgan der LAG ist und mit dieser Empfehlung das LKG faktisch an sich selbst be-

richten müsste. Der Fachbeirat geht davon aus, dass von Seiten des G-BA bei der Überführung der QSKH-Verfahren in die DeQS-RL auf die konsistente Zuordnung der Funktionen gemäß des allgemeinen Teils 1 der Richtlinie Wert gelegt und für Krankenhäuser sowie Vertragsärzte ein sektorengleiches Vorgehen vorgesehen wird.

2. Es wird vom IQTIG vorgeschlagen, dass die Festlegung der Zusammensetzung der Fachkommissionen in der Verantwortung der LAGen als zuständige Stellen erfolgen sollte. Ferner wird eine Veröffentlichung empfohlen. Es bestehen Zweifel, ob hiermit die angestrebten Ziele: „Reduktion der Heterogenität in der Vorgehensweise zwischen den beauftragten Stellen“ und „Erhöhung der Transparenz und Nachvollziehbarkeit der Entscheidungen“ erreicht werden wird. Warum soll von den bisherigen Vorgaben der DeQS-RL abgewichen werden, die in den themenspezifischen Bestimmungen vorgibt, wie sich leistungsbereichsbezogen die jeweiligen Fachkommissionen zusammensetzen? Es erscheint zweckmäßig, weiterhin eine Mindestbesetzung vorzugeben und lediglich zusätzliche Expertisen landesspezifisch in der Verantwortung der LAG zuzulassen.
3. Die Verkürzung des Zeitraumes bis zum Abschluss des Bewertungsprozesses primär durch die Straffung der Abläufe in den umsetzenden Strukturen auf Landesebene und bei den Leistungserbringern genau in der Urlaubszeit der Sommerferien erreichen zu wollen, wird als realitätsfern und schwierig umsetzbar eingeschätzt (Abschnitt 8.1). In Baden-Württemberg werden teilweise die bisherigen strukturierten Dialoge bereits im Mai begonnen, da die Krankenhausauswertungen von der Geschäftsstelle erstellt werden. Sobald letzte Informationen von Seiten des IQTIG Ende April vorliegen, können frühzeitig erste Auswertungen durch das IQTIG fertig gestellt werden. Für einen frühzeitigeren Beginn nachgeordneter Prozesse (wie die schriftlichen Stellungnahmeverfahren, Bewertungen, Zielvereinbarungen etc.) wäre die Übermittlung der Auswertungen sämtlicher Leistungsbereiche erst Ende Juni von Seiten des IQTIG kontraproduktiv und deshalb sollten die Auswertungen und deren Übermittlung durch das IQTIG vorverlagert werden beispielsweise durch sukzessive fertigberechnete, leistungsbereichsbezogene Auswertungen.
4. Qualität der Bewertungen nicht dem Formalismus opfern: Kritisch ist zu sehen, dass das Nicht-Einhalten von Verfahrensregeln den Bewertungsprozess unterbrechen soll. Zwar wird in diesen Fällen die Einstufung als „auffällig“ (A) mit der erklärenden Ziffer „0“ (= Verzicht auf Einreichung einer Stellungnahme oder Stellungnahme nicht fristgerecht eingereicht oder Stellungnahme entsprach nicht den Anforderungen) ergänzt, jedoch wird mit „A“ die Klassifikation „Qualitätsdefizit“ vergeben. Das Akzeptieren der Feststellung des Defizites von Seiten des Leistungserbringers durch Nicht-Abgabe einer erläuternden Stellungnahme sollte nicht gleichgesetzt werden mit dem Nicht-Einhalten von gesetzten formalen Anforderungen an eine Stellungnahme. Es handelt sich hier zwar auch um ein Qualitätsdefizit, jedoch bezüglich des Qualitäts-/ Kommunikationsmanagements des Leistungserbringers und nicht der medizinischen Leistungserbringung. Letztere ist jedoch für Patienten / die Öffentlichkeit im Rahmen der Transparenz der Qualität der Leistungserbringer von Interesse. Daher sollte für diese Fallkonstellationen die Klassifikation „Organisationsdefizit“ eingeführt werden.
5. Die empfohlene Nachberechnung eines informativen Indikatorwertes im Rahmen des Bewertungsprozesses wird als zu komplex und als erneute Quelle für Heterogenität betrachtet. Ferner wird hierdurch die Gefahr für Zeitverlust im Prozess

und erheblichen Ressourcenverbrauch gesehen. Die vorgeschlagene technische Unterstützung durch ein web-basiertes Nachberechnungsportal ist hilfreich, ersetzt jedoch nicht die Bewertung durch die Fachkommissionen, ob ein vom Leistungserbringer genannter Einflussfaktor in der partiellen Nachberechnung Berücksichtigung finden soll.

Ergänzend zum Umsetzungs- und Evaluierungskonzept sollte zwingend vor Umsetzung ein Kommunikations- und Schulungskonzept erstellt werden. Gründe hierfür: Die bewerteten Leistungserbringer, aber auch die Öffentlichkeit, die ja die Transparenz nutzen soll, um Entscheidungen treffen zu können, sollten verstehen und nachvollziehen können, wie die Bewertungen zu Stande kommen (Transparenzgebot). Die Kenntnis könnte möglicherweise die Entscheidungsfindung beeinflussen. Beim Leistungserbringer stellt Nachvollziehbarkeit die Grundlage für Akzeptanz dar. Ohne diese wird keine Motivation zur geforderten Qualitätsverbesserung zu erreichen sein. Andererseits müssen die umsetzenden Stellen und einbezogenen Experten in die vorgeschlagene komplexe Vorgehensweise eingearbeitet werden, um sich dem angestrebten Ziel, die Heterogenität zu reduzieren, ansatzweise nähern zu können.

Mit freundlichem Gruß



Dr. Ingo Bruder



Institut für Qualitätssicherung und  
Transparenz im Gesundheitswesen

# Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern

Würdigung der Stellungnahmen zum Vorbericht

Erstellt im Auftrag des  
Gemeinsamen Bundesausschusses

Stand: 26. Mai 2020

---

# Impressum

**Thema:**

Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern. Würdigung der Stellungnahmen zum Vorbericht

**Ansprechpartnerin:**

Andrea Wolf

**Auftraggeber:**

Gemeinsamer Bundesausschuss

**Datum des Auftrags:**

18. Januar 2018

**Datum der Abgabe:**

26. Mai 2020

**Herausgeber:**

IQTIG – Institut für Qualitätssicherung  
und Transparenz im Gesundheitswesen

Katharina-Heinroth-Ufer 1  
10787 Berlin

Telefon: (030) 58 58 26-0  
Telefax: (030) 58 58 26-999

[info@iqtig.org](mailto:info@iqtig.org)

<https://www.iqtig.org>

# Inhaltsverzeichnis

Abkürzungsverzeichnis.....	4
Verzeichnis der stellungnahmeberechtigten Organisationen .....	5
Stellungnahmeverfahren zum Vorbericht „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“ .....	7
1 Einleitung.....	8
2 Analyse der Ausgangssituation .....	10
3 Methodischer Hintergrund.....	12
4 Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und -förderung .....	15
5 Statistische Auswertungsmethodik.....	17
6 Fachliche Bewertung .....	22
7 Modul Qualitätsförderung .....	38
8 Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens .....	40
9 Empfehlungen zur Evaluation des neuen Konzepts .....	42
Übergeordnete Themenbereiche .....	43
Literatur.....	45

## Abkürzungsverzeichnis

Abkürzung	Bedeutung
DeQS-RL	Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung
EJ	Erfassungsjahr
EJ +1	Jahr, das ein Jahr nach dem Erfassungsjahr endet (usw. EJ +2, EJ +3)
G-BA	Gemeinsamer Bundesausschuss
IQTIG	Institut für Qualitätssicherung und Transparenz im Gesundheitswesen
KI	künstliche Intelligenz
LAG	Landesarbeitsgemeinschaft(en) für sektorenübergreifende Qualitätssicherung
LQS	Landesgeschäftsstelle(n) für Qualitätssicherung
plan. QI-RL	Richtlinie zu planungsrelevanten Qualitätsindikatoren
Qesü-RL	Richtlinie zur einrichtungs- und sektorenübergreifenden Qualitätssicherung
QSKH-RL	Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern
QS-Verfahren	Qualitätssicherungsverfahren
SGB V	Fünftes Buch Sozialgesetzbuch

# Verzeichnis der stellungnahmeberechtigten Organisationen

## Organisationen nach § 137a Abs. 7 SGB V

- Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V. (AWMF)
- Bundesärztekammer (BÄK)
- Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS)
- Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e. V. (DGGG) und der Deutschen Gesellschaft für Senologie e. V. (DGS)
- Deutsche Gesellschaft für HNO-Heilkunde, Kopf- und Hals-Chirurgie e. V. (DGHNO-KHC)
- Deutsche Gesellschaft für Innere Medizin e. V. (DGIM)
- Deutsche Gesellschaft für Kardiologie – Herz- und Kreislaufforschung e. V. (DGK)
- Deutsche Gesellschaft für Kinder- und Jugendmedizin e. V. (DGKJ e. V.)
- Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS)
- Deutsche Gesellschaft für NeuroIntensiv- und Notfallmedizin (DGNI)
- Deutsche Gesellschaft für Neuroradiologie e. V. (DGNR)
- Deutsche Gesellschaft für Orthopädie und Unfallchirurgie e. V. (DGOU), Deutsche Gesellschaft für Orthopädie und Orthopädische Chirurgie e. V. (DGOOC) und Deutsche Gesellschaft für Unfallchirurgie e. V. (DGU)
- Deutsche Gesellschaft für Thorax-, Herz- und Gefäßchirurgie e. V. (DGTHG)
- Deutsche Krankenhausgesellschaft (DKG)
- Deutscher Pflegerat e. V. (DPR)
- Deutsches Netzwerk Versorgungsforschung e. V. (DNVF)
- Gesellschaft für Qualitätsmanagement in der Gesundheitsversorgung e. V. (GQMG)
- GKV-Spitzenverband (GKV-SV)
- Kassenärztliche Bundesvereinigung (KBV)
- Kassenzahnärztliche Bundesvereinigung (KZBV)
- Patientenvertretung: maßgebliche Organisationen nach § 140f SGB V (PatV)

## Organisationen auf Landesebene

- Ärztekammer Sachsen-Anhalt, Projektgeschäftsstelle Qualitätssicherung (AEKSA QS)
- Bayerische Arbeitsgemeinschaft für Qualitätssicherung in der stationären Versorgung (BAQ) bei der Bayerischen Krankenhausgesellschaft
- Geschäftsstelle Qualitätssicherung Hessen (GQH) bei der Hessischen Krankenhausgesellschaft
- Geschäftsstelle Qualitätssicherung Nordrhein-Westfalen bei den Ärztekammern Nordrhein und Westfalen-Lippe (QS-NRW)
- Landesgeschäftsstelle für einrichtungs- und sektorenübergreifende Qualitätssicherung der Sächsischen Landesärztekammer, Projektgeschäftsstelle Qualitätssicherung Sachsen (SLAEK QS)

- Landesgeschäftsstelle für Qualitätssicherung Berlin (LQS BE)
- Landesgeschäftsstelle für Qualitätssicherung Mecklenburg-Vorpommern (LQS MV)
- Landesgeschäftsstelle für Qualitätssicherung Rheinland-Pfalz (LQS RP)
- Landesgeschäftsstelle für Qualitätssicherung Saarland (LQS SL)
- Landesgeschäftsstelle für Qualitätssicherung Thüringen (LQS TH)
- Niedersächsische Krankenhausgesellschaft e. V., Projektgeschäftsstelle Qualitätssicherung (NKGeV QS)
- Qualitätsbüro Bremen (QB Bremen)
- Qualitätssicherung im Gesundheitswesen Baden-Württemberg (QiG BW GmbH)

## **Stellungnahmeverfahren zum Vorbericht „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“**

Der Vorbericht „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“ wurde am 10. Oktober 2019 an die gemäß § 137a Abs. 7 SGB V zu beteiligenden Institutionen und Organisationen sowie an alle Vertreterinnen und Vertreter der Landesarbeitsgemeinschaften für sektorenübergreifende Qualitätssicherung (LAG) und der Landesgeschäftsstellen für Qualitätssicherung (LQS) mit der Möglichkeit versendet, schriftlich zu den dort enthaltenen Entwicklungsergebnissen Stellung zu nehmen. Die Einsendefrist für Stellungnahmen endete am 21. November 2019. Insgesamt gingen 32 Stellungnahmen von 35 Organisationen ein, wobei fünf Landesgeschäftsstellen<sup>1</sup> eine gemeinsame Stellungnahme abgegeben haben (siehe Verzeichnis der stellungnahmeberechtigten Organisationen).

Die in den eingegangenen Stellungnahmen vorgebrachten fachlichen Hinweise und Argumente zum vorgelegten Konzept wurden sorgfältig geprüft und gegebenenfalls wurden Anpassungen für den Abschlussbericht vorgenommen. Im vorliegenden Dokument werden die zentralen Aspekte aus den Stellungnahmen zusammengefasst und es wird erläutert, wie das Institut mit den vorgebrachten Hinweisen umgegangen ist. Die Hinweise aus den Stellungnahmen sind nach der Reihenfolge der Kapitel im Vorbericht geordnet und durch ein Kapitel zu übergeordneten Themen ergänzt.

---

<sup>1</sup> Gemeinsame Stellungnahme der Landesgeschäftsstellen Berlin, Mecklenburg-Vorpommern, Saarland, Rheinland-Pfalz und Thüringen (gem. StN LQS BE, MV, SL, RP, TH).

# 1 Einleitung

In einigen Stellungnahmen wird die Entscheidung des IQTIG thematisiert, Empfehlungen für ein richtlinienunabhängiges Stellungnahmeverfahren zu geben, das sowohl für die QSKH-RL als auch für die DeQS-RL und die Qesü-RL umsetzbar sein sollte. So gebe es zwar Überlappungen zwischen der Beauftragung vom 18. Januar 2018 („Kriterien für die Qualitätsbewertung sowie für die Einleitung und Durchführung von Qualitätssicherungsmaßnahmen nach der Qesü-RL“) und der Beauftragung vom 18. Mai 2018 („Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern“), aber auch nicht näher bezeichnete Unterschiede (KBV, S. 3).

In ähnlicher Weise wird die Bearbeitung der Beauftragung in Form von richtlinienunabhängigen Empfehlungen als kritisch gesehen, da sich die Richtlinien (Qesü-RL, DeQS-RL, QSKH-RL, plan. QI-RL) unterscheiden würden, z. B. in Form von Zuständigkeiten, Abläufen und sprachlichem Duktus (KZBV, S. 8). Vor allem sei aber für die plan. QI-RL ein hoher Grad an Justiziabilität nötig, nicht jedoch für die anderen Richtlinien (KZBV, S. 8).

**IQTIG:** In den zitierten Stellungnahmen wird auf Unterschiede zwischen den genannten Richtlinien (Qesü-RL, DeQS-RL, QSKH-RL, plan. QI-RL) hingewiesen. Aus Sicht des IQTIG besteht allerdings eine grundlegende Gemeinsamkeit zwischen diesen Richtlinien in dem Ziel, die Qualität der Gesundheitsversorgung leistungserbringerbezogen festzustellen und zu steigern. Während im Rahmen der DeQS-RL und der QSKH-RL an die Qualitätsmessung Maßnahmen zur Steigerung der Versorgungsqualität bei den einzelnen Leistungserbringern anschließen können, ist der Handlungsanschluss an die Qualitätsmessung im Rahmen der plan. QI-RL das Zurverfügungstellen der Ergebnisse für die Landesplanungsbehörden, damit diese die Ergebnisse der Qualitätsmessung in ihre Planungsentscheidungen mit einbeziehen können.

Zukünftig sollen die Ergebnisse der Qualitätsmessungen Patientinnen und Patienten gezielt für Auswahlentscheidungen zur Verfügung gestellt werden (vgl. Beauftragung „Veröffentlichung von einrichtungsbezogenen vergleichenden risikoadjustierten Übersichten über die Qualität in maßgeblichen Bereichen der stationären Versorgung“, sog. Qualitätsportal). Damit müssen aus Sicht des IQTIG und im Einklang mit der Beauftragung (siehe Punkt I. „Hintergrund der Beauftragung“ des Beschlusses über eine Beauftragung des IQTIG zur Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern, G-BA 2018) sowie im Einklang mit der internationalen fachwissenschaftlichen Literatur auch die Qualitätsmessungen grundsätzlich höheren methodischen Anforderungen genügen, als dies in der Vergangenheit der Fall war. Daher wurde

auf Basis der Methodischen Grundlagen V1.1 sowie auf Basis der fachwissenschaftlichen Literatur eine einheitliche Methodik für die indikatorbasierte Qualitätsmessung entwickelt.

In einigen Stellungnahmen wird ausgeführt, dass die Bewertung der Dokumentationsqualität ein Bestandteil des Strukturierten Dialogs gemäß QSKH-RL sei und daher auch Bestandteil der Weiterentwicklung sein müsse (z. B. GKV-SV, S. 13; BAQ, S. 2; GQH, S. 2).

**IQTIG:** Die leistungserbringerbezogene Bewertung der Dokumentationsqualität mittels sog. Auffälligkeitskriterien soll gemäß § 9 Abs. 7 QSKH-RL auch im Rahmen des Strukturierten Dialogs erfolgen. Wie für die Durchführung des Strukturierten Dialogs zur Bewertung der Versorgungsqualität beinhaltet die QSKH-RL auch für die Bewertung der Auffälligkeitskriterien keine konkreten diesbezüglichen Vorgaben. Der Bedarf für eine einheitliche Methodik für den Umgang mit den Auffälligkeitskriterien wird daher auch vom IQTIG gesehen. Allerdings wird dieser Aspekt des Strukturierten Dialogs in der Beauftragung nicht erwähnt. Daraus kann geschlossen werden, dass dieser Aspekt zum Zeitpunkt der Auftragserteilung als weniger wichtig angesehen wurde als andere, in der Beauftragung explizit erwähnte Aspekte. Daher wurde der Fokus der Entwicklung, wie im Abschlussbericht erläutert, auf die in der Beauftragung erwähnten Themen gelegt. Die Entwicklung einer solchen Methodik hat für das IQTIG jedoch hohe Priorität.

## 2 Analyse der Ausgangssituation

In zwei Stellungnahmen wird die vom IQTIG und in der Beauftragung identifizierte Erweiterung der Verwendung von Qualitätsmessungen der gesetzlich verpflichtenden Qualitätssicherung (siehe Punkt I. „Hintergrund der Beauftragung“ in G-BA 2018) im sog. „*accountability*“-Kontext (z. B. Auswahlentscheidungen) thematisiert. So wurde etwa kritisiert, dass das IQTIG die Verwendung von Qualitätsindikatoren einseitig und ausschließlich im Kontext „*accountability*“ sehe (KZBV, S. 11). Von anderer Seite wurde die Funktion von Qualitätsindikatoren weiterhin nur als Aufgreifkriterien für ein späteres Peer-Review oder Audit („*improvement*“) gesehen (BÄK, S. 8; DGK, S. 2).

**IQTIG:** Die Verwendung von Qualitätsmessungen für „*improvement*“-Zwecke (d. h. Verbesserung der Versorgungsqualität durch eigene Motivation und Anstrengungen der Leistungserbringer) und für „*accountability*“-Zwecke verfolgen das gleiche Ziel einer Steigerung der Versorgungsqualität. Beide sind komplementäre und sinnvolle Strategien der Qualitätsentwicklung im Gesundheitswesen. Im Einklang mit der fachwissenschaftlichen Literatur stellt die Verwendung im „*accountability*“-Kontext jedoch höhere methodische Anforderungen an die Qualitätsmessung. Sollen Qualitätsmessungen also für beide Zwecke bzw. Strategien eingesetzt werden, müssen sie den höheren Anforderungen des „*accountability*“-Kontextes genügen.

In ihrer Stellungnahme stellt die KBV (S. 6) die Frage, wie repräsentativ die ausgewählten Indikatoren in der Heterogenitätsanalyse von Abschnitt 2.5 für die Gesamtmenge der Qualitätsindikatoren der QSKH-RL seien. Ebenfalls fehle aus Sicht der KBV die Erwähnung der Indikatoren aus den Verfahren nach DeQS-RL, insbesondere eine Erläuterung, ob das angewendete Modell auf Sozialdatenindikatoren, Follow-up-Indikatoren oder das Index-Konzept der Einrichtungsbefragung im Verfahren *Vermeidung nosokomialer Infektionen – postoperative Wundinfektionen (QS WI)* theoretisch übertragbar sei.

**IQTIG:** Für die Heterogenitätsanalysen wurde aus methodischen Gründen von einer Zufallsauswahl von Indikatoren zugunsten einer gezielten Stichprobe abgesehen. Die Auswahl wurde anhand der hohen absoluten Anzahl an rechnerisch und qualitativ auffälligen Leistungserbringern getroffen, die für diese drei Qualitätsindikatoren zu verzeichnen sind. Dabei wurde aus Gründen der Repräsentativität bewusst jeweils ein Indikations-, Prozess- und Ergebnisindikator gewählt. Für viele Qualitätsindikatoren würde die Datenbasis von Auffälligkeiten eines Erfassungsjahres nicht ausreichen, um aussagekräftige statistische Modelle mit vergleichbarer Komplexität berechnen zu können. Unter diesen Einschränkungen ist auch die Generalisierbarkeit auf andere Indikatoren

zu betrachten: Prinzipiell sind ähnliche Modelle für Follow-up-Indikatoren (mit und ohne Sozialdaten) und Index-Konstruktionen aus der Einrichtungsbefragung denkbar. Theoretisch sind auch komplexere, indikatorübergreifende Modelle denkbar, in denen Unterschiede zwischen den Qualitätsindikatoren als Modellparameter spezifiziert werden. Für solche Modellierungen ist es wichtig, dass die Anzahl der Beobachtungen (also Bewertungen im Strukturierten Dialog) und die Modellkomplexität, die proportional zur Anzahl an Bewertungsstellen ist, für hinreichend genaue Schätzungen der Modellparameter in einem ausgewogenen Verhältnis stehen. Eine Analyse der Indikatoren der Verfahren nach DeQS-RL ist zum jetzigen Zeitpunkt nicht möglich, da die beiden ehemaligen Verfahren nach Qesü-RL erst zum Erfassungsjahr (EJ) 2019 in die DeQS-RL überführt wurden.

In mehreren Stellungnahmen der Stellen auf Landesebene (BAQ, S. 3; GQH, S. 1) wird Bedauern darüber zum Ausdruck gebracht, dass in der Heterogenitätsanalyse in Abschnitt 2.5 Themen wie Strukturen der Bundesländer (z. B. Anzahl der zu betreuenden Krankenhäuser, Flächenländer/Stadtstaaten, Personalausstattung der Geschäftsstellen) keine Berücksichtigung finden. Es stelle sich auch die Frage, inwiefern die Unterschiede durch fehlende Güte der Qualitätsindikatoren erklärt werden können (GQH, S.1).

**IQTIG:** Auf dem Weg zu einer Reduktion in der Heterogenität der Bewertungen des Strukturierten Dialogs ist es selbstverständlich wichtig, vergleichbare strukturelle Voraussetzungen in den Bewertungsstellen zu schaffen. Dass die strukturellen Voraussetzungen für den Strukturierten Dialog nicht überall die gleichen sind, zeigen frühere Arbeiten (Winkler-Komp et al. 2014) und wird in Abschnitt 2.3 des Berichts aufgegriffen. Ganz bewusst sind diese strukturellen Merkmale der einzelnen Bewertungsstellen nicht Teil der statistischen Modellierung in Abschnitt 2.5. Ziel dieser Analyse ist es, den Umfang, nicht die Ursachen, der Heterogenität bei der Bewertung von Leistungserbringern mit vergleichbaren Ergebnissen zu quantifizieren. Die Analyse macht keine Aussagen darüber, ob die Vorgehensweise einzelner Landesstellen positiv oder negativ zu sehen ist, gerade weil dies unter unterschiedlichen strukturellen Voraussetzungen geschieht.

Dass Unterschiede in der Güte der Qualitätsindikatoren auch zur Heterogenität von Bewertung beitragen, ist eine Hypothese, die sich im Rahmen der vorliegenden Analyse nicht untersuchen lässt. Dies würde die systematische Prüfung und Bewertung der Güte der Qualitätsindikatoren vorwegnehmen, die nachgelagert zu diesem Projekt, ein essenzieller Bestandteil der Weiterentwicklung der bestehenden QS-Verfahren ist.

### 3 Methodischer Hintergrund

#### Funktion von Stellungnahmen

Die empfohlene Beschränkung der fachlichen Bewertung im Stellungnahmeverfahren auf das dem Qualitätsindikator zugrunde liegende Qualitätsmerkmal wurde in einigen Stellungnahmen thematisiert. Während dies einerseits als „nachvollziehbar“ (GKV-SV, S. 16) oder als „methodisch nicht zu beanstanden“ (DNVF, S. 2) beschrieben wird, wird dies von anderer Stelle als „höchst problematisch“ (DKG, S. 16) bezeichnet oder „nicht mitgetragen“, beispielsweise, weil die Sorge besteht, dass dadurch zugrunde liegende Strukturen und Prozesse nicht betrachtet werden könnten (BÄK, S. 8 f.). In ähnlicher Weise wird teilweise die Befürchtung zum Ausdruck gebracht, dass mit der empfohlenen Methodik erkannte Qualitätsdefizite nicht für qualitätssteigernde Maßnahmen berücksichtigt werden könnten (DKG, S. 16; ähnlich: GKV-SV, S. 16). Ebenso besteht die Sorge, dass durch die Empfehlungen des IQTIG verhindert werde, dass sich die Leistungserbringer ausreichend mit ihren Qualitätsergebnissen auseinandersetzen (BAQ, S. 1).

**IQTIG:** Die in diesen Stellungnahmen ausgedrückte Sorge geht nach Lesart des IQTIG auf ein Verständnis von Qualitätsindikatoren als Aufgreifkriterien für einen anschließenden Peer-Review-Prozess zurück. Aus den in Abschnitt 2.2 des Berichts dargelegten Gründen erfüllen jedoch Qualitätsmessungen, bei denen Indikatoren eher als Aufgreifkriterien fungieren, die Anforderungen an die indikatorbasierte Qualitätsmessung nicht.

Wie in Abschnitt 3.3 des Berichts ausgeführt, betrachtet das IQTIG Unterschiede zwischen Bewertungskriterien, die durch Qualitätsindikatoren explizit formalisiert sind, auf der einen Seite und Bewertungskriterien auf der anderen Seite, die implizit im bisherigen Strukturierten Dialog zur Anwendung kommen, als ein erhebliches methodisches Problem. Aus Sicht des IQTIG müssen zu jedem Zeitpunkt des Bewertungsprozesses möglichst explizite, transparente Kriterien zur Bewertung der Leistungserbringung angelegt werden. Diese Kriterien werden in einem aufwendigen Prozess der Qualitätsindikatorenentwicklung abgeleitet, begründet und durch den Gemeinsamen Bundesausschuss (G-BA) rechtsverbindlich und allgemeingültig beschlossen. Die Anwendung impliziter und damit nicht transparenter Bewertungskriterien, die über die in den Indikatoren formalisierten Kriterien hinausgehen, stellt nach sorgfältiger Analyse einen entscheidenden Grund für die in der Beauftragung bemängelte bewertungsstellenabhängige Heterogenität und niedrige „Effizienz“ im Vorgehen dar, und sollte daher so weit wie möglich vermieden werden. Daher wurde eine explizite Methodik für das Stellungnahmeverfahren entwickelt.

Das IQTIG empfiehlt, dass auf Basis der Stellungnahmen der Leistungserbringer identifizierte Mängel an zugrunde liegenden Strukturen und Prozessen unbedingt für die Entscheidung über qualitätssteigernde Maßnahmen berücksichtigt werden sollten (siehe Abschlussbericht, Kapitel 7). Eine Auseinandersetzung mit ggf. unzureichenden Qualitätsergebnissen sollte sinnvollerweise auf einen abgeschlossenen Bewertungsprozess aufbauen. Die Kritik, durch die Empfehlungen des IQTIG werde zukünftig eine Betrachtung zugrunde liegender Prozesse und Strukturen für die Qualitätsförderung nicht mehr möglich sein, wird daher nicht geteilt. Es ist das explizite Anliegen des Instituts, dass eine solche umfassende Betrachtung im Rahmen der Qualitätsförderung für diejenigen Leistungserbringer vorgenommen wird, deren Versorgungsqualität im ersten Schritt, d. h. im Rahmen der Qualitätsbewertung anhand der durch die Qualitätsindikatoren vorgegebenen Bewertungsmaßstäbe, als unzureichend bewertet wurde.

### Qualitätsmessung mittels Qualitätsindikatoren

In einigen Stellungnahmen wird eine grundsätzliche Skepsis gegenüber der Messung der Qualität der Gesundheitsversorgung mittels Qualitätsindikatoren ausgedrückt. So wird beispielsweise aufgrund vermuteter methodischer Mängel der Qualitätsindikatoren die Notwendigkeit für ein anschließendes Peer-Review der Indikatorergebnisse gesehen (DKG, S. 11; DGHNO, S. 1). In ähnlicher Weise wird argumentiert, dass in der Medizin Komplikationen nicht systematisch auftraten, was Grenzen der Standardisierungsmöglichkeiten bedeute (DGK, S. 1). An anderer Stelle wird die Messung von Qualität auf Basis von Aggregatmaßen, wie Qualitätsindikatoren ohne Einzelfallbewertungen, „nicht mitgetragen“ (DGK, S. 4). Auch anhand einer nicht näher spezifizierten Literaturlage wird die Qualitätsmessung mittels Qualitätsindikatoren kritisiert (DGVS, S. 6 f.).

**IQTIG:** Während in einigen Stellungnahmen konkrete Argumente gegen die Aussagekraft von Qualitätsindikatoren angeführt werden (z. B. mangelnde methodische Güte, Komplexität des Versorgungsgeschehen), bleiben andere Stellungnahmen vage, was die genauen Gründe für die Kritik angeht. Die unbestreitbare Komplexität der Gesundheitsversorgung stellt aus Sicht des IQTIG hohe methodische Anforderungen an die Qualitätsmessung. Um diesen Anforderungen gerecht zu werden, entwickelt und bewertet das IQTIG Qualitätsindikatoren anhand eines umfassenden Systems von Eignungskriterien. Aus dem gleichen Grund können auch keine Rückschlüsse auf die gesamte Versorgungsqualität anhand eines einzelnen Qualitätsindikators gezogen werden. Gerade aufgrund der Komplexität der Gesundheitsversorgung erfordert es ein entsprechendes Set an Qualitätsindikatoren, um die Qualität in einem bestimmten Bereich der Versorgung möglichst umfassend abzubilden. Die eher grundsätzlichen Vorbehalte gegenüber der Messung der Versorgungsqualität

mittels Qualitätsindikatoren in einigen Stellungnahmen sieht das IQTIG jedoch als unbegründet an.

Zwei Organisationen merkten an, dass die Güte der Qualitätsindikatoren verbessert werden müsse, bevor das Konzept angewandt werden könne (BAQ, S. 4; GQH, S. 2 ff.). Auch die vorgeschlagene statistische Methodik setze valide und hochwertige Qualitätsindikatoren voraus. Die derzeitigen Indikatoren dürften lediglich als Aufgreifkriterien verwendet werden (BAQ, S. 2; GQH, S. 3). Eine andere Organisation begrüßte die beabsichtigte sukzessive Prüfung der Qualitätsindikatoren anhand der Eignungskriterien, um Indikatoren bei Nichterfüllung der Kriterien entweder weiterzuentwickeln oder auszusetzen (KZBV, S. 17). Weiterhin solle ein konkreter Zeitplan für die Überarbeitung der Qualitätsindikatoren vorgelegt (GKV-SV, S. 6) bzw. eine Abschätzung der potenziell zu entfernenden Qualitätsindikatoren durchgeführt werden (DNVF, S. 4). Eine andere Organisation hob hervor, dass die Gleichsetzung von Qualitätsindikatorergebnissen mit qualitativen Auffälligkeiten grundsätzlich abzulehnen sei (QS-NRW, S. 2). Dagegen wurde ausgeführt, dass bei Qualitätsindikatoren mit hoher Güte Nachberechnungen nicht mehr nötig seien (SLAEK QS, S. 3).

**IQTIG:** Die Güte der Qualitätsindikatoren ist von großer Bedeutung für die Effektivität und die Akzeptanz der gesetzlich verpflichtenden Qualitätssicherung. Nur methodisch hochwertige Sets von Indikatoren erlauben eine zuverlässige Abbildung der Versorgungsqualität und eine solche vertrauenswürdige Messung ist die Grundlage für die Akzeptanz jedes Verfahrens zur Messung und Darstellung der Versorgungsqualität. Daher erachtet das IQTIG die Weiterentwicklung der Indikatoren als eine zentrale Aufgabe. Allerdings kann das vorliegende Konzept für den Strukturierten Dialog auch unabhängig von der Weiterentwicklung der bestehenden Indikatoren umgesetzt werden. Ein konkreter Zeitplan für die Weiterentwicklung aller Qualitätsindikatoren und eine Abschätzung möglicher auszusetzender Qualitätsindikatoren geht über den Rahmen dieser Beauftragung hinaus.

## 4 Eckpunkte des Rahmenkonzepts für die Qualitätsbewertung und -förderung

### Trennung von Qualitätsbewertung und -förderung

Die empfohlene Trennung der Prozesse im Rahmen der Qualitätsbewertung und der Prozesse im Rahmen der Qualitätsförderung wird auch aufgrund der sich dadurch eröffnenden Möglichkeit, Patientinnen und Patienten deutlich früher über die Qualität der Versorgung zu informieren (PatV, S. 12), in vielen Stellungnahmen begrüßt (PatV, S. 12; DKG, S. 17; KZBV, S. 11; KBV, S. 7; SLAEK QS, S. 2). Von anderen wird die Trennung der beiden Prozesse dagegen kritisch hinterfragt (GQH, S. 2; NKGeV QS, S. 2) bis abgelehnt, da es den Dialog zwischen Leistungserbringern und LAG bzw. Bundesstelle verhindere (BÄK, S. 9).

**IQTIG:** Das IQTIG hält den Dialog zwischen Leistungserbringer und LAG bzw. Bundesstelle für ein entscheidendes Element der gesetzlich verpflichtenden Qualitätssicherung. Gerade für eine Ursachenanalyse von Qualitätsdefiziten und für die Auswahl geeigneter Maßnahmen zur Qualitätssteigerung ist er unverzichtbar. Wie in Abschnitt 4.4 des Abschlussberichts dargelegt, ist jedoch für die indikatorgestützte Qualitätsmessung im ersten Schritt ein über eine schriftliche Stellungnahme hinausgehender Dialog methodisch nicht notwendig. Die Empfehlungen des IQTIG zielen darauf ab, den Dialog im Rahmen der Qualitätsförderung zwischen Leistungserbringern und LAG bzw. Bundesstelle auf diejenigen Leistungserbringer zu fokussieren, bei denen im ersten Schritt ein Qualitätsdefizit festgestellt wurde.

### Freiwilligkeit von Stellungnahmen

Von einigen stellungnehmenden Organisationen wurde die Sorge geäußert, dass die empfohlene Freiwilligkeit bezüglich der Abgabe einer Stellungnahme dazu führen könnte, dass sich Leistungserbringer, deren Indikatorergebnis einen hinreichenden Hinweis auf ein Qualitätsdefizit aufweist, nicht mehr ausreichend mit den potenziellen Defiziten ihrer Versorgung auseinandersetzen (BAQ, S. 1; SLAEK QS, S. 3).

**IQTIG:** Grundsätzlich geht das IQTIG von einer intrinsischen Motivation der Leistungserbringer aus, ihre Versorgungsqualität auf Basis der Rückmeldeberichte zu hinterfragen und ggf. zu verbessern. Sollte ein Leistungserbringer, dessen Indikatorergebnis einen hinreichenden Hinweis auf ein Qualitätsdefizit aufweist, auf die Abgabe einer Stellungnahme verzichten, soll zukünftig gemäß der empfohlenen Methodik die Bewertung mit „Qualitätsdefizit“ erfolgen. In diesem Fall sollte die LAG bzw. die Bundesstelle in Erwägung ziehen, zusam-

men mit dem Leistungserbringer qualitätssteigernde Maßnahmen umzusetzen. Eine mögliche Maßnahme könnte sein, den Leistungserbringer beispielsweise mittels eines indicatorspezifischen Fragebogens aufzufordern, sich mit dem Indikatorergebnis selbstkritisch auseinanderzusetzen. Diese Möglichkeit geht also mit der neuen Methodik nicht verloren, sondern wird lediglich nicht mehr als Teil der Qualitätsbewertung gesehen. Ein Ziel, das mit dieser Methodik verfolgt wird, besteht gerade darin, mehr Ressourcen für solche qualitätssteigernden Maßnahmen zur Verfügung zu stellen.

### **Einholung von Stellungnahmen**

Während die Empfehlung, Stellungnahmen zu Indikatorergebnissen auf Basis bundeseinheitlicher statistischer Kriterien einzuholen, von einer Organisation begrüßt wird (GKV-SV, S. 26), wird dies von einer anderen kritisch gesehen (KZBV, S. 18) und auf bestehende Richtlinien verwiesen. Auch der Wegfall der Möglichkeit, in bestimmten Fällen sog. Hinweise zu versenden, statt eine Stellungnahme anzufordern, wurde aus Sorge vor einem höheren Aufwand kritisch gesehen (KBV, S. 11).

**IQTIG:** Der G-BA hat das Institut beauftragt, Empfehlungen zu entwickeln, wie bewertungsstellenabhängige Heterogenität vermieden und das Verfahren effizienter gestaltet werden kann. Die Anwendung bundeseinheitlicher statistischer Kriterien für die Einholung von Stellungnahmen ist geeignet, beide Ziele ohne Zielkonflikte zu erreichen. Auf der einen Seite führt die Anwendung fachwissenschaftlich hergeleiteter und bundeseinheitlicher Kriterien zur Elimination bewertungsstellenabhängiger Heterogenität bei der Einholung von Stellungnahmen. Auf der anderen Seite führt die Implementation statistischer Kriterien zu einer deutlichen Zeit- und Aufwandseinsparung, da die Entscheidung nicht mehr durch ein Expertengremium getroffen werden muss.

Der Wegfall der Möglichkeit, Hinweise zu versenden, ist aus Sicht des Instituts folgerichtig. Wenn ein Indikatorergebnis gemäß der bundeseinheitlichen Auswertungsmethodik einen hinreichenden Hinweis auf ein Qualitätsdefizit darstellt, ist eine Stellungnahme anzufordern. Andernfalls sollten die Rückmeldeberichte selbst ausreichende Hinweise auf die Versorgungsqualität bieten.

## 5 Statistische Auswertungsmethodik

Mehrere Stellungnahmen kritisieren, dass Kapitel 5 für Nichtstatistiker bzw. für Leserinnen und Leser „ohne akademische Vorkenntnisse in den Methoden der Statistik“ unverständlich sei (LQS Hessen, S. 1, ähnlich DGIM, S. 2; KZBV, S. 13; BÄK, S. 8; DKG, S. 18;).

**IQTIG:** Das IQTIG empfiehlt eine neue statistische Methodik, die den Anspruch hat, dem komplexen Entscheidungsproblem der quantitativen Auffälligkeitseinstufung im Rahmen eines zweistufigen Prozesses und unter Berücksichtigung fallzahlabhängiger Unsicherheit gerecht zu werden. Ziel von Kapitel 5 ist dabei die Beschreibung, Begründung und Diskussion der dafür entwickelten Methoden. Vor diesem Hintergrund kann in diesem Kapitel ein gewisser Grad an Komplexität nicht vermieden werden, sodass sich das Kapitel eher an Leserinnen und Leser mit Kenntnissen der statistischen Fachterminologie und Methodik richtet. Jedoch wurde darauf geachtet, dass die Beschreibung auch für einen erweiterten Kreis von Leserinnen und Lesern mit statistischen Kenntnissen, wie z. B. Epidemiologen und Gesundheitsökonominnen, verständlich ist. Eine gewisse Komplexität in der Beschreibung ist für eine wissenschaftlich fundierte Bearbeitung dieses komplexen Themas unvermeidlich. Um die Auswirkungen der empfohlenen statistischen Methodik allgemeinverständlich zu vermitteln, wurde zusätzlich zum bereits bestehenden Abschnitt 5.6 des Abschlussberichts (Zusammenfassung der Empfehlungen) mit Abschnitt 5.5 ein zusätzlich illustrierender Text mit Beispiel ergänzt.

Eine Organisation kritisiert die Verwendung des Begriffs „Kompetenzparameter“, da dies der Behauptung gleichkäme, „dass sich mit Qualitätsindikatoren nicht nur vergangene, sondern auch zukünftig zu erwartende Qualität messen lasse“ (BÄK, S. 4).

**IQTIG:** Mit dem Begriff „Kompetenzparameter“ wird der dem jeweiligen Indikator zugrunde liegende (latente) Parameter bezeichnet, der üblicherweise in der Inferenzstatistik mittels eines Punktschätzers (z. B. beobachtete Rate) geschätzt wird. Damit wird zum Ausdruck gebracht, dass der beobachtete Punktschätzer nur eine mit Unsicherheit behaftete Schätzung für den eigentlich interessierenden Parameter darstellt (siehe auch Kap. 15, IQTIG 2019). Um dies zu erläutern, wurde die Einführung des „Kompetenzparameters“ im Abschlussbericht umfassend erweitert.

Eine Organisation fragt, ab wann das IQTIG von einer „kleinen Fallzahl“ spricht und welche Methodik ab welcher Fallzahl angewandt werden soll und bittet das IQTIG um konkrete Vorschläge zum Umgang am Beispiel eines Transplantationsverfahrens (GKV-SV, S. 7).

**IQTIG:** Die im Bericht empfohlene Auswertungsmethodik ist insbesondere deshalb gewählt worden, weil sie die Berücksichtigung fallzahlabhängiger Unsicherheit erlaubt und auch bei kleinen Fallzahlen anwendbar ist. Sie sollte daher für alle QS-Verfahren und Leistungserbringer, unabhängig von ihrem jeweiligen Behandlungsvolumen, angewandt werden. Durch diese einheitliche Auswertungsmethodik ist eine Unterscheidung zwischen QS-Verfahren oder Leistungserbringern mit kleiner und großer Fallzahl nicht notwendig. Es kann jedoch sinnvoll sein, das Signifikanzniveau  $\alpha$  QS-Verfahren-spezifisch zu wählen und bei der Wahl die Fallzahlen oder die bisherige Anzahl an durchgeführten Stellungnahmeverfahren zu berücksichtigen. Um dies zu illustrieren, wurde die Evaluation der empfohlenen Auswertungsmethodik (Abschnitt 10.3) um das Beispiel des Verfahrens *Nierentransplantation* ergänzt. In der Zusammenfassung und Empfehlung (Abschnitt 5.6) weist das IQTIG darauf hin, dass das Signifikanzniveau  $\alpha$  übergangsweise so gewählt werden könnte, dass die Anzahl an quantitativen Auffälligkeiten in etwa der Anzahl an bisher durchgeführten Stellungnahmeverfahren entspricht. Dies würde zu QS-Verfahren-spezifischen Signifikanzniveaus führen, da die Anzahl an durchgeführten Stellungnahmeverfahren stark zwischen den QS-Verfahren variiert, wie an der Betrachtung der zwei QS-Verfahren *Hüftendoprothesenversorgung* und *Nierentransplantation* deutlich wird. Abgesehen von der Wahl eines bestimmten Signifikanzniveaus empfiehlt das IQTIG jedoch eine über alle QS-Verfahren und Leistungserbringer einheitliche statistische Auswertungsmethodik.

In mehreren Stellungnahmen wurde angeregt, die quantitative Auffälligkeitseinstufung auf die Daten von drei statt zwei Erfassungsjahren zu beziehen (LQS Hessen, S. 3; gem. StN LQS BE, MV, SL, RP, TH, S. 5).

**IQTIG:** Die Berücksichtigung von drei statt nur zwei Erfassungsjahren für die quantitative Auffälligkeitseinstufung erscheint grundsätzlich sinnvoll. Um Erfahrung mit der Anwendung der neuen Auswertungsmethodik zu gewinnen und einen sinnvollen Umgang mit den Fällen aus dann vorangegangenen Erfassungsjahren im Stellungnahmeverfahren zu finden, empfiehlt das IQTIG jedoch, die Methodik zunächst auf die Auswertung zweier Erfassungsjahre zu beschränken und ggf. in einem nächsten Schritt zu erweitern.

Es ist jedoch wichtig, zwischen den Argumenten für die Befürwortung der drei Jahre zu differenzieren. Aus Sicht des IQTIG sollte es sich dabei nicht um eine praktisch motivierte Überlegung handeln, weil „die Qualitätsbewertung im Erfassungsjahr +1 zu einem Zeitpunkt erfolgt, wo Verbesserungsmaßnahmen erst spät greifen können“ (gem. StN LQS BE, MV, SL, RP, TH, S. 5). Dies ist aus Sicht des IQTIG irrelevant für die Bewertung der Qualitätsergebnisse, kann aber sehr sinnvoll bei z. B. der Vereinbarung von qualitätsfördernden Maßnah-

men berücksichtigt werden. Die Argumentation für eine Erweiterung der Datenbasis sollte daher eher eine verbesserte Evidenzgrundlage für die Bewertung sein.

Eine Organisation merkt an, dass bei der qualitativen Bewertung einer quantitativen Auffälligkeit die Fallzahl der jeweiligen Einrichtung mit hoher Wahrscheinlichkeit berücksichtigt wird und daher aus den modellierten Ergebnissen nicht geschlossen werden kann, dass die qualitative Einstufung in Wahrheit fallzahlunabhängig und die Nichtberücksichtigung von statistischer Unsicherheit falsch sei (DNVF, S. 2).

**IQTIG:** Das IQTIG untersucht in diesem Bericht weder den Zusammenhang zwischen der Fallzahl eines Leistungserbringers und dessen zugrunde liegendem Kompetenzparameter, noch zwischen der Fallzahl und dem Ergebnis der qualitativen Auffälligkeitsbewertung. Stattdessen wird bei der Entwicklung der statistischen Methodik davon ausgegangen, dass der zugrunde liegende Kompetenzparameter eines Leistungserbringers unabhängig von dessen Fallzahl ist. Selbst wenn der zugrunde liegende Kompetenzparameter fallzahlabhängig wäre, sollte die fallzahlabhängige Stochastizität der beobachteten Ergebnisse bei der quantitativen Auffälligkeitseinstufung berücksichtigt werden.

In einer Stellungnahme wird kritisiert, dass sich der empfohlene Methodikansatz der künstlichen Intelligenz (KI) bedient und dadurch eine Plausibilisierung von computergenerierten Auffälligkeitseinstufungen erschwert werden (DGNHO, S. 1).

**IQTIG:** Die Methodik der Auffälligkeitseinstufung bleibt weiterhin ein zweistufiges Verfahren, in welchem eine quantitative Auffälligkeitseinstufung zu einem Stellungnahmeverfahren führt. In diesem Stellungnahmeverfahren wird über ein standardisiertes Verfahren unter Einbeziehung fachlicher Expertise die Validität der Qualitätsaussage eines Indikators für einen konkreten Leistungserbringer geprüft. Darüber hinaus basiert auch die vom IQTIG empfohlene quantitative Auffälligkeitseinstufung auf einem System von transparenten Regeln, anhand derer jede Einstufungsentscheidung nachvollzogen und begründet werden kann. Die Berücksichtigung von Elementen wie Stochastik und Unsicherheit ist kein KI-System, sondern ein fachlich fundierter statistischer Zugang zu dem komplexen Entscheidungsproblem.

Eine Organisation kritisiert, dass bei der statistisch relevanten Auffälligkeitseinstufung Qualitätsdefizite eher toleriert werden, wenn wenige Patientinnen und Patienten betroffen sind, was ethisch kaum zu tolerieren sei (DGK, S. 3).

**IQTIG:** Die verschiedenen Auffälligkeitseinstufungen werden über die Lösung eines Entscheidungsproblems hergeleitet, wobei sich je nach den zugrunde liegenden Aufwand-Nutzen-Relationen unterschiedliche Einstufungsmethoden als optimal erweisen. Für eine umfassende Betrachtung wären diese gesundheitsökonomisch motivierten Überlegungen entsprechend mit weiteren Zielkriterien der Qualitätssicherung abzuwägen. Diese Abwägung wird jedoch nicht im vorliegenden Bericht getroffen, weshalb auch die Konsequenzen der verschiedenen Einstufungsmethoden nicht hinsichtlich z. B. ethischer Aspekte diskutiert werden. In Abschnitt 5.3.5 wird jedoch kritisch diskutiert, dass es im Rahmen dieser Einstufungsmethodik Konstellationen geben kann, in denen ein Leistungserbringer unabhängig vom Ergebnis nicht quantitativ auffällig werden kann und entsprechende Vorgehensweisen für diese Situation dargelegt, um dem Problem Abhilfe zu schaffen. Eine Grundproblematik bleibt jedoch bei den ethischen Konsequenzen jeglicher Auswertungsmethodik unbenommen: Es kann nicht jedem Fall mit unerwünschtem Qualitätsereignis nachgegangen werden.

In derselben Stellungnahme werden die vorgestellten statistischen Methoden für die Einstufung von Leistungserbringern mit kleinen Fallzahlen als nicht anwendbar kritisiert (DGK, S. 3).

**IQTIG:** Die empfohlene statistische Methodik ist unter anderem deshalb gewählt worden, weil sie die fallzahlabhängige Stochastizität der beobachteten Indikatorergebnisse angemessen berücksichtigt und daher auch für Leistungserbringer mit kleinen Fallzahlen anwendbar ist. Darüber hinaus wird vorgeschlagen, die Einstufung auf Basis mehrerer Erfassungsjahre vorzunehmen, um durch eine größere Datengrundlage den Einfluss zufälliger Faktoren auf die beobachteten Indikatorergebnisse, insbesondere auf jene von Leistungserbringern mit kleinen Fallzahlen, zu verringern. Anhand eines theoriegeleiteten Vergleichs wird dargelegt, dass dies in der Erwartung zu besseren Klassifikationsergebnissen führt als eine Einstufung basierend auf nur einem Erfassungsjahr.

In einer Stellungnahme wurde kritisiert, dass das Signifikanzniveau als zusätzliche „Stellschraube“ für die Einholung von Stellungnahmen bezeichnet wird und dass stattdessen die Einstufung als Qualitätsdefizit grundsätzlich zum Stellungsverfahren führen müsse (DGK, S. 4).

**IQTIG:** Über das Signifikanzniveau wird der Schwellenwert für die statistische Evidenz für die Entscheidung „hinreichender Hinweis auf ein Qualitätsdefizit“ festgelegt. Anders als im aktuellen Verfahren, in dem beim Vorliegen einer quantitativen Auffälligkeit auch Hinweise versendet werden können, empfiehlt das IQTIG, dass zukünftig jede quantitative Auffälligkeitseinstufung zur Aufnahme eines Stellungsverfahrens führen sollte.

Eine Organisation hinterfragt, warum in Abschnitt 10.3 die statistisch relevante Auffälligkeitseinstufung als gleichwertige Option zur statistisch signifikanten Auswertungsmethodik evaluiert wird, obwohl sich in Kapitel 5 für die Verwendung der statistisch signifikanten Auffälligkeitseinstufung ausgesprochen wurde (KBV, S. 12).

**IQTIG:** Abschnitt 10.3 wurde als Würdigung dahingehend angepasst und enthält nun nur noch die Evaluation der empfohlenen statistischen Methodik. Das beispielhaft ausgewählte QS-Verfahren *Hüftendoprothesenversorgung* wurde darüber hinaus um das Verfahren der Nierentransplantationen als Verfahren mit kleineren Fallzahlen ergänzt.

## 6 Fachliche Bewertung

### Bewertungsschema

Die Kategorien des vorgeschlagenen Bewertungsschemas für die Ergebnisse der fachlichen Bewertung wurden in vielen Stellungnahmen angesprochen. Während eine stellungnehmende Organisation insgesamt keine fachlichen Einwände sah (DGTHG, S. 1), suggeriere das Schema nach Auffassung einer anderen Organisation fälschlich „die Sicherheit einer Tatbestandsfeststellung“ (BÄK, S. 6). In einer anderen Stellungnahme wurde angemerkt, dass selbst nach Abschluss des Strukturierten Dialogs nur von „Hinweisen“ auf das Fehlen oder auf das Vorhandensein eines Qualitätsdefizits gesprochen werden könne (GQH, S. 3). In einer Stellungnahme wurden die Begrifflichkeiten „Qualitätsdefizit“ bzw. „kein Hinweis auf Qualitätsdefizit“ als richtige und klare Benennung hervorgehoben (PatV, S. 4 und 12), in zwei anderen Stellungnahmen wurde hingegen die Befürchtung geäußert, die Formulierungen „kein Hinweis auf Qualitätsdefizit“ (DGK, S. 3; DKG, S. 19) und „kein hinreichender Hinweis auf Qualitätsdefizit“ (DGK, S. 3; DKG, S. 19; KZBV, S. 13) stellten die Leistungserbringer unter den Generalverdacht schlechter Qualität.

**IQTIG:** Das Bewertungsschema für die Ergebnisse der fachlichen Bewertung stellt einen Kompromiss dar, da es für unterschiedliche Zwecke und Adressaten verwendet wird: Einerseits sollen die Kategorien eines solchen Schemas die Bewertungsergebnisse detailliert beschreiben, um z. B. Leistungserbringer und qualitätsfördernde Stellen auf die richtigen Maßnahmen zur Verbesserung hinzuweisen, andererseits sollen die Kategorien für Zwecke der Rechenschaft gegenüber der Allgemeinheit (*accountability*) eine eindeutige und verständliche Aussage treffen. Dies wurde im Text des Abschlussberichts nun deutlicher dargestellt.

Aus den im Bericht erläuterten Gründen ist das Nichtverfehlen des Referenzbereichs eines Indikators nicht gleichzusetzen mit Sicherheit über das Erreichen des Referenzbereichs. Das bedeutet, dass bei Abwesenheit eines „Hinweises auf ein Qualitätsdefizit“ offenbleibt, ob zureichende Qualität vorliegt. Die Bezeichnung „kein Hinweis auf Qualitätsdefizit“ ist daher nach Auffassung des IQTIG sachgerecht und unterstellt keineswegs schlechte Versorgungsqualität in den Einrichtungen.

Eine stellungnehmende Organisation befürwortete ausdrücklich die Beschränkung der oberen Ebene des Bewertungsschemas auf drei Kategorien (QS-NRW, S. 3). Unterschiedliche Auffassungen wurden dagegen bezüglich der Differenzierung der Einstufungsergebnisse mittels Unterkategorien geäußert. Eine stellungnehmende Organisation vertrat die Ansicht, dass dem erhöhten Aufwand für die

Vergabe von Unterkategorien kein erkennbarer Nutzen gegenüberstehe (DGTHG, S. 4), in einer weiteren Stellungnahme wurde dies für die Differenzierung der Kategorie „Qualitätsdefizit“ ebenfalls infrage gestellt (DKG, S. 21). Aus Sicht einer stellungnehmenden Organisation sei fraglich, ob die Differenzierung der Kategorie „Qualitätsdefizit“ von den Adressaten verstanden werde (DNVF, S. 3). Andere Organisationen sprachen sich dagegen für eine Differenzierung aus (BAQ, S. 2; GQH, S. 3; NKGeV QS, S. 2).

**IQTIG:** Vor- und Nachteile der Differenzierung der Kategorie „Qualitätsdefizit“ mittels Ziffern wurden im Abschlussbericht ausführlicher erläutert. Das IQTIG schlägt weiterhin eine solche differenzierte Einstufung vor, da sie nach dem vorgeschlagenen Konzept ohne nennenswerten zusätzlichen Aufwand vorgenommen werden kann (siehe neuen Abschnitt 6.6 im Abschlussbericht).

Aus Sicht mehrerer stellungnehmender Organisationen sollten Datenfehler nicht der Kategorie „Qualitätsdefizit“ zugeordnet werden (BAQ, S. 2; DKG, S. 20; GQH, S. 3, NKGeV QS, S. 2). Eine andere stellungnehmende Organisation beurteilte dagegen die Differenzierung eines „Qualitätsdefizits“ gemäß dem vorgeschlagenen Algorithmus als nachvollziehbar (GKV-SV, S. 23).

**IQTIG:** Bei der fachlichen Bewertung sollen Defizite der Versorgungsqualität nicht durch angegebene oder tatsächliche Datenfehler verdeckt werden können. Die Zuordnung von Datenfehlern, die vom Leistungserbringer zu verantworten sind, zur Hauptkategorie „Qualitätsdefizit“ soll Anreize vermeiden, die Qualitätsbewertung durch Geltendmachung von Datenfehlern zu umgehen.

Von mehreren der stellungnehmenden Organisationen wurde kritisiert, dass bei formal unzureichender Stellungnahme eines Leistungserbringers die Bewertung „Qualitätsdefizit“ erfolgen soll. Formfehler seien nicht als Qualitätsdefizit zu bezeichnen (DKG, S. 11, 19 f. und 23 f.; GQH, S. 3; KBV, S. 4 und 9; KZBV, S. 13; QB Bremen, S. 2; NKGeV QS, S. 2; QS-NRW, S. 3). Dies sei auch für Patientinnen und Patienten irreführend (DKG, S. 11 und 19 f.). Zudem solle nach Auffassung mehrerer stellungnehmender Organisationen im Bewertungsschema zwischen der Nichtabgabe einer Stellungnahme und formal ungültigen Stellungnahmen unterschieden werden (AEKSA QS, S. 2; DGGG, S. 2; DKG, S. 24; QS-NRW, S. 3; QiG BW GmbH, S. 2).

**IQTIG:** Nach Auffassung des IQTIG ist eine Nichteinhaltung der formalen Vorgaben des Stellungnahmeverfahrens nicht gleichbedeutend mit einer Bestätigung der Indikатораussage. Andererseits sollte es – sowohl aus Gründen der Effizienz des Verfahrens als auch der Fairness – nicht möglich sein, die Qualitätsbewertung durch nicht kriterienkonforme oder nicht fristgerecht eingereichte Stellungnahmen zu umgehen oder den Bewertungsprozess zu verzögern. Um Anreize für Formfehler zu vermeiden, wird daher empfohlen, in allen

Fällen, in denen der betreffende Leistungserbringer keine adäquaten Informationen zur Verfügung gestellt hat, die die Indikatoraussage entkräften, die Bewertung „Qualitätsdefizit“ zu vergeben. Die Information, ob eine Stellungnahme zur fachlichen Bewertung vorlag, ist im vorgeschlagenen Bewertungsschema durch Vergabe von Ziffern ersichtlich.

Mehrere stellungnehmende Organisationen wiesen darauf hin, dass die Nichtabgabe einer Stellungnahme durch den Leistungserbringer nicht mit einem Qualitätsdefizit gleichzusetzen sei (BÄK, S. 6; DGGG, S. 1 f.; DKG, S. 11; GQH, S. 3; KZBV, S. 13 und 18; QS-NRW, S. 3). Von einer stellungnehmenden Organisation wurde vorgeschlagen, die Nichtabgabe stattdessen als „Organisationsdefizit“ zu bezeichnen (QiG BW GmbH, S. 2). In einer weiteren Stellungnahme wird in diesem Zusammenhang angeführt, dass zwischen der versäumten Abgabe und dem bewussten Verzicht auf Abgabe einer Stellungnahme unterschieden werden müsse (gem. StN LQS BE, MV, RP, SL, TH, S. 6). Eine andere stellungnehmende Organisation sah dagegen die Einstufung einer Nichtabgabe als „Qualitätsdefizit“ als nachvollziehbar an, da dies der Effizienz des Verfahrens diene (GKV-SV, S. 18).

**IQTIG:** Ein grundsätzlicher Unterschied zwischen einer formal unzureichenden Stellungnahme (z. B. aufgrund von Verstößen gegen den Datenschutz) und einer versäumten Frist für die Abgabe einer Stellungnahme besteht aus Sicht des IQTIG nicht, da in beiden Fällen keine gültigen Informationen für eine fachliche Bewertung zur Verfügung stehen. Für beide Konstellationen wird daher weiterhin eine gemeinsame Unterkategorie (A0) empfohlen. Nachvollziehbar ist jedoch, dass der bewusste Verzicht auf eine Stellungnahme keinen Formfehler bedeutet, sondern die Bestätigung der Indikatoraussage durch den Leistungserbringer. Diese Konstellation wurde daher im Abschlussbericht jetzt der Unterkategorie A3 zugeordnet.

Zwei der stellungnehmenden Organisationen vertraten die Ansicht, dass auch bei einem Indikatorergebnis innerhalb des Referenzbereichs ein Hinweis auf ein Qualitätsdefizit vorliegen könne. Die Bezeichnung der Kategorie U0 solle entsprechend geändert werden (NKGeV QS, S. 2; QB Bremen, S. 3).

**IQTIG:** Bei der in den beiden Stellungnahmen geäußerten Auffassung bleibt offen, auf welchen Informationen anderweitige Hinweise auf ein Qualitätsdefizit basieren würden. Das IQTIG versteht diese Äußerung als Verweis darauf, dass neben Qualitätsindikatoren auch andere Methoden zur Qualitätsbewertung existieren, die Hinweise auf Qualitätsdefizite geben können und dass ein einzelner Qualitätsindikator nur eins von vielen möglichen Qualitätsmerkmalen abbilden kann. Nach dem Verständnis des IQTIG überschreitet diese Betrachtung jedoch den Rahmen der indikatorbasierten Qualitätssicherung. Mittels der gesetzlich verpflichtenden indikatorbasierten Qualitätssiche-

Die Qualität der Versorgung kann nur für die Merkmale der Versorgung abgebildet werden, für die Indikatoren existieren. Qualitätsdefizite in der Versorgung können auch bestehen, obwohl die bestehenden Indikatoren der indikatorbasierten Qualitätssicherung keinen Hinweis auf ein Qualitätsdefizit geben.

### Methodik und Ablauf der fachlichen Bewertung

Der vorgeschlagene Ablauf der fachlichen Bewertung wurde von den stellungnehmenden Organisationen unterschiedlich bewertet. In mehreren Stellungnahmen wurde der geschilderte Bewertungsalgorithmus als fachlich nachvollziehbar (DGTHG, S. 3 f.; DKG, S. 21; DNFV, S. 3) und verständlich (PatV, S. 9) beurteilt. Als positiv gewertet wurde auch das Prinzip, mittels Nachberechnungen eine möglichst objektive Bewertungsbasis zu schaffen (PatV, S. 8; GKV-SV, S. 20 f.), sowie die daraus resultierende Ähnlichkeit zum Vorgehen bei planungsrelevanten Qualitätsindikatoren (PatV, S. 8; QS-NRW, S. 5).

Viele stellungnehmende Organisationen äußerten andererseits die Sorge, dass die Nachberechnungen im Rahmen der fachlichen Bewertung mit einer hohen Komplexität einhergingen. Dies könne zu unerwünschter Heterogenität zwischen den bewertenden Stellen führen (BAQ, S. 2; QB Bremen, S. 2; LQS-Sachsen, S. 3; QiG BW GmbH, S. 2 f.; QS-NRW, S. 3) und erfordere ein Schulungskonzept bei Einführung (GKV, S. 9; KBV, S. 9; NKGeV QS, S. 2; QiG BW GmbH, S. 3). Zudem seien die Bewertungsergebnisse weniger nachvollziehbar (ÄK-SA, S. 1; BÄK, S. 6; BAQ, S. 2; DPR, S. 2; GQH, S. 3; LAG Bremen, S. 2; SLAEK QS, S. 3; NKGeV QS, S. 2). Befürchtet wurde darüber hinaus, dass die Nachberechnung mit einem hohen Aufwand einhergehe (BÄK, S. 5 f.; BAQ, S. 2; QiG BW GmbH, S. 2 f.; KZBV, S. 15; QS-NRW, S. 3; QB Bremen, S. 2; gem. StN LQS BE, MV, RP, SL, TH, S. 8 f. und S. 14 f.); dieser wird von einigen stellungnehmenden Organisationen auch deshalb vermutet, da angenommen wird, dass durch die Einführung von Nachberechnungen mehrfache Sitzungen der Fachkommission erforderlich würden (AEKSA QS, S. 1; gem. StN LQS BE, MV, SL, RP, TH, S. 10 und 15).

**IQTIG:** Das vorgeschlagene Konzept für die fachliche Bewertung verfolgt das Ziel, die Regeln für die Beurteilung von Stellungnahmen eindeutig und transparent zu beschreiben. Dadurch wird die Komplexität, mit der diese Beurteilung aufgrund der Komplexität der zu beurteilenden Sachverhalte grundsätzlich einhergeht, sichtbar. Bei einem rein heuristischen Vorgehen bleibt die Komplexität der Beurteilung dagegen verborgen. Die Auffassung, dass bei Anwendung klar festgelegter Regeln zur Nachberechnung die Bewertungsergebnisse schwerer nachzuvollziehen seien als bei einem Vorgehen, das ausschließlich auf einer heuristischen Beurteilung beruht, ist daher aus Sicht des IQTIG nicht nachvollziehbar. Richtig ist, dass die angestrebte Steigerung der Objektivität und Nachvollziehbarkeit der Beurteilung nur dann zu erwarten ist, wenn die expliziten Beurteilungsregeln einheitlich und konsequent angewendet

werden. Im Abschlussbericht (siehe Abschlussbericht, Abschnitt 11.5) wurde die Empfehlung einer strukturierten Begleitung der Umsetzung daher nochmals deutlicher formuliert.

Der Abschlussbericht wurde um einen Abschnitt ergänzt, in dem der Aufwand, der sich bei Umsetzung des vorgeschlagenen Konzepts für die fachliche Bewertung ergibt, ausführlich diskutiert wird (siehe Abschlussbericht, Abschnitt 6.6). Das IQTIG geht dabei von der in Kapitel 3 des Abschlussberichts methodisch hergeleiteten Prämisse aus, dass das Stellungnahmeverfahren der Bestätigung oder Entkräftung der Indikатораussage dient. Wenn fundierte Aussagen über die Erfüllung des Qualitätsmerkmals eines Indikators getroffen werden sollen, ist der Aufwand, der durch eine strukturierte Analyse der Einflüsse auf ein Indikatorergebnis entsteht, nicht vermeidbar. Würden stattdessen die Indikatorergebnisse nur als Aufgreifkriterien genutzt und das Stellungnahmeverfahren als ein Bewertungsinstrument verstanden, durch das Qualität umfassender als mit einem Qualitätsindikator bewertet werden soll, könnte auf die Validierung der Indikатораussage verzichtet werden. Für eine solche anschließende umfassende Qualitätsbewertung wie beispielsweise bei einem Audit-Verfahren sind jedoch ebenfalls explizite Bewertungskriterien festzulegen, um ein objektives Verfahren zu gewährleisten. Denn auch in einem solchen Fall ist allein schon aus Transparenzgründen ein expliziter Kriterienkatalog sowie eine Methodik für die Bewertung notwendig. Daher ist auch bei einer angemessen standardisierten Vorgehensweise im Rahmen einer umfassenderen Bewertung mit erhöhten Aufwänden gegenüber einem gering standardisierten Vorgehen zu rechnen.

Von einigen stellungnehmenden Organisationen wurden Änderungen der formalen Rahmenbedingungen der fachlichen Bewertung angeregt. Es wurde vorgeschlagen, der LAG bzw. Bundesstelle die Möglichkeit zu geben, Präzisierungen der Stellungnahmen vom Leistungserbringer anzufordern (DGTHG, S. 6; DKG, S. 3 und 21; KZBV, S. 17 f.) und bei verspäteter Abgabe einer Stellungnahme durch den Leistungserbringer Fristverlängerungen und Erinnerungsschreiben einzusetzen (gem. StN LQS BE, MV, SL, RP, TH, S. 11; KZBV, S. 17 f.). Außerdem sollten nach Auffassung einer stellungnehmenden Organisation Leistungserbringer mit dem Bewertungsergebnis „Qualitätsdefizit“ die Möglichkeit zu einer Erwidern erhalten (DKG, S. 3 und 11 f.).

**IQTIG:** Das vorgeschlagene Konzept der fachlichen Bewertung sieht die Möglichkeit vor, innerhalb der Stellungnahmefrist Erinnerungen zu versenden sowie Ergänzungen anzufordern und zu berücksichtigen. Aus Gründen der Effizienz des Verfahrens und der gewünschten Verkürzung der Verfahrensdauer seitens des G-BA empfiehlt das IQTIG jedoch weiterhin, auf Fristverlängerungen und auf Präzisierungen bzw. Erwidern durch den Leistungserbringer nach

Fristende zu verzichten. Dies dient auch der Fairness zwischen den Leistungserbringern und entspricht dem Vorgehen in anderen Verfahren, in denen ein besonderes Augenmerk auf Gleichbehandlung liegt, beispielsweise Bewerbungen, Prüfungen und Ausschreibungen. Es wird empfohlen, bei der Einführung des Konzepts Übergangsregelungen einzusetzen, um den Beteiligten eine rechtzeitige Anpassung an die geänderten Rahmenbedingungen zu ermöglichen.

In einer Stellungnahme wurde geäußert, dass eine Neuberechnung von Indikatorwerten alle Leistungserbringer umfassen müsse, um eine bundesweite Vergleichbarkeit der Werte zu gewährleisten (QS-NRW, S. 2).

**IQTIG:** Die Neuberechnung des Indikatorwerts dient ausschließlich der Beurteilung, ob für einen bestimmten Leistungserbringer das Bewertungsergebnis des Indikators zutreffend ist. Eine Veröffentlichung oder vergleichende Darstellung der nachberechneten Indikatorwerte ist aus den genannten methodischen Gründen nicht sinnvoll (siehe dazu Abschnitt 6.2 im Bericht). Dies würde eine vollständige Erhebung der bisher nicht im Indikator berücksichtigten Einflussfaktoren bei allen Leistungserbringern erfordern. Anpassungen der Datenerhebung, z. B. durch Ergänzung um die Erfassung weiterer häufiger Einflussfaktoren, sollten stattdessen im Rahmen der Weiterentwicklung des jeweiligen Indikators erfolgen.

### Stellungnahmen – Einflussfaktoren

Mehrere stellungnehmende Organisationen baten um nähere Ausführungen, nach welchen Kriterien entschieden werden sollte, ob ein in der Stellungnahme des Leistungserbringers angeführter Einflussfaktor von diesem zu verantworten ist oder nicht (DKG, S. 23; GKV-SV, S. 8; SLAEK QS, S. 3; gem. StN LQS BE, MV, SL, RP, TH, S. 6).

**IQTIG:** Kriterien, wann ein Einflussfaktor auf das Indikatorergebnis vom Leistungserbringer zu verantworten ist, lassen sich nicht sinnvoll allgemeingültig formulieren, da derselbe Einflussfaktor je nach Leistungserbringer und betroffenem Qualitätsindikator in unterschiedlichem Maße für das Indikatorergebnis relevant sein kann. Der Standardisierung sind hier aus methodischen Gründen Grenzen gesetzt. Die medizinisch-fachliche Expertise kann nicht durch explizite Kriterien ersetzt werden. Die Beurteilung, ob ein Einflussfaktor vom Leistungserbringer beeinflusst werden konnte und daher von ihm zu verantworten ist, beruht daher wie bisher auf der fachlichen Expertise der Fachkommission und der LAG bzw. Bundesstelle.

## Umgang mit Datenfehlern

In mehreren Stellungnahmen wurde die Bewertung von Datenfehlern im Rahmen der fachlichen Bewertung angesprochen. Aus Sicht zweier stellungnehmender Organisationen sei unklar, wie zwischen Datenfehlern unterschieden werden könne, die vom Leistungserbringer zu verantworten sind, und solchen, die nicht vom Leistungserbringer zu verantworten sind (DKG, S. 20 f.; GKV-SV, S. 21 f.). Von einer stellungnehmenden Organisation wurde vorgeschlagen, dass alle in der Stellungnahme eines Leistungserbringers angegebenen Datenfehler mit der Fachkommission besprochen werden sollten (gem. StN LQS BE, MV, SL, RP, TH, S. 9).

**IQTIG:** Grundsätzlich sind die Leistungserbringer für die ordnungsgemäße Dokumentation der Daten für die externe Qualitätssicherung verantwortlich. Dies schließt auch die rechtzeitige einrichtungsinterne Qualitätskontrolle des Dokumentationsprozesses sowie der verwendeten Hilfsmittel ein. Beispielsweise kann erwartet werden, dass bei technischen Problemen bei der QS-Dokumentation, die der Leistungserbringer frühzeitig erkennen konnte, rechtzeitig Maßnahmen ergriffen wurden, um diese Probleme zu beheben. Es sind jedoch seltene Konstellationen denkbar, in denen z. B. ein Fehler in den QS-Daten aus technischen Gründen nicht rechtzeitig vom Leistungserbringer erkannt werden konnte. Die Beurteilung der Verantwortlichkeit für Datenfehler erfolgt also durch Beantwortung der Frage, ob der Leistungserbringer bei der Erstellung der QS-Dokumentation seiner Sorgfaltspflicht nachgekommen ist und dem Auftreten von Fehlern in angemessenen Maße entgegengewirkt hat. Eine abschließende Festlegung a priori, welche Datenfehler vom Leistungserbringer zu verantworten bzw. nicht zu verantworten sind, ist aus Sicht des IQTIG nicht sinnvoll. Diese Beurteilung soll daher durch die LAG bzw. Bundesstelle, ggf. unter Beratung durch die Fachkommission, erfolgen. Es ist davon auszugehen, dass viele der in den Stellungnahmen der Leistungserbringer angeführten Gründe für Datenfehler eher technischer Natur sind und ohne medizinisches Fachwissen beurteilt werden können, sodass eine Diskussion der Datenfehler durch die Fachkommission meist verzichtbar ist.

Von zwei Organisationen wurde des Weiteren kritisch gesehen, dass in den Stellungnahmen angegebene Datenfehler nicht verlässlich daraufhin geprüft werden könnten, ob sie tatsächlich vorliegen und das Indikatorergebnis beeinflusst haben (GKV-SV, S. 13; gem. StN LQS BE, MV, SL, RP, TH, S. 8). Des Weiteren wurde zu Bedenken gegeben, dass eine Nichtberücksichtigung von Datenfehlern, die ein Leistungserbringer in seiner Stellungnahme geltend gemacht hat, zu Konflikten führen könne (GKV-SV, S. 21).

**IQTIG:** Eine generelle Prüfung der dokumentierten QS-Daten für alle Leistungserbringer mit einem Indikatorergebnis außerhalb des Referenzbereichs und

für alle Verfahren nach DeQS-RL wäre mit einem erheblichen Aufwand insbesondere für Leistungserbringer, prüfende Stellen sowie Auswertungsstellen verbunden. Die umfangreichen Regelungen der plan. QI-RL zur Datenvalidierung lassen sich daher nach Auffassung des IQTIG nicht auf sämtliche Verfahren nach DeQS-RL übertragen. Wie bisher soll die Beurteilung von Datenfehlern auf Basis der Stellungnahmen erfolgen.

Im vorgeschlagenen Bewertungsschema werden Datenfehler, die in der Verantwortung der Leistungserbringer liegen, der Kategorie „A – Qualitätsdefizit“ zugeordnet. Ob ein vom Leistungserbringer benannter Datenfehler, der von ihm zu verantworten ist, tatsächlich vorgelegen und zum Indikatorendergebnis beigetragen hat, ist somit nur für die Einstufung in die Unterkategorien (Ziffern) des Bewertungsschemas von Bedeutung, die der Qualitätsverbesserung dienen. Es ist anzunehmen, dass für die Leistungserbringer ein geringerer Anreiz als bisher besteht, Datenfehler als Ursache für das Verfehlen des Referenzbereichs geltend zu machen, wenn diese ohnehin vom Leistungserbringer zu verantworten sind. Nach Auffassung des IQTIG ergibt sich daraus ein eher reduziertes Konfliktpotenzial in Bezug auf die Frage, ob ein Datenfehler tatsächlich vorgelegen hat.

### Referenzbereiche für die fachliche Bewertung

Mehrere stellungnehmende Organisationen schilderten es als nachvollziehbar, dass auf Basis perzentilbasierter Referenzbereiche keine Qualitätsdefizite benannt werden können (DGTHG, S. 4; GQH, S. 3; PatV, S. 9). Während in zwei Stellungnahmen gefordert wurde, deshalb eine Festlegung fester Referenzbereiche für alle Indikatoren vorzunehmen (DGTHG, S. 4; KBV, S. 4; PatV, S. 9), wurde in zwei anderen Stellungnahmen die Meinung geäußert, dass auch für Indikatoren mit perzentilbasierten Referenzbereichen im Stellungnahmeverfahren eine Qualitätsbewertung erfolgen könne (GQH, S. 3; gem. StN LQS BE, MV, SL, RP, TH, S. 9). In einer Stellungnahme wurde die Frage aufgeworfen, ob sich das Konzept, in der fachlichen Bewertung die Indikатораussage zu validieren, überhaupt auf Indikatoren mit perzentilbasierten Referenzbereichen anwenden lasse (GKV-SV, S. 22 f.).

**IQTIG:** In der externen Qualitätssicherung werden indikatorbasierte Qualitätsmessungen mit dem Zweck eingesetzt, Aussagen über das Erreichen festgelegter Standards zu treffen. Dies erfordert die Verwendung fester Referenzbereiche als Vergleichswerte, die daher für alle solchermaßen eingesetzten Indikatoren festgelegt werden sollten. Die Funktion eines Stellungnahmeverfahrens bei indikatorbasierten Qualitätsmessungen muss auf diesen Verwendungszweck abgestimmt sein und besteht in der Prüfung der Validität der Indikатораussage (siehe Kapitel 3 des Berichts).

Indikatoren, für die kein fester, kriteriumsbezogener Referenzbereich festgelegt wurde, eignen sich nicht für die Feststellung eines Qualitätsdefizits, sondern können allenfalls als Auslösekriterium für andere Instrumente der Qualitätsbeurteilung verwendet werden. Das IQTIG versteht die Organisationen, die eine Verwendung perzentilbasierter Referenzbereiche für angemessen halten, so, dass sie das Stellungnahmeverfahren nach DeQS-RL als ein solches, eigenes Instrument der Qualitätsbewertung ansehen. Für diese Methode der Qualitätsbewertung sind jedoch bisher keine Anforderungen an die Versorgung (z. B. in Form eines einheitlichen Kriterienkatalogs) definiert und durch den G-BA beschlossen. Im Gegensatz zu den einheitlichen Anforderungen an die Versorgung, die vom G-BA in Form der Qualitätsindikatoren und ihrer Referenzbereiche beschlossen sind, würde ein solches Vorgehen erhebliche Heterogenität zulassen. Da eine möglichst objektive Bewertung der Versorgungsqualität vorgenommen und unerwünschte Heterogenität verringert werden soll, wird dieses Vorgehen im vorliegenden Konzept nicht empfohlen.

Von einer stellungnehmenden Organisation wurde eine Klarstellung erbeten, wie groß bei der Berechnung des bereinigten Referenzbereichs das Verbesserungspotenzial der Leistungserbringer geschätzt werde und ob es dabei zu einer „Aufweichung“ des originalen Referenzbereichs kommen könne (GKV-SV, S. 21).

**IQTIG:** Die Breite eines Referenzbereichs, d. h. der Abstand der Referenzbereichsgrenze vom Ende der Indikatorkala, entspricht der Summe zweier Anteile: Der eine Anteil repräsentiert nicht vom Leistungserbringer zu verantwortende Faktoren und der andere Anteil eine Toleranz suboptimaler Qualität. Beide Anteile haben definitionsgemäß einen Wert größer oder gleich Null. Der bereinigte Referenzbereich, der durch den Abzug des Anteils der nicht vom Leistungserbringer zu verantwortenden Faktoren ermittelt wird, kann daher nicht breiter („großzügiger“) sein als der originale Referenzbereich.

Die Schätzung des Verbesserungspotenzials einer Einrichtung ist nicht Funktion des Referenzbereichs. Auch Einrichtungen mit Indikatorwerten innerhalb des originalen Referenzbereichs oder innerhalb des bereinigten Referenzbereichs können Verbesserungspotenzial aufweisen.

Des Weiteren wurde die Frage aufgeworfen, ob der bereinigte Referenzbereich „nicht konsequenterweise dann für alle Einrichtungen eingesetzt werden“ müsse (GKV-SV, S. 20).

**IQTIG:** Der bereinigte Referenzbereich wird für alle Leistungserbringer, die eine Stellungnahme abgeben, auf gleiche Weise berechnet (siehe Kapitel 9 des Anhangs zum Abschlussbericht). Das IQTIG versteht die Frage der stellungnehmenden Organisation so, dass vorgeschlagen wird, anhand des bereinigten Referenzbereichs auch diejenigen Leistungserbringer zu bewerten, für die keine

fachliche Bewertung einer Stellungnahme erfolgt. Für diese Leistungserbringer liegen jedoch keine neuen Informationen aus der fachlichen Bewertung vor, sodass davon ausgegangen werden muss, dass die nicht vom Leistungserbringer zu verantwortenden Einflussfaktoren in dem Ausmaß vorliegen, das im originalen Referenzbereich berücksichtigt ist.

### Kriterien für Stellungnahmen

Mehrere stellungnehmende Organisationen beurteilten die im Bericht empfohlenen Kriterien für Stellungnahmen als nachvollziehbar und sachgerecht (DGTHG, S. 4; GKV-SV, S. 23, SLAEK QS, S. 3) und sahen eine formale und inhaltliche Standardisierung als sinnvoll an (KZBV, S. 11).

Von mehreren stellungnehmenden Organisationen wurde die Anforderung an den Leistungserbringer, in seiner Stellungnahme möglichst übergeordnete Gründe für das Verfehlen des Referenzbereichs zu analysieren, als Widerspruch zu der Anforderung wahrgenommen, der Stellungnahme Vorgangsnummern beizufügen, und um Klarstellung gebeten (BÄK, S. 5; GKV-SV, S. 23 f.; gem. StN LQS BE, MV, SL, RP, TH, S. 11).

**IQTIG:** Der Einsatz von Qualitätsindikatoren hat das Ziel, die Versorgungsqualität eines Leistungserbringers über alle Behandlungsfälle zu beschreiben, um Hinweise auf Verbesserungsmaßnahmen in den Behandlungsstrukturen und -prozessen des Leistungserbringers zu geben. Gutachten über einzelne Behandlungsfälle sind nicht das Ziel der indikatorbasierten Qualitätsbewertung. Bei dem Verfassen seiner Stellungnahme soll der Leistungserbringer daher ebenfalls eine möglichst behandlungsfallübergreifende Analyse vornehmen. Da Prozess- und Ergebnisindikatoren auf den aggregierten Daten von Behandlungsfällen basieren, lassen sich auch übergeordnete Gründe für das Verfehlen eines Referenzbereichs immer auf mehrere oder alle Behandlungsfälle eines Leistungserbringers beziehen. Die Angabe der Vorgangsnummern der betreffenden Behandlungsfälle dient dabei nicht einer detaillierten Einzelfallanalyse, sondern der Abschätzung (in Form der Nachberechnung), in welchem Ausmaß der angeführte Grund das Indikatorergebnis verändert hat. Im Abschlussbericht wurden diese Zusammenhänge nun etwas deutlicher dargestellt.

Mehrere stellungnehmende Organisationen wiesen darauf hin, dass auch der Aufwand bedacht werden müsse, der für die Leistungserbringer durch das vorgeschlagene Konzept entsteht (DGGG, S. 2; DGIM, S. 2). So wurde von einem erhöhten Aufwand für die Leistungserbringer ausgegangen, der aus den Vorgaben für die Anfertigung der Stellungnahmen resultiere (BÄK, S. 5; KZBV, S. 5). Insbesondere wurde die Vermutung geäußert, dass die Dokumentation von Vorgangsnummern in den Stellungnahmen zu einem hohen Aufwand führe (BÄK, S. 5; GKV-SV, S. 23 f.; gem. StN LQS BE, MV, SL, RP, TH, S. 11).

**IQTIG:** Bisher werden in den Stellungnahmen der Leistungserbringer oft nur diejenigen Behandlungsfälle aufgeführt, die ein interessierendes Ereignis (z. B. eine Komplikation) aufgewiesen haben. Diese werden anhand der zugehörigen Vorgangsnummern identifiziert. Liegt für diese Behandlungsfälle ein Einflussfaktor auf das Indikatoreergebnis vor, der vom Leistungserbringer nicht zu verantworten war, so wird die Information über das Vorliegen dieses Einflussfaktors auch für alle anderen Behandlungsfälle des Leistungserbringers benötigt. Andernfalls kommt es zu einer Verzerrung der Bewertung. Dies wird in Abschnitt 6.5.2 des Abschlussberichts nun ausführlicher erläutert. Der damit verbundene Mehraufwand für die Leistungserbringer ist nicht zu umgehen, wenn eine Verzerrung durch die selektive Betrachtung von bestimmten Fällen vermieden werden soll. Aus Sicht des IQTIG bewegt sich dieser Mehraufwand in einem vertretbaren Rahmen, da nicht etwa für alle klinischen Charakteristika der Patientinnen und Patienten eine Zuordnung der Vorgangsnummern erfolgen muss, sondern nur für Faktoren, die vom Leistungserbringer nicht zu verantworten und gleichzeitig für das Indikatoreergebnis relevant sind.

In einer Stellungnahme wurde angeführt, dass Vorgangsnummern nicht für alle Qualitätsindikatoren sinnvoll angegeben werden könnten (gem. StN LQS BE, MV, SL, RP, TH, S. 12).

**IQTIG:** Es ist richtig, dass die Angabe von Vorgangsnummern in bestimmten Konstellationen nicht sinnvoll oder nicht möglich ist, beispielsweise weil ein struktureller Einflussfaktor sich auf alle Behandlungsfälle gleichermaßen bezieht oder weil sich die Ergebnisse mancher Indikatoren nicht bestimmten Behandlungsfällen zuordnen lassen. Sachverhalte, für die keine Nachberechnung möglich ist, können bei der fachlichen Bewertung im heuristischen Beurteilungsschritt berücksichtigt werden.

Zwei stellungnehmende Organisationen baten um detailliertere Vorgaben als im Vorbericht, welche Anforderungen ein „schriftlicher Beleg“ in der Stellungnahme eines Leistungserbringers erfüllen müsse (DKG, S. 23) bzw. „in welcher Form“ der Nachweis durch einen Leistungserbringer erfolgen solle, dass ein Indikatoregebnis nicht von ihm zu verantworten sei (KBV, S. 9).

**IQTIG:** Belege ohne Personenbezug (Personal und Patientinnen und Patienten) können vom Leistungserbringer einer Stellungnahme beigelegt werden, um diese für die LAG bzw. Bundesstelle und die Fachkommission nachvollziehbar zu machen. In welchem Ausmaß Belege erforderlich sind, hängt davon ab, inwieweit der betreffende Sachverhalt anhand der Unterlagen plausibel nachvollzogen werden kann. Dies entspricht im Wesentlichen dem bisherigen Vorgehen, bei dem der Umfang der Stellungnahmen ebenfalls im Ermessen und in der Verantwortung des Leistungserbringers liegt. Eine Festlegung zusätzlicher,

für alle Belege geltender, inhaltlicher Anforderungen ist aus Sicht des IQTIG nicht erforderlich.

Mehrere stellungnehmende Organisationen unterstützten die Textvorschläge für die Einholung von Stellungnahmen der Leistungserbringer im Anhang des Vorberichts (DGTHG, S. 4; DKG, S. 23; PatV, S. 8). In zwei Stellungnahmen wurde dagegen vorgeschlagen, bei der Anforderung von Stellungnahmen der Leistungserbringer keine generischen, sondern indikatorspezifische Texte zu verwenden (gem. StN LQS BE, MV, SL, RP, TH, S. 17; SLAEK QS, S. 4). Darüber hinaus wurde vorgeschlagen, dass die LAG jeweils eigene Formulierungen bei der Anforderung von Stellungnahmen ergänzen können sollten (gem. StN LQS BE, MV, SL, RP, TH, S. 17).

**IQTIG:** Mit dem Ziel eines einheitlichen, möglichst objektiven Stellungnahmeverfahrens sollte auch das Einholen der Stellungnahmen von den Leistungserbringern bundesweit einheitlich erfolgen. Da der Sachverhalt, der zur Anforderung einer Stellungnahme führt, nämlich ein Indikatorergebnis außerhalb des Referenzbereichs, bundesweit derselbe ist, ergibt sich auch keine Notwendigkeit regional unterschiedlicher Texte für die Anforderung. Das IQTIG teilt die Auffassung, dass die Anschreiben für die Zielgruppe gut verständlich und nachvollziehbar sein sollten. Bei einer Umsetzung des Konzepts können die Textvorschläge unter Berücksichtigung der bisher genutzten Formulierung nochmals diesbezüglich geprüft werden. Unter anderem dazu empfiehlt das IQTIG eine formative Evaluation der Empfehlungen.

Die Verwendung indikatorspezifischer Formulierungen ist grundsätzlich möglich, geht aber mit erhöhtem Aufwand einher, da für jedes quantitativ auffällige Indikatorergebnis der passende Textbaustein verwendet werden muss. Zudem ergibt sich ein erhöhter Pflegeaufwand für die Textbausteine bei Änderungen an den Qualitätsindikatoren. Da auf der anderen Seite das Verfehlen des Referenzbereichs eines Indikators im Grundsatz bei allen Indikatoren auf gleiche Weise untersucht wird, nämlich durch Analyse der Einflussfaktoren auf das Indikatorergebnis, bieten indikatorspezifische Formulierungen aus Sicht des IQTIG nur einen sehr geringen Mehrwert. Aus Aufwand-Nutzen-Überlegungen werden diese daher nicht empfohlen.

Im Rahmen des vorliegenden Konzepts können indikatorspezifische Fragebögen allerdings im Modul zur Qualitätsförderung eingesetzt werden. So kann beispielsweise ein solcher Fragebogen ein Mittel sein, um bei einem Leistungserbringer, bei dem ein Qualitätsdefizit im ersten Schritt festgestellt wurde, eine selbstkritische Auseinandersetzung mit der eigenen Versorgungsqualität anzuregen (siehe auch Kapitel 7).

In zwei Stellungnahmen wurde kritisiert, dass im Vorbericht eine Einreichung der Stellungnahmen der Leistungserbringer in digitaler Form empfohlen wurde (KZBV, S. 20; gem. StN LQS BE, MV, SL, RP, TH, S. 12).

**IQTIG:** Für den digitalen Übermittlungsweg nimmt das IQTIG eine höhere Effizienz bei der Verarbeitung der Stellungnahmen an. Es ist richtig, dass aus methodischer Sicht eine Verarbeitung in Papierform ebenfalls möglich ist. Im Abschlussbericht wurde daher auf eine Festlegung des Übermittlungswegs verzichtet. Für ggf. erforderliche Listen von Vorgangsnummern in den Stellungnahmen wird jedoch weiterhin eine digitale Übermittlung empfohlen, um Übertragungsfehler zu vermeiden.

### Umsetzung und Aufwand

Eine technische Unterstützung der Nachberechnung des informativen Indikatorwerts, z. B. durch eine Software, wurde von mehreren stellungnehmenden Organisationen als essenziell erachtet (BÄK, S. 7; DGGG, S. 3; DKG, S. 21). Es wurden genauere Informationen zur Umsetzung einer solchen Software gewünscht (GKV-SV, S. 8; KBV, S. 9; KZBV, S. 18; QB Bremen, S. 2; NKGeV QS, S. 2) und insbesondere die Frage aufgeworfen, ob bei einer Durchführung von Nachberechnungen mittels eines Web-Portals mit zeitlichen Verzögerungen zu rechnen ist (QB Bremen, S. 2; NKGeV QS, S. 2; gem. StN LQS BE, MV, SL, RP, TH, S. 10 f.).

**IQTIG:** Eine webbasierte technische Unterstützung der Nachberechnung, die die Ergebnisse im Echtzeit-Betrieb bereitstellt, ist grundsätzlich möglich, d. h. durch den Berechnungsvorgang entstehen während der Beratungen in den Fachkommissionssitzungen keine Verzögerungen. Die grundsätzliche Umsetzbarkeit dieses Vorgehens und einer solchen Software zeigen die Erfahrungen aus dem Verfahren der planungsrelevanten Qualitätsindikatoren. Der Einsatz einer zentralen, webbasierten Software hätte dabei gegenüber lokalen Softwareimplementierungen den Vorteil einer größeren Einheitlichkeit und Effizienz. Weitergehende Spezifikationen, Anforderungen an die Software sowie organisatorische Fragen zum Aufbau und Betrieb können erst im Rahmen einer Umsetzungsplanung geklärt werden und überschreiten den Rahmen des Abschlussberichts.

In zwei Stellungnahmen wurde die Auffassung geäußert, dass das Konzept des Stellungnahmeverfahrens an den bisherigen Strukturierten Dialog im stationären Bereich angelehnt sei, und es wurde um Klarstellung gebeten, inwieweit es überhaupt im vertragsärztlichen Bereich angewendet werden könne (BÄK, S. 9; KBV, S. 3 f.).

**IQTIG:** Das IQTIG versteht die Frage der stellungnehmenden Organisationen so, dass zum einen Bedenken an der Gültigkeit der Methodik für den vertragsärztlichen Bereich und zum anderen Zweifel an der praktischen Umsetzbarkeit im vertragsärztlichen Bereich geäußert werden. Das vorgeschlagene Konzept berücksichtigt zwar die Analyseergebnisse zum Strukturierten Dialog im stationären Bereich, wurde aber nicht von diesem abgeleitet. Wie in Kapitel 3 und

den folgenden Kapiteln des Abschlussberichts ausführlich dargestellt, wurde das Konzept vielmehr aus den methodischen Anforderungen an Qualitätsmessungen und der methodischen Funktion von Stellungnahmen hergeleitet. Das grundsätzliche Vorgehen in einem Stellungnahmeverfahren ist dabei unabhängig von Charakteristika der Leistungserbringer wie beispielsweise der Zugehörigkeit zu einem Versorgungssektor.

Richtig ist, dass das Verfassen einer Stellungnahme für kleinere Einrichtungen, etwa eine Einzelpraxis, relativ zu den verfügbaren Ressourcen einen deutlich höheren Aufwand bedeutet als im Krankenhausbereich. Allerdings geht jedes denkbare Verfahren zur Prüfung der Aussagekraft eines Indikatorergebnisses mit Aufwänden bei den betreffenden Leistungserbringern einher. Die Entscheidung, ein Stellungnahmeverfahren durchzuführen, bzw. die Entscheidung eines Leistungserbringers zur Verfassung einer Stellungnahme stellt damit immer einen Trade-off zwischen dem Gewinn an Validität bzw. Fairness der Messergebnisse und dem benötigten Aufwand dar. Dies wurde in Abschnitt 11.2 des Abschlussberichts nun deutlicher thematisiert.

Eine Organisation vertrat die Auffassung, dass aus Aufwandsgründen statt schriftlicher Stellungnahmen auch telefonische Gespräche für die fachliche Bewertung eingesetzt werden sollten (KBV, S. 10).

**IQTIG:** Das Stellungnahmeverfahren hat die Funktion, bei einem Verfehlen des Referenzbereichs durch einen Leistungserbringer zu prüfen, ob eine besondere Konstellation den Hinweis aus dem Indikatorergebnis entkräftet (siehe Abschnitt 6.2 des Abschlussberichts). Wie im Bericht hergeleitet wurde, erfordert diese Prüfung die Nutzung expliziter und transparenter Beurteilungsregeln, wenn sie einheitlich und fair erfolgen soll, und die Beurteilungsregeln müssen wiederum auf standardisierten Informationen in den Stellungnahmen aufbauen. Für diese benötigten standardisierten Informationen (siehe Abschnitt 6.5.2 des Abschlussberichts) ergibt sich aus Sicht des IQTIG gerade keine Aufwandsreduktion durch eine mündliche Übermittlung, stattdessen aber ein erhöhtes Risiko von Übermittlungsfehlern z. B. von Fallzahlen und Vorgangsnummern. Darüber hinaus ist zu berücksichtigen, dass eine mündlich eingeholte und von der LAG bzw. Bundesstelle dokumentierte Stellungnahme gleichermaßen wie eine primär schriftliche Stellungnahme anschließend einer fachlichen Bewertung durch die Fachkommission unterzogen würde. Es ergibt sich also auch im folgenden Bewertungsprozess keine Aufwandsreduktion. Mündliche Stellungnahmen werden daher im Konzept des IQTIG nicht empfohlen.

## Berufung von Fachkommissionen

Die vom IQTIG vorgeschlagenen Kriterien zur Zusammensetzung der Fachkommissionen sowie zur fachlichen und formalen Eignung von Expertinnen und Experten wurden in den Stellungnahmen kontrovers diskutiert. Einige stellungnehmende Organisationen bewerteten die Kriterien für die Zusammensetzung der Fachkommissionen als angemessen (DGGG, S. 3) und befürworteten insbesondere die interdisziplinäre Zusammensetzung der Fachkommissionen (DGVS, S. 7; DGTHG, S. 6). Positiv beurteilt wurde auch, dass die Auswahl der Personen durch die LAG bzw. Bundesstelle und nicht durch die Interessenvertretungen erfolgen solle. Dies verringere das Risiko interessegeleiteter Bewertungen im Stellungnahmeverfahren (GQH, S. 4). Von einer stellungnehmenden Organisation wurde die Vorgabe bundeseinheitlicher Kriterien zur Zusammensetzung der Fachkommissionen ausdrücklich befürwortet (DKG, S. 24).

Andere Organisationen vertraten dagegen die Auffassung, dass die Stellen auf Landesebene Freiheit bei der Zusammenstellung der Fachkommissionen bräuchten (DGNR, S. 2) und dass die vorgeschlagenen Kriterien zu weit in die Regelungshoheit auf Landesebene eingriffen (GKV-SV, S. 24). Es wurde in mehreren Stellungnahmen dafür plädiert, die bestehenden Regelungen der DeQS-RL zur Zusammensetzung der Fachkommissionen beizubehalten (BAQ, S. 2; QB Bremen, S. 3; gem. StN LQS BE, MV, SL, RP, TH, S. 13; KBV, S. 9; KZBV, S. 16; QS-NRW, S. 4; QiG BW GmbH, S. 2). Mehrere stellungnehmende Organisationen äußerten die Vermutung, dass die Besetzung von Fachkommissionen über ein Bewerbungsverfahren wenig erfolgversprechend sei (gem. StN LQS BE, MV, SL, RP, TH, S. 13) und auf potenzielle Fachexpertinnen und -experten abschreckend wirke (BAQ, S. 2; GQH, S. 4; QS-NRW, S. 4).

**IQTIG:** Die Kriterien für die Zusammensetzung von Fachkommissionen sollten sich nach der Aufgabe der Expertinnen und Experten im Stellungnahmeverfahren richten. Diese besteht in der fachlich-inhaltlichen Beurteilung der in den Stellungnahmen angeführten Sachverhalte. Eine angemessene Zusammensetzung der Fachkommissionen für diesen Zweck ist nach Auffassung des IQTIG dann am ehesten zu gewährleisten, wenn die Besetzung allein nach fachlichen Gesichtspunkten durch die LAG bzw. Bundesstelle auf Basis geeigneter Bewerberinnen und Bewerber erfolgt. Die im Bericht empfohlenen Kriterien stellen dabei Rahmenbedingungen dar, die einerseits eine Einheitlichkeit zwischen den verschiedenen Ländern fördern sollen, und andererseits den LAG bzw. der Bundesstelle ausreichend Freiheit bei der Besetzung lassen, wenn dies aus besonderen Gründen erforderlich ist und sofern dies transparent gemacht wird. Die Sorge, dass sich keine Bewerberinnen und Bewerber für die Arbeit in den Fachkommissionen finden, teilt das IQTIG nicht.

Die vom IQTIG vorgeschlagenen Kriterien zur fachlichen und formalen Eignung der Expertinnen und Experten wurden von einigen stellungnehmenden Organisationen als grundsätzlich nachvollziehbar beurteilt (DGTHG, S. 5; KZBV, S. 15). Einige der vorgeschlagenen Details wie beispielsweise die Anforderungen an die Dauer der klinischen Tätigkeit wurden von verschiedenen stellungnehmenden Organisationen dagegen kritisiert (DGGG, S. 3; DGTHG, S. 5; DPR; S. 2; gem. StN LQS BE, MV, SL, RP, TH, S. 13) .

**IQTIG:** Bei den im Bericht empfohlenen Vorgaben zur Aktualität der klinischen Expertise handelt es sich nicht um eine vollständige Beschreibung der erforderlichen fachlichen Qualifikation für die Arbeit in der Fachkommission. Es bleibt die Aufgabe der LAG bzw. Bundesstelle, angemessen qualifizierte Expertinnen und Experten für die Fachkommissionen zu berufen. Die empfohlenen Vorgaben stellen vielmehr Rahmenbedingungen dar, die einen einheitlichen Standard gewährleisten sollen und von denen bei entsprechender Begründung abgewichen werden kann.

In mehreren Stellungnahmen wurden Fragen zu formalen Verfahrensaspekten gestellt. Es wurden Klarstellungen und Festlegungen erbeten, welche Verantwortlichkeiten den Geschäftsstellen und Lenkungsorganen der LAG in den Stellungsverfahren zugewiesen werden sollten (PatV, S. 10) und auf welcher fachlichen Grundlage und mit welcher Legitimation diese von den Empfehlungen der Fachkommission abweichen dürften (DGTHG, S. 3; GKV-SV, S. 12 f.; GQH, S. 4; SLAEK QS, S. 3).

**IQTIG:** Beim Konzept des IQTIG zur Weiterentwicklung des Strukturierten Dialogs bzw. Stellungsverfahrens handelt es sich um wissenschaftliche Empfehlungen, die beschreiben, welche Methoden und Prozesse bei der fachlichen Bewertung angewendet werden sollten, damit diese fachlich angemessen und möglichst objektiv und fair erfolgt. Davon zu unterscheiden ist die formale Ausgestaltung der Verantwortlichkeiten und Entscheidungsprozesse der beteiligten Stellen. Empfehlungen dazu sind nicht Bestandteil des Berichts.

## 7 Modul Qualitätsförderung

Mehrere Stellungnahmen wünschen, dass ausführlichere Empfehlungen für das „Modul Qualitätsförderung“ gegeben werden sollten (BÄK, S. 4 bzw. 8; DNVF, S. 4; GKV-SV, S. 25; GQMG, S. 3, QS-NRW, S. 4). Beispielweise wurde gebeten, Maßnahmen der Qualitätsförderung nicht nur aufzuzählen, sondern auch herauszuarbeiten, welche Maßnahmen bei welchen Qualitätsdefiziten hilfreich seien (DNVF, S. 4). Zudem seien die Ursachen von Qualitätsdefiziten als Voraussetzung für eine strukturierte Berichterstattung systematisch zu kategorisieren (GKV-SV, S. 25). Dieselbe stellungnehmende Organisation würdigte die Empfehlungen für den klar strukturierten Vorschlag bezüglich des Vorgehens bei der Einleitung von Verbesserungsmaßnahmen. Dieser Vorschlag erscheint als erste Grundlage für ein möglichst homogenes Vorgehen bei länderbezogenen Verfahren geeignet (GKV-SV, S. 25).

**IQTIG:** Die Beauftragung nennt explizit die Vereinheitlichung und Effizienzsteigerung hinsichtlich der Qualitätsbewertung als Ziel. Der Begriff Qualitätsförderung taucht im Beauftragungstext allerdings nicht auf. Das bedeutet, dass eine Weiterentwicklung der Qualitätsförderungskomponente des Strukturierten Dialogs nicht vorgesehen wurde. Aufgrund der hohen Bedeutung des Strukturierten Dialogs für die Qualitätsförderung entwickelte das IQTIG trotzdem konkrete Empfehlungen für das Einleiten von qualitätsfördernden Maßnahmen und erste Empfehlungen für die Identifikation von Verbesserungsmaßnahmen sowie zum Abschluss von Zielvereinbarungen.

Eine Organisation hob hervor, dass mögliche Verbesserungsmaßnahmen mit Abschluss der Bewertung (insbesondere bei negativen Bewertungen) benannt werden sollten (DKG, S. 25). Für eine andere stellungnehmende Organisation blieb unklar, wie Verbesserungsmaßnahmen identifiziert werden sollen. Zudem wurde ein Ergänzungsbedarf des Berichts bei der „Verzahnung“ der Module Qualitätsbewertung und Qualitätsförderung gesehen ohne darzustellen, welche Art der Vernetzung wünschenswert wäre (GKV-SV, S. 25).

**IQTIG:** Die Trennung von Qualitätsbewertung und Qualitätsförderung ist ein zentrales Element des Rahmenkonzepts, um einerseits ein standardisiertes und möglichst objektives Vorgehen bei der Qualitätsbewertung sicherzustellen und andererseits den Erfolg der Qualitätsverbesserungsmaßnahmen nicht dadurch zu gefährden, dass die Ursachenanalyse wegen einer noch nicht abgeschlossenen Bewertung verzerrt wird. Das IQTIG empfiehlt im Anschluss an eine Vergabe der Qualitätsbewertung „Qualitätsdefizit“ eine von der LAG bzw. Bundesstelle initiierte und durch die Fachkommission unterstützte Ursachenanalyse des Leistungserbringers (siehe Abschlussbericht Kapitel 7). Ziel

dieser Ursachenanalyse ist die Identifikation möglicher Verbesserungsmaßnahmen. Gegebenenfalls können bereits im Rahmen der Qualitätsbewertung erste Annahmen hinsichtlich der Ursachen und möglicher Verbesserungsansätze festgehalten werden. Dennoch sollte aus Sicht des IQTIG zunächst eine umfangreiche Ursachenanalyse unterstützt von der LAG bzw. Bundesstelle und Mitgliedern der Fachkommission durchgeführt werden, bevor mögliche Verbesserungsmaßnahmen benannt werden. Im Rahmen dieser Ursachenanalyse können von den LAG bzw. der Bundesstelle derzeit bereits eingesetzte indikatorspezifische Fragebögen zur Ursache von Qualitätsdefiziten weiterhin zur Anwendung kommen. Eine Verbindung der Module ist aus Sicht des IQTIG dadurch sichergestellt, dass das Ergebnis der Indikatorbewertung zum Anlass für eine gezielte Ursachenanalyse herangezogen wird, um geeignete Verbesserungsmaßnahmen einzuleiten.

## 8 Möglichkeiten zur Verkürzung des Stellungnahmeverfahrens

Zwei stellungnehmende Organisationen hoben hervor, dass die Verkürzung der Datenannahmefrist auf den 31. Januar für die Leistungserbringer im stationären Bereich (DKG, S. 26) bzw. die einrichtungsbezogenen Daten (KBV, S. 11) schwierig umzusetzen sei. Demgegenüber wurde von anderen Organisationen eine Anpassung entsprechend den Empfehlungen des IQTIG begrüßt, um die nachgelagerten Prozesse im Jahresverlauf früher beginnen zu können (NKGeV QS, S. 1; gem. StN LQS BE, MV, SL, RP, TH, S. 16; QB Bremen, S. 3). Weiterhin wurde hervorgehoben, dass die Jahresauswertungen (vor dem Hintergrund der Umstellung auf das Entlassungsdatum und der prospektiven Rechenregeln) in Form der Rückmeldeberichte seitens der Bundesstelle früher an die Landesebene zurückzuspielen seien (BAQ, S. 3; QiG BW GmbH, S. 2; SLAEK QS, S. 4 und 7; gem. StN LQS BE, MV, SL, RP, TH, S. 15ff, GQH, S. 4, NKGeV QS, S. 1; QS-NRW, S. 4).

Zudem wurde angemerkt, dass der empfohlene Zeitraum, innerhalb dessen eine Stellungnahme abgegeben werden könne, für die Leistungserbringer zu kurz sei bzw. nicht verkürzt werden solle (DKG, S. 25, KBV, S. 11; AWMF, S. 2; Sachsen, S. 5; DGGG&DGS Mammachirurgie, S. 3; DGHNO, S. 2; DGTHG, S. 6; QB Bremen, S. 3; QiG BW GmbH, S. 2; SLAEK QS, S. 4). Auch die Sommerurlaubszeit und die damit im Zusammenhang stehende reduzierte personelle Ausstattung bei den LAG, den Fachkommissionen und den Leistungserbringern müsse bei der Festlegung des Beginns und der Dauer des Bewertungszeitraums berücksichtigt werden (DKG, S.4 und 26; KZBV, S. 18; BAQ, S. 3; QB Bremen, S. 3; QiG BW GmbH, S. 2, NKGeV QS, S. 1; gem. StN LQS BE, MV, SL, RP, TH, S. 15; GQH, S. 4).

Mehrere Organisationen äußerten sich hingegen positiv bezüglich der Vereinheitlichung und transparenten Darstellung von Fristen für die Qualitätsbewertung (DGIM, S. 2, DGTHG, S. 6) bzw. hinsichtlich der Verfahrensverkürzung mit einem früheren Abschluss des Stellungnahmeverfahrens, und der damit einhergehenden Möglichkeit, über die Bewertung der Versorgungsqualität bereits im Jahr nach der Erfassung der Daten (EJ +1) zu berichten (KBV, S. 12, GKV-SV, S. 26). Dies wurde als wichtiger Fortschritt bewertet (GKV-SV, S. 26). Zudem sei eine frühere Berichterstattung auch deshalb zu begrüßen, da dadurch das interne Qualitätsmanagement die Qualitätsdefizite frühzeitiger beheben könne (PatV, S. 13).

**IQTIG:** Fristen stellen wichtige Rahmenbedingungen für ein Verfahren und deren Beteiligte (Leistungserbringer, LAG bzw. Bundesstelle und Fachkommissionen) dar. Aufgrund der miteinander verknüpften Prozesse und der Vielzahl an beteiligten Akteuren geht eine Veränderung von Fristen immer mit einer hohen Komplexität und mit Zielkonflikten einher. Beispielsweise ist einerseits eine frühzeitigere Berichterstattung der Versorgungsqualität für Patientinnen

und Patienten, Kostenträger und Leistungserbringer wünschenswert. Andererseits bedeutet der dafür notwendige zeitnahe Dateneingang bei der Bundesstelle für die Leistungserbringer eine kürzere Frist für Korrekturen der Dokumentationen. Vor diesem Hintergrund erscheint eine Priorisierung der Ziele unter Berücksichtigung der unterschiedlichen Perspektiven der einzelnen Akteure notwendig.

Das IQTIG hat Empfehlungen für eine Verfahrensverkürzung formuliert, deren Ziel eine frühzeitigere Berichterstattung der Versorgungsqualität (noch im EJ +1) ist. Sofern eine frühzeitigere Berichterstattung gewünscht ist, sind die Empfehlungen des Rahmenkonzepts unter Berücksichtigung gewisser Bedingungen (z. B. eine IT-seitige Unterstützung der Nachberechnung sowie ein früherer Start des Stellungnahmeverfahrens deutlich vor der Sommerferienzeit) dazu geeignet, die Ergebnisse zur Versorgungsqualität im EJ +1 zu berichten.

## 9 Empfehlungen zur Evaluation des neuen Konzepts

Zwei Organisationen fragen nach dem weiteren Vorgehen hinsichtlich einer summarischen Evaluation (DNVF, S. 5; GKV, S. 27).

**IQTIG:** Das IQTIG empfiehlt grundsätzlich, dass eine summarische Evaluation immer auf einem Evaluationsplan basiert. Dieser Evaluationsplan wird separat beauftragt und veröffentlicht. Im Rahmen der Erstellung des Evaluationsplans wird mit dem Auftraggeber u. a. erörtert, welche Dimensionen und Endpunkte für die Entscheidungsfindung relevant sind.

Mehrere Organisationen geben Anregungen, welche Stakeholder bzw. Fachgesellschaften bei einer Evaluation miteinbezogen werden sollen (DGHNO, S. 2; DKG, S. 26; QS-NRW, S. 5).

**IQTIG:** Die in den Stellungnahmen genannten potenziell zu Beteiligten erscheinen sinnvoll und plausibel. Im Rahmen der Erstellung des Evaluationsplans werden diese und auch weitere Vorschläge geprüft.

In einer Stellungnahme wird nach der zeitlichen Dimension einer Evaluation durch das IQTIG gefragt (DGIM, S. 3).

**IQTIG:** Die formative, d. h. begleitende Evaluation kann je nach Umfang jährlich berichtet werden. Die summarische Evaluation sollte sinnvoll erst dann durchgeführt werden, wenn das Verfahren hinreichend etabliert ist. Dies ist frühestens nach 2 bis 3 Jahren der Fall. Für die Evaluation selbst ist mit einer Projektdauer von ca. 12 bis 24 Monaten zu rechnen.

## Übergeordnete Themenbereiche

### Abstimmung mit dem Verfahren gemäß plan. QI-RL

Eine Organisation hat, Unterschiede zwischen dem im Bericht empfohlenen Verfahren für die DeQS-RL und dem Verfahren gemäß plan. QI-RL aufzuzeigen. Insbesondere Unterschiede hinsichtlich der Begriffe „unzureichende Qualität“ (plan. QI-RL) und „Qualitätsdefizit“ sowie der Bezeichnung angeführter Gründe in den Stellungnahmen als „Ausnahmetatbestände“ (plan. QI-RL) und „Einflussfaktoren“ seien darzustellen. Darüber hinaus wurde die Frage aufgeworfen, ob unterschiedliche statistische Vorgehensweisen gemäß plan. QI-RL (frequentistisch) und dem vorliegenden Konzept (bayesianisch) notwendig seien (GKV-SV, S. 12).

**IQTIG:** Das vorgeschlagene Stellungsnahmeverfahren zu den quantitativen Ergebnissen der datengestützten Qualitätssicherung kann prinzipiell in allen indikatorbasierten Verfahren zur Anwendung kommen. Sowohl die Bewertung „unzureichende Qualität“ nach dem plan. QI-RL-Verfahren als auch die Bewertung „Qualitätsdefizit“ im vorliegenden Konzept bedeuten, dass die mittels der Indikatoren (einschließlich Referenzbereiche) operationalisierten Anforderungen an die Versorgungsqualität nicht erreicht wurden. Sowohl der Begriff „Ausnahmetatbestand“ aus der plan. QI-RL als auch der im vorliegenden Konzept verwendete Begriff „Einflussfaktor“ bezeichnet von Leistungserbringern in Stellungnahmen angeführte Gründe für das Nichterreichen des Referenzbereichs in einem Indikator. Die vom Leistungserbringer angeführten Gründe werden erst dann als Qualitätsdefizit bzw. als unzureichende Qualität klassifiziert, wenn geklärt wurde, ob die Gründe vom Leistungserbringer zu verantworten sind (siehe Bewertungsschema in Abschnitt 6.1.4 des Abschlussberichts). Daher wurde der neutrale Begriff „Einflussfaktor“ gewählt.

Als statistische Methodik zur Auffälligkeitseinstufung kommt beim Verfahren gemäß plan. QI-RL ein frequentistisches Verfahren und bei dem hier vorgeschlagenen Vorgehen ein bayesianisches Verfahren zum Einsatz. Beiden Verfahren gemein ist, dass statistische Unsicherheit bei der Auffälligkeitseinstufung berücksichtigt wird. In beiden Verfahren stellt das Signifikanzniveau einen Schwellenwert dar, ab dem von einem hinreichenden Hinweis auf ein Qualitätsdefizit ausgegangen werden kann. Wird beispielsweise in der hier empfohlenen bayesianischen Auswertungsmethodik bei einem Ratenindikator  $\alpha = 0,05$  gewählt, entspricht dies praktisch der frequentistischen Vorgehensweise gemäß plan. QI-RL. Es kann gezeigt werden, dass die vorgeschlagene bayesianische Version mit Jeffrey-Priori prinzipiell das bayesianische Analogon zum frequentistischen Vorgehen mit mid-p-Werten darstellt (Brown et al. 2001). Die hier vorgeschlagene bayesianische Einstufungsmethodik führt bei den 11 Indikatoren der plan. QI-RL und dem derzeitigen Signifikanzniveau zu

praktisch identischen Ergebnissen im Vergleich zur frequentistischen Methode gemäß plan. QI-RL. Um dies zu illustrieren, wurde die Funnelplot-App (<https://iqtig.shinyapps.io/funnelplot/>) entsprechend erweitert, sodass die Unterschiede zwischen den verschiedenen statistischen Methoden zur Auffälligkeitseinstufung dargestellt werden können.

Die bayesianische Variante wird hier empfohlen, da sie eine Auswertungsmethodik bietet, die flexibel an die Vielzahl an Anforderungen der verschiedenen Indikatorarten (z. B. bei Patientenbefragungen) angepasst werden kann und einfacher interpretierbar ist als die frequentistische Variante. Grundsätzlich könnte die empfohlene Methodik auch für das plan. QI-RL-Verfahren angewendet werden, wodurch eine methodische Einheitlichkeit über alle Verfahren sichergestellt werden könnte.

Aus Sicht des IQTIG stellt die damals entwickelte statistische Methodik bei der Auswertung von Indikatoren nach der plan. QI-RL einen wichtigen Schritt in Richtung Berücksichtigung von Unsicherheit dar. Jedoch waren zum Zeitpunkt der Beauftragung der Entwicklung für die plan. QI-RL im Jahr 2016 die auszuwertenden Indikatorarten durch die QSKH-RL vorgegeben. Im Zuge der statistischen Weiterentwicklung der Auswertungen bei z. B. Patientenbefragungen, gezielten Stichproben, Strukturvariablen etc. wurde immer deutlicher, dass eine frequentistische Auswertungsmethodik nur sehr schwer auf die Vielzahl an möglichen Konstruktionen erweiterbar ist – ein Beispiel dafür ist die derzeit vorgeschlagene Auswertungsmethodik bei Patientenbefragungen (vgl. z. B. IQTIG 2018). Aus Sicht des IQTIG muss es methodisch daher zu einer Vereinheitlichung der Auswertungsmethodik aller Richtlinien kommen, sodass Auswertungen von Indikatoren immer als Teil der in Kapitel 5 vorgestellten Taxonomie abgebildet werden können (Herangehensweise, Stichprobenart, Berechnungsart, Bewertungsart).

### **Frequenz der Rückmeldungen an die Leistungserbringer**

Eine Organisation merkte an, dass es für das Qualitätsmanagement der Leistungserbringer vorteilhaft sei, dass Rückmeldungen über die erzielten Leistungen nicht nur ein Mal pro Jahr erfolgten, sondern mindestens halbjährlich. Dies würde eine rasche Korrektur von fehlerhaften Prozessen erlauben bzw. anderweitige rasche Verbesserungen erzielen können (AWMF, S. 3).

**IQTIG:** Mit Überführung der QS-Verfahren in die DeQS-RL ist zukünftig in der Regel eine quartalsweise Auswertung der Indikatorergebnisse, die über die fallbezogene Dokumentation erfasst werden, vorgesehen, sodass eine zeitnahe Verbesserung der Prozesse und/oder Strukturen erfolgen kann.

## Literatur

- Brown, LD; Cai, TT; DasGupta, A (2001): Interval Estimation for a Binomial Proportion. *Statistical Science* 16(2): 101-133. DOI: 10.1214/ss/1009213286.
- G-BA [Gemeinsamer Bundesausschuss] (2018): Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des IQTIG: Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern. [Stand:] 18.01.2018. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/39-261-3196/2018-01-18\\_IQTIG-Beauftragung\\_Weiterentwicklung-strukturierter-Dialog.pdf](https://www.g-ba.de/downloads/39-261-3196/2018-01-18_IQTIG-Beauftragung_Weiterentwicklung-strukturierter-Dialog.pdf) (abgerufen am: 27.02.2019).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018): Entwicklung von Patientenbefragungen im Rahmen des Qualitätssicherungsverfahrens *Perkutane Koronarintervention und Koronarangiographie*. Abschlussbericht. Stand: 15.12.2018. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2018/IQTIG\\_Patientenbefragung\\_QS-PCI\\_Abschlussbericht\\_2018-12-15\\_barrierefrei.pdf](https://iqtig.org/downloads/berichte/2018/IQTIG_Patientenbefragung_QS-PCI_Abschlussbericht_2018-12-15_barrierefrei.pdf) (abgerufen am: 20.05.2020).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2019): Methodische Grundlagen V1.1. Stand: 15.04.2019. Berlin: IQTIG. URL: [https://iqtig.org/dateien/dasiqtig/grundlagen/IQTIG\\_Methodische-Grundlagen-V1.1\\_barrierefrei\\_2019-04-15.pdf](https://iqtig.org/dateien/dasiqtig/grundlagen/IQTIG_Methodische-Grundlagen-V1.1_barrierefrei_2019-04-15.pdf) (abgerufen am: 08.05.2019).
- Winkler-Komp, G; Misselwitz, B; Kupfernagel, F; van Emmerich, C; Döbler, K (2014): Maßnahmen zur Qualitätssicherung in Krankenhäusern: Strukturiertes Dialog – Strukturen und Prozesse. Ergebnis einer Umfrage des Gemeinsamen Bundesausschusses bei den auf Landesebene beauftragten Stellen und der Institution nach § 137 a SGB V. *Das Krankenhaus* 2014(3): 198-205.

**Begleitende Kommentierung zum Beschluss des G-BA über die Freigabe zur Veröffentlichung des IQTIG-Abschlussberichts „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern (Stufe 1 und Stufe 2)“ gemäß Beauftragung vom 18. Januar 2018**

Der Gemeinsame Bundesausschuss hat am 16. Juni 2022 die Freigabe zur Veröffentlichung des IQTIG-Abschlussberichts „Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern (Stufe 1 und Stufe 2)“ gemäß Beauftragung vom 18. Januar 2018 beschlossen. Der Abschlussbericht wurde vom IQTIG primär aus den Erfahrungen mit dem Strukturierten Dialog im Rahmen der stationären datengestützten Qualitätssicherung gemäß der Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern (QSKH-RL) entwickelt, die zum 1. Januar 2021 außer Kraft getreten ist. Das IQTIG schlägt im Wesentlichen folgende Weiterentwicklungen vor:

- klare Trennung von Qualitätsbewertung und Qualitätsförderung;
- Berechnung und ggf. Nachberechnung von Auffälligkeiten durch ein mathematisch-statistisches Verfahren;
- alleiniger Fokus auf eine Verifizierung bzw. Falsifizierung der rechnerischen Ergebnisse, d.h. eine statistisch signifikante Abweichung des Indikatorergebnisses vom Referenzbereich gilt primär als Qualitätsdefizit, bis der Leistungserbringer im Rahmen des schriftlichen Stellungnahmeverfahrens darlegen kann, dass kein Qualitätsdefizit vorliegt;
- Verfahren zur Einstufung der Indikatorergebnisse.

Der G-BA würdigt die bisherigen Entwicklungsarbeiten des IQTIG in diesem Abschlussbericht und den hiermit verbundenen Aufwand. Gleichwohl wird der G-BA eine Umsetzung des Abschlussberichts nicht weiterverfolgen, da noch relevanter Weiterentwicklungsbedarf besteht:

- Weitgehend unberücksichtigt blieben Regelungen der Richtlinie zur einrichtungs- und sektorenübergreifenden Qualitätssicherung, die am 1. Januar 2019 von der Richtlinie für datengestützte Qualitätssicherung (DeQS-RL) abgelöst wurde. Dies ist insofern relevant, da dem Bericht eine Beauftragung vom 17. Mai 2018 zur Entwicklung von Bewertungskriterien und -kategorien für die sektorenübergreifenden Verfahren QS PCI und QS WI subsummiert wurde. Diese Beauftragung wurde mit dem Abschlussbericht jedoch nicht erfüllt.
- Die methodische Güte der Qualitätsindikatoren und somit die Spezifität und Sensitivität der Qualitätsbewertung in den datengestützten QS-Verfahren der DeQS-RL erscheint als Grundlage für das Weiterentwicklungskonzept des IQTIG derzeit nicht in jedem Fall ausreichend. In der Würdigung der Stellungnahmen konstatiert das IQTIG, dass sich das Verfahren nicht für Qualitätsindikatoren mit perzentilbasierten Referenzbereichen eigne. Damit ist es für einen Teil der derzeitigen Qualitätsindikatoren der DeQS-Verfahren nicht anwendbar.
- Die mit der Trennung von Qualitätsbewertung und Qualitätsförderung angestrebte Aufwandsreduktion ist vom Ziel her nachvollziehbar, es besteht jedoch die Annahme, dass hierdurch im Verfahren an anderer Stelle Mehraufwand generiert wird, der wahrscheinlich nicht geleistet werden kann bzw. zu Lasten der Ressourcen für die

Patientenbehandlung geht. Konkret betrifft dies die Vorgaben für die Anfertigung der Stellungnahmen und die damit verbundenen Prozesse.

- Das vom IQTIG vorgeschlagene Stellungnahmeverfahren führt nicht zur Beurteilung der Qualität auf der (systemischen) Ebene der Leistung oder der Fachabteilung. Betrachtet wird lediglich das einzelne Indikatorergebnis. Dies hat die Konsequenz, dass für die Qualitätsförderung und das Qualitätsmanagement praktisch keine Informationsgrundlage geschaffen wird.
- Unter Würdigung der Stellungnahmen konstatiert das IQTIG, dass Kriterien, wann ein Einflussfaktor auf das Indikatorergebnis vom Leistungserbringer zu verantworten sei, nicht sinnvoll allgemeingültig formuliert werden könnten. Begründet wird dies damit, dass derselbe Einflussfaktor je nach Leistungserbringer und betroffenem Qualitätsindikator in unterschiedlichem Maße für das Indikatorergebnis relevant sein könne. Daher könnte die erwartete Homogenisierung (und Aufwandsreduktion) durch die automatische Auslösung der Stellungnahmeverfahren ggf. mit einer Steigerung der Heterogenität, des Aufwands und der Komplexität u. a. durch Nachberechnungen von Indikatorergebnissen und deren Prüfung durch die Fachkommissionen einhergehen. Eine fachliche Bewertung hierzu ist jedoch nicht erfolgt.

Darüber hinaus strebt der G-BA eine Weiterentwicklung, insbesondere der datengestützten Qualitätssicherung, an. Diese wird sich auch auf die zukünftige Konzeption des Stellungnahmeverfahrens im Rahmen der Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung (DeQS-RL) auswirken. Sobald der G-BA Beschlüsse zur Weiterentwicklung der datenbasierten Qualitätssicherung gefasst hat, wird er auch über das Vorgehen zur Umgestaltung der Qualitätsbewertung (wie qualitative Beurteilung), Durchführung von qualitätssichernden Maßnahmen (z.B. Ursachenanalyse von Einzelergebnissen in Verbindung mit einer übergreifenden Betrachtung; Vorgaben zur Durchführung qualitätsverbessernder Maßnahmen und Maßnahmen zur Einhaltung von Qualitätsanforderungen) sowie Veröffentlichung der Ergebnisse der Qualitätssicherung und der qualitätssichernden Maßnahmen beraten und notwendige Weiterentwicklungsaufträge veranlassen.