

**Dossier zur Nutzenbewertung
gemäß § 35a SGB V**

Ipilimumab (YERVOY®)

Bristol-Myers Squibb GmbH & Co. KGaA

Module 3B und 4B

*Behandlung von fortgeschrittenen (nicht resezierbaren
oder metastasierten) Melanomen bei Erwachsenen
[, die zuvor noch keine Therapie erhalten haben]*

Zusatzdokument zur Nachforderung fehlender
Unterlagen/Angaben
(gemäß 5. Kapitel, § 17 Abs. 1 VerfO)

Inhaltsverzeichnis

Inhaltsverzeichnis	2
Tabellenverzeichnis	4
Abbildungsverzeichnis	6
Abkürzungsverzeichnis	7
Informationen zu den nachzureichenden Angaben (gemäß 5. Kapitel, § 17 Abs. 1 VerfO)	9
3.2.5 Beschreibung der Informationsbeschaffung für Abschnitt 3.2.....	9
4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel.....	9
4.3.1.2.2 Verzerrungspotenzial auf Studienebene.....	10
4.3.1.2.2 Matrix der Endpunkte.....	11
4.3.1.3.1 <Endpunkt xxx> - RCT.....	11
4.3.1.3.1 Mortalität – Bewertung des Verzerrungspotenzials und Operationalisierung.....	11
4.3.1.3.1 Mortalität – Tabelle „Ergebnisse für den Endpunkt Mortalität aus den eingeschlossenen Studien mit Ipilimumab Monotherapie 3mg/kg bei (chemo-)therapienaiven Patienten.....	13
4.3.1.3.1 Mortalität – Metaanalyse.....	14
4.3.1.3.2 Gesundheitsbezogene Lebensqualität – H2hH.....	14
4.3.1.3.2 Gesundheitsbezogene Lebensqualität – Tabelle „Ergebnisse für die Änderungen von Baseline zu Woche 12 für die gesundheitsbezogene Lebensqualität anhand EORTC QLQ-C30 aus RCT mit dem zu bewertenden Arzneimittel“.....	15
4.3.1.3.2 Gesundheitsbezogene Lebensqualität – Metaanalyse.....	17
4.3.1.3.4 Verträglichkeit – H2hH.....	17
4.3.1.3.4 Verträglichkeit – Metaanalyse.....	21
4.3.1.3.5 Subgruppenanalysen – Ergebnisse zu den Einzelstudien.....	21
4.3.1.3.5 Subgruppenanalysen – Metaanalysen.....	24
4.3.2.1 Indirekter Vergleich auf Basis randomisierter kontrollierter Studien.....	24
4.3.2.1.1 Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche.....	24
4.3.2.1.2 Charakteristika der Studien für indirekte Vergleiche.....	26
4.3.2.1.3.1 Mortalität – Tabelle „Ergebnisse für Mortalität aus RCT für indirekte Vergleiche“.....	27
4.3.2.1.3.1 Mortalität – Homogenität der Ergebnisse, gepoolte Ergebnisse (Metaanalyse).....	28
4.3.2.1.3.3 Mortalität – Tabelle „Ergebnisse für Verträglichkeit aus Studien für indirekte Vergleiche“.....	28
4.3.2.1.3.3 Verträglichkeit – Homogenität der Ergebnisse, gepoolte Ergebnisse (Metaanalyse).....	28
4.3.2.3 Weitere Untersuchungen.....	29
4.3.2.3.1 Ergebnis der Informationsbeschaffung – Weitere Untersuchungen.....	29
4.3.2.3.2 Charakteristika der weiteren Untersuchungen.....	30
Anhang A.....	30

Anhang B.....	32
Anhang F.....	33
Anhang G.....	38

Tabellenverzeichnis

	Seite
Tabelle Z-1 [Tabelle 4-33]: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel – Ergänzung der Einzelstudien der Phase II/III.....	10
Tabelle Z-2 [Tabelle 4-34]: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel – Ergänzung der Einzelstudien der Phase II/III	11
Tabelle Z-3 [Tabelle 4-35]: Operationalisierung Gesamtüberleben (OS) – Ergänzung der Einzelstudien der Phase II/III	12
Tabelle Z-4 [Tabelle 4-36]: Bewertung des Verzerrungspotenzials für Mortalität in eingeschlossenen Studien mit dem zu bewertenden Arzneimittel – Ergänzung der Einzelstudien der Phase II/III	12
Tabelle Z-5 [Tabelle 4-38]: Ergebnisse für den Endpunkt Mortalität aus den eingeschlossenen Studien mit Ipilimumab Monotherapie 3m/kg bei (chemo-)therapienaiven Patienten – Ergänzung für die Einzelstudien CA184-004 und CA184-022	13
Tabelle Z-6 [Tabelle 4-41]: Operationalisierung von gesundheitsbezogene Lebensqualität – Ergänzung der Einzelstudien der Phase II/III	14
Tabelle Z-7 [Tabelle 4-42]: Bewertung des Verzerrungspotenzials für gesundheitsbezogene Lebensqualität in den RCT mit dem zu bewertenden Arzneimittel – Ergänzung der Einzelstudien der Phase II/III	15
Tabelle Z-8 [Tabelle 4-43]: Ergebnisse für die Änderungen von Baseline zu Woche 12 für die gesundheitsbezogene Lebensqualität anhand EORTC QLQ-C30 aus RCT mit dem zu bewertenden Arzneimittel – Angaben nach Einzelstudien 020/ 022	16
Tabelle Z-9 [Tabelle 4-47]: Operationalisierung von Verträglichkeit – Ergänzung der Einzelstudien der Phase II/III	17
Tabelle Z-10 [Tabelle 4-48]: Bewertung des Verzerrungspotenzials für Verträglichkeit in RCT mit dem zu bewertenden Arzneimittel – Ergänzung der Einzelstudien der Phase II/III	19
Tabelle Z-11: [Tabelle 4-52] Anzahl der Patienten mit UE zum Vergleich Ipilimumab 3mg vs. DTIC aus dem H2hH – Ergänzung der Einzelstudien der Phase II/III	20
Tabelle Z-12: Subgruppenanalyse für das Gesamtüberleben in RCT mit dem zu bewertenden Arzneimittel	22
Tabelle Z-13: Verweise zur Charakterisierung der eingeschlossenen Studien – RCT für den indirekten Vergleich	26
Tabelle Z-14: Verweise zur Charakterisierung der Intervention – RCT für den indirekten Vergleich	27
Tabelle Z-15: Verweise zur Charakterisierung der Studienpopulation – RCT für den indirekten Vergleich	27
Tabelle Z-16 [4-100 (Anhang)]: Studiendesign und -methodik für Studie CA184-004.....	34
Tabelle Z-17 [4-102 (Anhang)]: Studiendesign und -methodik für Studie CA184-024.....	35
Tabelle Z-18 [Tabelle 4-104 (Anhang)]: Studiendesign und -methodik für Studie <CA184-338>	36

Tabelle Z-19 [Tabelle 4-105 (Anhang)]: Studiendesign und -methodik für Studie MDX010-08	37
Tabelle Z-20 (Anhang 4-G): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie "CA184-004"	39
Tabelle Z-21 (Anhang 4-G): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie "CA184-022"	51
Tabelle Z-22: (Anhang 4-G): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie "MDX010-08"	67
Tabelle Z-23 (Anhang 4-G): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie "MDX010-020"	80
Tabelle Z-24 (Anhang 4-G): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie "CA184-024"	96
Tabelle Z-25 (Anhang 4-G): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie "CA184-338"	106

Abbildungsverzeichnis

	Seite
Abbildung Z-1: Forest Plot der Subgruppenanalyse für "Gesamtüberleben" (ITT) in RCT mit dem zu bewertenden Arzneimittel	23

Abkürzungsverzeichnis

004, 022, 008, 020, 024	Abkürzung für die klinischen Studien CA184-004, CA184-022, MDX010-008, MDX010-020 und CA184-024
Abs.	Absatz
B-MS	Bristol-Myers Squibb
CHMP	Ausschuss für Humanarzneimittel der Europäischen Arzneimittelagentur (<i>Committee for Medical Products for Human Use</i>)
CSR	Studienbericht (<i>Clinical Study Report</i>)
DTIC	Dacarbazin
ECOG	Eastern Cooperative Oncology Group Performance Status
EMA	European Medicines Agency
EORTC QLQ-C30	Fragebogen zur Messung der Lebensqualität der <i>European Organisation for Research and Treatment of Cancer</i>
G-BA	Gemeinsamer Bundesausschuss
H2hH	Abkürzung für die verwendete Methode des direkter Vergleichs einer Metaanalyse gegen eine historische Kontrolle (<i>Head to historical Head</i>)
i.d.R.	in der Regel
IPI	Ipilimumab
ir	immunvermittelt
ITT	Patienten, die einer Behandlungsoption zugeteilt wurden (<i>Intention to Treat</i>)
k. A.	keine Angaben
kg	Kilogramm
KI	Konfidenzintervall
mg	Milligramm
MW	Mittelwert
N	Anzahl
OS	Gesamtüberleben (<i>Overall survival</i>)
Phase II/III	gepoolte Daten chemotherapienaiver Patienten aus RCT (CA184-004, CA184-022, MDX010-008, MDX010-020), die mit Ipilimumab 3mg/kg (Induktionsregime/ Monotherapie) untersucht wurden
pk	pharmakokinetisch(e)
PS	Propensity Score

QoL	Lebensqualität (<i>Quality of Life</i>)
RCT	Randomisiert kontrollierte Studie (<i>Randomized controlled trial</i>)
s.	siehe
SD	Standardabweichung
SUE	schwerwiegende (<i>serious</i>) UE
TASUE	therapieassoziierte schwerwiegende (<i>serious</i>) UE
TAUE	therapieassoziierte (<i>drug-related</i>) UE
UE	unerwünschte Ereignisse (<i>treatment-emergent adverse event</i>)
UEA	UE, die zum Therapieabbruch des Patienten geführt haben
VerfO	Verfahrensordnung
z. B.	zum Beispiel
ZVT	Zweckmäßige Vergleichstherapie

Informationen zu den nachzureichenden Angaben (gemäß 5. Kapitel, § 17 Abs. 1 VerfO)Erläuterung zum grundsätzlichen Vorgehen:

Informationen zu den Nachforderungen werden im Folgenden chronologisch beantwortet, wobei der Wortlaut der Nachforderung des G-BA in *kursiv* der Antwort durch B-MS vorangestellt wird. Tabellen und Abbildungen dieses Zusatzdokuments werden fortlaufend nummeriert, handelt es sich um eine Tabelle, die gegenüber des Dossiers vom 29.11.2013 mit Ergänzungen und/ oder Anpassungen nachgereicht wird, ist auch die ursprüngliche Tabellenummer [in Klammern] angegeben.

3.2.5 Beschreibung der Informationsbeschaffung für Abschnitt 3.2

Beschreibung der Informationsbeschaffung für Abschnitt 3.2.4 fehlt oder ist nicht zuordenbar. Bitte reichen Sie fehlende Angaben nach.

Für die Angaben in Abschnitt 3.2.4 wurden keine gesonderten Quellen recherchiert. Die Angabe der Anzahl der Patienten mit therapeutisch bedeutsamem Zusatznutzen entspricht der Herleitung in Abschnitt 3.2.3. Das Ausmaß des Zusatznutzens von Ipilimumab ergibt sich aus den Angaben in Modul 4 Abschnitt 4.4.3.

4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden ArzneimittelErläuterung zu den nachzureichenden Angaben für Modul 4.3.1.:

Eine zentrale Nachforderung des G-BA vom 07.01.2014 bezieht sich auf die Darstellung der Studiendaten der in die Nutzenbewertung eingeschlossene „Phase II/III“, genauer die Nachforderung, diese gepoolten Daten nach Einzelstudien separat darzustellen.

Bei den Studiendaten „Phase II/III“ handelt es sich um eine zulassungsrelevante Patientenpopulation, die der EMA gepoolt vorgelegt wurden. In dem am 29.11.2013 vorgelegten Nutzendossier wurden ebenfalls nur gepoolte Daten vorgelegt, zum einen um konsistent mit der EMA-Einreichung zu bleiben, zum anderen um Fehlschlüsse aus den geringen Patientenzahlen in den Einzelstudien zu vermeiden.

Einzig die Patientencharakteristika chemotherapienaiven Patienten, die in die „Phase II/III“ zusammengefasst wurden, sind zur Darstellung der Vergleichbarkeit nach Einzelstudien separat ausgewiesen worden (s. Tabellen 4-18 S. 95f., 4-19 S. 97f., 4-23 S. 105f., 4-24 S. 107f.).

Im Folgenden kommen wir der Nachforderung zur Darstellung der Daten auf Einzelstudienebene nach sofern machbar.

4.3.1.2.2 Verzerrungspotenzial auf Studienebene

In Tabelle „Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel“ (im vorliegenden Dossier Tabelle 4-33, Seite 121) ist für jede Studie separat das Verzerrungspotenzial auf Studienebene einzuschätzen.

Bitte reichen Sie für jede der unter dem Kürzel „Phase II/III“ zusammengefassten Studien (Studie CA184-004, CA184-022, MDX010-008 und MDX010-020) die entsprechenden Angaben als ergänzende Tabelle nach.

Tabelle Z-1 [Tabelle 4-33]: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel – Ergänzung der Einzelstudien der Phase II/III

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung		Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial auf Studienebene
			Patient	Behandler			
Studie 004	ja	ja	ja	ja	nein	nein	niedrig
Studie 022	ja	ja	ja	ja	nein	nein	niedrig
Studie 008	ja	ja	ja	ja	nein	nein	niedrig
Studie 020	ja	ja	ja	ja	nein	nein	niedrig
Studie 332	nein	nein	nein	nein	nein	ja	hoch
Studie 338	nein	nein	nein	nein	nein	ja	hoch
Studie 024	ja	ja	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Für die Studien 004, 022, 008, 020 und 024 ergeben sich keine Hinweise auf eine mögliche Verzerrung der Ergebnisse der Studie. Da die Punkte zum Verzerrungspotenzial auf alle fünf Studien zutreffen, werden sie im Folgenden zusammengefasst diskutiert. Die Randomisierungssequenz für die Zuteilung der Gruppen wurde computergeneriert und die Zuteilung zu den Gruppen erfolgte verdeckt. Sowohl die Patienten, der Behandlungsarzt als auch das medizinische Personal und die Personen, die mit der Datenauswertung beauftragt waren, waren gegenüber der tatsächlichen Medikation der Patienten verblindet. Die Ergebnisse sind gemäß dem statistischen Analyseplan dargestellt. Es wurden auch keine sonstigen, das Verzerrungspotential beeinflussenden Faktoren identifiziert.

Die Studien 332 und 338 sind als einarmige Beobachtungsstudien nicht randomisiert und offen durchgeführt worden. Durch die Art der Studien sind die Standards bezüglich Monitoring und Durchführung der Studien auf einem geringeren Level als es bei klinischen Studien der Fall ist. Somit wird das Verzerrungspotential dieser Studien als hoch eingestuft.

4.3.1.2.2 Matrix der Endpunkte

In Tabelle „Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel“ (im vorliegenden Dossier Tabelle 4-34, Seite 122) sind für jede Studie separat die Endpunkte anzugeben.

Bitte reichen Sie für jede der unter dem Kürzel „Phase II/III“ zusammengefassten Studien (Studie CA184-004, CA184-022, MDX010-008 und MDX010-020) die entsprechenden Angaben als ergänzende Tabelle nach.

Tabelle Z-2 [Tabelle 4-34]: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel – Ergänzung der Einzelstudien der Phase II/III

Studie	Gesamtüberleben	gesundheitsbezogene Lebensqualität	Verträglichkeit
Studie 004	ja	nein	ja
Studie 022	ja	ja	ja
Studie 008	ja	nein	ja
Studie 020	ja	ja	ja
Studie 332	ja	nein	nein
Studie 338	ja	nein	ja
Studie 024	ja	ja	ja

4.3.1.3.1 <Endpunkt xxx> - RCT

Anforderung:

Zunächst soll für **jede Studie** das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die **Ergebnisse der einzelnen Studien** zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. **Anschließend** sollen die Ergebnisse, wenn möglich und sinnvoll, in einer **Meta-Analyse** zusammengefasst und beschrieben werden.

4.3.1.3.1 Mortalität – Bewertung des Verzerrungspotenzials und Operationalisierung

In Tabelle „Bewertung des Verzerrungspotenzials für Mortalität in eingeschlossenen Studien mit dem zu bewertenden Arzneimittel (im vorliegenden Dossier Tabelle 4-36, S. 123) ist für jede Studie separat das Verzerrungspotenzial für den Endpunkt Mortalität einzuschätzen.

Bitte reichen Sie für jede der unter dem Kürzel „Phase II/III“ zusammengefassten Studien (Studie CA184-004, CA184-022, MDX010-008 und MDX010-020) die entsprechenden Angaben als ergänzende Tabelle nach.

Tabelle Z-3 [Tabelle 4-35]: Operationalisierung Gesamtüberleben (OS) – Ergänzung der Einzelstudien der Phase II/III

Studien	Operationalisierung
Studie 004	Der Endpunkt Gesamtüberleben (OS) ist definiert als Zeit ab dem Zeitpunkt der Randomisierung in der jeweiligen RCT bzw. für die Studien 332/338 der Zeitpunkt der ersten Gabe Ipilimumab bis zum Versterben aus jedem Grund. Patienten, die noch nicht verstorben sind, wurden zum letzten bekannten Zeitpunkt, an dem sie gelebt haben, zensiert. Patienten, die zum jeweiligen Zeitpunkt der Primäranalyse noch lebten, wurden zu diesem Zeitpunkt zensiert. Die Analyse erfolgte auf Basis der ITT Population. Die Daten wurden mit einem Cox Proportional Hazard Modell ausgewertet, stratifiziert für die Klassen des Propensity Scores. Im Modell wurden das Hazard Ratio mit 95% KI für die Gesamtgruppe und pro Propensity Score (PS) Klasse sowie das mediane Überleben und die Kaplan-Meier Plots mit Anzahl Patienten unter Risiko für die Propensity Score Klassen dargestellt.
Studie 022	
Studie 008	
Studie 020	
Studie 332	
Studie 338	
Studie 024	

Tabelle Z-4 [Tabelle 4-36]: Bewertung des Verzerrungspotenzials für Mortalität in eingeschlossenen Studien mit dem zu bewertenden Arzneimittel – Ergänzung der Einzelstudien der Phase II/III

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
Studie 004	niedrig	ja	ja	nein	Ja	hoch
Studie 022	niedrig	ja	ja	nein	Ja	hoch
Studie 008	niedrig	ja	ja	nein	Ja	hoch
Studie 020	niedrig	ja	ja	nein	Ja	hoch
Studie 332	hoch	nein	ja	nein	Ja	hoch
Studie 338	hoch	nein	ja	nein	Ja	hoch
Studie 024	niedrig	ja	ja	nein	Ja	hoch

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial des Endpunktes Gesamtüberleben wurde in den RCT (Studien 004, 022, 008, 020 und 024) als hoch eingestuft. Da die Punkte zum Verzerrungspotenzial auf alle fünf Studien zutreffen, werden sie im Folgenden zusammengefasst diskutiert. Die Untersucher und Erheber der Endpunkte waren verblindet gegenüber der Behandlung der Patienten, die Auswertung wurde unter Berücksichtigung der *drop-outs* mittels ITT-Prinzips durchgeführt und alle Ergebnisse wurden gemäß der geplanten Analysen adäquat berichtet. Allerdings wurden die Studien als Vergleich von Ipilimumab gegen eine historische Kontrolle DTIC durchgeführt und nicht als RCT. Zwar wurden die Studienpopulationen durch eine

Propensity Score Analyse angeglichen, allerdings nur für bekannte Baseline Charakteristika. Somit ist eine Verzerrung nicht auszuschließen. Das gilt ebenfalls für die Studien 332 und 338.

Zudem lag für die Beobachtungsstudien 332 und 338 auf Studienebene bereits ein hohes Verzerrungspotential vor, so dass auf Endpunktebene auch deshalb ein hohes Verzerrungspotential angenommen wird.

4.3.1.3.1 Mortalität – Tabelle „Ergebnisse für den Endpunkt Mortalität aus den eingeschlossenen Studien mit Ipilimumab Monotherapie 3mg/kg bei (chemo-) therapienaiven Patienten

In der Tabelle (vorliegend Tabelle 4-38, Seite 129) werden für die Studien CA184-004 und CA184-022 nur gepoolte Ergebnisse dargestellt. Bitte reichen Sie die entsprechenden Angaben getrennt nach Studie für jede der Studien CA184-004 und CA184-022 als ergänzende Tabelle nach.

Tabelle Z-5 [Tabelle 4-38]: Ergebnisse für den Endpunkt Mortalität aus den eingeschlossenen Studien mit Ipilimumab Monotherapie 3m/kg bei (chemo-) therapienaiven Patienten – Ergänzung für die Einzelstudien CA184-004 und CA184-022

Studie	Behandlungsarm	N	Medianes Follow Up, in Monaten (25 %-Quantil, 75 %-Quantil)	Medianes Gesamtüberleben, in Monaten (95 % KI ^b)	1-Jahresüberlebensrate, in % (95% KI ^b)	2-Jahresüberlebensrate, in % (95% KI ^b)
CA184-004/022	chemotherapienaiv	25 ^a	14,6 (8,5-58,7)	22,6 (9,5-60,7)	68% (0,45-0,82)	45% (0,25-0,64)
CA184-004	chemotherapienaiv	17	14,6 (8,5-48,2)	24,0 (5,1-k.A.)	70% (0,42-0,86)	45% (0,20-0,66)
CA184-022	chemotherapienaiv	8	17,2 (9,1-61,5)	22,6 (1,1-k.A.)	63% (0,23-0,86)	47% (0,12-0,76)
MDX010-020	chemotherapienaiv	13	12,0 (8,2-39,3)	12,2 (8,1-k.A.)	62% (0,31-0,82)	35% (0,11-0,61)

k.A.= keine Angabe, KI= Konfidenzintervall, N= Anzahl

a: Im Modul 4 vom 29.11.2013 wurde entsprechend der im Bewertungsberichts des Committee for Medicinal Products for Human Use (CHMP) zur Indikationserweiterung des Medikaments Yervoy vom 24.10.2013 zugänglichen Daten die Ergebnisse der therapienaiven-Patienten (15 Patienten) ausgewiesen. Da für den Zusatznutzen die chemotherapienaiven Patienten dieser beiden Studien (25 Patienten) eingeschlossen werden, werden hier die Ergebnisse der chemotherapienaiven Patienten der Studien 004/022 gepooled sowie nach Einzelstudien ausgewiesen.

b: KI berechnet gemäß Log-Log Methode

Quellen: interne Angaben zur Nachforderung

4.3.1.3.1 Mortalität – Metaanalyse

In der Metaanalyse (vorliegend Abbildung 13, Seite 136) werden für die Studien CA184-004, CA184-022, MDX010-008 und MDX010-020 die bereits gepoolten Ergebnisse („Phase II/III“) dargestellt. Bitte reichen Sie die Metaanalyse unter separater Darstellung jeder der eingeschlossenen Studien nach.

Bei den Studiendaten „Phase II/III“ handelt es sich um die zulassungsrelevante Patientenpopulation, die der EMA gepoolt vorgelegt wurden. Hiermit in Konsistenz und die kleinen Patientenzahlen reflektierend, wurden mit dem am 29.11.2013 vorgelegten Nutzendossier nur die gepoolten Daten vorgelegt.

Für die Bewertung des Zusatznutzens wurde dabei pro Studie ein Behandlungsarm in die Analyse einbezogen. Zur Angleichung der Populationen der mit Ipilimumab-behandelten Patienten mit den DTIC-behandelten Patienten aus der Studie 024 wurde eine Propensity Score Analyse durchgeführt. Diese ist dann sinnvoll, wenn ausreichend viele Patienten für diese Analysen zur Verfügung standen. Der zum Teil sehr kleine Stichprobenumfang pro Ipilimumab-Studie war ein weiterer Grund, auch für diese Analyse die Phase II und III Studien zu poolen, um auf Basis des Pools die Propensity Score Analyse durchzuführen und die für die Herleitung des Zusatznutzens notwendigen Analysen durchzuführen.

Daher liegen keine Einzelergebnisse zu diesem Endpunkt für die Studien 004, 022, 008 und 020 vor.

4.3.1.3.2 Gesundheitsbezogene Lebensqualität – H2hH

Anforderung:

*Zunächst soll für **jede Studie** das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die **Ergebnisse der einzelnen Studien** zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. **Anschließend** sollen die Ergebnisse, wenn möglich und sinnvoll, in einer **Meta-Analyse** zusammengefasst und beschrieben werden.*

Tabelle Z-6 [Tabelle 4-41]: Operationalisierung von gesundheitsbezogene Lebensqualität – Ergänzung der Einzelstudien der Phase II/III

Studie	Operationalisierung
Studie 022 Studie 020	Der Endpunkt gesundheitsbezogene Lebensqualität wurde mit dem Fragebogen EORTC QLQ-C30 zum Zeitpunkt Woche 12 als Änderung zu Baseline erhoben. Die Analyse erfolgte auf Basis einer eingeschränkten ITT Population. Fehlende Werte wurden nicht ersetzt.

Tabelle Z-7 [Tabelle 4-42]: Bewertung des Verzerrungspotenzials für gesundheitsbezogene Lebensqualität in den RCT mit dem zu bewertenden Arzneimittel – Ergänzung der Einzelstudien der Phase II/III

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
Studie 022	hoch	ja	nein	nein	ja	hoch
Studie 020	hoch	ja	nein	nein	ja	hoch

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial des Endpunktes Gesundheitsbezogene Lebensqualität wurde in den zwei RCT (Studien 022 und 020) als hoch eingestuft. Da die Punkte zum Verzerrungspotenzial auf beide Studien zutreffen, werden sie im Folgenden zusammengefasst diskutiert. Die Untersucher und Erheber der Endpunkte waren verblindet gegenüber der Behandlung der Patienten, die Auswertung wurde ohne Berücksichtigung der *drop-outs* durchgeführt. Alle Ergebnisse wurden ausschließlich deskriptiv gemäß der geplanten Analysen berichtet.

4.3.1.3.2 Gesundheitsbezogene Lebensqualität – Tabelle „Ergebnisse für die Änderungen von Baseline zu Woche 12 für die gesundheitsbezogene Lebensqualität anhand EORTC QLQ-C30 aus RCT mit dem zu bewertenden Arzneimittel“

In der Tabelle (vorliegend Tabelle 4-43, Seite 138) werden für die Studien CA184-022 und MDX010-020 nur gepoolte Ergebnisse dargestellt. Hinsichtlich der Ergebnisse der Einzelstudien wird auf Modul 5 verwiesen. Alle Methoden, Datenanalysen und Ergebnisse sind zusammenfassend in den Modulen 1 bis 4 zu beschreiben. Verweise auf Modul 5 sind nicht ausreichend.

Bitte reichen Sie die entsprechenden Angaben getrennt nach Studie für jede der Studien CA184-022 und MDX010-020 als ergänzende Tabelle nach.

Die Einzel-Ergebnisse beider Studien sind (erst) im folgenden Abschnitt „Meta-Analyse“ zusammenzufassen und in geeigneter Weise darzustellen (i.d.R. als Forest-Plot).

Tabelle Z-8 [Tabelle 4-43]: Ergebnisse für die Änderungen von Baseline zu Woche 12 für die gesundheitsbezogene Lebensqualität anhand EORTC QLQ-C30 aus RCT mit dem zu bewertenden Arzneimittel – Angaben nach Einzelstudien 020/ 022

	Ipilimumab Total		Ipilimumab CA184-022		Ipilimumab MDX010-20		DTIC	
	N	MW (SD)	N	MW (SD)	N	MW (SD)	N	MW (SD)
Global	14	-11,3 (16,5)	5	-5,0 (12,6)	9	-14,8 (18,1)	146	-6,45 (23,50)
Physisch	14	-5,5 (7,6)	5	-5,3 (8,7)	9	-5,6 (7,5)	147	-8,48 (20,71)
Rolle	14	-11,9 (22,1)	5	-23,3 (25,3)	9	-5,6 (18,6)	147	-10,88 (27,91)
Kognitiv	14	-6,0 (16,8)	5	-6,7 (25,3)	9	-5,6 (11,8)	147	-3,97 (17,59)
Emotional	14	0,0 (16,0)	5	1,7 (7,0)	9	-0,9 (19,7)	147	1,97 (20,98)
Sozial	14	-7,1 (18,2)	5	-6,7 (14,9)	9	-7,4 (20,6)	147	-6,35 (26,05)
Fatigue	14	8,7 (17,0)	5	8,9 (19,9)	9	8,6 (16,5)	147	10,77 (23,03)
Nausea-Erbrechen	14	7,1 (23,3)	5	3,3 (7,5)	9	9,3 (29,0)	147	8,16 (19,45)
Schmerz	14	6,0 (16,8)	5	10,0 (14,9)	9	3,7 (18,2)	147	5,33 (29,90)
Dyspnoe	14	-2,4 (20,5)	5	0,0 (0,0)	9	-3,7 (26,1)	144	9,26 (21,39)
Appetitlosigkeit	14	7,1 (29,8)	5	0,0 (23,6)	9	11,1 (33,3)	147	9,98 (31,31)
Schlafstörungen	14	0,0 (26,1)	5	-13,3 (29,8)	9	7,4 (22,2)	146	-0,23 (27,54)
Verstopfung	14	-7,1 (19,3)	5	-6,7 (14,9)	9	-7,4 (22,2)	145	8,28 (25,01)
Diarrhoe	13	2,6 (25,3)	5	-6,7 (14,9)	8	8,3 (29,5)	147	1,36 (16,03)
Finanzieller Einfluß	14	7,1 (19,3)	5	0,0 (0,0)	9	11,1 (23,6)	144	-0,23 (22,12)
Quellen: s. Dossier S. 138								

4.3.1.3.2 Gesundheitsbezogene Lebensqualität – Metaanalyse

Es liegt keine Metaanalyse der Studien CA184-022 und MDX010-020 vor. Bitte reichen Sie die Metaanalyse unter geeigneter Darstellung (i.d.R. als Forest-Plot) nach.

Daten zur Lebensqualität liegen nur in einer sehr begrenzten Stichprobengröße für die Ipilimumab Studien vor. Die Daten werden nur deskriptiv dargestellt und interpretiert, da die geringe Stichprobengröße der gepoolten Daten keine Propensity Score Analyse zulässt. Eine nicht-adjustierte Analyse wurde nicht durchgeführt. Es werden nur die gepoolten Ergebnisse der beiden Studien dargestellt. Daher wird hier keine weitere Metaanalyse nachgereicht.

4.3.1.3.4 Verträglichkeit – H2hH

Anforderung:

*Zunächst soll für **jede Studie** das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die **Ergebnisse der einzelnen Studien** zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. **Anschließend** sollen die Ergebnisse, wenn möglich und sinnvoll, in einer **Meta-Analyse** zusammengefasst und beschrieben werden.*

Im Folgenden werden die relevanten Tabellen aktualisiert, an den zusammenfassenden Bewertungen ändert diese Aktualisierung nichts.

Tabelle Z-9 [Tabelle 4-47]: Operationalisierung von Verträglichkeit – Ergänzung der Einzelstudien der Phase II/III

Studien	Endpunkt	Operationalisierung
UE, SUE, UEA		
Studie 004		Dichotomer Endpunkt (ja/nein) bezüglich des Auftretens eines UE, SUE bzw. UEA innerhalb des Beobachtungszeitraums, der durch den jeweiligen verblindeten Prüfarzt dokumentiert wurde. Die Analyse basiert auf der Sicherheitspopulation, in der alle Patienten eingeschlossen sind, denen mindestens eine Dosis einer Studienmedikation verabreicht wurde. Die Gruppenzuteilung erfolgte auf Basis der Studienmedikation, nicht der Randomisierung.
Studie 022		
Studie 008		
Studie 020		
Studie 024		

Studien	Endpunkt	Operationalisierung
338		Dichotomer Endpunkt (ja/nein) bezüglich des Auftretens eines UE, SUE bzw. UEA retrospektiv erhoben seit Beginn der Behandlung. Die UE wurden durch den jeweiligen Prüfarzt dokumentiert. Die Analyse basiert auf der Sicherheitspopulation, in der alle Patienten eingeschlossen sind, denen mindestens eine Dosis einer Studienmedikation verabreicht wurde.
TAUE, irUE, irSUE		
Studie 004 Studie 022 Studie 008 Studie 020 Studie 024		Dichotomer Endpunkt (ja/nein) bezüglich des Auftretens eines immunvermittelten UE oder SUE innerhalb des Beobachtungszeitraums, der durch den jeweiligen verblindeten Prüfarzt dokumentiert wurde. Die Klassifizierung als immunvermittelt wurde über ein Programm nach einer vordefinierten Liste der Gruppenbezeichnungen festgelegt. Die Analyse basiert auf der Sicherheitspopulation, in der alle Patienten eingeschlossen sind, denen mindestens eine Dosis einer Studienmedikation verabreicht wurde. Die Gruppenzuteilung erfolgte auf Basis der Studienmedikation, nicht der Randomisierung.
338		Dichotomer Endpunkt (ja/nein) bezüglich des Auftretens eines immunvermittelten UE oder SUE seit Behandlungsbeginn, retrospektiv erhoben vom jeweiligen Prüfarzt. Die Klassifizierung als immunvermittelt wurde über ein Programm nach einer vordefinierten Liste der Gruppenbezeichnungen festgelegt. Die Analyse basiert auf der Sicherheitspopulation, in der alle Patienten eingeschlossen sind, denen mindestens eine Dosis einer Studienmedikation verabreicht wurde.

Tabelle Z-10 [Tabelle 4-48]: Bewertung des Verzerrungspotenzials für Verträglichkeit in RCT mit dem zu bewertenden Arzneimittel – Ergänzung der Einzelstudien der Phase II/III

Studie	Endpunkt	Verzerrungspotenzial auf Studienebene	Verblindung/Endpunktkriterherber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
Studie 004	UE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	SUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	UEA	Niedrig	Ja	Ja	Nein	Nein	Hoch
	irTAUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	irTASUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
Studie 022	UE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	SUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	UEA	Niedrig	Ja	Ja	Nein	Nein	Hoch
	irTAUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	irTASUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
Studie 008	UE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	SUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	UEA	Niedrig	Ja	Ja	Nein	Nein	Hoch
	irTAUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	irTASUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
Studie 020	UE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	SUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	UEA	Niedrig	Ja	Ja	Nein	Nein	Hoch
	irTAUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	irTASUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
Studie 024	UE	Niedrig	Ja	Ja	Nein	Nein	Hoch

Studie	Endpunkt	Verzerrungspotenzial auf Studienebene	Verblindung/Endpunktkriterien	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
	SUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	UEA	Niedrig	Ja	Ja	Nein	Nein	Hoch
	irTAUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
	irTASUE	Niedrig	Ja	Ja	Nein	Nein	Hoch
Studie 338	UE	Hoch	Nein	Ja	Nein	Ja	Hoch
	SUE	Hoch	Nein	Ja	Nein	Ja	Hoch
	UEA	Hoch	Nein	Ja	Nein	Ja	Hoch
	irTAUE	Hoch	Nein	Ja	Nein	Ja	Hoch
	irTASUE	Hoch	Nein	Ja	Nein	Ja	Hoch

Tabelle Z-11: [Tabelle 4-52] Anzahl der Patienten mit UE zum Vergleich Ipilimumab 3mg vs. DTIC aus dem H2hH – Ergänzung der Einzelstudien der Phase II/III

Endpunkt	IPI 3mg Phase II/III (n=78) ¹	IPI 3mg 022 (n=8) ²	IPI 3mg 004 (n=17) ²	IPI 3mg 008 (n=39) ³	IPI 3mg 020 (n=12) ⁴
Anzahl Patienten (%)					
UE	72 (92)	8 (100)	16 (94)	39 (100)	11 (92)
SUE	20 (26)	1 (13)	8 (47)	11 (28)	1 (8)
UEA	6 (8)	0 (0)	3 (18)	1 (3)	0 (0)

IPI: Ipilimumab, UE: unerwünschte Ereignisse, SUE: schwerwiegende UE, UEA: UE, die zum Therapieabbruch geführt haben

Quellen: 1) gemäß Tabelle 4-52 (S. 149) des Modul 4 vom 29.11.2013: chemotherapienaive Patienten, UE während des Induktionsregimes, gesamter Studienpool; 2) interne Informationen: chemotherapienaive Patienten, UE während des Induktionsregimes; 3) gemäß Tabelle 4-49 (S. 147) des Modul 4 vom 29.11.2013: chemotherapienaive Patienten, UE während der gesamten Beobachtungsdauer, Sicherheitspopulation 4) interne Informationen: chemotherapienaive Patienten, UE während des Induktionsregimes; Sicherheitspopulation

4.3.1.3.4 Verträglichkeit – Metaanalyse

Die Einzel-Ergebnisse der für den Endpunkt „Verträglichkeit“ herangezogenen Studien sind im Abschnitt „Meta-Analyse“ zusammenzufassen und in geeigneter Weise darzustellen (i.d.R. als Forest-Plot). Bitte reichen Sie die Metaanalyse unter geeigneter Darstellung nach.

Bei den Studiendaten „Phase II/III“ handelt es sich um die zulassungsrelevante Patientenpopulation, die der EMA gepoolt vorgelegt wurden.

Hiermit in Konsistenz und die kleinen Patientenzahlen reflektierend, wurden mit dem am 29.11.2013 vorgelegten Nutzendossier nur die gepoolten Daten vorgelegt. Für die Bewertung des Zusatznutzens wurde dabei pro Studie ein Behandlungsarm in die Analyse einbezogen. Zur Angleichung der Populationen der mit Ipilimumab-behandelten Patienten mit den DTIC-behandelten Patienten aus der Studie 024 wurde eine Propensity Score Analyse durchgeführt. Diese ist dann sinnvoll, wenn ausreichend viele Patienten für diese Analysen zur Verfügung stehen. Der zum Teil sehr kleine Stichprobenumfang pro Ipilimumab-Studie war auch bei der Verträglichkeitsauswertung ein weiterer Grund, für diese Analyse die Phase II und III Studien zu poolen, um auf Basis des Pools die Propensity Score Analyse durchzuführen und die für die Herleitung des Zusatznutzens notwendigen Analysen durchzuführen.

Daher liegen keine Einzelergebnisse zu diesem Endpunkt für die Studien 004, 022, 008 und 020 vor.

4.3.1.3.5 Subgruppenanalysen – Ergebnisse zu den Einzelstudien

Anforderung:

Die Ergebnisse in den Subgruppen sind zunächst für jede einzelne Studie in tabellarischer Form darzustellen. Diese Anforderung gilt sowohl für Subgruppenanalysen auf Basis individueller Patientendaten als auch für solche auf Basis aggregierter Daten.

Bitte reichen Sie die fehlenden Subgruppenanalysen (i.e. Einzelstudien-Ergebnisse) nach.

Die Subgruppenanalysen sind im Kapitel 4.3.1.3.5 des Moduls 4 dargestellt. Diese basieren wie die Hauptanalysen des Endpunkts Mortalität auf den Propensity Score Analysen, die pro Subgruppe durchgeführt wurden. Im Folgenden sind die Ergebnisse der Subgruppenanalysen in tabellarischer Form und als Forestplot dargestellt.

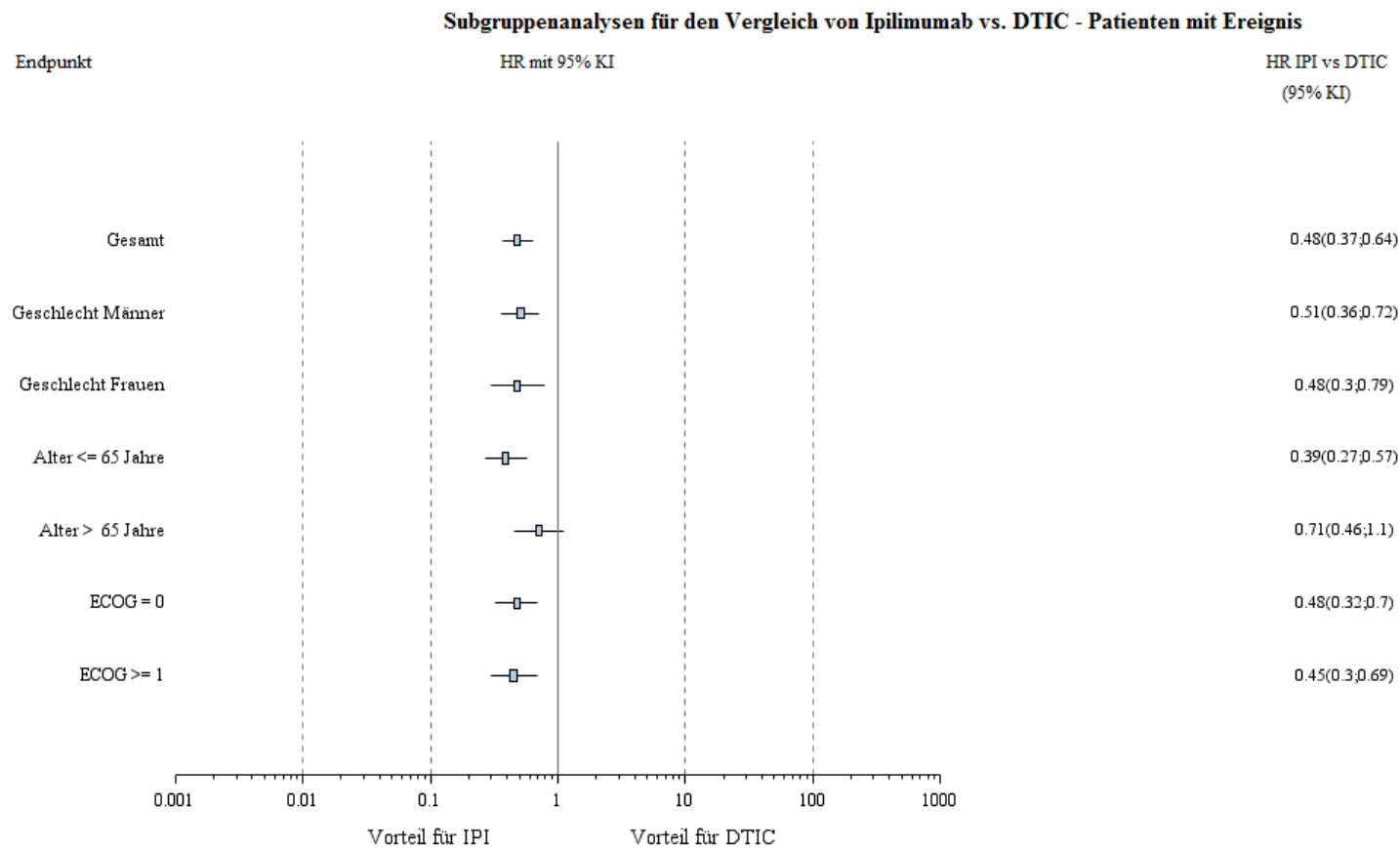
Tests auf Interaktion zwischen Subgruppe und Behandlung waren nicht möglich, da die Propensity Score Analyse pro Subgruppe durchgeführt wurde und somit die Analysen zwischen z.B. Männern und Frauen nicht direkt vergleichbar waren.

Die im Modul 4.3.1.3.5 (S. 145f.) ausgewiesenen Ergebnisse werden im Folgenden tabellarisch sowie als Forest Plot ausgewiesen.

Tabelle Z-12: Subgruppenanalyse für das Gesamtüberleben in RCT mit dem zu bewertenden Arzneimittel

Subgruppenanalyse für Gesamtüberleben (ITT)			
Hazard Ratio (IPI vs. DTIC)			
Subgruppen-Variable	Schätzung	95%-KI	p-Wert
Geschlecht			
Männer	0,51	0,36; 0,72	p<0,01
Frauen	0,48	0,30; 0,79	p<0,01
Alter			
<65 Jahre	0,39	0,27; 0,57	p<0,001
≥65 Jahre	0,71	0,46; 1,10	p=0,12
ECOG			
0	0,48	0,32; 0,70	p<0,001
≥1	0,45	0,30; 0,69	p<0,001

Abbildung Z-1: Forest Plot der Subgruppenanalyse für "Gesamtüberleben" (ITT) in RCT mit dem zu bewertenden Arzneimittel



Quelle: eigene Darstellung

4.3.1.3.5 Subgruppenanalysen – Metaanalysen

Anforderung:

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Ergebnisse mithilfe einer Meta-Analyse quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analyse (als Forest-Plot) dar.

Bitte reichen Sie die fehlenden Metaanalysen in geeigneter Darstellung nach.

Bei den Studiendaten „Phase II/III“ handelt es sich um die zulassungsrelevante Patientenpopulation, die der EMA gepoolt vorgelegt wurden.

Wie oben beschrieben wurden hiermit in Konsistenz und die kleinen Patientenzahlen reflektierend mit dem am 29.11.2013 vorgelegten Nutzendossier ebenfalls nur die gepoolten Daten vorgelegt. Für die Bewertung des Zusatznutzens wurde dabei pro Studie ein Behandlungsarm in die Analyse einbezogen. Zur Angleichung der Populationen der mit Ipilimumab-behandelten Patienten mit den DTIC-behandelten Patienten aus der Studie 024 wurde eine Propensity Score Analyse pro Subgruppe durchgeführt. Diese ist dann sinnvoll, wenn ausreichend viele Patienten für diese Analysen zur Verfügung waren. Der zum Teil sehr kleine Stichprobenumfang pro Ipilimumab-Studie war ein weiterer Grund, auch für diese Analyse die Phase II und III Studien zu poolen, um auf Basis des Pools die Propensity Score Analyse durchzuführen und die für die Herleitung des Zusatznutzens notwendigen Analysen durchzuführen.

Daher liegen keine Einzelergebnisse zu den Subgruppen für die Studien 004, 022, 008 und 020 vor, die somit auch nicht dargestellt werden können.

4.3.2.1 Indirekter Vergleich auf Basis randomisierter kontrollierter Studien

Erläuterung zu den nachzureichenden Angaben für Modul 4.3.2.1:

Bzgl. des indirekten Vergleichs fordert der G-BA sowohl weitere Angaben zur Informationsbeschaffung als auch zu den Patienten-Charakteristika nach. B-MS beschreibt dazu im folgenden Absatz 4.3.2.1.1 die Informationsbeschaffung und ergänzt Verweise auf die bereits vorliegenden Angaben im Modul 4.3.1.

Ferner bittet der G-BA auch für den indirekten Vergleich, die Ergebnisse nach Einzelstudien separat darzustellen. Die im Kapitel 4.3.1.3 dieses Zusatzdokuments ausgeführten Begründungen treffen auch für den indirekten Vergleich zu. Es werden daher keine Ergebnisse nach Einzelstudien nachgereicht.

4.3.2.1.1 Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche

Für Abschnitt 4.3.2.1 werden die Studien CA184-004, CA184-022, MDX010-008, MDX010-020, CA184-024, CA184-332 und CA184-338 des pharmazeutischen Unternehmers

herangezogen. Diese ergeben sich nicht aus Abschnitt 4.3.2.1.1 bzw. wird nicht auf ggf. in vorherigen Abschnitten dargestellten entsprechenden Angaben oder Tabellen verwiesen:

4.3.2.1.1.1 „Studien des pharmazeutischen Unternehmers – Studien für indirekte Vergleiche mit Tabelle „Liste der Studien des pharmazeutischen Unternehmers – RCT für indirekte Vergleiche“ mit der Tabelle „Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT für indirekte Vergleiche“.

- 4.3.2.1.1.2 Studien aus der bibliografischen Literaturrecherche – Studien für indirekte Vergleiche mit den Flussdiagrammen der entsprechenden Suchen (bitte achten Sie darauf, die Flussdiagramme mit der entsprechenden Beschreibung („Abbildung ...“) zu versehen).

- 4.3.2.1.1.3 Studien aus der Suche in Studienregistern – Studien für indirekte Vergleiche mit der Tabelle „Relevante Studien aus der Suche in Studienregistern – RCT für indirekte Vergleiche“.

- 4.3.2.1.1.4 Resultierender Studienpool – Studien für indirekte Vergleiche mit der Tabelle „Studienpool – RCT für indirekte Vergleiche“.

Bitte reichen Sie fehlende Angaben, sowie ggf. die entsprechenden Abschnitte für Anhang 4-A und Anhang 4-B (s.u.) nach.

Ziel der dreischrittigen Informationsbeschaffung (zur Methodik s. Modul 4.2.3 S. 30) war es, die bestverfügbare Evidenz für die Zusatznutzenbewertung von Ipilimumab im relevanten Anwendungsgebiet sowohl gegenüber DTIC als auch gegenüber Vemurafenib als perspektivisch mögliche zweckmäßige Vergleichstherapie (ZVT, zur Diskussion der ZVT s. Modul 3.1). zu identifizieren.

Dabei wird im ersten Schritt (s. Modul 4.3.1.1, insbesondere Tabelle 4-12 und Schlussfolgerungen, S. 75) ausgeschlossen, dass RCT von Ipilimumab gegenüber DTIC und/oder Vemurafenib vorliegen.

In einem zweiten Schritt wird festgestellt, dass auf Basis von RCT kein indirekter Vergleich von Ipilimumab gegenüber DTIC über einen Brückenkompator möglich ist (s. Modul 4.3.1.1, insbesondere Tabelle 4-13 und Schlussfolgerungen, S. 75ff.).

Im dritten Schritt wird – als bestverfügbare Evidenz für die Zusatznutzenbewertung von Ipilimumab vs. DTIC – für einen Vergleich aller verfügbaren Studiendaten von Ipilimumab im hinzukommenden Indikationsgebiet vs. einer historischen Kontrolle DTIC (bezeichnet als „H2hH“) folgender Studienpool identifiziert: CA184-004, CA184-022, MDX010-008, MDX010-020, CA184-024, CA184-332 und CA184-338 (s. Modul 4.3.1.1, insbesondere Tabelle 4-15 S. 83ff.).

Für den indirekten Vergleich von Ipilimumab gegen Vemurafenib wird dieser H2hH wiederholt als best-verfügbare Evidenz berücksichtigt. Auf eine erneute Beschreibung der Informationsbeschaffung wird daher in Modul 4.3.2.1.1 verzichtet und an dieser Stelle auf Modul 4.2.3 in Verbindung mit 4.3.1.1 verwiesen.

In den unter Abschnitt 4.3.2.1 vorgelegten Recherchen wird daher geprüft, ob für einen indirekten Vergleich von Ipilimumab gegenüber Vemurafenib eine RCT im Anwendungsgebiet, die Vemurafenib mit dem Brückenkomparator DTIC vergleicht, vorliegt. Es kann eine Studie identifiziert werden: BRIM3 (s. Modul 4.3.2.1.1).

In Modul 4.3.2.1 wird dann der indirekte Vergleich von Ipilimumab gegenüber Vemurafenib auf Basis der BRIM3 und der H2hH (auf Basis der Studien CA184-004, CA184-022, MDX010-008, MDX010-020, CA184-024, CA184-332 und CA184-338) vorgelegt.

4.3.2.1.2 Charakteristika der Studien für indirekte Vergleiche

Für Abschnitt 4.3.2.1 werden die Studien CA184-004, CA184-022, MDX010-008, MDX010-020, CA184-024, CA184-332 und CA184-338 des pharmazeutischen Unternehmers herangezogen. Zu diesen werden in Abschnitt 4.3.2.1.2 nicht die entsprechenden Informationen dargestellt bzw. wird nicht auf ggf. in vorherigen Abschnitten dargestellte entsprechende Angaben oder Tabellen verwiesen:

Die Verweise werden nachgereicht.

- *Tabelle „Charakterisierung der eingeschlossenen Studien – RCT für den indirekten Vergleich“.*

Tabelle Z-13: Verweise zur Charakterisierung der eingeschlossenen Studien – RCT für den indirekten Vergleich

	Tabellenbezeichnung (Modul 4)	Seitenzahl (Modul 4)
CA184-004	4-16	88-91
CA184-022	4-16	88-91
CA184-024	4-16	88-91
CA184-332	4-16	88-91
CA184-338	4-16	88-91
MDX010-008	4-16	88-91
MDX010-020	4-16	88-91

- *Tabelle „Charakterisierung der Interventionen – RCT für den indirekten Vergleich“.*

Tabelle Z-14: Verweise zur Charakterisierung der Intervention – RCT für den indirekten Vergleich

	Tabellenbezeichnung (Modul 4)	Seitenzahl (Modul 4)
CA184-004	4-17	93-94
CA184-022	4-17	93-94
CA184-024	4-17	93-94
CA184-332	4-17	93-94
CA184-338	4-17	93-94
MDX010-008	4-17	93-94
MDX010-020	4-17	93-94

- *Tabelle „Charakterisierung der Studienpopulationen – RCT für den indirekten Vergleich“.*

Tabelle Z-15: Verweise zur Charakterisierung der Studienpopulation – RCT für den indirekten Vergleich

	Tabellenbezeichnung (Modul 4)	Seitenzahl (Modul 4)
CA184-004	4-18	95-96
CA184-022	4-19	97-98
CA184-024	4-20	99-100
CA184-332	4-21	101-102
CA184-338	4-22	103-104
MDX010-008	4-23	105-106
MDX010-020	4-24	107-108

Bitte reichen Sie fehlende Angaben nach.

Erfolgt in Zuordnung zum jeweiligen Tabellenblatt.

4.3.2.1.3.1 Mortalität – Tabelle „Ergebnisse für Mortalität aus RCT für indirekte Vergleiche“

In der Tabelle (vorliegend Tabelle 4-67, Seite 170) werden für die Studien CA184-004, CA184-022, MDX010-008 und MDX010-020 nur gepoolte Ergebnisse („Phase II/III“) dargestellt. Bitte reichen Sie die entsprechenden Angaben getrennt nach Studie für jede der Studien CA184-004, CA184-022, MDX010-008 und MDX010-020 als ergänzende Tabelle nach.

Gemäß obiger Ausführungen (s. z.B. die Informationen zu 4.3.1.3.1) wurde keine Propensity Score Analyse auf Basis der Einzelstudien durchgeführt. Daher liegen keine Einzelergebnisse zu diesem Endpunkt für die Studien 004, 022, 008 und 020 vor.

4.3.2.1.3.1 Mortalität – Homogenität der Ergebnisse, gepoolte Ergebnisse (Metaanalyse)

In der Metaanalyse (vorliegend Abbildung 15, Seite 171) werden für die Studien CA184-004, CA184-022, MDX010-008 und MDX010-020 die bereits gepoolten Ergebnisse („Phase II/III“) dargestellt. Bitte reichen Sie zur Untersuchung der Homogenität die Metaanalyse unter separater Darstellung jeder der eingeschlossenen Studien, sowie ggf. Diskussion beobachteter Heterogenität, nach.

Gemäß obiger Ausführungen (s. z.B. die Informationen zu 4.3.1.3.1) wurde keine Propensity Score Analyse auf Basis der Einzelstudien durchgeführt. Daher liegen keine Einzelergebnisse zu diesem Endpunkt für die Studien 004, 022, 008 und 020 vor. Somit kann auch keine Untersuchung auf Heterogenität der Studien erfolgen.

4.3.2.1.3.3 Mortalität – Tabelle „Ergebnisse für Verträglichkeit aus Studien für indirekte Vergleiche“

In der Tabelle (vorliegend Tabellen 4-73, Seite 177) werden für die Studien CA184-004, CA184-022, MDX010-008 und MDX010-020 nur gepoolte Ergebnisse („Phase II/III“) dargestellt. Bitte reichen Sie die entsprechenden Angaben getrennt nach Studie für jede der Studien CA184-004, CA184-022, MDX010-008 und MDX010-020 als ergänzende Tabelle nach.

Gemäß obiger Ausführungen (s. z.B. die Informationen zu 4.3.1.3.1) wurde keine Propensity Score Analyse auf Basis der Einzelstudien durchgeführt. Daher liegen keine Einzelergebnisse zu diesem Endpunkt für die Studien 004, 022, 008 und 020 vor.

4.3.2.1.3.3 Verträglichkeit – Homogenität der Ergebnisse, gepoolte Ergebnisse (Metaanalyse)

In der Metaanalyse (vorliegend Abbildung 18, Seite 179) werden für die Studien CA184-004, CA184-022, MDX010-008 und MDX010-020 die bereits gepoolten Ergebnisse („Phase II/III“) dargestellt. Bitte reichen Sie zur Untersuchung der Homogenität die Metaanalyse unter separater Darstellung jeder der eingeschlossenen Studien, sowie ggf. Diskussion beobachteter Heterogenität, nach.

Gemäß obiger Ausführungen (s. z.B. die Informationen zu 4.3.1.3.1) wurde keine Propensity Score Analyse auf Basis der Einzelstudien durchgeführt. Daher liegen keine Einzelergebnisse zu diesem Endpunkt für die Studien 004, 022, 008 und 020 vor. Somit kann auch keine Untersuchung auf Heterogenität der Studien erfolgen.

4.3.2.3 Weitere Untersuchungen

Allgemeine Erläuterung zu den nachzureichenden Angaben für Modul 4.3.2.3:

B-MS hat in Modul 4.3.2.3 des am 29.11.2013 eingereichten Nutzendossiers Ergebnisse einer einarmigen, auf individuellen Patientendaten beruhenden Metaanalyse vorgelegt, die im Rahmen des Europäischen Krebs Kongresses 2013 (European Cancer Congress 2013) vorgestellt wurden.

Dabei handelt es sich um eine Subgruppenanalyse der Patienten, die Ipilimumab in der in Deutschland zugelassenen Dosierung von 3mg (Induktionsregime mit einer Ipilimumabgabe alle drei Wochen) erhalten hatten.

Diese Vorlage erfolgte in Ergänzung zur Zusatznutzenbewertung von Ipilimumab gegenüber DTIC bzw. Vemurafenib, die in den Modulen 4.3.1 sowie 4.3.2.1 geführt wurden mit dem Ziel, diese Evidenz bzgl. des langfristigen Therapieerfolgs von Ipilimumab zu teilen und somit auch Transparenz über weitere im hinzukommenden Indikationsgebiet publizierte Daten zu geben.

Die Darstellung erfolgte in Konsistenz mit den kongresspublizierten Daten und damit nicht gemäß der Anforderungen zur Informationsbeschaffung der Verfahrensordnung. B-MS hat diese Daten daher nicht in der Herleitung des Zusatznutzens verwendet. Im Folgenden werden keine weiteren Angaben zu dieser Untersuchung nachgereicht.

4.3.2.3.1 Ergebnis der Informationsbeschaffung – Weitere Untersuchungen

Für Abschnitt 4.3.2.3 werden die Studien CA184-004, CA184-022, CA184-045, CA184-332, CA184-338 und MDX010-20 des pharmazeutischen Unternehmers, sowie die Studien 02C0106, 03C0109 und 04C0083 herangezogen. Diese ergeben sich nicht aus Abschnitt 4.3.2.3.1:

- *4.3.2.3.1.1 „Studien des pharmazeutischen Unternehmers – Weitere Untersuchungen mit Tabelle „Liste der Studien des pharmazeutischen Unternehmers – Weitere Untersuchungen“ mit der Tabelle „Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – Weitere Untersuchungen“.*
- *[optional: 4.3.2.3.1.2 Studien aus der bibliografischen Literaturrecherche – Weitere Untersuchungen mit den Flussdiagrammen der entsprechenden Suchen (bitte achten Sie darauf, die Flussdiagramme mit der entsprechenden Beschreibung („Abbildung ...“) zu versehen)].*
- *4.3.2.3.1.3 Studien aus der Suche in Studienregistern – Weitere Untersuchungen mit der Tabelle „Relevante Studien aus der Suche in Studienregistern – Weitere Untersuchungen“.*
- *4.3.2.3.1.4 Resultierender Studienpool – Weitere Untersuchungen mit der Tabelle „Studienpool – Weitere Untersuchungen“.*

Es wurde keine Suche in Studienregistern durchgeführt, sowie die Recherche für diesen Abschnitt nicht in Anhang 4-B hinterlegt. Eine Bewertung der Vollständigkeit des Studienpools ist nicht möglich.

Bitte reichen Sie fehlende Angaben, sowie den entsprechenden Abschnitt für Anhang 4-B (s.u.) nach.

Weitere Angaben (ergänzend zu den Angaben in Modul 4.3.2.3.1, S. 184) zur Informationsbeschaffung werden nicht aufgeführt, da die in 4.3.2.3 vorgestellte Untersuchung zur Evidenz des Langzeitüberlebens unter Ipilimumab nicht in die Zusatznutzenbewertung einbezogen wurde.

4.3.2.3.2 Charakteristika der weiteren Untersuchungen

Für Abschnitt 4.3.2.3 werden neben den Studien CA184-004, CA184-022, CA184-045, CA184-332, CA184-338 und MDX010-20 (mit Verweis auf Abschnitt 4.3.1) die Studie CA184-045 des pharmazeutischen Unternehmers, sowie die Studien 02C0106, 03C0109 und 04C0083 herangezogen. Zu den Studien CA184-045, 02C0106, 03C0109, 04C0083 werden in Abschnitt 4.3.2.3.2 nicht die entsprechenden Informationen dargestellt:

- *Tabelle „Charakterisierung der eingeschlossenen Studien – Weitere Untersuchungen“.*
- *Tabelle „Charakterisierung der Interventionen – Weitere Untersuchungen“.*
- *Tabelle „Charakterisierung der Studienpopulationen – Weitere Untersuchungen“.*

Bitte reichen Sie fehlende Angaben, sowie den entsprechenden Abschnitt für Anhang 4-F und Anhang 4-G (s.u.) nach.

Weitere Angaben (ergänzend zu den Angaben in Modul 4.3.2.3.1, S. 184f.) zu den Charakteristika werden nicht aufgeführt, da die in 4.3.2.3 vorgestellte Untersuchung zur Evidenz des Langzeitüberlebens unter Ipilimumab nicht in die Zusatznutzenbewertung einbezogen wurde.

Anhang A

Anhang A fehlt ggf. für Abschnitt „4.3.2.1 Studien für indirekte Vergleiche“.

Bitte reichen Sie eine entsprechende bibliographische Literaturrecherche für Abschnitt 4.3.2.1 für Ipilimumab oder eine entsprechende Begründung nach.

In Anhang 4A wurde eine Zweiteilung vorgenommen in „Bibliographische Literaturrecherche für Ipilimumab“ und „Bibliographische Literaturrecherche für den indirekten Vergleich (ZVT)“.

Insofern wird die bibliographische Literaturrecherche für den Abschnitt „4.3.2.1 Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien“ bereits berichtet, s. auch Antwort zur Nachforderung 4.3.2.1.1 dieses Dokuments.

Anhang B***Anhang B fehlt für Abschnitt „4.3.2.3 Weitere Untersuchungen“.***

Bitte reichen Sie eine entsprechende Suche in Studienregistern für Abschnitt 4.3.2.3 nach.

Die Notwendigkeit einer Suche in Studienregistern für Abschnitt 4.3.2.3 entfällt, da die in 4.3.2.3 vorgestellte Untersuchung zur Evidenz des Langzeitüberlebens unter Ipilimumab nicht in die Zusatznutzenbewertung einbezogen wurde (s. auch die allgemeinen Erläuterungen unter 4.3.2.3 dieses Dokumentes).

Die nachfolgend unter Anhang 4B aufgeführten Nachforderungen zu den Studien CA184-004, CA184-024, CA184-332, CA184-338 und MDX010-08 werden unseres Erachtens versehentlich doppelt genannt, diese beziehen sich auf Anhang 4F und werden daher ausschließlich unter dem entsprechenden Abschnitt dieses Dokuments berücksichtigt.

Studie CA184-004:

- *Angaben zu den Items 7 b, 8 a, 9, 10, 11 b und 14 b fehlen.*

Bitte reichen Sie die fehlenden Angaben nach.

Studie CA184-024:

- *Angaben zu den Items 9, 10, 11 b und 14 b fehlen.*

Bitte reichen Sie die fehlenden Angaben nach.

Studie CA184-332:

- *Flow-Chart (Patientenfluss) fehlt (S. 303). Bitte reichen Sie das entsprechende Flow-Chart nach.*

Studie CA184-338:

- *Angaben zu den Items 4 b, 11 b und 12 b fehlen.*

Bitte reichen Sie die fehlenden Angaben nach.

- *Flow-Chart (Patientenfluss) fehlt (S. 308). Bitte reichen Sie das entsprechende Flow-Chart nach.*

Studie MDX010-08:

- *Angaben zu den Items 7 b, 9, 11 b, 12 und 14 b fehlen.*

Bitte reichen Sie die fehlenden Angaben nach.

Anhang FGrundsätzliche Erläuterung zur Umsetzung der Nachforderung:

Es werden nur die zusätzlichen Angaben gemäß Nachforderungen aufgeführt, für die vollständigen Tabellen wird auf das Dossier verwiesen.

Die fehlenden Angaben wurden, sofern zutreffend, aus den bereits eingereichten CSR bzw. dessen vorliegende Anlagen unter Angabe der Seitenzahl entnommen. Für Informationen, die in diesen Berichten nicht hinterlegt sind, wurden B-MS intern zur Verfügung gestellte Informationen ergänzt bzw. eine entsprechendes Nicht-Zutreffen vermerkt.

Anhang F fehlt für die Studien CA184-045, 02C0106, 03C0109, 04C0083.

Bitte reichen Sie „Tabelle 4-xx (Anhang): Studiendesign und -methodik“ für die Studien CA184-045, 02C0106, 03C0109 und 04C0083 nach.

Die unter Anhang 4F nachgeforderten Unterlagen zu den Studien CA184-045, 02C0106, 03C019 und 04C0083 werden nicht aufgeführt, da diese Studien nicht in die Nutzenbewertung einbezogen wurden.

Studie CA184-004:

- *Angaben zu den Items 7 b, 8 a, 9, 10, 11 b und 14 b fehlen.*

Bitte reichen Sie die fehlenden Angaben nach.

Tabelle Z-16 [4-100 (Anhang)]: Studiendesign und -methodik für Studie CA184-004

Item ^a	Charakteristikum	Studieninformation
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	Laut Studienprotokoll der Studie CA184-004 war einzig eine pharmakokinetische (pk) Zwischenanalyse der Serumkonzentration Ipilimumab mit 125-150 Patienten der Studien CA184-004, CA184-007 und CA184-008 geplant (CSR bzw. Amendment: S. 936, 1110, 3334). Dafür sollte eine Datenbanksperre bezüglich Dosierungsdauer, Anzahl der erhaltenen Dosen, Probenahmezeiten sowie verschiedener Kovariablen einschließlich aber nicht ausschließlich Alter, Geschlecht und Gewicht vorgenommen werden. Gemäß B-MS interner Informationen war diese Zwischenanalyse vorgesehen für den Fall einer frühen regulatorischen Einreichung (2008). Da die Einreichung aber erst 2010 erfolgte, sind keine pk-Zwischenanalysen sondern nur eine Gesamt-pk-Analyse der oben genannten Studien durchgeführt worden, die separat berichtet wurden.
8a	Methode zur Generierung der zufälligen Zuteilung	Die Patienten wurden über eine zentrale Randomisierungsstelle unter Beachtung einer vorangegangenen Immuntherapie den Studienarmen zugeteilt (CSR S. 55f.).
9	Randomisierung, Geheimhaltung der Behandlungsfolge (allocation concealment) Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war	Die Zuteilung wurde über einen interaktiven Sprachserver (IVRS) durchgeführt. Über das IVRS wird der Studieneintritt des Patienten registriert und eine individuelle Identifikationsnummer zugeteilt (CSR S. 55f.).
10	Randomisierung, Durchführung Wer führte die Zuteilung durch, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?	Sobald ein Patient als geeignet befunden wurde, rief das jeweilige Studienzentrum bei der zentralen Randomisierungsstelle an, um eine Gruppenzuteilung für den Patienten zu erhalten. Die Einteilung erfolgte zufällig in einem 1:1-Verhältnis auf die beiden Studienarme. Stratifiziert wurde nach vorangegangener Immuntherapie (ja/nein) (CSR S. 55f.).
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	Die Studienarme unterschieden sich lediglich in der Dosierung der Studienmedikation. In beiden Studienarmen wurde Ipilimumab in vier separaten Dosen in denselben Zeitintervallen von drei Wochen (Woche 1, 4, 7 und 10) verabreicht. Die Studienmedikation wurde in 10-ml-Ampullen als klare, farblose Flüssigkeit dargereicht (CSR S. 55).
14b	Informationen, warum die Studie endete oder beendet wurde	Die Studie wurde gemäß dem Studienprotokoll beendet.

Item ^a	Charakteristikum	Studieninformation
a: nach CONSORT 2010.		

Studie CA184-024:

- Angaben zu den Items 9, 10, 11 b und 14 b fehlen. Reichen Sie die fehlenden Angaben nach.

Tabelle Z-17 [4-102 (Anhang)]: Studiendesign und -methodik für Studie CA184-024

Item ^a	Charakteristikum	Studieninformation
9	Randomisierung, Geheimhaltung der Behandlungsfolge (allocation concealment) Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war	Randomisierung über ein zentralisiertes Randomisierungsschema (CSR S. 57). Die Patienten wurden in einem Verhältnis von 1:1 den Studienarmen zugeteilt. Stratifiziert wurde nach: M-Stadium zu Studienbeginn (M0 vs. M1a vs. M1b vs. M1c), ECOG-Performance-Status wie bei der Randomisierung erfasst (0 vs. 1) und Studienzentrum.
10	Randomisierung, Durchführung Wer führte die Zuteilung durch, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?	Laut Studienprotokoll ist das jeweilige Studienzentrum dazu angewiesen, ein Interaktives Voice-Response-System anzurufen, um eine Patientenummer zu erhalten. Um den Patienten anzumelden, werden Daten zur Geburt, Geschlecht und Datum der Einwilligungserklärung benötigt. Zusätzliche Informationen zur Anmeldung eines Patienten werden in einem separaten IVRS-Handbuch bereit gestellt (p.2075 Studienprotokoll).
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	DTIC-Administration war in beiden Studienarmen identisch (CSR S. 57). Placebo und Ipilimumab wurden in 90minütigen Infusionen als Kochsalzlösung verabreicht.
14b	Informationen, warum die Studie endete oder beendet wurde	Die Studie wurde gemäß dem Studienprotokoll beendet.
a: nach CONSORT 2010.		

Studie CA184-332:

- Flow-Chart (Patientenfluss) fehlt (S. 303). Bitte reichen Sie das entsprechende Flow-Chart nach.

Da es sich um Beobachtungsstudie handelt, wird kein Flow-Chart (Patientfluss) vorgelegt.

Studie CA184-338:

- Angaben zu den Items 4 b, 11 b und 12 b fehlen. Bitte reichen Sie die fehlenden Angaben nach.

- *Flow-Chart (Patientenfluss) fehlt (S. 308). Bitte reichen Sie das entsprechende Flow-Chart nach.*

Tabelle Z-18 [Tabelle 4-104 (Anhang)]: Studiendesign und -methodik für Studie <CA184-338>

Item ^a	Charakteristikum	Studieninformation
4b	Studienorganisation und Ort der Studiendurchführung	Gemäß CSR (S. 6 und S. 33) wurden die Daten retrospektiv von Patienten gesammelt, die zwischen dem 01. April 2011 und dem 30. September 2012 bei einer Zytokin-Arbeitsgruppe oder einem angeschlossenen Institut die Behandlung mit 3 mg/kg Ipilimumab begonnen haben. Die Institute befinden sich in den Vereinigten Staaten.
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	Nicht relevant, da es sich um eine einarmige retrospektive Beobachtungsstudie handelt.
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	Durch eine zusätzliche Analyse soll die Inzidenzrate aller mit der Behandlung in Zusammenhang stehenden unerwünschten Ereignisse bestimmt werden (CSR S. 43).
a: nach CONSORT 2010.		

Da es sich um Beobachtungsstudie handelt, wird kein Flow-Chart (Patientfluss) vorgelegt.

Studie MDX010-08:

- *Angaben zu den Items 7 b, 9, 11 b, 12 und 14 b fehlen.
Bitte reichen Sie die fehlenden Angaben nach.*

Tabelle Z-19 [Tabelle 4-105 (Anhang)]: Studiendesign und -methodik für Studie MDX010-08

Item ^a	Charakteristikum	Studieninformation
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	Laut Studienprotokoll soll eine erste Überprüfung der Daten stattfinden, sobald alle Patienten 12 Monate an der Studie teilnehmen oder sobald alle Patienten die Studie beendet haben.
9	Randomisierung, Geheimhaltung der Behandlungsfolge (allocation concealment) Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war	Die Randomisierung erfolgte über den Sponsor der Studie (S.35). Laut Amendment 2 des Studienprotokolls (S.1224) vergibt der Projektleiter nach Abgabe der Einverständniserklärung an jeden Patienten eine Screeningnummer. Geeignete Patienten werden einer von zwei Behandlungsgruppen zugeteilt. Die Biostatistikgruppe von Medarex stellt eine Randomisierungsliste bereit. Um die Behandlungszuteilung auszubalancieren und den Selektionsbias zu reduzieren, wird die Randomisierung innerhalb jeder Studie statifiziert. Hierzu wird eine Randomisierungsblockgröße von 2 bis 4 verwendet. Die Randomisierungsliste enthält das Studienzentrum, die Patientenummer und die Behandlungsgruppe. Die Biostatistikgruppe wird eine allgemeine Randomisierungsliste bereit stellen und eine Liste je Studienzentrum.
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	Keine Verblindung, daher nicht relevant.
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	Keine weiteren; s. auch Angaben zu den sekundären Endpunkten bzgl. weiterer Analysen
14b	Informationen, warum die Studie endete oder beendet wurde	Die Studie wurde gemäß dem Studienprotokoll beendet.
a: nach CONSORT 2010.		

Anhang G

Anhang G fehlt für die Studien CA184-045, 02C0106, 03C0109, 04C0083.

Bitte reichen Sie die ausgefüllten Bewertungsbögen zur Einschätzung von Verzerrungsaspekten für die Studien CA184-045, 02C0106, 03C0109 und 04C0083 nach.

Die unter Anhang 4G nachgeforderten Unterlagen zu den Studien CA184-045, 02C0106, 03C0109 und 04C0083 werden nicht aufgeführt, da diese Studien nicht in die Nutzenbewertung einbezogen wurden.

Anhang G ist für jede eingeschlossene Studie separat auszufüllen (Abschnitt A: Studienebene, Abschnitt B: Nr. 1 bis 4 jeweils für jeden relevanten Endpunkt).

*Bitte reichen Sie für jede der unter dem Kürzel „**Phase II/III**“ zusammengefassten Studien (Studie CA 184-004, CA184-022, MDX010-008 und MDX010-020) einen entsprechend ausgefüllten separaten Bewertungsbogen zur Einschätzung von Verzerrungsaspekten nach.*

Bitte beachten Sie hierbei jeweils die für den Bewertungsbogen der gepoolten Studien „Phase II/ III“ identifizierte Mängel.

Studien „Phase II/III“ (S. 336 bis 347):

- *A: Einstufung des Verzerrungspotenzials auf Studienebene (S. 336): Angaben zum Kriterium fehlen (obligat). Bitte machen sie jeweils Angaben zum Kriterium.*

- *B: Endpunkt Mortalität (S. 343): Nr. 3 nicht angekreuzt.*

Bitte kreuzen Sie jeweils das Zutreffende an und machen ggf. Angaben zum Kriterium.

- *B: Endpunkt Verträglichkeit (S. 346): Nr. 3 fälschlicherweise als Nr. 7 benannt und nicht angekreuzt.*

Bitte kreuzen Sie jeweils das Zutreffende an und machen ggf. Angaben zum Kriterium.

Tabelle Z-20 (Anhang 4-G): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie "CA184-004"

Studie: CA184-004

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genauere Benennung der Quelle	Kürzel
Final Clinical Study Report for Study CA184004: An Exploratory Study to Determine Potential Predictive Markers of Response and/or Toxicity in Patients with Unresectable Stage III or IV Malignant Melanoma Randomized and Treated with Ipilimumab (MDX-010/BMS-734016) at Two Dose	CA184-004

Kriterien zur Bewertung des Verzerrungspotenzials der Ergebnisse:

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A: Kriterien des Verzerrungspotenzials der Ergebnisse auf Studienebene:

Einstufung als randomisierte Studie

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anhaltspunkte vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

Die Studie wurde randomisiert durchgeführt. Es liegen keine Hinweise vor, die dem widersprechen.

1.

für randomisierte Studien:

Adäquate Erzeugung der Randomisierungssequenz

ja: Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

unklar: Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

nein: Die Erzeugung der Zuteilungssequenz war nicht adäquat.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:

Zeitliche Parallelität der Gruppen

ja: Die Gruppen wurden zeitlich parallel verfolgt.

unklar: Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

nein: Die Gruppen wurden nicht zeitlich parallel verfolgt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien:

Verdeckung der Gruppenzuteilung („allocation concealment“)

ja: Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierten Arzneimitteln/Arzneimittelbehältern
- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

unklar: Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau.

nein: Die Gruppenzuteilung erfolgte nicht verdeckt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Studien wurden randomisiert und die Zuteilung erfolgte verdeckt, allerdings wurde nur ein Arm der Studie in die Bewertung aufgenommen.

für nicht randomisierte Studien:

Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren

ja: Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anhaltspunkte dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

unklar: Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau.

nein: Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und Behandlern

Patient

ja: Die Patienten waren verblindet.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Die Studien wurden verblindet für die Patienten durchgeführt, allerdings wurde nur ein Arm der Studie in die Bewertung aufgenommen.

Behandler bzw. Weiterbehandler

ja: Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, den primären Behandler (Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Behandler nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).

Beispiele zu a und b:

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter

- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:

- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*
- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.*
Zulässige Gründe sind:
 - *erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
 - *Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*
 - *geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben*
- *Prüfen, ob statistisch nicht signifikante Ergebnisse weniger ausführlich dargestellt sind.*
- *Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.*

Anzumerken ist, dass Anhaltspunkte für eine ergebnisgesteuerte Darstellung eines Endpunkts das Verzerrungspotenzial für die Ergebnisse der übrigen Endpunkte erhöhen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anhaltspunkten dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, ist das Verzerrungspotenzial für die anderen Endpunkte u. U. erhöht. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Erhöhung des Verzerrungspotenzials der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen.

Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und das Verzerrungspotenzial anderer Endpunkte nicht berührt.

- ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.
- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Daten der Studien wurden nach a-priori spezifiziertem statistischen Analyseplan berichtet, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

5. Fehlen sonstiger (endpunktübergreifender) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien
- intransparenter Patientenfluss
- Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:
 - Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).
 - Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (ansonsten ggf. im Nachhinein von der Biometrie durchzuführen).
 - Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.
 - Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studien in die Bewertung aufgenommen.

Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Die Studien an sich wurden qualitativ hochwertig als doppel-blinde, randomisierte, multizentrische Studien durchgeführt. Allerdings wurden nur

Patienten mit Behandlung mit Ipilimumab im neuen Anwendungsgebiet selektiert. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

B: Kriterien des Verzerrungspotenzials der Ergebnisse pro Endpunkt

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: Mortalität

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass

diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studien basierten auf der ITT Population, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

- ja**
- nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studie in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

- niedrig:** Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.
- hoch:** Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Endpunkt: gesundheitsbezogene Lebensqualität

Es wurden keine Daten zur gesundheitsbezogenen Lebensqualität in der Studie erhoben. Daher erfolgte keine Bewertung des Verzerrungspotenzials für diesen Endpunkt.

Endpunkt: Verträglichkeit**1. Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen

Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studien basierten auf der Sicherheitspopulation, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen*z. B.*

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

 ja **nein**Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studien in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Tabelle Z-21 (Anhang 4-G): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie "CA184-022"

Studie: CA184-022

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genauere Benennung der Quelle	Kürzel
FINAL CLINICAL STUDY REPORT FOR CA184022: A Randomized, Double-Blind, Multi-center, Phase II Fixed Dose Study of Multiple Doses of Ipilimumab (MDX-010) Monotherapy in Patients with Previously Treated Unresectable Stage III or IV Melanoma	CA184-022

Kriterien zur Bewertung des Verzerrungspotenzials der Ergebnisse:

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A: Kriterien des Verzerrungspotenzials der Ergebnisse auf Studienebene:

Einstufung als randomisierte Studie

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anhaltspunkte vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

Die Studie wurde randomisiert durchgeführt. Es liegen keine Hinweise vor, die dem widersprechen.

1.

für randomisierte Studien:**Adäquate Erzeugung der Randomisierungssequenz**

ja: Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

unklar: Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

nein: Die Erzeugung der Zuteilungssequenz war nicht adäquat.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:**Zeitliche Parallelität der Gruppen**

ja: Die Gruppen wurden zeitlich parallel verfolgt.

unklar: Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

nein: Die Gruppen wurden nicht zeitlich parallel verfolgt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien:**Verdeckung der Gruppenzuteilung („allocation concealment“)**

ja: Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierten Arzneimitteln/Arzneimittelbehältern

- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

unklar: Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau.

nein: Die Gruppenzuteilung erfolgte nicht verdeckt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Studien wurden randomisiert und die Zuteilung erfolgte verdeckt, allerdings wurde nur ein Arm der Studie in die Bewertung aufgenommen.

für nicht randomisierte Studien:

Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren

ja: Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anhaltspunkte dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

unklar: Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau.

nein: Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und Behandlern

Patient

ja: Die Patienten waren verblindet.

- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Die Studien wurden verblindet für die Patienten durchgeführt, allerdings wurde nur ein Arm der Studie in die Bewertung aufgenommen.

Behandler bzw. Weiterbehandler

ja: Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, den primären Behandler (Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass die Behandler nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).

Beispiele zu a und b:

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter

- *Subgruppen*

- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:

- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*
- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.
Zulässige Gründe sind:*
 - *erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
 - *Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*
 - *geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben*
- *Prüfen, ob statistisch nicht signifikante Ergebnisse weniger ausführlich dargestellt sind.*
- *Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.*

Anzumerken ist, dass Anhaltspunkte für eine ergebnisgesteuerte Darstellung eines Endpunkts das Verzerrungspotenzial für die Ergebnisse der übrigen Endpunkte erhöhen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anhaltspunkten dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, ist das Verzerrungspotenzial für die anderen Endpunkte u. U. erhöht. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Erhöhung des Verzerrungspotenzials der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen.

Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und das Verzerrungspotenzial anderer Endpunkte nicht berührt.

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Daten der Studien wurden nach a-priori spezifiziertem statistischen Analyseplan berichtet, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

5. Fehlen sonstiger (endpunktübergreifender) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien*
- *intransparenter Patientenfluss*
- *Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:*
 - *Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).*
 - *Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (ansonsten ggf. im Nachhinein von der Biometrie durchzuführen).*
 - *Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.*
 - *Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studien in die Bewertung aufgenommen.

Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Die Studien an sich wurden qualitativ hochwertig als doppel-blinde, randomisierte, multizentrische Studien durchgeführt. Allerdings wurden nur Patienten mit Behandlung mit Ipilimumab im neuen Anwendungsgebiet selektiert. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

B: Kriterien des Verzerrungspotenzials der Ergebnisse pro Endpunkt

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: Mortalität

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten

ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studien basierten auf der ITT Population, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen*z. B.*

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

 ja **nein**Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studie in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Endpunkt: gesundheitsbezogene Lebensqualität**1. Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante

Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.

- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studien basierten auf der Population der Patienten, die den Fragebogen ausgefüllt haben. Es wurden keine fehlenden Werte ergänzt. Zudem wurde nur ein Arm der Studien in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studie in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Die Analysen der Studien basierten auf der Population der Patienten, die den Fragebogen ausgefüllt haben. Es wurden keine fehlenden Werte ergänzt. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Endpunkt: Verträglichkeit

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).

- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studien basierten auf der Sicherheitspopulation, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studien in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Tabelle Z-22: (Anhang 4-G): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie "MDX010-08"

Studie: MDX010-08

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genauere Benennung der Quelle	Kürzel
Final Clinical Study Report for Study MDX010-08: A RANDOMIZED STUDY COMPARING MDX-010 ALONE OR IN COMBINATION WITH DTIC IN THE TREATMENT OF PATIENTS WITH CHEMOTHERAPY NAÏVE METASTATIC MELANOMA	MDX010-08

Kriterien zur Bewertung des Verzerrungspotenzials der Ergebnisse:

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A: Kriterien des Verzerrungspotenzials der Ergebnisse auf Studienebene:

Einstufung als randomisierte Studie

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anhaltspunkte vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

Die Studie wurde randomisiert durchgeführt. Es liegen keine Hinweise vor, die dem widersprechen.

1.

für randomisierte Studien:**Adäquate Erzeugung der Randomisierungssequenz**

- ja:** Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).
- unklar:** Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.
- nein:** Die Erzeugung der Zuteilungssequenz war nicht adäquat.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:**Zeitliche Parallelität der Gruppen**

- ja:** Die Gruppen wurden zeitlich parallel verfolgt.
- unklar:** Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.
- nein:** Die Gruppen wurden nicht zeitlich parallel verfolgt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien:**Verdeckung der Gruppenzuteilung („allocation concealment“)**

- ja:** Eines der folgenden Merkmale trifft zu:
- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
 - Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierte Arzneimitteln/Arzneimittelbehältern
 - Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

unklar: Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau.

nein: Die Gruppenzuteilung erfolgte nicht verdeckt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Studien wurden randomisiert und die Zuteilung erfolgte verdeckt, allerdings wurde nur ein Arm der Studie in die Bewertung aufgenommen.

für nicht randomisierte Studien:

Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren

ja: Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anhaltspunkte dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

unklar: Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau.

nein: Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und Behandlern

Patient

ja: Die Patienten waren verblindet.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Die Studien wurden verblindet für die Patienten durchgeführt, allerdings wurde nur ein Arm der Studie in die Bewertung aufgenommen.

Behandler bzw. Weiterbehandler

ja: Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, den primären Behandler (Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Behandler nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).

Beispiele zu a und b:

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter

- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*

- Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)
- Cut-off-points bei Dichotomisierung
- statistischer Verfahren

Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:

- Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).
- Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.
Zulässige Gründe sind:
 - erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung
 - Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe
 - geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben
- Prüfen, ob statistisch nicht signifikante Ergebnisse weniger ausführlich dargestellt sind.
- Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.

Anzumerken ist, dass Anhaltspunkte für eine ergebnisgesteuerte Darstellung eines Endpunkts das Verzerrungspotenzial für die Ergebnisse der übrigen Endpunkte erhöhen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anhaltspunkten dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, ist das Verzerrungspotenzial für die anderen Endpunkte u. U. erhöht. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Erhöhung des Verzerrungspotenzials der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen.

Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und das Verzerrungspotenzial anderer Endpunkte nicht berührt.

- ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.
- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Daten der Studien wurden nach a-priori spezifiziertem statistischen Analyseplan berichtet, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

5. Fehlen sonstiger (endpunktübergreifender) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien
- intransparenter Patientenfluss
- Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:
 - Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).
 - Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (ansonsten ggf. im Nachhinein von der Biometrie durchzuführen).
 - Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.
 - Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studien in die Bewertung aufgenommen.

Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Die Studien an sich wurden qualitativ hochwertig als doppel-blinde, randomisierte, multizentrische Studien durchgeführt. Allerdings wurden nur Patienten mit Behandlung mit Ipilimumab im neuen Anwendungsgebiet selektiert. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

B: Kriterien des Verzerrungspotenzials der Ergebnisse pro Endpunkt

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: Mortalität

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten

ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studien basierten auf der ITT Population, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen*z. B.*

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

 ja **nein**Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studie in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Endpunkt: _gesundheitsbezogene Lebensqualität_____

Es wurden keine Daten zur gesundheitsbezogenen Lebensqualität in der Studie erhoben. Daher erfolgte keine Bewertung des Verzerrungspotenzials für diesen Endpunkt.

Endpunkt: _Verträglichkeit_____**1. Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein

Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studien basierten auf der Sicherheitspopulation, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen
- unplausible Angaben
- Anwendung inadäquater statistischer Verfahren

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studien in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Tabelle Z-23 (Anhang 4-G): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie "MDX010-020"

Studie: MDX010-020

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genauere Benennung der Quelle	Kürzel
Clinical Study Report for Study MDX010-20 A Randomized, Double-blind, Multicenter Study Comparing MDX-010 Monotherapy, MDX-010 in Combination with a Melanoma Peptide Vaccine, and Melanoma Vaccine Monotherapy in HLA-A*0201-Positive Patients with Previously Treated Unresectable Stage III or IV Melanoma	MDX010-020

Kriterien zur Bewertung des Verzerrungspotenzials der Ergebnisse:

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A: Kriterien des Verzerrungspotenzials der Ergebnisse auf Studienebene:

Einstufung als randomisierte Studie

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anhaltspunkte vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

Die Studie wurde randomisiert durchgeführt. Es liegen keine Hinweise vor, die dem widersprechen.

1.

für randomisierte Studien:**Adäquate Erzeugung der Randomisierungssequenz**

ja: Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

unklar: Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

nein: Die Erzeugung der Zuteilungssequenz war nicht adäquat.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:**Zeitliche Parallelität der Gruppen**

ja: Die Gruppen wurden zeitlich parallel verfolgt.

unklar: Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

nein: Die Gruppen wurden nicht zeitlich parallel verfolgt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien:**Verdeckung der Gruppenzuteilung („allocation concealment“)**

ja: Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierten Arzneimitteln/Arzneimittelbehältern

- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

unklar: Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau.

nein: Die Gruppenzuteilung erfolgte nicht verdeckt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Studien wurden randomisiert und die Zuteilung erfolgte verdeckt, allerdings wurde nur ein Arm der Studie in die Bewertung aufgenommen.

für nicht randomisierte Studien:

Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren

ja: Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anhaltspunkte dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

unklar: Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau.

nein: Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und Behandlern

Patient

ja: Die Patienten waren verblindet.

- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Die Studien wurden verblindet für die Patienten durchgeführt, allerdings wurde nur ein Arm der Studie in die Bewertung aufgenommen.

Behandler bzw. Weiterbehandler

ja: Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, den primären Behandler (Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass die Behandler nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).

Beispiele zu a und b:

- Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter

- Subgruppen*

- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:

- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*
- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.*

Zulässige Gründe sind:

- 2. erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
 - 3. Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*
- *geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben*
 - *Prüfen, ob statistisch nicht signifikante Ergebnisse weniger ausführlich dargestellt sind.*
 - *Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.*

Anzumerken ist, dass Anhaltspunkte für eine ergebnisgesteuerte Darstellung eines Endpunkts das Verzerrungspotenzial für die Ergebnisse der übrigen Endpunkte erhöhen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anhaltspunkten dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, ist das Verzerrungspotenzial für die anderen Endpunkte u. U. erhöht. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Erhöhung des Verzerrungspotenzials der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen.

Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und das Verzerrungspotenzial anderer Endpunkte nicht berührt.

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Daten der Studien wurden nach a-priori spezifiziertem statistischen Analyseplan berichtet, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

5. Fehlen sonstiger (endpunktübergreifender) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien*
- *intransparenter Patientenfluss*
- *Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:*
 - *Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).*
 - *Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (ansonsten ggf. im Nachhinein von der Biometrie durchzuführen).*
 - *Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.*
 - *Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studien in die Bewertung aufgenommen.

Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Die Studien an sich wurden qualitativ hochwertig als doppel-blinde, randomisierte, multizentrische Studien durchgeführt. Allerdings wurden nur Patienten mit Behandlung mit Ipilimumab im neuen Anwendungsgebiet selektiert. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

B: Kriterien des Verzerrungspotenzials der Ergebnisse pro Endpunkt

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: Mortalität

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten

ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studien basierten auf der ITT Population, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen*z. B.*

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

 ja **nein**Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studie in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Endpunkt: gesundheitsbezogene Lebensqualität**1. Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante

Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.

- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studien basierten auf der Population der Patienten, die den Fragebogen ausgefüllt haben. Es wurden keine fehlenden Werte ergänzt. Zudem wurde nur ein Arm der Studien in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studie in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Die Analysen der Studien basierten auf der Population der Patienten, die den Fragebogen ausgefüllt haben. Es wurden keine fehlenden Werte ergänzt. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Endpunkt: Verträglichkeit

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugewiesene Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien,

statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studien basierten auf der Sicherheitspopulation, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studien in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Die Vergleichsgruppe für die Bewertung des Zusatznutzens war eine historische Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Studie CA184-024:

- A: Einstufung des Verzerrungspotenzials auf Studienebene (S. 349): Angaben zum Kriterium fehlen (obligat). Bitte reichen Sie die Angaben zum Kriterium nach.
- B: Endpunkt Mortalität (S. 356): Nr. 3 fälschlicherweise als Nr. 11 benannt und nicht angekreuzt. Bitte kreuzen Sie das zutreffende an und machen ggf. Angaben zum Kriterium.
- B: Endpunkt Verträglichkeit (S. 359): Nr. 3 fälschlicherweise als Nr. 15 benannt und nicht angekreuzt. Bitte kreuzen Sie das zutreffende an und machen ggf. Angaben zum Kriterium.

Tabelle Z-24 (Anhang 4-G): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie "CA184-024"

Studie: CA184-024**Tabelle: Liste der für die Bewertung herangezogenen Quellen**

Genauere Benennung der Quelle	Kürzel
Final Clinical Study Report for Study CA184024 A Multi-Center, Randomized, Double-Blind, Two-Arm, Phase III Study in Patients with Untreated Stage III (Unresectable) or IV Melanoma Receiving Dacarbazine Plus 10 mg/kg of Ipilimumab (MDX-010) vs. Dacarbazine with Placebo	Studie 024

Kriterien zur Bewertung des Verzerrungspotenzials der Ergebnisse:

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A: Kriterien des Verzerrungspotenzials der Ergebnisse auf Studienebene:**Einstufung als randomisierte Studie**

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anhaltspunkte vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

Die Studie wurde randomisiert durchgeführt. Es liegen keine Hinweise vor, die dem widersprechen.

B: Kriterien des Verzerrungspotenzials der Ergebnisse pro Endpunkt

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: Mortalität

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studie wurde verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studie in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei

einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studie basierten auf der ITT Population, allerdings wurde nur ein Arm der Studie in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studie in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Zudem wurden nur Patienten mit Behandlung mit DTIC als historischer Vergleich zu Ipilimumab im neuen Anwendungsgebiet selektiert, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Endpunkt: gesundheitsbezogene Lebensqualität**1. Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studien in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr

genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studien basierten auf der Population der Patienten, die den Fragebogen ausgefüllt haben. Es wurden keine fehlenden Werte ergänzt. Zudem wurde nur ein Arm der Studien in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen
- unplausible Angaben
- Anwendung inadäquater statistischer Verfahren

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studie in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Die Analysen der Studien basierten auf der Population der Patienten, die den Fragebogen ausgefüllt haben. Es wurden keine fehlenden Werte ergänzt. Der Studienarm DTIC diente als Vergleichsgruppe für die Bewertung des Zusatznutzens im Sinne einer historischen Kontrolle, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Endpunkt: Verträglichkeit

1. Verblindung der Endpunkterheber

*Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.*

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Studien wurden verblindet für den Behandler durchgeführt, allerdings wurde nur ein Arm der Studie in die Bewertung aufgenommen.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Analysen der Studie basierten auf der Sicherheitspopulation, allerdings wurde nur ein Arm der Studie in die Bewertung aufgenommen.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*

- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Es wurde nur ein Arm der Studie in die Bewertung aufgenommen

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Auf Studienebene lag ein hohes Verzerrungspotenzial vor. Zudem wurden nur Patienten mit Behandlung mit DTIC als historischer Vergleich zu Ipilimumab im neuen Anwendungsgebiet selektiert, so dass für die Nutzenherleitung davon ausgegangen werden muss, dass ein hohes Verzerrungspotenzial vorliegt.

Studie CA184-338:

B: Endpunkt Verträglichkeit (S. 383): „Es lagen keine Daten zur Verträglichkeit vor“ – dennoch wurde Abschnitt B für den Endpunkt ausgefüllt. Bitte Prüfen Sie dies und streichen Sie ggf. die Angabe „Es lagen keine Daten zur Verträglichkeit vor“.

Tabelle Z-25 (Anhang 4-G): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie "CA184-338"

Studie: CA184-338**Tabelle: Liste der für die Bewertung herangezogenen Quellen**

Genauere Benennung der Quelle	Kürzel
IPILIMUMAB (BMS-734016/MDX-010) INTERIM ABBREVIATED SUMMARY FOR OBSERVATIONAL STUDY PROTOCOL CA184338 ADVANCED MELANOMA	Studie 338

Endpunkt: Verträglichkeit

~~Es lagen keine Daten zur Verträglichkeit vor.~~

Dieser Satz wird gestrichen, da Daten zur Verträglichkeit vorliegen.