

IQWiG-Berichte – Nr. 370

**Regorafenib –
Addendum zum Auftrag A15-43**

Addendum

Auftrag: A16-06
Version: 1.0
Stand: 26.02.2016

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Regorafenib – Addendum zum Auftrag A15-43

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

09.02.2016

Interne Auftragsnummer:

A16-06

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Im Mediapark 8 (KölnTurm)
50670 Köln

Tel.: +49 (0)221 – 35685-0

Fax: +49 (0)221 – 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

ISSN: 1864-2500

An dem Addendum beteiligte Mitarbeiter des IQWiG¹:

- Thomas Kaiser
- Katharina Biester
- Charlotte Guddat

Schlagwörter: Regorafenib, Kolorektale Tumoren, Nutzenbewertung

Keywords: Regorafenib, Colorectal Neoplasms, Benefit Assessment

¹ Aufgrund gesetzlicher Datenschutzbestimmungen haben Mitarbeiter das Recht, ihrer Namensnennung nicht zuzustimmen.

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	iv
Abbildungsverzeichnis	v
Abkürzungsverzeichnis.....	vi
1 Hintergrund.....	1
2 Bewertung.....	2
2.1 Datenverfügbarkeit.....	2
2.2 Ergebnisse	3
2.3 Gesamtaussage zum Zusatznutzen	7
3 Literatur	8
Anhang A – Vergleich der Ergebnisse zu den Mittelwertdifferenzen aus den MMRM-Analysen mit den standardisierten Mittelwertendifferenzen	9

Tabellenverzeichnis

Tabelle 1: Ergebnisse zur Symptomatik und zur gesundheitsbezogenen Lebensqualität – Studie CORRECT, Regorafenib + BSC vs. Placebo + BSC.....	4
Tabelle 2: Positive und negative Effekte aus der Bewertung von Regorafenib + BSC im Vergleich zu Placebo + BSC	7

Abbildungsverzeichnis

	Seite
Abbildung 1: LS-MD mit statistisch signifikantem Gruppenunterschied – Symptomskalen ..	10
Abbildung 2: Vom pU berechnete SMD – Symptomskalen	10
Abbildung 3: Selbst berechnete SMD – Symptomskalen	10
Abbildung 4: LS-MD mit statistisch signifikantem Gruppenunterschied – Skalen zur gesundheitsbezogenen Lebensqualität	11
Abbildung 5: Vom pU berechnete SMD – Skalen zur gesundheitsbezogenen Lebensqualität	11
Abbildung 6: Selbst berechnete SMD – Skalen zur gesundheitsbezogenen Lebensqualität ...	11

Abkürzungsverzeichnis

Abkürzung	Bedeutung
BSC	best supportive care
EORTC-QLQ	European Organization for Research and Treatment of Cancer – Quality of Life Questionnaire
G-BA	Gemeinsamer Bundesausschuss
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KI	Konfidenzintervall
LS-MD	Least-Squares-Mittelwertdifferenz
MMRM	Mixed Model for Repeated Measures
pU	pharmazeutischer Unternehmer
SD	Standardabweichung
SE	Standardfehler
SGB	Sozialgesetzbuch
SMD	Standardisierte Mittelwertdifferenz
UE	unerwünschtes Ereignis

1 Hintergrund

Der Gemeinsame Bundesausschuss (G-BA) hat das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) am 09.02.2016 mit einer ergänzenden Bewertung zum Auftrag A15-43 (Regorafenib – Nutzenbewertung gemäß § 35a SGB V) beauftragt.

In seinem Dossier [1] hatte der pharmazeutische Unternehmer (pU) zum Nachweis des Zusatznutzens von Regorafenib in der Indikation „metastasiertes Kolorektalkarzinom“ Ergebnisse aus den Studien CORRECT und CONCUR vorgelegt. Dazu gehörten auch Auswertungen zur gesundheitsbezogenen Lebensqualität und zu Symptomen. Diese waren jedoch nicht verwertbar, da die Daten vom pU nicht angemessen ausgewertet worden waren [2].

Mit seiner schriftlichen Stellungnahme hat der pU veränderte Analysen zur gesundheitsbezogenen Lebensqualität und zu Symptomen vorgelegt [3]. Der G-BA hat das IQWiG mit der Bewertung dieser Analysen beauftragt.

Die Verantwortung für die vorliegende Bewertung und das Bewertungsergebnis liegt ausschließlich beim IQWiG. Die Bewertung wird dem G-BA übermittelt. Über den Zusatznutzen beschließt der G-BA.

2 Bewertung

2.1 Datenverfügbarkeit

Neue Auswertungen wurden zu den Studien CORRECT und CONCUR vorgelegt

Der pU hat für die Nutzenbewertung von Regorafenib gegenüber der zweckmäßigen Vergleichstherapie best supportive care (BSC) 2 Studien vorgelegt, die in der Dossierbewertung A15-43 bewertet wurden [1,2]:

- die Studie CORRECT; diese Studie wurde auch im ersten Nutzenbewertungsverfahren 2013 zu Regorafenib vorgelegt; in diese Studie wurden 760 Patienten eingeschlossen; die Studie wurde in Asien, Australien, Nordamerika, Osteuropa und Westeuropa durchgeführt.
- die Studie CONCUR; diese Studie wurde im aktuellen Nutzenbewertungsverfahren erstmals vorgelegt; in diese Studie wurden 75 Patienten eingeschlossen (relevante Teilpopulation); die Studie wurde ausschließlich in Asien durchgeführt.

Sowohl in der Studie CORRECT als auch in der Studie CONCUR wurden Symptome und die gesundheitsbezogene Lebensqualität mithilfe des EORTC-QLQ-C30 (European Organization for Research and Treatment of Cancer – Quality of Life Questionnaire) erhoben. Bereits im ersten Nutzenbewertungsverfahren zu Regorafenib hatte der pU für die Studie CORRECT keine sachgerechte Auswertung hierzu vorgelegt, da er in seinen Analysen die hohen Ausfallraten nicht angemessen berücksichtigt hatte [2,4]. In seinem Dossier zum aktuellen Bewertungsverfahren hatte der pU zwar die Ausfallraten in seinen Analysen angemessen berücksichtigt (Analysen basierend auf einem gemischten Modell für Messwiederholungen [MMRM]). Die Auswertungen entsprachen jedoch nicht den Vorgaben des Manuals des Fragebogens [2], während der pU diese Vorgaben im ersten Bewertungsverfahren noch berücksichtigt hatte [5].

Mit den Stellungnahmen zum aktuellen Bewertungsverfahren hat der pU Analysen vorgelegt, in denen sowohl die hohen Ausfallraten als auch die Vorgaben des Manuals des Fragebogens berücksichtigt wurden [3].

Nur die Daten zur Studie CORRECT sind für die vorliegende Bewertung relevant

Wie in der Dossierbewertung A15-43 beschrieben [2], sind die Daten zur gesundheitsbezogenen Lebensqualität aus der Studie CONCUR unabhängig von der Art der Auswertung inhaltlich nicht verwertbar, da sich die Ergebnisse im Bereich einzelner unerwünschter Ereignisse wesentlich von denen der CORRECT-Studie unterscheiden. Dies gilt auch für mehrere der mit dem EORTC-QLQ-C30 erhobenen Symptome [3]. Überdies unterscheidet sich, wie ebenfalls in der Dossierbewertung A15-43 bereits beschrieben [2], die mediane Behandlungsdauer zwischen den Behandlungssarmen in der Studie CONCUR ca. um den Faktor 1,5 (Angabe für die Gesamtpopulation, Angaben für die Teilpopulation fehlen),

ohne dass der pU dies in den vorgelegten Auswertungen adressiert hätte (z. B. mithilfe von Überlebenszeitanalysen).

Die Ergebnisse zur gesundheitsbezogenen Lebensqualität und zu Symptomen der Studie CONCUR können zusammenfassend die der Studie CORRECT weder bestätigen noch infrage stellen. Nachfolgend erfolgt die Bewertung daher ausschließlich auf Basis der Ergebnisse der Studie CORRECT.

Responderanalysen fehlen; Analysen des pU zur Bewertung der Relevanz der Effekte ungeeignet

Ergebnisse zu Symptomen und zur gesundheitsbezogenen Lebensqualität, die mittels Fragebögen erhoben werden, bedürfen einer Betrachtung dahin gehend, ob die beobachteten Gruppenunterschiede relevant sind. Idealerweise erfolgt dies durch Responderanalysen auf Basis validierter Responsekriterien [6]. Solche Analysen hat der pU nicht vorgelegt.

Ersatzweise betrachtet der pU die standardisierte Mittelwertdifferenz (SMD; Hedges' g) und leitet daraus seine Aussagen zur Relevanz der Effekte ab. Die Betrachtung der SMD stellt zwar im Falle nicht vorhandener Responderanalysen eine geeignete Möglichkeit zur Relevanzbewertung dar. Das konkrete Vorgehen des pU zur Berechnung der SMD ist jedoch inadäquat (siehe auch Anhang A):

- Die Berechnungen des pU führen zu inkonsistenten Ergebnissen zwischen den Ausgangsanalysen (MMRM) und den von ihm ermittelten SMD zur Relevanzbewertung, da die vom pU verwendete Methode ungeeignet ist.
- Eine geeignete Methode zur Berechnung der SMD führt dem gegenüber zu konsistenten Ergebnissen zwischen den Ausgangsanalysen (MMRM) und den SMD.

Zur Bewertung der Relevanz der Effekte wurden daher SMD mit der geeigneten Methode selbst berechnet.

2.2 Ergebnisse

Die nachfolgende Tabelle 1 zeigt die Ergebnisse zu Symptomen und zur gesundheitsbezogenen Lebensqualität aus der Studie CORRECT.

Tabelle 1: Ergebnisse zur Symptomatik und zur gesundheitsbezogenen Lebensqualität – Studie CORRECT, Regorafenib + BSC vs. Placebo + BSC

Studie Kategorie Skala	Regorafenib + BSC			Placebo + BSC			Regorafenib vs. Placebo LS-MD ^b [95 %-KI]; p-Wert SMD [95 %-KI] ^c
	N ^a	Werte Studien- beginn MW (SD)	Änderung Studien- ende LS-MW (SE)	N ^a	Werte Studien- beginn MW (SD)	Änderung Studien- ende LS-MW (SE)	
CORRECT							
Symptomskalen (EORTC QLQ-C30)^d							
Fatigue	478	35,6 (25,0)	9,33 (1,79)	243	32,2 (23,6)	7,08 (1,89)	2,26 [0,52; 4,00]; 0,011 Hedges' g: 0,20 [0,05; 0,36]
Übelkeit und Erbrechen	478	8,5 (17,6)	2,24 (0,68)	243	6,8 (14,7)	3,00 (0,82)	-0,75 [-2,03; 0,52]; 0,248
Schmerz	479	27,2 (28,1)	7,99 (1,64)	243	26,2 (29,1)	3,57 (1,80)	4,42 [2,42; 6,43]; < 0,001 Hedges' g: 0,34 [0,18; 0,50]
Atemnot	476	20,2 (26,7)	6,22 (1,44)	243	17,3 (24,5)	4,88 (1,61)	1,34 [-0,64; 3,32]; 0,184
Insomnie	477	23,6 (28,2)	2,44 (1,55)	243	25,2, (29,9)	1,59 (1,73)	0,84 [-1,29; 2,98]; 0,439
Appetitverlust	478	24,9 (33,1)	12,33 (2,06)	243	20,3 (28,4)	7,18 (2,22)	5,15 [2,83; 7,46]; < 0,001 Hedges' g: 0,34 [0,19; 0,50]
Obstipation	478	15,9 (25,5)	2,55 (0,85)	243	16,5 (26,8)	4,23 (1,11)	-1,68 [-3,65; 0,28]; 0,093
Diarrhö	477	12,7 (22,9)	5,95 (0,89)	240	12,1 (21,9)	-1,17 (1,16)	7,12 [5,07; 9,16]; < 0,001 Hedges' g: 0,54 [0,38; 0,70]

(Fortsetzung)

Tabelle 1: Ergebnisse zur Symptomatik und zur gesundheitsbezogenen Lebensqualität – Studie CORRECT, Regorafenib + BSC vs. Placebo + BSC (Fortsetzung)

Studie Kategorie Skala	Regorafenib + BSC			Placebo + BSC			Regorafenib vs. Placebo LS-MD ^b [95 %-KI]; p-Wert SMD [95 %-KI] ^c
	N ^a	Werte Studien- beginn MW (SD)	Änderung Studien- ende LS-MW (SE)	N ^a	Werte Studien- beginn MW (SD)	Änderung Studien- ende LS-MW (SE)	
Skalen zur gesundheitsbezogenen Lebensqualität (EORTC QLQ-C30)^e							
Globaler Gesundheitsstatus	476	62,6 (21,6)	-7,83 (1,5)	240	64,7 (22,4)	-6,23 (1,59)	-1,61 [-3,1; -0,11]; 0,035 Hedges' g: -0,17 [-0,32; -0,01]
Physische Funktion	477	78 (19,7)	-6,91 (1,57)	243	79,7 (19,6)	-4,46 (1,65)	-2,44 [-3,91; -0,98]; 0,001 Hedges' g: -0,26 [-0,41; -0,10]
Rollenfunktion	478	74,5 (29,3)	-13,27 (2,29)	243	77,6 (27,3)	-7,78 (2,4)	-5,49 [-7,57; -3,41]; < 0,001 Hedges' g: -0,41 [-0,56; -0,25]
Emotionale Funktion	477	78,2 (20,8)	-2,37 (1,33)	241	79,3 (20,0)	0,5 (1,43)	-2,87 [-4,38; -1,37]; < 0,001 Hedges' g: -0,30 [-0,45; -0,14]
Kognitive Funktion	477	88,7 (15,8)	-4,46 (1,16)	241	87,3 (16,6)	-2,15 (1,25)	-2,31 [-3,63; -1,00]; < 0,001 Hedges' g: -0,27 [-0,43; -0,12]
Soziale Funktion	477	77,3 (25,7)	-7,45 (1,82)	241	80,5 (24,3)	0,19 (1,96)	-7,64 [-9,63; -5,64]; < 0,001 Hedges' g: -0,59 [-0,75; -0,43]
Finanzielle Schwierigkeiten	476	15,6 (26,2)	0,22 (0,53)	240	13,6 (23,8)	-0,57 (0,82)	0,79 [-0,92; 2,50]; 0,363
<p>a: Anzahl der Patienten, die in der Auswertung zur Berechnung des Effektschätzers berücksichtigt wurden, die Werte bei Studienanfang können auf anderen Patientenzahlen basieren.</p> <p>b: Mittlerer Behandlungseffekt der Änderungen zu Baseline über die Zeit, basierend auf einem gemischten Modell für Messwiederholungen (MMRM). Der Faktor Visit geht als zufälliger Effekt in das Modell ein. Das Modell beinhaltet die Faktoren Baseline-Wert, Behandlung, den Tag des Visits nach Randomisierung und die Interaktion aus Behandlung und Tag.</p> <p>c: eigene Berechnung</p> <p>d: positive Veränderungen entsprechen einer Verschlechterung (auf einer Skala von 0 bis 100)</p> <p>e: negative Veränderungen entsprechend einer Verschlechterung (auf einer Skala von 0 bis 100)</p> <p>BSC: best supportive care; EORTC: European Organization for Research and Treatment of Cancer; KI: Konfidenzintervall; LS: least squares; MD: Mittelwertdifferenz; MMRM: Mixed Model for Repeated Measures; MW: Mittelwert; N: Anzahl ausgewerteter Patienten; QLQ-C30: Quality of Life Questionnaire-C30; SD: Standardabweichung; SE: Standardfehler; SMD: standardisierte Mittelwertdifferenz; vs.: versus</p>							

Symptome

Für die Skalen Fatigue, Schmerz, Appetitverlust und Diarrhö zeigt sich jeweils ein statistisch signifikantes Ergebnis zuungunsten von Regorafenib. Mangels Responderanalysen wurde für die Beurteilung der Relevanz des jeweiligen Effekts die SMD (Hedges' g) herangezogen:

- Für die Skala Diarrhö liegt das 95 %-Konfidenzintervall (KI) der SMD vollständig oberhalb der Irrelevanzschwelle von 0,2. Dies wird als relevanter Effekt interpretiert. Daher ergibt sich ein Anhaltspunkt für einen geringeren Nutzen von Regorafenib im Vergleich mit BSC für diesen Endpunkt.
- Für die Skalen Fatigue, Schmerz und Appetitverlust liegt das 95 %-KI der SMD jeweils nicht vollständig oberhalb der Irrelevanzschwelle von 0,2. Damit lässt sich jeweils nicht ableiten, dass der Effekt relevant ist. Daher ergibt sich jeweils kein Anhaltspunkt für einen geringeren Nutzen von Regorafenib gegenüber BSC, ein Zusatznutzen ist für diese Endpunkte damit nicht belegt.

Für die Skalen Übelkeit und Erbrechen, Atemnot, Insomnie sowie Obstipation zeigt sich jeweils kein statistisch signifikanter Gruppenunterschied. Es gibt daher jeweils keinen Anhaltspunkt für einen Zusatznutzen von Regorafenib im Vergleich mit BSC, ein Zusatznutzen von Regorafenib ist für diese Endpunkte nicht belegt.

Gesundheitsbezogene Lebensqualität

Für die Skalen globaler Gesundheitszustand, Rollenfunktion, physische, emotionale, kognitive und soziale Funktion zeigt sich jeweils ein statistisch signifikantes Ergebnis zuungunsten von Regorafenib. Mangels Responderanalysen wurde für die Beurteilung der Relevanz des jeweiligen Effekts die SMD (Hedges' g) herangezogen:

- Für die Skalen Rollenfunktion und soziale Funktion liegt das 95 %-KI der SMD jeweils vollständig unterhalb der Irrelevanzschwelle von -0,2. Dies wird jeweils als relevanter Effekt interpretiert. Daher ergibt sich jeweils ein Anhaltspunkt für einen geringeren Nutzen von Regorafenib im Vergleich mit BSC für diese Endpunkte.
- Für die Skalen globaler Gesundheitsstatus sowie physische, emotionale und kognitive Funktion liegt das 95 %-KI der SMD jeweils nicht vollständig unterhalb der Irrelevanzschwelle von -0,2. Damit lässt sich jeweils nicht ableiten, dass der Effekt relevant ist. Daher ergibt sich jeweils kein Anhaltspunkt für einen geringeren Nutzen von Regorafenib gegenüber BSC, ein Zusatznutzen ist für diese Endpunkte damit nicht belegt.

Für die Skala finanzielle Schwierigkeiten zeigt sich kein statistisch signifikanter Gruppenunterschied. Es gibt daher keinen Anhaltspunkt für einen Zusatznutzen von Regorafenib im Vergleich mit BSC; ein Zusatznutzen von Regorafenib ist für diesen Endpunkt nicht belegt.

2.3 Gesamtaussage zum Zusatznutzen

Durch die vom pU nachgereichten Analysen ergeben sich im Vergleich zur Dossierbewertung A15-43 zusätzliche negative Effekte von Regorafenib für die gesundheitsbezogene Lebensqualität (Rollenfunktion und soziale Funktion) mit dem Ausmaß „nicht quantifizierbar“. Auch zum Symptom Diarrhö zeigt sich ein solcher negativer Effekt, wobei für „Diarrhö“ bereits ein negativer Effekt von Regorafenib aus der Beurteilung der schweren unerwünschten Ereignisse (UE) in der Dossierbewertung A15-43 festgestellt wurde.

Tabelle 2 fasst die Resultate, die in die Gesamtaussage zum Ausmaß des Zusatznutzens einfließen, unter Berücksichtigung der Dossierbewertung A15-43 und des vorliegenden Addendums zusammen.

Tabelle 2: Positive und negative Effekte aus der Bewertung von Regorafenib + BSC im Vergleich zu Placebo + BSC

Positive Effekte	Negative Effekte
<ul style="list-style-type: none"> ▪ Anhaltspunkt für einen Zusatznutzen – Ausmaß: beträchtlich (Mortalität: Gesamtüberleben) 	<ul style="list-style-type: none"> ▪ Anhaltspunkt für einen größeren Schaden – Ausmaß: erheblich (schwere / schwerwiegende Nebenwirkungen: schwere UE mit CTCAE-Grad ≥ 3; darunter Diarrhö, Exanthem, Hand-Fuß-Syndrom, jeweils mit CTCAE-Grad 3 – Ausmaß: erheblich; Fatigue mit CTCAE-Grad 3 – Ausmaß: beträchtlich) ▪ Anhaltspunkt für einen geringeren Nutzen – Ausmaß: nicht quantifizierbar (Symptome: Diarrhö) ▪ Anhaltspunkt für einen geringeren Nutzen – Ausmaß: nicht quantifizierbar (gesundheitsbezogene Lebensqualität: Rollenfunktion, soziale Funktion)
CTCAE: Common Terminology Criteria for Adverse Events; UE: unerwünschtes Ereignis	

Im Unterschied zur Dossierbewertung A15-43 stehen dem positiven Effekt auf die Gesamtmortalität nicht nur schwere unerwünschte Nebenwirkungen gegenüber, sondern insbesondere auch ein geringerer Nutzen im Bereich der gesundheitsbezogenen Lebensqualität (Rollenfunktion und soziale Funktion). Daraus ergibt sich eine Änderung der Einschätzung dahin gehend, dass der Zusatznutzen im Bereich der Gesamtmortalität insgesamt infrage gestellt wird. Der Zusatznutzen für Regorafenib im Vergleich zu BSC bei Patienten mit metastasiertem Kolorektalkarzinom ist daher nicht belegt.

Das Vorgehen zur Ableitung einer Gesamtaussage zum Zusatznutzen stellt einen Vorschlag des IQWiG dar. Über den Zusatznutzen entscheidet der G-BA.

3 Literatur

1. Bayer Vital. Regorafenib (Stivarga): Dossier zur Nutzenbewertung gemäß §35a SGB V; Modul 4 A; Behandlung von erwachsenen Patienten mit metastasiertem Kolorektalkarzinom, die zuvor mit verfügbaren Therapien behandelt wurden oder die für diese nicht geeignet sind; diese Therapien umfassen Fluoropyrimidin-basierte Chemotherapie, eine Anti-VEGF-Therapie und eine Anti-EGFR-Therapie; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 29.09.2015 [Zugriff: 22.02.2016]. URL: https://www.g-ba.de/downloads/92-975-1197/2015-09-29_Modul4A_Regorafenib.pdf.
2. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Regorafenib: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A15-43 [online]. 23.12.2015 [Zugriff: 22.02.2016]. (IQWiG-Berichte; Band 352). URL: https://www.iqwig.de/download/A15-43_Regorafenib_Nutzenbewertung-35a-SGB-V.pdf.
3. Bayer Vital. Stellungnahme zum IQWiG-Bericht Nr. 352: Regorafenib; Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A15-43. [Demnächst verfügbar unter: <https://www.g-ba.de/informationen/nutzenbewertung/195/#tab/beschluesse> im Dokument "Zusammenfassende Dokumentation"].
4. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Addendum zum Auftrag A13-37 (Regorafenib): Auftrag A14-09 [online]. 26.02.2014 [Zugriff: 01.04.2014]. (IQWiG-Berichte; Band 207). URL: https://www.iqwig.de/download/A14-09_Addendum-zum-Auftrag-A13-37_Regorafenib.pdf.
5. Bayer Vital. Regorafenib (Stivarga): Dossier zur Nutzenbewertung gemäß §35a SGB V; Modul 4 A; Behandlung von erwachsenen Patienten mit metastasiertem Kolorektalkarzinom, die zuvor mit verfügbaren Therapien behandelt wurden oder die für diese nicht geeignet sind; diese Therapien umfassen Fluoropyrimidin-basierte Chemotherapie, eine Anti-VEGF-Therapie und eine Anti-EGFR-Therapie; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 30.09.2013 [Zugriff: 22.02.2016]. URL: https://www.g-ba.de/downloads/92-975-385/2013-09-30_Modul4_Regorafenib_mCRC.pdf.
6. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 4.2. Köln: IQWiG; 2015. URL: https://www.iqwig.de/download/IQWiG_Methoden_Version_4-2.pdf.

Anhang A – Vergleich der Ergebnisse zu den Mittelwertdifferenzen aus den MMRM-Analysen mit den standardisierten Mittelwertdifferenzen

Ausgangspunkt für die Betrachtungen sind die jeweiligen Least-Squares-Mittelwertdifferenzen (LS-MD) der einzelnen Symptom- bzw. Lebensqualitätsskalen. Diese sind das Ergebnis der MMRM-Analysen des pU unter Berücksichtigung der hohen Ausfallraten. Die MMRM-Analysen zeigen für 4 der 8 Symptomskalen (Fatigue, Schmerz, Appetitverlust und Diarrhö) und bei 6 der 7 Skalen zur gesundheitsbezogenen Lebensqualität (globaler Gesundheitsstatus, Rollenfunktion sowie physische, emotionale, kognitive und soziale Funktion) statistisch signifikante Gruppenunterscheide, und zwar ausschließlich zuungunsten von Regorafenib und mit zumeist sehr kleinen p-Werten (siehe Tabelle 1).

Der pU berechnet zu den einzelnen Skalen jeweils SMD. Solche Analysen dienen der Beurteilung der Effektgröße (Relevanzbetrachtung), sie sollen jedoch grundsätzlich die Aussagen zur statistischen Signifikanz nicht ändern. Bei Betrachtung der vom pU vorgelegten SMD fallen jedoch bezüglich der statistischen Signifikanz erhebliche Diskrepanzen zu den MMRM-Analysen auf: nur bei 2 der 10 Skalen mit statistisch signifikantem Gruppenunterschied auf Basis der MMRM-Analysen zeigt auch die vom pU berechnete SMD ein statistisch signifikantes Ergebnis (Diarrhö und soziale Funktion).

Zur Berechnung der SMD bedarf es einer gepoolten Standardabweichung (SD). Es lässt sich rückermitteln, dass der pU für deren Berechnung die beiden Standardfehler (SE) der mittleren Veränderung seit Studienbeginn pro Behandlungsarm verwendet hat. Aus den vorliegenden Daten und vom pU vorgelegten Informationen ist letztlich nicht ersichtlich, worauf die deutliche Diskrepanz zwischen seiner Berechnung der SMD und den LS-MD beruht. Möglicherweise ist bei der Berechnungsvariante des pU eine Abhängigkeit zwischen den LS-MW-Schätzern der beiden Behandlungsarme nicht beachtet worden und hierfür verantwortlich.

Wird die gepoolte SD dagegen unter Verwendung des SE der LS-MD (also des Gruppenunterschieds basierend auf den MMRM-Analysen) bestimmt, sind die Signifikanzaussagen zwischen LS-MD und SMD konsistent. Es muss davon ausgegangen werden, dass die Schätzung inklusive SE für den Gruppenunterschied durch den MMRM-Ansatz tragfähig ist, denn anderenfalls wären diese Analysen insgesamt nicht zu verwerten. Daher müssen auch die LS-MD mit zugehörigen SE zur Berechnung der SMD verwendet werden, was dann auch zu konsistenten Ergebnissen zwischen den Gruppenunterschieden basierend auf den MMRM-Analysen und den SMD führt.

Nachfolgend werden die Ergebnisse zu den LS-MD den vom pU ermittelten sowie den auf Basis des SE der LS-MD selbst berechneten SMD visuell gegenübergestellt. Sowohl für die Symptomskalen (Abbildung 1 bis Abbildung 3) als auch für die Skalen zur gesundheitsbezogenen Lebensqualität (Abbildung 4 bis Abbildung 6) zeigt sich, dass die Berechnungen des pU (jeweils 2. Abbildung) zu deutlichen Diskrepanzen zu den LS-MD

(jeweils 1. Abbildung) führen. Die eigenen Berechnungen (jeweils 3. Abbildung) sind hingegen mit den LS-MD konsistent.

Symptomskalen

Regorafenib vs. Placebo
Symptomatik (EORTC QLQ-C30), LS-MD

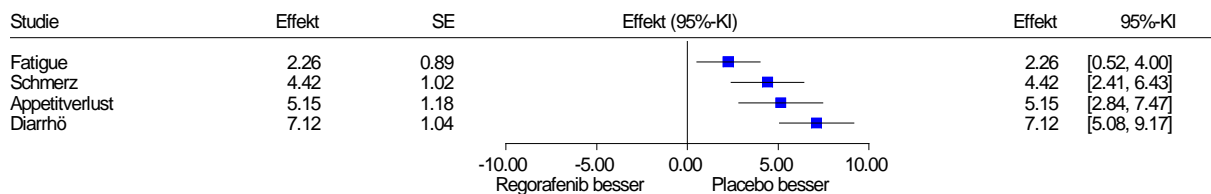


Abbildung 1: LS-MD mit statistisch signifikantem Gruppenunterschied – Symptomskalen

Regorafenib vs. Placebo
Symptomatik (EORTC QLQ-C30), SMD Variante 1

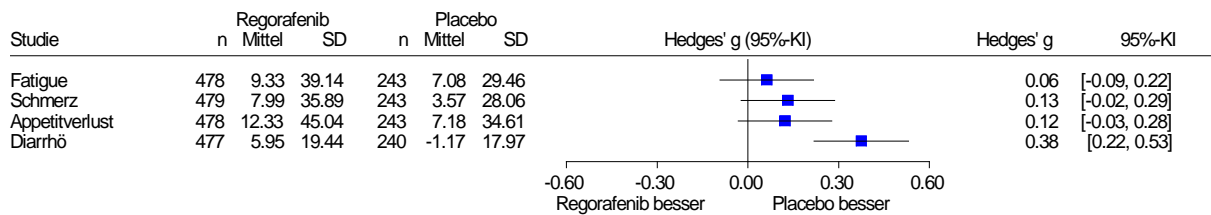


Abbildung 2: Vom pU berechnete SMD – Symptomskalen

Regorafenib vs. Placebo
Symptomatik (EORTC QLQ-C30), SMD Variante 2

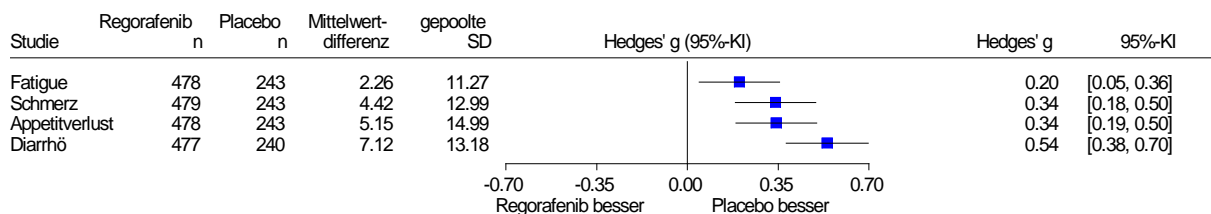


Abbildung 3: Selbst berechnete SMD – Symptomskalen

Skalen zur gesundheitsbezogenen Lebensqualität

Regorafenib vs. Placebo
 Gesundheitsbezogene Lebensqualität (EORTC QLQ-C30), LS-MD

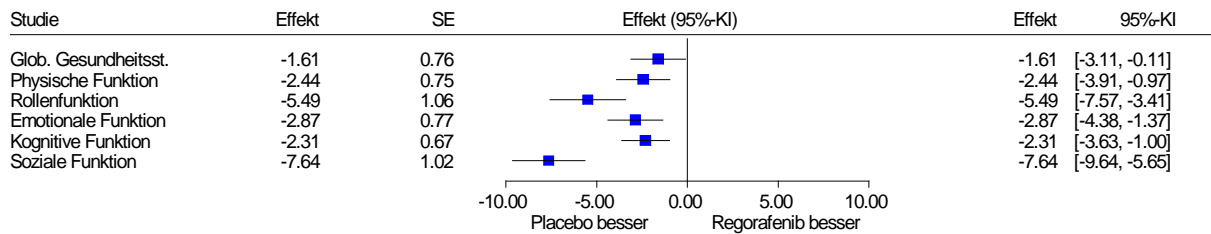


Abbildung 4: LS-MD mit statistisch signifikantem Gruppenunterschied – Skalen zur gesundheitsbezogenen Lebensqualität

Regorafenib vs. Placebo
 Gesundheitsbezogene Lebensqualität (EORTC QLQ-C30), SMD Variante 1

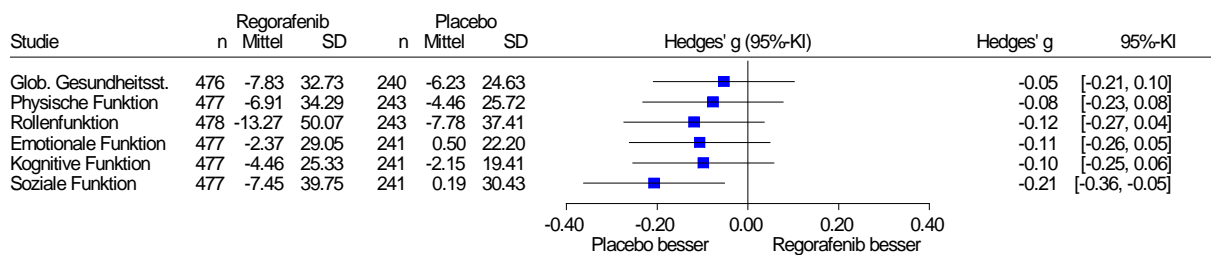


Abbildung 5: Vom pU berechnete SMD – Skalen zur gesundheitsbezogenen Lebensqualität

Regorafenib vs. Placebo
 Gesundheitsbezogene Lebensqualität (EORTC QLQ-C30), SMD Variante 2

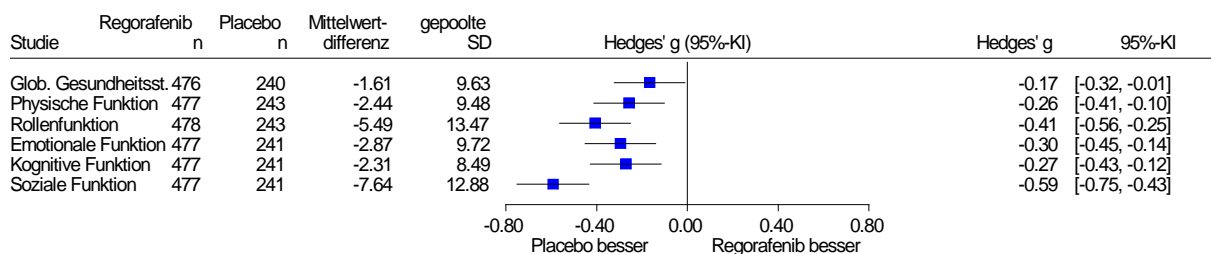


Abbildung 6: Selbst berechnete SMD – Skalen zur gesundheitsbezogenen Lebensqualität