

IQWiG-Berichte – Nr. 678

Treatment Switching in onkologischen Studien

Arbeitspapier

Auftrag: GA14-04
Version: 1.0
Stand: 30.10.2018

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

Thema:

Treatment Switching in onkologischen Studien

Auftraggeber:

Bearbeitung im Rahmen des Generalauftrags

Datum des Auftrags:

12.02.2014

Interne Auftragsnummer:

GA14-04

Anschrift des Herausgebers:

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

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

ISSN: 1864-2500

Dieser Bericht wurde ohne die Beteiligung externer Sachverständiger erstellt.

Für die Inhalte des Berichts ist allein das IQWiG verantwortlich.

Mitarbeiterinnen und Mitarbeiter des IQWiG

- Michael Köhler
- Lars Beckmann
- Ulrich Grouven
- Charlotte Guddat
- Volker Vervölgyi
- Siw Waffenschmidt

Schlagwörter: Tumoren, Treatment Switching, Dateninterpretation – Statistische, Randomisierte Kontrollierte Studien

Keywords: Neoplasms, Treatment Switching, Data Interpretation – Statistical, Randomized Controlled Trials

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	v
Abbildungsverzeichnis	vi
Abkürzungsverzeichnis	vii
Kurzfassung	viii
1 Hintergrund	1
2 Fragestellung	3
3 Projektbearbeitung	4
4 Methoden	5
4.1 Teilziel 1: Zusammenfassung und Bewertung der zugrunde liegenden Annahmen der Methoden zur Berücksichtigung von Treatment Switching in onkologischen Studien	5
4.1.1 Informationsbeschaffung	5
4.1.2 Informationssynthese und -analyse	5
4.2 Teilziel 2: Überprüfung der Umsetzung der Methoden zur Berücksichtigung von Treatment Switching anhand von Beispielen	6
4.2.1 Informationsbeschaffung	6
4.2.2 Datenextraktion	6
4.2.3 Informationsanalyse	7
5 Ergebnisse	8
5.1 Teilziel 1: Zusammenfassung und Bewertung der zugrunde liegenden Annahmen der Methoden zur Berücksichtigung von Treatment Switching in onkologischen Studien	8
5.1.1 Ergebnis der Informationsbeschaffung	8
5.1.2 Einleitung	8
5.1.3 Naive Methoden	10
5.1.3.1 Intention to treat.....	10
5.1.3.2 Ausschluss der Behandlungswechsler	11
5.1.3.3 Zensierung zum Zeitpunkt des Wechsels	11
5.1.3.4 Überlebenszeitanalyse mit zeitabhängigen Variablen	12
5.1.4 Komplexere Methoden	12
5.1.4.1 Inverse Probability of Censoring Weighting (IPCW)	12
5.1.4.2 Rank-preserving structural Failure Time.....	17
5.1.4.3 Weitere Methoden	22
5.1.5 Allgemeine Anforderungen	22

5.2	Teilziel 2: Überprüfung der Umsetzung der Methoden zur Berücksichtigung von Treatment Switching anhand von Beispielen.....	24
5.2.1	Ergebnis der Informationsbeschaffung	24
5.2.2	Studiencharakteristika	27
5.2.3	Methoden zur Berücksichtigung des Treatment Switchings	38
6	Diskussion.....	49
7	Fazit.....	53
8	Literatur	54

Tabellenverzeichnis

Tabelle 1: Übersicht über Studien in frühen Nutzenbewertungen mit möglichem Behandlungswechsel von Kontrolle zu Prüfintervention.....	26
Tabelle 2: Allgemeine Angaben zu Studien aus Nutzenbewertungen mit Treatment Switching (Behandlungswechsel von Kontrolle zu Prüfintervention).....	28
Tabelle 3: Angaben zu Methodik und zugrunde liegenden Annahmen zur Berücksichtigung von Treatment Switching (Wechsel der Behandlung von Kontrolle zu Interventionsbehandlung).....	39
Tabelle 4: Ableitung eines Zusatznutzens bei Dossierbewertungen, in die Studien mit Treatment Switching eingeschlossen wurden	47

Abbildungsverzeichnis

Abbildung 1: Szenarien für einen vorzeitigen Behandlungswechsel in onkologischen RCTs: (a) einfaches Szenarium; (b) komplexeres Szenarium mit zusätzlichen Folgetherapien	9
Abbildung 2: Ergebnis der Informationsbeschaffung	25

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AFT	Accelerated Failure Time
ALK	anaplastische Lymphomkinase
BSC	Best supportive Care
CD4	Cluster-of-Differentiation-4-Rezeptor
CONSORT	Consolidated Standards of Reporting Trials
DGHO	Deutsche Gesellschaft für Hämatologie und Onkologie
ECOG	Index zur Lebensqualität der Eastern Cooperative Oncology Group
FACT-G	Functional Assessment of Cancer Therapy-General
FSKI	Functional Assessment of Cancer Therapy Kidney Symptom Index
IDMC	Independent Data Monitoring Committee
IPCW	Inverse Probability of Censoring Weights
IPE	Iterative Parameter Estimation
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
MSE	Mean squared Error
MSM	Marginal structural Models
MSKCC	Risikoscore des Memorial Sloan-Kettering Cancer Center
NICE	National Institute for Health and Care Excellence
PFS	progressionsfreies Überleben
PP	per Protocol
pU	pharmazeutischer Unternehmer
RCT	randomisierte kontrollierte Studie
RPSFT	Rank-preserving structural Failure Time
SGB	Sozialgesetzbuch
SNM	Structural nested Model
VEGF	vaskulärer endothelialer Wachstumsfaktor

Kurzfassung

Hintergrund

In randomisierten kontrollierten Studien zur Prüfung onkologischer Arzneimittel haben Patientinnen und Patienten als Besonderheit des Studiendesigns nicht selten die Möglichkeit, von einem Behandlungsarm in einen anderen zu wechseln. Dieser Vorgang wird als Cross-over oder Treatment Switching bezeichnet. Treatment Switching bedeutet hier einen bedarfsgesteuerten Wechsel von Patientinnen und Patienten von einer Behandlung zu einer anderen. Ein solcher Wechsel kann aus mehreren Gründen erfolgen, etwa weil Patientinnen und Patienten einer Kontrollgruppe bei frühzeitigem Nachweis der Wirksamkeit einer Prüfintervention ein Behandlungsvorteil nicht vorenthalten werden soll. Im Rahmen dieser Untersuchung wird der Begriff auf den Wechsel von einer Kontrollintervention oder Placebo zur Prüfintervention begrenzt.

Für alle beobachteten Endpunkte einer Studie führt das Treatment Switching zu einer Verzerrung des Behandlungseffektes. Wie groß die Differenz zwischen dem beobachteten und dem wahren Behandlungseffekt ist, kann dann in der Regel nicht bestimmt werden, da keine Patientenpopulation ohne Treatment Switching beobachtet wurde. In der Literatur wird eine Reihe von Methoden zu Überlebenszeitanalysen vorgeschlagen. Jeder dieser Methoden liegen jedoch bestimmte Annahmen zugrunde, von deren Richtigkeit die Aussagekraft der Analysen abhängt.

Fragestellung

Die Fragestellung der vorliegenden Untersuchung gliedert sich in folgende 2 Teilziele:

- Teilziel 1: Zusammenfassung und Bewertung der statistischen Methoden zur Berücksichtigung von Treatment Switching und der zugrundeliegenden Annahmen,
- Teilziel 2: Bewertung, inwiefern solche Methoden in Studien der pharmazeutischen Unternehmer in bisher abgeschlossenen Dossierbewertungen angewendet wurden, und, falls ja, die Überprüfung, ob die zugrundeliegenden Annahmen dargelegt wurden und sofern möglich Berücksichtigung fanden.

Methoden

Teilziel 1: Zusammenfassung und Bewertung der statistischen Methoden zur Berücksichtigung von Treatment Switching und der zugrunde liegenden Annahmen in onkologischen Studien

In diesem Projekt wurden in einem 1. Schritt die zugrunde liegenden Voraussetzungen der in der Literatur beschriebenen statistischen Methoden vollständig anhand der Originalpublikationen zusammengefasst. Es wurde überprüft, inwieweit und mit welchen Informationen die Voraussetzungen überprüfbar sind. Es wurde am Beispiel des Endpunktes Gesamtüberleben zusammengeführt, inwieweit Verletzungen der Voraussetzungen zu systematischen Verzerrungen führen. Hieraus wurde abgeleitet, wie die Anwendung der

verschiedenen statistischen Methoden zur Berücksichtigung von Treatment Switching in der Nutzenbewertung bzgl. der Ergebnissicherheit beurteilt werden kann.

Teilziel 2: Überprüfung der Umsetzung der Methoden zur Berücksichtigung von Treatment Switching anhand von Beispielen

Dieses Teilziel erfasste die Umsetzung der in Teilziel 1 beschriebenen Methoden in aktuellen onkologischen klinischen Studien. Insbesondere sollte beschrieben werden, welche Methoden in der Praxis angewendet wurden und inwieweit die zugrunde liegenden Annahmen dieser Methoden berücksichtigt wurden. Die möglichen Auswirkungen auf die Aussagesicherheit der in die Bewertung eingeschlossenen Studien wurden diskutiert.

Zur Bearbeitung der vorliegenden Fragestellung wurden auf Basis der bis März 2014 abgeschlossenen Aufträge zur frühen Nutzenbewertung nach § 35a SGB V zu onkologischen Arzneimitteln diejenigen Studien identifiziert, die der pharmazeutische Unternehmer (pU) oder das IQWiG in die Bewertung eingeschlossen hat. Aus diesen wurden diejenigen Studien ausgewählt, in denen ein Treatment Switching (im Sinne eines Wechsels von der Kontrollintervention zur Prüflintervention) möglich war. Diese Studien bildeten die Stichprobe für die Datenerhebung. Die extrahierten Daten wurden mithilfe deskriptiver statistischer Verfahren ausgewertet. Zudem erfolgte eine abschließende zusammenfassende Bewertung der Informationen.

Ergebnisse

Teilziel 1: Zusammenfassung und Bewertung der zugrunde liegenden Annahmen der Methoden zur Berücksichtigung von Treatment Switching in onkologischen Studien

Für die Adjustierung von Effektschätzern bei Treatment Switching stehen verschiedene Methoden zur Verfügung. Sie können in naive und komplexe Methoden unterteilt werden.

Naive Methoden umfassen die Intention-to-treat (ITT)-Analyse, den Ausschluss der Behandlungswechsler, die Zensierung zum Zeitpunkt des Wechsels und die Überlebenszeitanalyse mit zeitabhängigen Variablen.

Komplexe Methoden umfassen u. a. die Inverse-Probability-of-Censoring-Weighting (IPCW)-Methode und die Rank-preserving-structural-Failure-Time (RPSFT)-Methode.

Zusammenfassend lässt sich aus der vorliegenden Literatur schließen, dass es aktuell keine validierte statistische Methode gibt, die mit ausreichender Sicherheit die Analyse des Endpunktes Gesamtüberleben in Studien mit einem Therapiewechsel ermöglicht. Alle vorliegenden Methoden sind anfällig für Verzerrungen, deren Richtung und Ausmaß in der Regel nicht ohne zusätzliche Analysen einzuschätzen sind. Andererseits ist es für die Ableitung eines (Zusatz-)Nutzens notwendig, dass ein beobachteter Effekt mit ausreichender Sicherheit nur auf eine einzige Einflussgröße, d. h. auf die interessierende Therapie, zurückzuführen ist.

Aufgrund der beschriebenen Aspekte sollten die Ergebnisse der ITT-Analyse immer dargestellt werden. Weitere Ergebnisse aus Analysen, die den Behandlungswechsel explizit berücksichtigen, können ergänzend dargestellt werden. Allerdings sind für eine Bewertung dieser Analysen zusätzliche Informationen notwendig, die über die im CONSORT-Statement geforderten Angaben hinausgehen. Die notwendigen Informationen müssen ggf. zusätzlich für interessierende Subgruppen ausreichend beschrieben werden. Die Ergebnissicherheit und das Ausmaß des Behandlungseffektes werden in der Gesamtschau der vorliegenden Ergebnisse bewertet.

Teilziel 2: Überprüfung der Umsetzung der Methoden zur Berücksichtigung von Treatment Switching anhand von Beispielen

Seit der Einführung der frühen Nutzenbewertung von Arzneimitteln wurden am IQWiG bis Februar 2014 64 Dossierbewertungen durchgeführt, davon 21 zu onkologischen Fragestellungen. Von den entsprechenden pUs wurden darin insgesamt 20 randomisierte kontrollierte Studien (RCTs) zu direkten oder indirekten Vergleichen eingeschlossen. Bei 11 dieser Studien wurde den Patientinnen und Patienten der Kontrollgruppe unter bestimmten Voraussetzungen ein Wechsel von der Kontroll- zur Prüftherapie ermöglicht.

Allen Studien ist gemein, dass die pUs bezüglich eines Behandlungswechsels grundsätzlich nur Überlegungen zum Endpunkt Gesamtüberleben anstellen, obwohl mitunter weitere Endpunkte davon betroffen sind. In 63 % der Studien erfolgte die durch den Behandlungswechsel notwendige Entblindung nicht zu einem präspezifizierten Zeitpunkt, sondern aufgrund der signifikant positiven Studienergebnisse.

Sofern besondere Methoden wie IPCW oder RPSFT für die Berücksichtigung von Treatment Switching zum Einsatz kamen, waren diese in allen Studien a priori geplant oder es fehlen entsprechende Angaben dazu. Welche Annahmen im jeweiligen Fall zu treffen sind und ob diese erfüllt sind, wird in keinem der vorgelegten Dossiers diskutiert. Damit ist die Validität der dadurch gewonnenen Ergebnisse fraglich. Es zeigt sich somit, dass die hier betrachteten Dossiers, in denen diese Methoden zur Anwendung kamen, bezogen auf das Treatment Switching methodisch unvollständig sind. Diese komplexen Methoden wurden allerdings in nur 3 Studien (27 %) a priori geplant und angewendet. In anderen Studien wurden zusätzlich zur ITT-Analyse beispielsweise Behandlungswechsler zensiert (2 Studien [18 %]) oder von der Auswertung ausgeschlossen (1 Studie [9 %]), oder es wurden nach der Entblindung keine weiteren Analysen zum Gesamtüberleben mehr durchgeführt bzw. es wurden nur Datenschnitte dargestellt, die noch nicht vom Treatment Switching betroffen waren (4 Studien [36 %]).

Charakteristisch für alle Bewertungen ist, dass ein Zusatznutzen nur dann abgeleitet werden konnte, wenn Auswertungen vorlagen, für die ein Behandlungswechsel zur Prüfintervention nicht berücksichtigt werden musste oder als nicht entscheidend für die Aussagekraft angesehen wurde.

Fazit

Die Anwendung statistischer Verfahren zur Berücksichtigung von Treatment Switching birgt teils selbst das Risiko einer verzerrten Auswertung oder ist an die Berücksichtigung von Annahmen zur Wirksamkeit des Arzneimittels gebunden, die in der Praxis nicht überprüfbar sind. Aktuelle Beispiele klinischer Studien zeigen, dass komplexe Studiendesigns mit einem Behandlungswechsel angewendet werden, in denen kein statistisches Verfahren mehr eine unverzerrte Auswertung der Studiendaten gewährleistet. Derzeit existiert keine statistische Methodik, die die Analyse patientenrelevanter Endpunkte unter dem Einfluss von Treatment Switching mit ausreichender Sicherheit ermöglicht.

1 Hintergrund

In randomisierten kontrollierten Studien zur Prüfung onkologischer Arzneimittel haben Patientinnen und Patienten als Besonderheit des Studiendesigns nicht selten die Möglichkeit, von einem Behandlungsarm in einen anderen zu wechseln. Dieser Vorgang wird als Cross-over oder Treatment Switching bezeichnet. Er sollte nicht verwechselt werden mit dem Design einer Cross-over-Studie, in der alle Patientinnen und Patienten einer Studie nach einer a priori festgelegten Behandlungsdauer obligatorisch den Behandlungsarm wechseln und die Reihenfolge der Behandlung randomisiert zugewiesen wird [1]. Vielmehr meint Treatment Switching einen bedarfsgesteuerten Wechsel von Patientinnen und Patienten von einer Behandlung zu einer anderen. Ein solcher Wechsel kann aus mehreren Gründen erfolgen, etwa weil Patientinnen und Patienten einer Kontrollgruppe bei frühzeitigem Nachweis der Wirksamkeit einer Prüfintervention ein Behandlungsvorteil nicht vorenthalten werden soll. Der Wechsel einer Patientin oder eines Patienten von einer Studienbehandlung zur anderen kann von verschiedenen Kriterien abhängig gemacht werden. Ein Treatment Switching kann etwa erfolgen, wenn eine Patientin oder ein Patient nicht auf die Behandlung anspricht, zu der sie oder er randomisiert wurde, oder trotz der Behandlung eine Krankheitsprogression erleidet [2].

Unter dem Begriff Treatment Switching kann ein Wechsel von der Kontrollgruppe zur Prüfintervention, ein Wechsel von der Prüfintervention zur Kontrollintervention, ein Wechsel von der Studienmedikation zu einer Behandlung außerhalb der protokollgemäßen Studienbehandlung oder ein Ende jeder Behandlung verstanden werden [2]. Im Rahmen dieser Untersuchung wird der Begriff auf den Wechsel von einer Kontrollintervention oder Placebo zur Prüfintervention begrenzt.

Für alle beobachteten Endpunkte einer Studie führt das Treatment Switching zu einer Verzerrung des Behandlungseffektes. In einem einfachen Modell kann das Treatment Switching (bei Wechsel von Kontrolle zu Intervention) für das Gesamtüberleben bei der Analyse der Intention-to-treat(ITT)-Population zu einer Unterschätzung des wahren Behandlungseffektes führen, allerdings nur in dem Fall, dass die Prüfintervention tatsächlich einen Vorteil gegenüber der Kontrollintervention hat. Patientinnen und Patienten aus der Kontrollgruppe profitieren jetzt vom Überlebensvorteil der Prüfintervention, werden aber als Patientinnen und Patienten der Kontrollgruppe analysiert [3]. In anderen Fällen und für andere Endpunkte kann die Richtung der Verzerrung unklar sein. Wie groß die Differenz zwischen dem beobachteten und dem wahren Behandlungseffekt ist, kann dann in der Regel nicht bestimmt werden, da keine Ergebnisse für eine randomisierte Patientenpopulation ohne Treatment Switching vorliegen. Daraus ergibt sich die Anforderung an die analytische Methodik einer Studie, derartige durch Treatment Switching bedingte Verzerrungen zu berücksichtigen.

Hierzu wird in der Literatur eine Reihe von Methoden zu Überlebenszeitanalysen vorgeschlagen. Übersichten dazu finden sich bei Morden 2011 [1], Watkins 2013 [3] und

Latimer 2014 [2]. Jeder dieser Methoden liegen jedoch bestimmte Annahmen zugrunde, von deren Richtigkeit die Aussagekraft der Analysen abhängt.

Die beschriebenen Methoden zu Überlebenszeitanalysen umfassen naive Methoden wie die ITT-Analyse und Per-Protocol(PP)-Analysen, bei denen Patientinnen und Patienten, die die Behandlung wechseln, entweder von der Analyse ausgeschlossen werden oder zum Zeitpunkt des Wechsels zensiert werden. Des Weiteren kommen auch komplexere Methoden basierend auf parametrischen Modellen des Zensierungsmechanismus zur Anwendung. Allen komplexeren Methoden liegen Voraussetzungen zugrunde. Dabei ist zu unterscheiden zwischen mittels Daten überprüfbareren Voraussetzungen und solchen, die nicht nachprüfbar sind. Nicht nachprüfbarere Voraussetzungen sind beispielsweise die Annahme, dass alle Störgrößen für den Effekt vorliegen (*No unmeasured Confounders*), und die Annahme, dass der Effekt einer Therapie unabhängig vom Zeitpunkt des Therapiebeginns ist (*Common Treatment Effect*).

Morden 2011 [1] untersuchte die Eigenschaften der verschiedenen Methoden zu Überlebenszeitanalysen in einer Simulationsstudie. Die Ausgangslage für alle untersuchten Szenarien war, dass nur ein Wechsel von dem Kontrollarm in den Interventionsarm erlaubt war. Die untersuchten Faktoren waren die Effektstärke und der Anteil der Patientinnen und Patienten, die die Therapie wechseln. Ein von der Therapie unabhängiger Prognosefaktor sorgte dafür, dass Patientinnen und Patienten mit einer schlechteren Prognose mit höherer Wahrscheinlichkeit wechselten als Patientinnen und Patienten mit einer besseren Prognose. Der Zeitpunkt des Wechsels war unabhängig von der zugrunde liegenden Überlebenszeit. Morden 2011 [1] untersuchte den Bias der einzelnen Methoden, jedoch nicht die Power bzw. den Fehler 1. Art. Latimer 2014 [2] führte die Analysen von Morden 2011 [1] weiter und präsentiert einen Analyserahmen bei Vorliegen von Treatment Switching.

2 Fragestellung

Die Fragestellung der vorliegenden Untersuchung gliedert sich in folgende 2 Teilziele:

- Teilziel 1: Zusammenfassung und Bewertung der statistischen Methoden zur Berücksichtigung von Treatment Switching und der zugrundeliegenden Annahmen,
- Teilziel 2: Bewertung, inwiefern solche Methoden in Studien der pharmazeutischen Unternehmen in bisher abgeschlossenen Dossierbewertungen angewendet wurden, und, falls ja, die Überprüfung, ob die zugrunde liegenden Annahmen dargelegt wurden und sofern möglich Berücksichtigung fanden.

3 Projektbearbeitung

Zeitlicher Verlauf des Projekts

Im Rahmen des Generalauftrags wurde das Thema Treatment Switching in onkologischen Studien bearbeitet. Das vorliegende Arbeitspapier wurde auf Basis einer Projektskizze erstellt. Bearbeitungsbeginn war im Juni 2014. Dieser Bericht wurde an den G BA übermittelt und 4 Wochen später auf der Website des IQWiG veröffentlicht.

4 Methoden

4.1 Teilziel 1: Zusammenfassung und Bewertung der zugrunde liegenden Annahmen der Methoden zur Berücksichtigung von Treatment Switching in onkologischen Studien

4.1.1 Informationsbeschaffung

Es existieren verschiedene vorgeschlagene Methoden, mit denen das Treatment Switching in der Auswertung berücksichtigt und so der verzerrende Einfluss verkleinert werden kann. Diese Methoden sind im Wesentlichen in den Arbeiten von Morden 2011 beschrieben [1].

Zur Überprüfung der Vollständigkeit der in Morden 2011 beschriebenen Methoden wurden Recherchen in folgenden Quellen durchgeführt:

- eine Suche in den Guidelines der Organisationen EMA, NICE, AHRQ und ggf. vergleichbarer Organisationen,
- eine fokussierte Literaturrecherche in PubMed und Web of Science,
- eine Anwendung der „Related Citations“-Funktion in PubMed anhand der als relevant bewerteten Publikationen sowie
- eine Sichtung der Literaturlisten relevanter Publikationen.

Veröffentlichungen, die einzelne, einmalig publizierte statistische Methoden zur Berücksichtigung von Treatment Switching zum Inhalt haben, wurden aufgrund ihrer eingeschränkten Relevanz für Nutzenbewertungen für Teilziel 1 nicht berücksichtigt.

4.1.2 Informationssynthese und -analyse

In diesem Projekt wurden in einem 1. Schritt die zugrunde liegenden Voraussetzungen der statistischen Methoden vollständig anhand der Originalpublikationen zusammengefasst. Es wurde überprüft, inwieweit und mit welchen Informationen die Voraussetzungen überprüfbar sind. Aufbauend auf den Ergebnissen der Informationsbeschaffung wurde zusammengeführt, inwieweit Verletzungen der Voraussetzungen zu systematischen Verzerrungen führen. Die Analysen beziehen sich, so weit nicht anders spezifiziert, auf den Endpunkt Gesamtüberleben für den Vergleich einer Prüfintervention mit einer Kontrolltherapie in einer randomisierten Studie.

In einem 2. Schritt sollten die Analysen von Morden 2011 durch eigene Simulationsstudien ergänzt werden, insbesondere bzgl. der empirischen Power und des empirischen Fehlers 1. Art, die bei den Analysen von Morden 2011 nicht geschätzt wurden. Des Weiteren sollten Szenarien simuliert werden, bei denen der Zeitpunkt des Wechsels in Abhängigkeit von der zugrunde liegenden Überlebenszeit modelliert wurde. Da zwischenzeitlich umfangreiche Simulationsstudien hierzu veröffentlicht wurden [2,4-6], wurde auf eigene Simulationen verzichtet.

Aus beiden Schritten sollte abgeleitet werden, wie die Anwendung der verschiedenen statistischen Methoden zur Berücksichtigung von Treatment Switching in der Nutzenbewertung bzgl. der Ergebnissicherheit beurteilt werden kann.

4.2 Teilziel 2: Überprüfung der Umsetzung der Methoden zur Berücksichtigung von Treatment Switching anhand von Beispielen

Dieses Teilziel erfasste die Umsetzung der in Teilziel 1 beschriebenen Methoden in aktuellen onkologischen klinischen Studien. Insbesondere sollte beschrieben werden, welche Methoden in der Praxis angewendet wurden und inwieweit die zugrunde liegenden Annahmen dieser Methoden berücksichtigt wurden. Die möglichen Auswirkungen auf die Aussagesicherheit der in die Bewertung eingeschlossenen Studien wurden diskutiert.

4.2.1 Informationsbeschaffung

Zur Bearbeitung der vorliegenden Fragestellung wurden auf Basis der bis März 2014 abgeschlossenen Aufträge zur frühen Nutzenbewertung nach § 35a SGB V zu onkologischen Arzneimitteln diejenigen Studien identifiziert, die der pharmazeutische Unternehmer (pU) oder das IQWiG in die Bewertung eingeschlossen hat. Aus diesen wurden diejenigen Studien ausgewählt, in denen ein Treatment Switching (im Sinne eines Wechsels von der Kontrollintervention zur Prüflintervention) möglich war. Diese Studien bildeten die Stichprobe für die Datenerhebung.

4.2.2 Datenextraktion

Die Extraktion der Daten aus den für die frühe Nutzenbewertung verfügbaren öffentlich zugänglichen Dokumenten (insbesondere Modul 4 und die Dossierbewertung) erfolgte auf Studienebene.

Falls in Modul 4 und der Dossierbewertung keine Angaben zu einem etwaigen Treatment Switching beschrieben waren, konnten zusätzlich die Vollpublikationen oder andere öffentlich verfügbare Dokumente zu den betreffenden Studien durchsucht werden.

Die extrahierten Daten wurden in eine Tabelle mit folgenden Spalten eingetragen:

- Allgemeine Angaben zu Studien:
 - Name der Studie
 - Dossier (Auftragsnr.)
 - Indikation
 - Behandlungsarme (Anzahl, Art der Intervention)
 - Zahl der randomisierten Patientinnen und Patienten pro Behandlungsarm
 - Beobachtungsdauer

- Behandlungsdauer
- primäre und patientenrelevante sekundäre Endpunkte der Studie
- Angaben zum Treatment Switching:
 - Art und Zeitpunkt des Treatment Switchings
 - Anzahl der Patientinnen und Patienten mit Treatment Switching
 - Methode zur Berücksichtigung von Treatment Switching (Art der Methode / nicht zutreffend / keine Angaben)
 - zugrunde liegende Annahmen berücksichtigt (ja / nein)
 - zugrunde liegende Annahmen verletzt (nicht verletzt / Beschreibung der Verletzung als Freitext / keine Aussage möglich [Annahmen nicht überprüfbar ja / nein])
 - Methoden zur Berücksichtigung von Treatment Switching a priori geplant (ja / nein)

Die Extraktion der Studiendaten erfolgte jeweils durch ein Mitglied der Projektgruppe; zur Qualitätssicherung der Extraktion wurde diese von einem anderen Mitglied überprüft.

4.2.3 Informationsanalyse

Die extrahierten Daten wurden mithilfe deskriptiver statistischer Verfahren ausgewertet. Zudem erfolgte eine abschließende zusammenfassende Bewertung der Informationen.

5 Ergebnisse

5.1 Teilziel 1: Zusammenfassung und Bewertung der zugrunde liegenden Annahmen der Methoden zur Berücksichtigung von Treatment Switching in onkologischen Studien

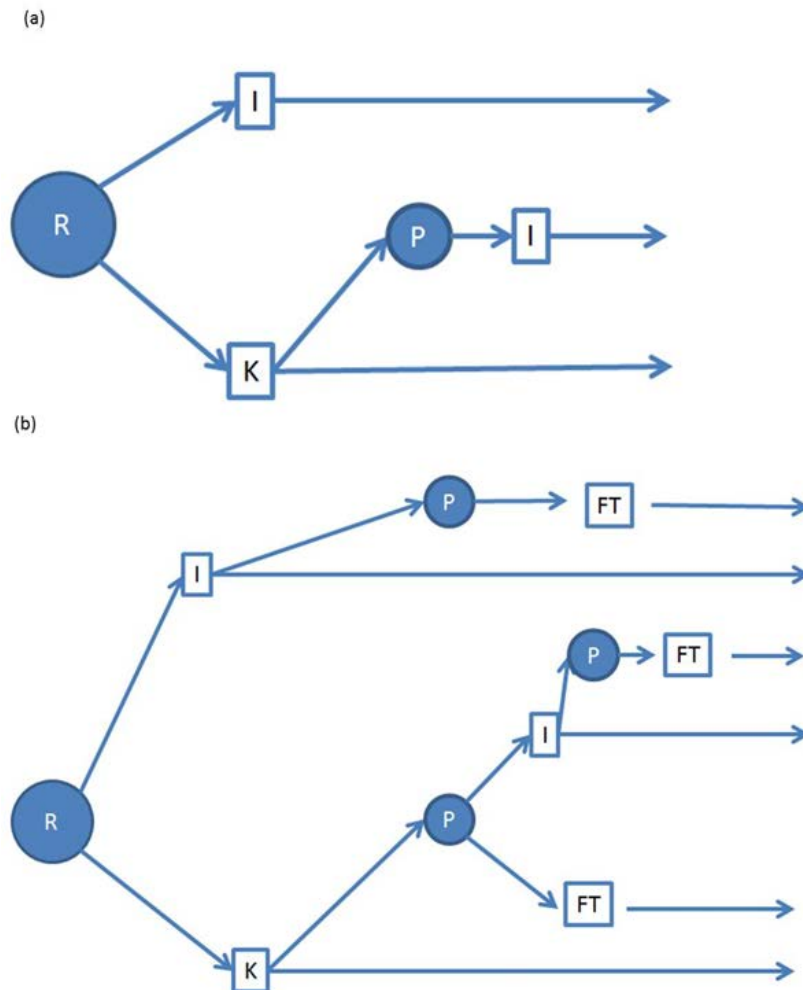
5.1.1 Ergebnis der Informationsbeschaffung

Es existieren verschiedene vorgeschlagene Methoden, mit denen das Treatment Switching in der Auswertung berücksichtigt und so der verzerrende Einfluss verkleinert werden kann. Die für Nutzenbewertungen relevanten Methoden sind im Wesentlichen in den Arbeiten von Morden 2011 beschrieben [1]. In diesem Arbeitspapier werden nur Veröffentlichungen berücksichtigt, die eine Methodik, eine Weiterentwicklung oder eine Anwendung dieser Methoden zum Inhalt haben. Die Überprüfung der Vollständigkeit von Morden 2011 erfolgte bis Mai 2018. Dabei wurden 53 zusätzliche relevante Veröffentlichungen identifiziert [2,3,5-55].

5.1.2 Einleitung

Im Folgenden wird davon ausgegangen, dass in einer 2-armigen randomisierten Studie (RCT) eine Intervention mit einer Kontrolle bzgl. des Endpunktes Gesamtüberleben verglichen werden soll. Das Ziel der Analyse ist ein Schätzer für den Unterschied in den Überlebenszeiten, der auf die zu vergleichenden Therapien, Intervention und Kontrolle, zurückzuführen ist.

Zunächst werden verschiedene Szenarien von Studiendesigns mit erlaubtem Behandlungswechsel dargestellt. Diese sind schematisch in Abbildung 1 dargestellt.



FT: Folgetherapie; I: Intervention; K: Kontrolle; P: Progression; R: Randomisierung

Abbildung 1: Szenarien für einen vorzeitigen Behandlungswechsel in onkologischen RCTs:
(a) einfaches Szenarium; (b) komplexeres Szenarium mit zusätzlichen Folgetherapien

Im einfachen Fall erfolgt der Behandlungswechsel, indem die Patientinnen und Patienten des Interventionsarms durchgängig bis zum Tod bzw. bis zum Ende des Follow-ups mit der Prüfintervention behandelt werden, während die Patientinnen und Patienten des Kontrollarms während des Studienverlaufs zur Prüfintervention wechseln (Abbildung 1, Szenarium a). Der Wechsel kann ohne einen klinischen Grund erfolgen, wenn beispielsweise ein primäres Studienziel erreicht wurde, oder nach einer Progression. In der Regel erfolgt dabei die Entblindung der Studienteilnehmer.

Das oben beschriebene Szenarium ist jedoch in der Onkologie in der Regel nicht anzutreffen. Die Patientinnen und Patienten im Interventionsarm werden ebenfalls bei Progression zu einer alternativen adjuvanten – kurativen und/oder palliativen – Folgetherapie wechseln (Abbildung 1, Szenarium b). Die Therapie kann bei weiterer Progression oder bei Nichtansprechen wiederum gewechselt werden. Die Patientinnen und Patienten im Kontrollarm können bei Progression zur Prüfintervention und, bei einer weiteren Progression oder bei

Nichtansprechen, ebenfalls zu weiteren, alternativen Folgetherapien wechseln oder direkt nach der 1. Progression auf alternative Folgetherapien gehen. In diesem Szenarium ergeben sich verschiedene Probleme bezüglich der Interpretation der Ergebnisse durch mögliche Verzerrungen, die im Weiteren diskutiert werden. Im Folgenden werden Überlegungen durchgeführt für Studien, in denen ein Behandlungswechsel nach einer Progression erfolgt. Dabei wird die Art der Diagnose der Progression nicht berücksichtigt.

5.1.3 Naive Methoden

5.1.3.1 Intention to treat

Bei der ITT-Analyse werden alle Studienteilnehmer gemäß ihrer Randomisierung ausgewertet. In dem in Abbildung 1a beschriebenen Szenarium wird ein Effekt dann unverzerrt geschätzt, wenn die Prüfintervention bei der späteren Verabreichung keinen Einfluss mehr auf das Überleben der Wechsler hat. Ist dies nicht der Fall, so wirkt der Behandlungswechsel als eine Störgröße (Confounder). Falls die Intervention auch bei den Wechslern einen positiven Einfluss hat, wird der wahre Effekt in der Regel unterschätzt, da die Überlebenszeiten der randomisierten Therapiearme sich angleichen, d. h., die Richtung der Verzerrung ist unter der Annahme eines positiven Effektes bekannt. Als Folge wird ein statistischer Test auf Unterschiede in den Überlebenszeiten an statistischer Power verlieren. Das Ausmaß der Verzerrung und des Powerverlusts hängt dabei u. a. von dem Anteil der Behandlungswechsler im Kontrollarm und von dem Einfluss der Prüfintervention auf die Überlebenszeit der Behandlungswechsler ab. Letzterer ist auch abhängig vom Zeitpunkt des Behandlungswechsels. In dem Fall, dass die Prüfintervention tatsächlich einen Vorteil gegenüber der Kontrollintervention hat, führt die ITT-Analyse zu einer Unterschätzung des wahren Behandlungseffektes unabhängig davon, ob der Behandlungswechsel durch ein klinisches Ereignis (Progression) oder unbeeinflusst davon erfolgt. Aufgrund der konservativen Richtung der Verzerrung wird der Fehler 1. Art eingehalten. Die Berücksichtigung weiterer Variablen wie zu Studienbeginn erhobener Charakteristika oder zeitabhängiger Variablen aus dem klinischen Verlauf ist nicht notwendig, falls eine konservative Effektschätzung ausreichend ist.

Die Interpretation einer ITT-Analyse ist jedoch eingeschränkt in Szenarien wie in Abbildung 1b beschrieben. Wenn es beispielsweise keinen positiven (oder negativen) Effekt für den Vergleich der Therapien Prüfintervention und Kontrolle gibt, so ergibt sich für den Vergleich der 2 Therapiearme ein positiver Effekt zugunsten der Prüfintervention, falls die Folgetherapien eine positive Wirkung im Vergleich zur Prüfintervention haben [8]. Ein beobachteter Effekt zugunsten der Intervention ergibt sich in dieser Situation durch die verspätete Verabreichung der wirksamen Folgetherapien im Kontrollarm. Aber auch wenn die Prüfintervention im Vergleich zur Kontrolle eine bessere Wirkung hätte, so würde dieser Effekt durch die Wirkung der Folgetherapien verzerrt. Ein therapiespezifischer Effekt kann mit einer ITT-Analyse nicht geschätzt werden.

Die ITT-Analyse sollte immer bei einer Bewertung dargestellt werden, da die Randomisierung aufrechterhalten wird. Wie oben beschrieben, ist die Richtung der Verzerrung nur unter der Annahme, dass die Prüfintervention einen gegenüber der Kontrolle positiven Effekt hat, bekannt. Allerdings muss diese Annahme mit geeigneten Daten belegt werden.

5.1.3.2 Ausschluss der Behandlungswechsler

Eine Methode zum Umgang mit Patientinnen und Patienten, die von der Kontrolle zur Prüfintervention wechseln, ist, diese aus der Überlebenszeitanalyse auszuschließen. Dieser Ansatz führt dann zu einer unverzerrten Schätzung des Behandlungseffekts von Prüfintervention und Kontrolle, wenn unabhängig von dem Behandlungswechsel die Verteilung der Überlebenszeit der ausgeschlossenen Patientinnen und Patienten gleich der Verteilung der Überlebenszeit der Patientinnen und Patienten ist, die nicht die Behandlung wechseln [3]. In dieser Situation würde der Ausschluss von Patientinnen und Patienten, die die Behandlung wechseln, ggf. nur zu einem Powerverlust führen. Der Ausschluss führt jedoch dazu, dass die Randomisierung nicht aufrechterhalten wird. Damit ist dieser Ansatz nicht robust gegenüber Selektionsbias in Situationen, in denen es Störgrößen gibt, die sowohl das Überleben als auch den Mechanismus des Behandlungswechsels beeinflussen. So können beispielsweise Patientinnen und Patienten selektiv ausgeschlossen werden, die aufgrund einer guten Prognose und einer daraus folgenden längeren Überlebenszeit eher im späteren Studienverlauf die Behandlung wechseln. In diesem Fall wäre der Behandlungseffekt zugunsten der Prüfintervention verzerrt. Würden andererseits eher Patientinnen und Patienten mit einer schlechteren Prognose vorzeitig wechseln, so wäre der Behandlungseffekt zuungunsten der Prüfintervention verzerrt.[9].

Bei einem Szenarium ähnlich dem in Abbildung 1b beschriebenen ist zu beachten, dass ein beobachteter Effekt außerdem von einer aktiven und wirksamen Folgetherapie beeinflusst sein kann.

5.1.3.3 Zensierung zum Zeitpunkt des Wechsels

Anstatt die Patientinnen und Patienten, die die Behandlung wechseln, aus der statistischen Auswertung auszuschließen, können diese zum Zeitpunkt des Wechsels zensiert werden. Dieser Ansatz führt – analog zu dem Ansatz, wechselnde Patientinnen und Patienten auszuschließen – dann zu einer unverzerrten Schätzung des Behandlungseffekts von Prüfintervention und Kontrolle, wenn unabhängig von dem Behandlungswechsel die Verteilung der Überlebenszeit der zensierten Patientinnen und Patienten gleich der Verteilung der Überlebenszeit der Patientinnen und Patienten ist, die nicht die Behandlung wechseln [3]. Beim Vorliegen von Störgrößen, die sowohl das Überleben als auch den Mechanismus des Behandlungswechsels beeinflussen, kommt es bei diesem Ansatz zu einer informativen Zensierung, d. h., die Zensierung ist nicht unabhängig vom Endpunkt und vom Therapiearm [10]. Da Überlebenszeitanalysen wie beispielsweise das Cox-Proportional-Hazards-Modell nicht informative Zensierungen voraussetzen, führt dieser Ansatz zu potenziell verzerrten Schätzungen. Analog zum Ansatz, die die Behandlung wechselnden Patientinnen und Patienten

aus der Analyse auszuschließen, kann die Verzerrung sowohl zugunsten als auch zuungunsten der Prüfintervention sein.

Bei einem Szenarium wie dem in Abbildung 1b beschriebenen ist zu beachten, dass ein beobachteter Effekt auch von einer aktiven und wirksamen Folgetherapie im Interventionsarm beeinflusst sein kann.

5.1.3.4 Überlebenszeitanalyse mit zeitabhängigen Variablen

Bei der Überlebenszeitanalyse mit zeitabhängigen Variablen werden die Patientinnen und Patienten gemäß ihrer Randomisierung ausgewertet. Der Wechsel von der Kontroll- zur Prüfintervention wird über eine zeitabhängige Variable berücksichtigt, d. h., es geht für jede Patientin und jeden Patienten die Information ein, ob er oder sie wechselt und, wenn ja, zu welchem Zeitpunkt. Dieser Ansatz ist eine Erweiterung des Ansatzes, bei dem Patientinnen und Patienten zum Zeitpunkt des Wechsels zensiert werden. Allerdings kann es auch hierbei zu verzerrten Ergebnissen kommen, wenn es einen Prädiktor gibt, der sowohl ein Risikofaktor für das Überleben als auch ein unabhängiger Prädiktor für den Zeitpunkt des Behandlungswechsels ist. In diesem Fall ist der Prädiktor eine zeitabhängige Störgröße (*Time-dependent Confounder*). Das Vorliegen einer solchen Störgröße führt in einem Cox-Proportional-Hazards-Modell zu verzerrten Ergebnissen, auch wenn die zeitabhängige Störgröße als solche mit in das Modell aufgenommen wird [11-13,56].

5.1.4 Komplexere Methoden

In den letzten Jahren wurden mehrere komplexere statistische Methoden vorgestellt, die sich mit dem Behandlungswechsel auseinandersetzen [1-3,14]. Die Methoden basieren auf Arbeiten zum Behandlungs- bzw. Studienabbruch aufgrund von fehlender Therapietreue (*Non-Compliance*) bei aktiven Folgetherapien [11,15-18]. Des Weiteren gibt es Überschneidungen mit Methoden zur Schätzung von Therapieeffekten in Situationen mit aktiven Begleitmedikationen [9,16,19,20].

Im Folgenden werden die 2 am meisten genannten Methoden beschrieben. Darüber hinaus werden die Anforderungen an die Darstellung der Analysen aufgezeigt, die für eine adäquate Interpretation der Ergebnisse notwendig sind, wie beispielsweise in Dossiers für die frühe Nutzenbewertung von Arzneimitteln.

5.1.4.1 Inverse Probability of Censoring Weighting (IPCW)

Methodik

Die Methode *Inverse Probability of Censoring Weighting* (IPCW) wurde von Robins und Finkelstein (2000) entwickelt und ist in den Bereich der *Marginal Structural Models* (MSMs) einzuordnen [10,57]. Die Motivation war die Auswertung einer RCT, in der 2 Therapien zur Prophylaxe der interstitiellen Lungenentzündung bei Aidspatienten verglichen wurden.

Bei der IPCW-Methode werden Patientinnen und Patienten zum Zeitpunkt des Behandlungswechsels zensiert. Dieser Ansatz führt zunächst zu verzerrten Schätzern aufgrund informativer Zensierung (siehe Abschnitt 5.1.3.3). Angenommen, es gibt eine Variable Prognose mit 2 Ausprägungen. Patientinnen und Patienten mit einer „guten“ Prognose haben ein längeres Gesamtüberleben als solche mit einer „schlechten“ Prognose. Patientinnen und Patienten, die eine schlechte Prognose haben, haben – in diesem Beispiel – eine höhere Wahrscheinlichkeit für einen Behandlungswechsel. Damit wäre die Überlebenszeitanalyse verzerrt. Für eine Korrektur dieser Verzerrung werden die Patientinnen und Patienten gemäß ihrer Prognose gewichtet. Im beschriebenen Beispiel werden eher Patientinnen und Patienten mit einer schlechten Prognose die Behandlung wechseln. Ereignisse dieser Patientinnen und Patienten, die nach dem Behandlungswechsel auftreten, werden aufgrund der Zensierung in der Auswertung nicht berücksichtigt. Die IPCW-Methode gibt daher nicht zensierten Patientinnen und Patienten mit einer schlechten Prognose größere Gewichte als nicht zensierten Patientinnen und Patienten mit einer guten Prognose. Ziel der Gewichtung ist es, den durch die Zensierung entstandenen Informationsverlust auszugleichen. Im Interventionsarm bekommen die Patientinnen und Patienten das Gewicht 1. Die gewichteten Patientinnen und Patienten bilden eine Pseudopopulation, die im Kontrollarm für die Variable Prognose und damit auch für die informative Zensierung adjustiert ist [10].

Die Prognose kann dabei zeitabhängig sein, d. h. sich im Studienverlauf ändern. Die Gewichte müssen über den gesamten Studienverlauf geschätzt werden, beispielsweise zu jedem Personenmonat oder jeder Personenwoche. Die Patientinnen und Patienten gehen mit ihren Gewichten in ein Cox-Proportional-Hazards-Modell mit zeitabhängigen Variablen zur Schätzung des Behandlungseffektes ein. Die wechselnden Patientinnen und Patienten werden, wie oben beschrieben, zum Zeitpunkt des Wechsels zensiert. Des Weiteren können Kaplan-Meier-Überlebenszeitkurven für die Pseudopopulation geschätzt werden. Dieser Ansatz liefert im betrachteten Beispiel dann einen unverzerrten Schätzer, wenn die Prognose die einzige zeitabhängige Störgröße sowohl für die Zensierung als auch für das Gesamtüberleben ist.

Zugrunde liegende Annahmen

Die IPCW-Methode liefert anhand der Pseudopopulation einen Schätzer für den Vergleich der 2 Behandlungsarme in der künstlichen Situation, dass keine Patientin und Patient die Behandlung wechselt. Der Schätzer ist dann unverzerrt, wenn die *No-unmeasured-Confounders*-Annahme erfüllt ist [10,14,26]. Diese Annahme bedeutet, dass alle Störgrößen, die sowohl die Zensierung – also den Mechanismus zum Behandlungswechsel – als auch das Gesamtüberleben beeinflussen, bei der Schätzung der Gewichte angemessen berücksichtigt werden. Dabei ist zu beachten, dass nicht nur zu Studienbeginn erhobene Charakteristika zu berücksichtigen sind, sondern auch zeitabhängige Störgrößen über den gesamten Studienverlauf. Zusätzlich zu den Charakteristika zu Studienbeginn wie Alter, Geschlecht und Schweregrade – beispielsweise ECOG- oder Karnofsky-Score – müssen auch zeitabhängige Störgrößen wie beispielsweise das Auftreten von Metastasen und unerwünschten Ereignissen, die Veränderung des Schweregrades, aber auch quantitative Laborparameter auf ihren Einfluss

auf den Behandlungswechsel und auf das Gesamtüberleben untersucht werden. Zeitabhängige Variablen müssen über den gesamten Studienverlauf erhoben werden. Sind nicht alle relevanten Variablen berücksichtigt, so kann das nicht berücksichtigte Confounding zu Verzerrungen führen, da dann die Annahme der *No unmeasured Confounders* verletzt ist [14,21,22].

Howe et al. (2010) geben 3 Kriterien an, damit die Annahme der *No unmeasured Confounders* erfüllt ist [22]. Erstens müssen ausreichend viele Prädiktoren des Zensierungsmechanismus und des Gesamtüberlebens beobachtet und angemessen operationalisiert werden. Das 2. Kriterium ist die *Positivity* und bedeutet, dass über den gesamten Studienverlauf für alle Kombinationen von Ausprägungen der Prädiktoren eine von null verschiedene Wahrscheinlichkeit besteht, nicht zensiert zu werden [10,21,23]. Die *Positivity* ist sowohl beeinflusst durch die Anzahl der Prädiktoren als auch durch die Anzahl der Patientinnen und Patienten im Kontrollarm. Kleine Fallzahlen und in hohem Maße stratifizierte Daten aufgrund vieler sowohl kategorialer als auch stetiger Prädiktoren können zu zufälliger *Non-Positivity* führen [21,22]. Kleine von null verschiedene Wahrscheinlichkeiten können zu extremen Gewichten und damit zu verzerrten Ergebnissen und höherer Unsicherheit führen [21-23], aber auch dazu, dass die Algorithmen zur Schätzung der Modellparameter nicht konvergieren und damit keine oder keine valide Effektschätzung liefern [24]. Auch die Verwendung stabilisierender Gewichte kann die Annahme der *Positivity* verletzen [23]. Das 3. Kriterium ist, dass kein Prädiktor berücksichtigt werden kann, der deterministisch sowohl den Zensierungsmechanismus als auch das Gesamtüberleben voraussagt. Auch hier kann es bei kleinen Stichproben dazu kommen, dass die Berücksichtigung fast deterministischer Prädiktoren zu verzerrten Ergebnissen führt [10,21-23].

Letztlich ist anzumerken, dass die Annahme der *No unmeasured Confounders* anhand der beobachteten Daten nicht zu überprüfen ist [21-23].

Schätzung der Modellparameter und des Behandlungseffektes

Die Schätzung der Gewichte ist nicht trivial und erfordert 3 sorgfältig zu planende Schritte: i) Auswahl der Zeitpunkte, zu denen Variablen erhoben werden; ii) Operationalisierung und Erhebung der potenziellen Prädiktoren; iii) Auswahl der Prädiktoren zur Schätzung der Gewichte anhand geeigneter Kriterien. Erstens müssen alle relevanten Prädiktoren des Zensierungsmechanismus und des Gesamtüberlebens über den Studienverlauf beobachtet und angemessen operationalisiert werden. Die Erhebung kann problematisch sein, wenn beispielsweise unerwünschte Ereignisse, die sowohl mit dem Behandlungswechsel als auch mit dem Gesamtüberleben assoziiert sind, nach einer Progression nicht weiter beobachtet werden. Die Operationalisierung kann insbesondere bei stetigen Variablen wie beispielsweise Laborparametern schwierig sein, wenn die Operationalisierung die *Positivity* und damit auch die Effektschätzung beeinflusst. Robins et al. (2000) und Howe et al. (2010) zeigen dies anhand von Therapiestudien mit HIV-infizierten Männern. Dabei wurde die Anzahl der Cluster-of-differentiation-4-Rezeptor (CD4)-positiven Lymphozyten als zeitabhängige Störgröße in der Analyse berücksichtigt [22,23,25].

Die Schätzung der (stabilisierenden) Gewichte erfolgt anhand von gepoolten logistischen Regressionen [14,21,22,26] oder über Cox-Proportional-Hazards-Modelle mit zeitabhängigen Variablen [10], wobei robuste oder Bootstrap-basierte Verfahren zur Schätzung der Varianzen zur Verfügung stehen. Cole und Hernan (2008) zeigen verschiedene Möglichkeiten für die Auswahl und Operationalisierung der Prädiktoren auf, wobei zwischen einer möglichen Verzerrung sowie einer erhöhten Unsicherheit in der Schätzung der Gewichte einerseits und der Anzahl der berücksichtigten Prädiktoren andererseits abgewogen wird [23]. Alternativ wird vorgeschlagen, diejenigen Prädiktoren auszuwählen, die die stärkste Assoziation mit der Gesamtmortalität [10] oder mit dem Behandlungswechsel zeigen [14]. Colleoni et al. (2011) untersuchten in einer RCT den Effekt von Letrozol gegenüber Tamoxifen bei Brustkrebspatientinnen [27]. Dabei wurden die Variablen für die Schätzung der Gewichte ausgewählt, welche sowohl mit dem Behandlungswechsel als auch mit dem Überleben assoziiert waren. Bisher liegen keine Untersuchungsergebnisse vor, welcher der Ansätze zur Auswahl der Prädiktoren und zur Schätzung der Gewichte in welcher Situation angemessen ist.

Ist die Anwendung der IPCW-Methode bereits auf Ebene der Studienplanung vorgesehen, so kann insbesondere der 1. Schritt, die Beobachtung der potenziellen Prädiktoren über den Studienverlauf, effektiv durchgeführt werden. Auch können die Modelle für die Schätzung der Gewichte und für die Schätzung des Behandlungseffektes a priori festgelegt werden. Wird die Methode jedoch post hoc durchgeführt, können wichtige Prädiktoren häufig gar nicht erhoben werden oder liegen nur unzureichend vor, wodurch die Aussagekraft der Analyse eingeschränkt ist. Da die Gewichte selbst zeitabhängig sind, ist sowohl die Auswahl der Variablen als auch die Schätzung der Gewichte aufwendig, da zu jedem Zeitpunkt die *Positivity* anhand geeigneter Kriterien überprüft werden muss. Das Modell zur Schätzung des Behandlungseffektes muss ebenfalls sorgfältig gewählt werden. Durch die Aufhebung der Randomisierung ist es erforderlich, ebenfalls bzgl. zu Studienbeginn erhobener Charakteristika für das Gesamtüberleben zu adjustieren. Hieraus können, falls diese Variablen ebenfalls zur Schätzung der Gewichte benutzt wurden, zusätzliche Probleme entstehen [23].

Bei der Interpretation der Ergebnisse muss beachtet werden, inwieweit andere Zensierungsmechanismen wie beispielsweise Abbruch wegen unerwünschter Ereignisse und Lost to Follow-up eine Rolle spielen. Robins und Finkelstein (2000) zeigen anhand eines Beispiels, wie weitere Zensierungsmechanismen berücksichtigt und interpretiert werden können [10].

Die beschriebene Methodik geht von einem Studiendesign wie dem in Abbildung 1a beschriebenen aus. Um ein Szenarium wie das in Abbildung 1b beschriebene mit der IPCW-Methode zu analysieren, müssten die Folgetherapien der Prüfintervention und der Kontrolle adäquat berücksichtigt werden. Robins und Greenland (1994) untersuchen einen Ansatz basierend auf der IPCW-Methode zur Korrektur von unterschiedlich angewendeten Begleittherapien [35]. Untersuchungen zur Berücksichtigung von Folgetherapien im Zusammenhang mit dem Behandlungswechsel wurden nicht identifiziert.

Ergebnisse aus Simulationsstudien

Howe und Kollegen zeigen in einer Simulationsstudie, dass kleine Stichproben, ein hohes Ausmaß des Selektionsbias durch den Behandlungswechsel, die Nichtberücksichtigung von relevanten Prädiktoren und eine falsche Modellierung Einfluss auf eine potenzielle Verzerrung haben [22]. Die IPCW-Methode erwies sich zur Schätzung der Überlebenszeitkurve als überlegen gegenüber einer Kaplan-Meier-Analyse hinsichtlich Bias und Mean squared Error (MSE). In Situationen, in denen die zugrunde liegenden Annahmen erfüllt waren, lieferte die IPCW-Methode gute Schätzungen mit kleinen MSEs. Bei Verletzungen der Voraussetzungen zeigten sich jedoch deutlich schlechtere Ergebnisse bzgl. Bias und MSE.

Latimer und Kollegen berichten, dass die IPCW-Methode zu deutlichen Verzerrungen und schlechten Überdeckungswahrscheinlichkeiten führt, wenn die Annahme der *No unmeasured Confounders* verletzt ist [5,24]. Die Verzerrung war dabei korreliert mit dem Anteil der Behandlungswechsler und kann sowohl zu Unter- als auch zu Überschätzungen eines wahren Therapieeffektes führen. Auch berichten die Autoren von Konvergenzproblemen bei der statistischen Analyse [24].

Des Weiteren ist ein ausgewogenes Verhältnis zwischen der Anzahl der wechselnden und der nicht wechselnden Patientinnen und Patienten erforderlich. Auch müssen in beiden Patientengruppen ausreichend viele Ereignisse beobachtet werden, um einerseits eine gute Schätzung der Gewichte und andererseits eine gute Effektschätzung zu erlauben. Insbesondere in kleinen Studien kann es vorkommen, dass zu wenige nicht wechselnde Patientinnen und Patienten übrig bleiben, um die Gewichte zu schätzen bzw. um die Überlebenszeitanalyse durchzuführen [5,22-24].

Neben den Simulationsstudien gibt es Anwendungsbeispiele, in denen die IPCW-Methode oder ähnliche Methoden wie beispielsweise das *Inverse Probability Treatment Weighting* für binäre Endpunkte bzgl. der Schätzung der Gewichte und der Auswahl der dazugehörigen Prädiktoren diskutiert werden [10,22,23,26,28-30]. Weitere Anwendungsbeispiele für IPCW-Analysen gibt es aus den Bereichen maligne Melanome [47], Prostatakarzinom [55] und chronische lymphomatische Leukämie oder kleinzelliges lymphozytisches Lymphom [58].

Anforderungen bei einer Bewertung

Um eine Einschätzung einer IPCW-Analyse vornehmen zu können, sind spezifische Angaben zu dem Zensierungsmechanismus, zu möglichen Prädiktoren, zur Schätzung der Gewichte und zu fehlenden Werten erforderlich.

- Der Zensierungsmechanismus, der zum Behandlungswechsel führt, muss eine klinisch begründbare Ursache haben. Wechseln Patientinnen und Patienten beispielsweise ausschließlich bei einer Krankheitsprogression, so können die Variablen, die plausibel mit einer Progression in Verbindung gebracht werden können, über den Studienverlauf beobachtet werden. Sollten Patientinnen und Patienten ohne eine klinisch begründbare

Ursache wechseln, so ist davon auszugehen, dass mangels geeigneter Prädiktoren die Methode nicht geeignet ist.

- Die Kriterien sowohl für die Auswahl der Zeitpunkte der Datenerhebungen als auch für die Auswahl der Variablen sowie für die Schätzung der Gewichte und für die Effektschätzung sind zu beschreiben. Angemessene statistische Methoden zur Schätzung der Gewichte wie auch zur Effektschätzung müssen begründet und nachvollziehbar beschrieben werden.
- Die Kriterien, die der Annahme der *No unmeasured Confounders* zugrunde liegen – insbesondere die *Positivity* –, sollten überprüft und entsprechende Sensitivitätsanalysen dargestellt werden.
- Aufgrund der aufwendigen Datenerhebung kann es zu fehlenden Werten kommen. Der Anteil der fehlenden Werte über den Studienverlauf muss berichtet werden. Sollten Ersetzungsstrategien angewendet werden, so ist deren Angemessenheit zu begründen.

Retrospektive Analysen mittels IPCW sind in der Regel nicht angemessen, da nicht ausgeschlossen werden kann, dass relevante Variablen nicht ausreichend erhoben wurden.

5.1.4.2 Rank-preserving structural Failure Time

Methodik

Die Motivation der Methode *Rank-preserving structural Failure Time* (RPSFT) besteht darin, den Therapieeffekt unter der Annahme, dass alle Patientinnen und Patienten über den gesamten Studienverlauf gemäß ihrer Randomisierung behandelt worden wären, zu schätzen [12,29,32]. Die Methode basiert auf einem *Accelerated-Failure-Time*(AFT)-Modell [31] und ist einzuordnen in den Bereich der *Structural nested Models* (SNMs). Die zugrunde liegende Idee ist die Konstruktion einer kontrafaktischen (*counterfactual*) Überlebenszeit T^{KF} für jede Patientin und jeden Patienten, d. h. derjenigen Überlebenszeit, die die Patientin oder der Patient erfahren hätte, wenn er oder sie ausschließlich unter der Kontrolle behandelt worden wäre. Dabei sind 3 Fälle zu unterscheiden.

Für eine Patientin oder einen Patienten aus dem Kontrollarm, die oder der nicht die Behandlung wechselt, entspricht die kontrafaktische Überlebenszeit T_K^{KF} der tatsächlich in der Studie beobachteten Überlebenszeit T_K , also

$$T_K^{KF} = T_K .$$

Eine Patientin oder ein Patient aus dem Interventionsarm mit beobachteter Überlebenszeit T_I hat eine um einen Faktor $\exp(\psi)$ veränderte Überlebenszeit gegenüber der Zeit, die sie oder er unter der Kontrollintervention erfahren hätte:

$$T_I^{KF} = T_I \times \exp(\psi) .$$

Der Faktor ψ wird als *Accelerated-Failure-Time*-Faktor (AFT-Faktor) bezeichnet und ist im Fall eines positiven Effektes der Prüfindervention negativ, d. h. $\psi < 0$. Für eine Patientin oder einen Patienten aus dem Kontrollarm, die oder der im Studienverlauf zur Prüfindervention wechselt, setzt sich die kontrafaktische Überlebenszeit aus 2 Zeiten zusammen. Dies ist zum einen die Zeit, in der die Patientin oder der Patient bis zum Wechsel mit der Kontrollintervention behandelt wurde, T_K , und zum anderen die Zeit, die er oder sie unter der Prüfindervention verbringt, T_I , korrigiert um den AFT-Faktor ψ , also

$$T_{KI}^{KF} = T_K + T_I \times \exp(\psi).$$

Aus den 3 Überlebenszeiten T_K^{KF} , T_{KI}^{KF} und T_I^{KF} wird eine kontrafaktische Population unter der Annahme, dass alle Patientinnen und Patienten unter der Kontrolle behandelt wurden, gebildet. Da die ursprüngliche Randomisierung aufrechterhalten bleibt, sollten sich die 2 Therapiearme – einmal die Prüfindervention mit der kontrafaktischen Überlebenszeit T_I^{KF} und einmal der Kontrollarm mit den kontrafaktischen Überlebenszeiten T_K^{KF} und T_{KI}^{KF} – nicht unterscheiden.

Zugrunde liegende Annahmen

Der RPSFT-Methode liegen 2 Annahmen zugrunde. Erstens ergibt sich aus der Konstruktion der RPSFT, dass der AFT-Faktor ψ konstant über den gesamten Studienverlauf ist, d. h. die Wirkung einer Therapie unabhängig davon ist, zu welchem Zeitpunkt im Studienverlauf – und damit auch im Krankheitsverlauf – eine Patientin oder ein Patient die entsprechende Therapie bekommt. Zweitens gilt der AFT-Faktor ψ für alle Patientinnen und Patienten. Die 2 Annahmen werden zusammengefasst unter dem Begriff des *Common Treatment Effects*. Daraus folgt, dass für 2 beliebige Patientinnen und Patienten, bei denen je ein Ereignis auftritt, die Reihenfolge des Auftretens dieses Ereignisses unabhängig von der Behandlung ist. Wenn also beide Patientinnen und Patienten mit der Prüfindervention behandelt würden und das Ereignis bei einer Patientin oder einem Patienten vorher auftritt als bei dem anderen, dann wäre dies auch der Fall, wenn beide mit der Kontrollintervention behandelt würden. Inwieweit diese Annahme plausibel ist, hängt vom untersuchten Kontext ab, d. h. von der untersuchten Indikation, den Wirkmechanismen der Therapien (Placebo oder aktive Kontrolle) und davon, wieweit prognostische Faktoren diese Annahme plausibel erscheinen lassen. Gerade im onkologischen Bereich ist diese Annahme des *Common Treatment Effects* kritisch zu sehen. Genauso wie die Annahme der *No unmeasured Confounders* ist die Annahme des *Common Treatment Effects* anhand empirischer Daten nicht zu verifizieren [1,32].

Schätzung der Modellparameter und des Behandlungseffektes

Unter der Annahme, dass das Modell korrekt spezifiziert ist, d. h. die Annahme des *Common Treatment Effects* zutrifft, entsprechen die kontrafaktischen Überlebenszeiten den erwarteten Überlebenszeiten, wenn alle Patientinnen und Patienten mit der Kontrolle behandelt worden wären. Insbesondere gibt es unter Beibehaltung der ursprünglichen Randomisierung keinen

Unterschied in den kontrafaktischen Überlebenszeiten zwischen den Therapiearmen. Ein Wert für den AFT-Faktor ψ kann nun geschätzt werden, indem über einen ausreichenden Bereich möglicher Werte dasjenige ψ gewählt wird, welches zur größten Übereinstimmung zwischen den Therapiearmen basierend auf den kontrafaktischen Überlebenszeiten führt. Dafür werden verschiedene Ansätze vorgeschlagen. Robins and Tsiatis (1991) schätzten den AFT-Faktor ψ mittels einer g-Schätzung, einer nicht parametrischen Technik des Schätzens und Testens aus dem Bereich der SNMs [12,32-34,59]. Bei diesem Ansatz können verschiedene parametrische, nicht parametrische oder semiparametrische Modelle angewandt werden [36,37]. Gemeinsam ist allen Ansätzen, dass der Bereich für ψ mit einem möglichst dichten Rastergitter überdeckt wird [8,32,34].

Wenn die ITT-Analyse kein statistisch signifikantes Ergebnis erbracht hat, ist die RPSFT-Methode ebenfalls nicht statistisch signifikant, vorausgesetzt es wurden für beide Methoden identische statistische Modelle verwendet [11,32,37]. Unter der Nullhypothese $\psi = 0$ entspricht die Wahrscheinlichkeit für den Fehler 1. Art der RPSFT-Methode derjenigen der ITT-Analyse [8,32,34]. Die statistische Power der g-Schätzung bleibt dabei mindestens so groß wie die der ITT-Analyse [11,32,34,37]. Eine Alternative zur g-Schätzung wurde von Korhonen et al. (2012) in einer Studie zur Therapie des metastasierten Nierenzellkarzinoms mit Everolimus vorgestellt [38]. In dieser Studie wurde die Randomisierung nach dem Risikoscore des Memorial Sloan-Kettering Cancer Center (MSKCC) zu Studienbeginn stratifiziert. Unter der Annahme des *Common Treatment Effects* wurde die stratifizierte Randomisierung in der Anwendung der RPSFT berücksichtigt. Dazu verwendeten die Autoren einen stratifizierten gewichteten zensierten Rangtest nach Fleming und Harrington [38,60].

Als Effektschätzer wird der AFT-Faktor ψ interpretiert als ein Skalierungsfaktor auf der entsprechenden Zeitebene. Robins und Tsiatis (1991) interpretieren $\exp(\psi)$ als einen Faktor, um den sich die Überlebenszeit einer Patientin oder eines Patienten gegenüber der Kontrolle verlängert, wenn sie oder er im gesamten Studienverlauf die Prüfintervention bekommen hätte [11,32]. Transformationen des AFT-Faktors ψ in ein Hazard Ratio (HR) sind möglich, basieren dann aber auf zusätzlichen parametrischen Verteilungsannahmen [37].

Bei der Schätzung des AFT-Faktors ψ können 2 Probleme auftreten. So kann es vorkommen, dass ψ nicht geschätzt werden kann, da der Algorithmus nicht konvergiert oder keine eindeutig identifizierbare Lösung für ψ hervorbringt [1,24,34,39]. Mark und Robins (1993) berichten, dass die Funktion zur Bestimmung des optimalen ψ über weite Bereiche des Rastergitters monoton sein kann [34]. Zu einer Stabilisierung der Schätzung des AFT-Faktors ψ können Sensitivitätsanalysen mittels Adjustierungen oder Gewichtungen beitragen, allerdings sind diese Methoden noch nicht ausreichend untersucht [3]. Ouwens et al. (2018) fanden in einer Simulationsstudie nur dann eindeutige Lösungen für den AFT-Faktor ψ , wenn der Anteil der Behandlungswechsler $< 40\%$ war [40].

Des Weiteren kann es bei den kontrafaktischen Überlebenszeiten zu einer informativen Zensierung kommen, da für eine Patientin oder einen Patienten aus dem Kontrollarm die Wahrscheinlichkeit, zum Ende der Studie zensiert zu werden, assoziiert ist mit der Länge der Zeit, die er oder sie nach einem Behandlungswechsel mit der Prüfintervention behandelt wird [8,13,32,37,38]. Um dem entgegenzuwirken, werden alle Patientinnen und Patienten re-zensiert, indem die Nachbeobachtungszeit bedingt auf den AFT-Faktor ψ angepasst wird. Bei der Re-Zensierung kann es zu einem Powerverlust bzw. einer erhöhten Unsicherheit kommen, da Ereignisse ggf. nicht berücksichtigt werden. Walker et al. (2004) präsentierten einen semiparametrischen Ansatz, bei dem in einem bivariaten Modell sowohl die Schätzung des AFT-Faktors ψ als auch die Re-Zensierung unter verschiedenen parametrischen Modellen berücksichtigt wird [37]. Allerdings kann es hierbei zu einem Informationsverlust kommen, da nicht nur die wechselnden, sondern alle Patientinnen und Patienten re-zensiert werden. Walker et al. (2004) präsentieren eine neue Methode zur Re-Zensierung [37,44]. White (2006) diskutiert kritisch den Mechanismus zur Re-Zensierung [61].

Parametrische Erweiterungen der RPSFT-Methode wurden von Branson und Whitehead (2002) vorgeschlagen [36], eine Modifikation wurde von Zhang und Chen (2016) vorgestellt [62]. Die *Iterative-Parameter-Estimation(IPE)*-Methode ersetzt die nicht parametrische g-Schätzung des AFT-Faktors ψ durch eine parametrische Likelihood-basierte Schätzung [36]. Bowden et al. (2015) erweiterten die RPSFT-Methodik um gewichtete Log-Rank-Statistiken mit dem Ziel eines Powergewinns [42]. Bennett et al. (2018) präsentierten eine Erweiterung, um mittels Adjustierungen Unsicherheiten in der Anwendung der RPSFT zu berücksichtigen [41].

Die ursprünglich formulierte RPSFT-Methode nimmt an, dass die Effektivität einer Therapie mit dem Ende der entsprechenden Behandlung aufhört. Daher entspricht die beschriebene Methodik dem Szenarium in Abbildung 1a, bei dem Folgetherapien nicht berücksichtigt werden. Im komplexeren Szenarium gemäß Abbildung 1b wird auch bei der Prüfintervention die Therapie nach einer Progression beendet und zu einer Folgetherapie gewechselt (siehe Abbildung 1). Für die RPSFT ergeben sich daraus bei der Interpretation bzgl. therapiespezifischer Effekte vergleichbare Probleme wie bei der ITT-Analyse. Latimer und Kollegen diskutieren in diesem Zusammenhang verschiedene Ansätze, um sowohl den Therapieabbruch als auch einen möglichen Wechsel von der Prüfintervention zur Kontrolle zu berücksichtigen [2,24,43]. Allerdings werden keine Folgetherapien berücksichtigt. Yamaguchi und Ohashi (2004) entwickelten einen auf der RPSFT-Methode basierenden Ansatz zur Berücksichtigung von Folgetherapien in RCTs, jedoch ohne Beachtung eines Behandlungswechsels [17,35].

Ergebnisse aus Simulationsstudien

Mark und Robins (1993) fanden in einer Simulationsstudie, dass die Wahrscheinlichkeit für den Fehler 1. Art der RPSFT-Methode mit derjenigen der ITT-Analyse identisch war [34]. Ein Verlust an statistischer Power gegenüber der ITT-Analyse ergab sich insbesondere bei hohen Anteilen von wechselnden Patientinnen und Patienten. Die Ergebnisse der RPSFT-Methode

waren weniger verzerrt als die der ITT-Analyse, allerdings waren die Ergebnisse mit einer größeren Unsicherheit behaftet. Des Weiteren berichten die Autoren von Problemen bei der Anwendung aufgrund von Monotonie, d. h., die Funktion zur Bestimmung des optimalen ψ war über weite Bereiche des Rastergitters monoton. Daraus ergaben sich Schwierigkeiten für die Schätzung des optimalen AFT-Faktors ψ .

In den Simulationen von Odoni und McNamee (2010) zeigte die RPSFT-Methode gute Ergebnisse bzgl. der Verzerrung über viele Situationen hinweg verglichen mit der ITT-Analyse [39]. In manchen Situationen erwies sich der Ansatz nach Walker [37,44] als besser bezüglich der Verzerrung. Allerdings simulierten die Autoren unter der Annahme des *Common Treatment Effects*.

In Morden 2011 simulierten die Autoren unter der Annahme des *Common Treatment Effects* [1]. Die 3 Ansätze RPSFT, IPE und die Methode nach Walker führten alle zu Überschätzungen des wahren AFT-Faktors ψ , wobei die IPE-Methode die geringste Verzerrung aufwies. Insbesondere bei kleinen Werten für ψ zeigten sich große Verzerrungen, für die Methode nach Walker zeigten sich insbesondere in Fällen mit großem Selektionsbias große Überschätzungen des wahren AFT-Faktors. Sowohl bei der RPSFT-Methode als auch bei der Methode nach Walker zeigten sich Probleme bei der Konvergenz der zugrunde liegenden Algorithmen.

In den Simulationen von Latimer und Kollegen erwiesen sich die Methoden RPSFT und IPE als nahezu gleichwertig [2,5,6,24]. Nur bei extremen Abweichungen von Standardverteilungen in den Simulationen erwies sich die nicht parametrische RPSFT robuster als die IPE. Allerdings zeigten beide Verfahren große Verzerrungen in Situationen, in denen unter Verletzung der Annahme des *Common Treatment Effects* simuliert wurde. Die Autoren empfehlen Bootstrap-Verfahren zur Schätzung von Konfidenzintervallen [5,6].

Die Gruppe um Latimer verwendete die Methoden RPSFT und IPE bei der Analyse verschiedener Studien zum malignen Melanom [45-47]. Schlumberger et al. (2016) präsentieren eine Anwendung der RPSFT mit Bootstrapping beim Schilddrüsenkarzinom [48], Demetri und Kollegen beim gastrointestinalen Stromatumor [49].

Anforderungen an eine Bewertung

- Die zugrunde liegende Annahme des *Common Treatment Effects* ist gerade im onkologischen Bereich kritisch zu sehen und, da die Annahme nicht verifizierbar ist, angemessen zu diskutieren.
- Der zeitliche Verlauf sowohl des Behandlungswechsels als auch der Re-Zensierung sollte dargestellt werden.
- Die Verwendung eines parametrischen, semiparametrischen oder nicht parametrischen Modells ist zu begründen. Werden mehrere Modelle angewendet, so sind alle Ergebnisse darzustellen und in einer Gesamtschau zu interpretieren.

- Die Verfahren zur Schätzung des AFT-Faktors ψ sind zu begründen. Sensitivitätsanalysen mittels Adjustierungen oder Gewichtungen sind darzustellen. Der potenzielle Einfluss des Rastergitters ist darzustellen, um die Plausibilität der Schätzung des konkreten ψ als optimale Lösung einschätzen zu können.
- In Studien, in denen Patientinnen und Patienten nach zu Studienbeginn erhobenen Charakteristika stratifiziert randomisiert werden, muss diese Stratifizierung in der Analyse entsprechend berücksichtigt werden [38].

5.1.4.3 Weitere Methoden

Die beiden vorgestellten komplexeren Methoden, die des *IPCW* und das *RPSFT*-Modell, können als Vertreter von Methoden aus dem Bereich der *Instrumental Variables* angesehen werden, die zur Identifizierung von kausalen Effekten entwickelt wurden, wenn auch teilweise für die Anwendung bei der Analyse nicht randomisierter Studien [25,33,35,50-52]. Mehrere Autoren beschäftigten sich mit der speziellen Problematik von Nichtunterlegenheitstests bei Behandlungswechseln in RCTs [53,54].

Latimer und Kollegen präsentieren einen *Two-Stage*-Algorithmus im Rahmen von SNMs [2,5,6,24], der eine simplifizierte Modifikation von Ansätzen von Robins und Greenland (1994) sowie Yamaguchi und Ohashi (2004) [17,35] darstellt. Die *Two-Stage*-Methode basiert auf dem *RPSFT*-Modell, berücksichtigt aber unter der Annahme der *No unmeasured Confounders*, dass die Randomisierung zum Zeitpunkt des Wechsels aufgehoben wird. Dabei werden mögliche Confounder zum Zeitpunkt des Behandlungswechsels in das AFT-Modell miteinbezogen. Der Vorteil gegenüber der *IPCW*-Methode ist, dass die berücksichtigten Variablen zur Adjustierung nur zu einem Zeitpunkt nach einer Progression erhoben werden müssen und nicht während des gesamten Studienverlaufs und des Follow-ups. Allerdings wird nur die Schätzung des Behandlungseffektes, aber nicht die Unsicherheit der Modellierung verbessert [5]. Der *Two-Stage*-Algorithmus fand Anwendung im Bereich des malignen Melanoms [47] und des Prostatakarzinoms [55].

5.1.5 Allgemeine Anforderungen

Zusammenfassend lässt sich aus der vorliegenden Literatur schließen, dass es aktuell keine validierte statistische Methode gibt, die mit ausreichender Sicherheit die Analyse des Endpunktes Gesamtüberleben in Studien mit einem Therapiewechsel ermöglicht [63]. Alle vorliegenden Methoden sind anfällig für Verzerrungen, deren Richtung und Ausmaß in der Regel nicht ohne zusätzliche Analysen einzuschätzen sind. Andererseits ist es für die Ableitung eines (Zusatz-)Nutzens notwendig, dass ein beobachteter Effekt mit ausreichender Sicherheit nur auf eine einzige Einflussgröße, d. h. auf die interessierende Therapie, zurückzuführen ist [64].

Aufgrund der beschriebenen Aspekte sollten die Ergebnisse der ITT-Analyse immer dargestellt werden.

Weitere Ergebnisse aus Analysen, die den Behandlungswechsel explizit berücksichtigen (wie z. B. IPCW oder RPSFT-Modelle), hinsichtlich der Frage, ob ein Behandlungseffekt vorliegt, oder bzgl. des Ausmaßes eines Behandlungseffektes können ergänzend dargestellt werden. Allerdings sind für eine Bewertung dieser Analysen zusätzliche Informationen notwendig, die über die im CONSORT-Statement geforderten Angaben hinausgehen. Dazu gehören:

- Die Ergebnisse müssen im Hinblick auf das Studiendesign und die damit einhergehenden Limitierungen der statistischen Methoden interpretiert werden.
- Angaben zur Anzahl der Therapie- und Studienabbrecher und zu den entsprechenden Gründen dafür
- Angaben zum Anteil der Wechsler sowie zum zugehörigen zeitlichen Verlauf (wie beispielsweise Kaplan-Meier-Kurven)
- Beschreibung des Mechanismus, nach dem gewechselt werden darf (wie beispielsweise die Krankheitsprogression)
- Diskussion des Mechanismus sowie der potenziellen Prädiktoren in Hinblick auf die Annahmen der jeweils benutzten statistischen Methode
- Beschreibung der Untersuchungen auf mögliche Prädiktoren sowohl für den Wechsel als auch für das Gesamtüberleben in der Teilpopulation der Wechsler
- Diskussion der möglichen Verzerrung und der Präzision der Effektschätzung
- Diskussion des möglichen Einflusses von Folgetherapien
- Darstellung der methodenspezifischen Anforderungen, wie oben beschrieben
- Beschreibung der eingehenden Daten sowie des Programms für die statistischen Analysen

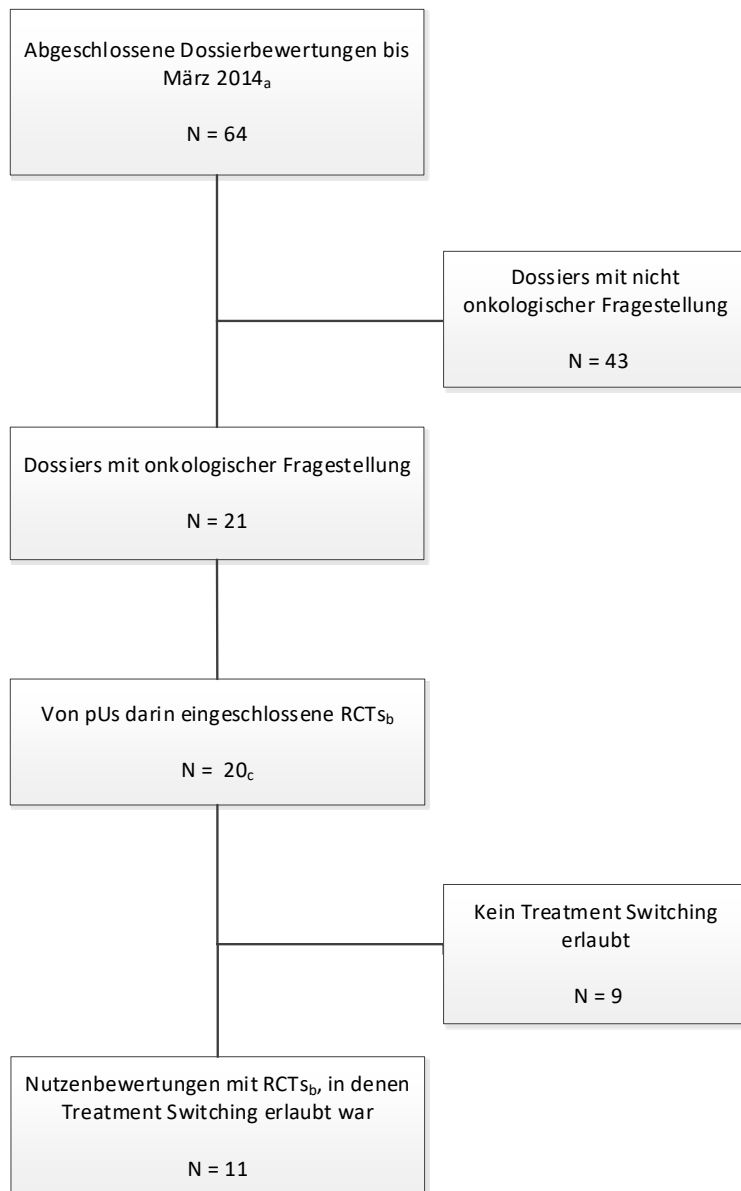
Die notwendigen Informationen müssen ggf. zusätzlich für interessierende Subgruppen ausreichend beschrieben werden. Die Ergebnissicherheit und das Ausmaß des Behandlungseffektes werden in der Gesamtschau der vorliegenden Ergebnisse bewertet.

5.2 Teilziel 2: Überprüfung der Umsetzung der Methoden zur Berücksichtigung von Treatment Switching anhand von Beispielen

5.2.1 Ergebnis der Informationsbeschaffung

Studien mit Treatment Switching in Dossierbewertungen des IQWiG

Seit der Einführung der frühen Nutzenbewertung von Arzneimitteln wurden am IQWiG bis Februar 2014 64 Dossierbewertungen durchgeführt, davon 21 zu onkologischen Fragestellungen. Von den entsprechenden pUs wurden darin insgesamt 20 RCTs zu direkten oder indirekten Vergleichen eingeschlossen. Bei 11 dieser Studien wurde den Patientinnen und Patienten der Kontrollgruppe unter bestimmten Voraussetzungen ein Wechsel von der Kontroll- zur Prüftherapie ermöglicht (siehe Abbildung 2).



a: ohne Orphan Drugs, ohne Addenda
b: randomisierte kontrollierte Studie
c: 2 dieser Studien wurden aufgrund der Neubewertung einzelner Arzneimittel jeweils 2-mal in Nutzenbewertungen eingeschlossen.

Abbildung 2: Ergebnis der Informationsbeschaffung

Tabelle 1 zeigt eine Zuordnung der eingeschlossenen Studien zu Nutzenbewertungen des IQWiG (Dossier- und Auftragsnummer) einschließlich der dort untersuchten Arzneimittel.

Tabelle 1: Übersicht über Studien in frühen Nutzenbewertungen mit möglichem Behandlungswechsel von Kontrolle zu Prüfintervention

Wirkstoff Zitat	Studie	G-BA- Dossiernummer(n)	IQWiG- Auftragsnummer(n)	Einschluss in Nutzenbewertung
Vemurafenib [65-68]	BRIM-3	2012-03-15-D-029	A12-08, A13-34	ja
Sorafenib ^a [69,70]	TARGET	2012-10-01-D-039	A12-14	nein
Everolimus ^a [69,70]	RECORD-1	2012-10-01-D-039	A12-14	nein
Crizotinib [71,72]	PROFILE1007	2012-11-15-D-040	A12-15	ja
Abirateron- acetat [73,74]	COU-AA-302	2013-01-15-D-056	A13-06	ja
Vandetanib [75-77]	D4200C00058	2012-03-15-D-030, 2013-03-15-D-059 ^b	A12-09, A13-09 ^b	ja
Pertuzumab [78,79]	CLEOPATRA (WO20698)	2013-04-01-D-057	A13-10	ja
Enzalutamid [80,81]	AFFIRM	2013-09-01-D-073	A13-33	ja
Dabrafenib [43,82]	BRF113683 (BREAK-3)	2013-10-01-D-076	A13-35	ja
Trastuzumab Emtansin [83,84]	EMILIA	2014-01-01-D-084	A14-01	ja
Radium-223- dichlorid [85,86]	BC1-06 (ALSYMPCA)	2014-01-01-D-094	A14-02	ja
<p>a: Diese Wirkstoffe waren nicht Gegenstand einer Dossierbewertung; die Studien TARGET und RECORD-1 wurden vom pU zum Zweck eines indirekten Vergleichs gegenüber dem zu bewertenden Arzneimittel Axitinib eingeschlossen.</p> <p>b: Der Wirkstoff Vandetanib wurde einer erneuten Nutzenbewertung unterzogen, da die erforderlichen Nachweise bei der Erstbewertung nicht vollständig waren.</p> <p>pU: pharmazeutischer Unternehmer</p>				

Von den 11 Studien, in denen ein Wechsel von der Kontroll- zur Prüfintervention möglich war, wurden 9 in die jeweilige Nutzenbewertung des IQWiG eingeschlossen. Die beiden Studien TARGET und RECORD-1 wurden vom pU im Rahmen der frühen Nutzenbewertung von Axitinib (Dossiernummer 2012-10-01-D-039, Auftrag A12-14) zum Zweck eines indirekten Vergleichs eingeschlossen. Schlussendlich führte er diesen indirekten Vergleich aufgrund mangelnder Ähnlichkeit der Studien aber nicht durch. Da zu den Studien TARGET und RECORD-1 aber eine Beschreibung des Designs und der Studiencharakteristika vorlag, werden sie an dieser Stelle eingeschlossen und beschrieben, obwohl sie für die damalige Nutzenbewertung nicht als relevant erachtet wurden.

5.2.2 Studiencharakteristika

Tabelle 2 zeigt die Charakteristika der Studien, in denen ein Wechsel der Behandlung von der Kontroll- zur Prüfintervention möglich war. Alle Angaben beziehen sich zeitlich auf den Stand der jeweiligen Bewertung. Beispielsweise kann eine in Tabelle 2 als laufend bezeichnete Studie mittlerweile abgeschlossen sein.

Tabelle 2: Allgemeine Angaben zu Studien aus Nutzenbewertungen mit Treatment Switching (Behandlungswechsel von Kontrolle zu Prüfintervention)

Name der Studie Intervention	Population	Studiendesign, Zeitraum	Interventionen (Zahl der randomisierten Patientinnen und Patienten)	Behandlungsdauer	primärer Endpunkt, patientenrelevante sekundäre Endpunkte	geplante Beobachtungsdauer für die patientenrelevanten Endpunkte
BRIM-3 Intervention: Vemurafenib	Behandlung von erwachsenen, nicht mit systemischen Krebstherapien vorbehandelten Patientinnen und Patienten mit BRAF-V600-Mutation und nicht resezierbarem (Stadium IIIC oder IV) metastasiertem Melanom	RCT, offen, parallel, aktiv kontrolliert Zeitraum: 01/2010–12/2010 (klinischer Cut-off) Interimsanalyse geplant nach Erreichen von 50 % der erwarteten Ereignisse ^a (98 von 196 Todesfällen); Datenauswertung zum 31.12.2010 Weitere Datenschnitte: ▪ 31.03.2011, ▪ 03.10.2011, ▪ 01.02.2012, ▪ 20.12.2012	Vemurafenib (N = 337) Dacarbazin (N = 338)	Behandlung geplant bis Tod, Toxizität oder Abbruch	Primäre Endpunkte: ▪ Gesamtüberleben, ▪ progressionsfreies Überleben Patientenrelevante sekundäre Endpunkte: ▪ Schmerz, ▪ gesundheitsbezogene Lebensqualität, ▪ unerwünschte Ereignisse	gesundheitsbezogene Lebensqualität und Schmerz bis Tag 1 des letzten Behandlungszyklus oder 28 Tage nach Progression Gesamtüberleben und unerwünschte Ereignisse: geplante Nachbeobachtung bis Tod oder Abbruch

(Fortsetzung)

Tabelle 2: Allgemeine Angaben zu Studien aus Nutzenbewertungen mit Treatment Switching (Behandlungswechsel von Kontrolle zu Prüfintervention) (Fortsetzung)

Name der Studie Intervention	Population	Studiendesign, Zeitraum	Interventionen (Zahl der randomisierten Patientinnen und Patienten)	Behandlungsdauer	primärer Endpunkt, patientenrelevante sekundäre Endpunkte	geplante Beobachtungsdauer für die patientenrelevanten Endpunkte
TARGET Intervention: Sorafenib	erwachsene Patientinnen und Patienten mit metastasiertem Nierenzellkarzinom nach einer vorangegangenen systemischen Therapie innerhalb der letzten 1 bis 8 Monate (zielgerichtete VEGF-Rezeptorinhibitoren ausgeschlossen)	RCT, doppelblind, parallel, placebokontrolliert Zeitraum: 12/2003–09/2006	Sorafenib + BSC (N = 451) Placebo + BSC (N = 452)	Behandlung geplant bis zum 1. Eintreten einer Progression, Tod, zur Toxizität oder zum Abbruch	Primärer Endpunkt: ▪ Gesamtüberleben Sekundäre Endpunkte: ▪ progressionsfreies Überleben, ▪ gesundheitsbezogene Lebensqualität (FKSI, FACT-G), ▪ unerwünschte Ereignisse	Nachbeobachtung Gesamtüberleben: bis zum Studienende (September 2006) oder Tod Nachbeobachtung (Endpunkte nicht spezifiziert): 30 Tage nach Behandlungsende
RECORD-1 Intervention: Everolimus	erwachsene Patientinnen und Patienten mit metastasiertem Nierenzellkarzinom mit klarzelliger Komponente nach vorangegangener Tumorprogression unter / nach vorangegangener Therapie mit Sunitinib und / oder Sorafenib (Vorbehandlungen mit mTOR-Inhibitoren ausgeschlossen)	RCT, doppelblind, parallel, placebokontrolliert Zeitraum: 11/2006–11/2008	Everolimus + BSC (N = 277) Placebo + BSC (N = 139)	Behandlung geplant bis zum 1. Eintreten einer Progression, Tod, zur Toxizität oder zum Abbruch	Primärer Endpunkt: ▪ progressionsfreies Überleben Sekundäre Endpunkte: ▪ Gesamtüberleben, ▪ gesundheitsbezogene Lebensqualität und Symptome (FKSI, EORTC-QLQ-C30), ▪ unerwünschte Ereignisse	Nachbeobachtung Gesamtüberleben: mindestens 2 Jahre nach Behandlungsbeginn oder bis zum Tod Andere Endpunkte: keine Angaben

(Fortsetzung)

Tabelle 2: Allgemeine Angaben zu Studien aus Nutzenbewertungen mit Treatment Switching (Behandlungswechsel von Kontrolle zu Prüflintervention) (Fortsetzung)

Name der Studie Intervention	Population	Studiendesign, Zeitraum	Interventionen (Zahl der randomisierten Patientinnen und Patienten)	Behandlungsdauer	primärer Endpunkt, patientenrelevante sekundäre Endpunkte	geplante Beobachtungsdauer für die patientenrelevanten Endpunkte
PROFILE100 7 Intervention: Crizotinib	erwachsene Patientinnen und Patienten mit nachgewiesener Translokation im ALK-Lokus und lokal fortgeschrittenem oder metastasierendem nicht kleinzelligen Bronchialkarzinom Die Krankheit musste progredient sein nach nur einer Vorbehandlung mit einer platinbasierten Kombinationschemotherapie, die eine Maintenance-Therapie beinhalten konnte.	RCT, offen, parallel, aktiv kontrolliert Zeitraum: 02/2010–02/2012, geplantes Studienende: 03/2013; Analysen zum Datenschnitt: 30.03.2012	Crizotinib (N = 173) Chemotherapie (N = 174) Davon: Pemetrexed (N = 99) Docetaxel (N = 72)	bis zum 1. Eintreten einer Progression, einer unzumutbaren Toxizität, einer Behandlungspause von mehr als 6 Wochen oder einer Verschlechterung der allgemeinen Krankheits-symptomatik	Primärer Endpunkt: ▪ progressionsfreies Überleben Sekundäre Endpunkte: ▪ Gesamtüberleben, ▪ Symptomatik, ▪ gesundheitsbezogene Lebensqualität, ▪ unerwünschte Ereignisse	Gesamtüberleben: bis zum Tod aufgrund jeder Ursache Gesundheitsbezogene Lebensqualität und Symptomatik: keine Angaben Unerwünschte Ereignisse: bis 28 Tage nach Behandlungsende

(Fortsetzung)

Tabelle 2: Allgemeine Angaben zu Studien aus Nutzenbewertungen mit Treatment Switching (Behandlungswechsel von Kontrolle zu Prüflintervention) (Fortsetzung)

Name der Studie Intervention	Population	Studiendesign, Zeitraum	Interventionen (Zahl der randomisierten Patientinnen und Patienten)	Behandlungsdauer	primärer Endpunkt, sekundäre Endpunkte	geplante Beobachtungsdauer für die patientenrelevanten Endpunkte
COU-AA-302 Intervention: Abirateron-acetat	erwachsene Männer mit metastasiertem kastrationsresistentem Prostatakarzinom; mindestens 1, aber nicht mehr als 2 fehlgeschlagene Chemotherapien, von denen mindestens 1 Docetaxel-haltig ist	RCT, doppelblind, parallel, placebokontrolliert Zeitraum: 04/2009–02/2014 3 Interimsanalysen und finale Auswertung geplant: ▪ 1. Datenschnitt: 20.12.2010, nach 116 Todesfällen geplant ▪ 2. Datenschnitt: 20.12.2011, nach 311 Todesfällen geplant ▪ 3. Datenschnitt: 22.05.2012, nach 425 Todesfällen geplant ▪ Finale Analyse: geplant nach 773 Todesfällen	Abirateron + Prednison + ADT (N = 546) Placebo + Prednison + ADT (N = 542)	Behandlung: bis Krankheitsprogression	Primäre Endpunkte: ▪ Gesamtüberleben, ▪ radiografisches progressionsfreies Überleben Sekundäre Endpunkte: ▪ Zeit bis zum Beginn einer Opiattherapie, ▪ gesundheitsbezogene Lebensqualität, ▪ unerwünschte Ereignisse	Beobachtung aller Endpunkte bis zum Ende der Behandlungsphase (letzte Visite 14 bis 28 Tage nach letzter Dosis Studienmedikation) Unerwünschte Ereignisse: bis 30 Tage nach der letzten Dosis der Studienmedikation Gesamtüberleben und Gebrauch von Opiattherapien: alle 3 Monate für bis zu 5 Jahre

(Fortsetzung)

Tabelle 2: Allgemeine Angaben zu Studien aus Nutzenbewertungen mit Treatment Switching (Behandlungswechsel von Kontrolle zu Prüflintervention) (Fortsetzung)

Name der Studie Intervention	Population	Studiendesign, Zeitraum	Interventionen (Zahl der randomisierten Patientinnen und Patienten)	Behandlungsdauer	primärer Endpunkt, patientenrelevante sekundäre Endpunkte	geplante Beobachtungsdauer für die patientenrelevanten Endpunkte
D4200C00058 Intervention: Vandetanib	erwachsene Patientinnen und Patienten mit messbarem, nicht resektablem, lokal fortgeschrittenem oder metastasiertem medullärem Schilddrüsenkarzinom	RCT, doppelblind, parallel, placebokontrolliert Zeitraum: 11/2006– andauernd, voraussichtlich bis 12/2016 Interimsanalyse: 31.07.2009 finale Analyse, sobald 50 % der randomisierten Patientinnen und Patienten verstorben	Vandetanib + BSC (N = 231) Placebo + BSC (N = 100) Relevante Teilpopulation mit symptomatischem und progressivem MTCb: Vandetanib + BSC (N = 126) Placebo + BSC (N = 60)	Behandlung bis zur objektiven Krankheitsprogression, dann Option einer offenen Behandlung mit Vandetanib	Primärer Endpunkt: ▪ progressionsfreies Überleben Sekundäre Endpunkte: ▪ Gesamtüberleben, ▪ gesundheitsbezogene Lebensqualität, ▪ Zeit bis zur Schmerzprogression, ▪ unerwünschte Ereignisse	Zeit bis Schmerzprogression und gesundheitsbezogene Lebensqualität: Ende der randomisierten Behandlungsphase Unerwünschte Ereignisse: bis 60 Tage nach letzter Einnahme der Studienmedikation Gesamtüberleben: Nachbeobachtung bis zum Tod

(Fortsetzung)

Tabelle 2: Allgemeine Angaben zu Studien aus Nutzenbewertungen mit Treatment Switching (Behandlungswechsel von Kontrolle zu Prüfintervention) (Fortsetzung)

Name der Studie Intervention	Population	Studiendesign, Zeitraum	Interventionen (Zahl der randomisierten Patientinnen und Patienten)	Behandlungsdauer	primärer Endpunkt, patientenrelevante sekundäre Endpunkte	geplante Beobachtungsdauer für die patientenrelevanten Endpunkte
CLEOPATRA WO20698 Intervention: Pertuzumab in Kombination mit Trastuzumab und Docetaxel	erwachsene Patientinnen und Patienten mit HER2-positivem, metastasiertem oder lokal rezidivierendem, inoperablem Brustkrebs, die zuvor noch keine Chemotherapie oder biologische Therapie zur Behandlung der metastasierten Erkrankung erhalten haben	RCT, doppelblind, parallel, placebokontrolliert Zeitraum: 3 Studienphasen: Studienphase 1: 02/2008–05/2011 1. Datenschnitt Studienphase 2: 05/2011–05/2012 2. Datenschnitt Studienphase 3: seit 05/2012, Analyse des Gesamtüberlebens nach 385 Todesfällen	Pertuzumab / Trastuzumab / Docetaxel (N = 402) Placebo / Trastuzumab / Docetaxel (N = 406) Teilpopulationen: metastasierter Brustkrebs (N = 787) lokal rezidivierender, inoperabler Brustkrebs (N = 19) keine Angabe (N = 2)	Behandlung bis Krankheitsprogression, nicht zu beherrschender Toxizität oder Studienende durch Sponsor	Primärer Endpunkt: ▪ progressionsfreies Überleben Sekundäre Endpunkte: ▪ Gesamtüberleben, ▪ unerwünschte Ereignisse	Unerwünschte Ereignisse: bis 4–6 Wochen nach letzter Gabe von Studienmedikation Nachbeobachtung Gesamtüberleben: bis zum Tod, Kontaktverlust, zur Rückziehung der Einwilligung oder zum Studienende durch Sponsor

(Fortsetzung)

Tabelle 2: Allgemeine Angaben zu Studien aus Nutzenbewertungen mit Treatment Switching (Behandlungswechsel von Kontrolle zu Prüfintervention) (Fortsetzung)

Name der Studie Intervention	Population	Studiendesign, Zeitraum	Interventionen (Zahl der randomisierten Patientinnen und Patienten)	Behandlungsdauer	primärer Endpunkt, patientenrelevante sekundäre Endpunkte	geplante Beobachtungsdauer für die patientenrelevanten Endpunkte
AFFIRM Intervention: Enzalutamid	erwachsene Männer mit metastasiertem kastrationsresistentem Prostatakarzinom, deren Erkrankung während oder nach einer Chemotherapie mit Docetaxel fortschreitet	RCT, doppelblind, parallel, placebokontrolliert Zeitraum: 09/2009–09/2011 (Datenschnitt der Interimsanalyse) Die Interimsanalyse nach dem 520. Todesfall führte zur Beendigung der Studie.	Enzalutamid + BSC (N = 800) Placebo + BSC (N = 399)	Behandlung bis zum Auftreten einer nicht akzeptablen Toxizität, bis zum Auftreten einer bestätigten Progression und Beginn einer systemischen antineoplastischen Therapie oder bis zum Tod	Primärer Endpunkt: ▪ Gesamtüberleben Sekundäre Endpunkte: ▪ skelettbezogene Komplikationen, ▪ Schmerz, ▪ gesundheitsbezogene Lebensqualität, ▪ unerwünschte Ereignisse	Unerwünschte Ereignisse: bis 30 Tage nach Absetzen der Studienmedikation Gesundheitsbezogene Lebensqualität: bis zum Ende der Behandlung Gesamtüberleben und Zeit bis zur 1. skelettbezogenen Komplikation: Langzeitnachbeobachtung, alle 12 Wochen bis zum Studienende

(Fortsetzung)

Tabelle 2: Allgemeine Angaben zu Studien aus Nutzenbewertungen mit Treatment Switching (Behandlungswechsel von Kontrolle zu Prüfintervention) (Fortsetzung)

Name der Studie Intervention	Population	Studiendesign, Zeitraum	Interventionen (Zahl der randomisierten Patientinnen und Patienten)	Behandlungsdauer	primärer Endpunkt, patientenrelevante sekundäre Endpunkte	geplante Beobachtungsdauer für die patientenrelevanten Endpunkte
BRF113683 BREAK-3 Intervention: Dabrafenib	erwachsene Patientinnen und Patienten mit histologisch gesichertem fortgeschrittenem (nicht resezierbar, Stadium III) oder metastasiertem (Stadium IV) Melanom und nachgewiesener BRAF-V600E-Mutation, ECOG-Performance-Status 0–1, keiner Vorbehandlung des fortgeschrittenen / metastasierten Melanoms mit Krebstherapien (außer IL-2), vorangegangene chirurgische Eingriffe und Radiotherapie waren erlaubt	RCT, offen, parallel, aktiv kontrolliert Zeitraum: 12/2010–09/2011 (Zeitraum der Randomisierung) primäre Analyse beim Auftreten von 102 PFS-Ereignissen und beim Einschluss von allen Patientinnen und Patienten Datenauswertung der primären Analyse: 19.12.2011 Weitere Auswertungen: ▪ 2. Datenschnitt: 25.06.2012, ▪ 3. Datenschnitt: 18.12.2012	Dabrafenib (N = 187) Dacarbazin (N = 63)	Behandlung geplant bis zur Krankheitsprogression, zum Tod, Abbruch oder zu einer intolerablen Toxizität	Primärer Endpunkt: ▪ progressionsfreies Überleben Sekundäre Endpunkte: ▪ Gesamtüberleben, ▪ Symptomatik, ▪ gesundheitsbezogene Lebensqualität, ▪ unerwünschte Ereignisse	Gesundheitsbezogene Lebensqualität, Symptomatik: bis 30 Tage nach Progression Unerwünschte Ereignisse: bis 28 Tage nach letztmaliger Einnahme der randomisierten Studienmedikation, Tag der 1. Einnahme von Dabrafenib für Cross-over-Patientinnen und -Patienten des Dacarbazinarmes, letzter Studienkontakt oder Tod Gesamtüberleben, progressionsfreies Überleben: Nachbeobachtung aller randomisierten Patientinnen und Patienten bis zum Tod

(Fortsetzung)

Tabelle 2: Allgemeine Angaben zu Studien aus Nutzenbewertungen mit Treatment Switching (Behandlungswechsel von Kontrolle zu Prüflintervention) (Fortsetzung)

Name der Studie Intervention	Population	Studiendesign, Zeitraum	Interventionen (Zahl der randomisierten Patientinnen und Patienten)	Behandlungsdauer	primärer Endpunkt, sekundäre Endpunkte	geplante Beobachtungsdauer für die patientenrelevanten Endpunkte
EMILIA Intervention: Trastuzumab Emtansin	erwachsene Patientinnen und Patienten mit HER2-positivem, inoperablem lokal fortgeschrittenem oder metastasiertem Brustkrebs, nach vorangegangener Therapie mit Trastuzumab und einem Taxan, die eine vorangegangene Therapie gegen die lokal fortgeschrittene oder metastasierte Erkrankung erhalten haben oder die ein Rezidiv während oder innerhalb von 6 Monaten nach Beendigung der adjuvanten Behandlung entwickelt haben	RCT, offen, parallel, aktiv kontrolliert Zeitraum: Studienstart 02/2009–laufend 1. Datenschnitt 01/2012 nach ca. 508 PFS-Ereignissen, 2. Datenschnitt 07/2012 nach 316 Todesfällen, geplant aufgrund der Ergebnisse des 1. Datenschnittes	Trastuzumab Emtansin (N = 495) Lapatinib + Capecitabin (N = 496) Relevante Teilpopulation mit Anthrazyklin-Vorbehandlung: Trastuzumab Emtansin (N = 303) Lapatinib + Capecitabin (N = 302)	Behandlung bis Krankheitsprogression, Tod oder Abbruch der Studienmedikation durch die Patientinnen und Patienten	Primäre Endpunkte: ▪ progressionsfreies Überleben, ▪ Gesamtüberleben Sekundäre Endpunkte: ▪ gesundheitsbezogene Lebensqualität, ▪ unerwünschte Ereignisse	Unerwünschte Ereignisse: bis 30 Tage nach letzter Gabe der Studienmedikation Gesamtüberleben, gesundheitsbezogene Lebensqualität: nach Beendigung der Studienbehandlung 3-monatliche Nachbeobachtung bis zum Tod

(Fortsetzung)

Tabelle 2: Allgemeine Angaben zu Studien aus Nutzenbewertungen mit Treatment Switching (Behandlungswechsel von Kontrolle zu Prüfintervention) (Fortsetzung)

Name der Studie Intervention	Population	Studiendesign, Zeitraum	Interventionen (Zahl der randomisierten Patientinnen und Patienten)	Behandlungsdauer	primärer Endpunkt, patientenrelevante sekundäre Endpunkte	geplante Beobachtungsdauer für die patientenrelevanten Endpunkte
BC1-06 ALSYMPCA Intervention: Radium-223- dichlorid	erwachsene Männer mit kastrationsresistentem Prostatakarzinom und symptomatischen Knochenmetastasen ohne bekannte viszerale Metastasen	RCT, doppelblind, parallel, placebokontrolliert Zeitraum: 06/2008–02/2011 (Rekrutierungszeitraum) Stichtag für 1. Interimsanalyse: 14.10.2010 Stichtag für 2. Interimsanalyse: 15.07.2011	Radium-223 + BSC (N = 614) Placebo + BSC (N = 307)	6 Injektionen im Abstand von jeweils 4 Wochen	Primärer Endpunkt: ▪ Gesamtüberleben Sekundäre Endpunkte: ▪ Zeit bis zum 1. symptomatischen skelettalen Ereignis, ▪ gesundheitsbezogene Lebensqualität, ▪ Nebenwirkungen	Gesundheitsbezogene Lebensqualität: bis Woche 20 nach Ende der Behandlung Unerwünschte Ereignisse: bis zu 12 Wochen nach der letzten Injektion Für Gesamtüberleben und skelettale Ereignisse: Nachbeobachtung für 3 Jahre
a: Die Analyse war ursprünglich als Interimsanalyse geplant und wurde wegen der Wirksamkeit von Vemurafenib als Hauptanalyse durchgeführt. ALK: anaplastische Lymphomkinase; BSC: Best supportive Care; ECOG: Eastern Cooperative Oncology Group; FACT-G: Functional Assessment of Cancer Therapy-General; FSKI: Functional Assessment of Cancer Therapy Kidney Symptom Index; RCT: randomisierte kontrollierte Studie; VEGF: vaskulärer endothelialer Wachstumsfaktor						

Alle eingeschlossenen Studien waren RCTs, entweder mit aktiver Vergleichstherapie oder placebokontrolliert. Falls die Studien placebokontrolliert waren, waren Intervention und Placebo jeweils mit Best supportive Care (BSC) kombiniert. Der primäre Endpunkt war entweder das progressionsfreie Überleben (PFS) oder das Gesamtüberleben (TARGET, AFFIRM und ASLYMPCA) oder beide Endpunkte waren koprimäre Endpunkte (COU-AA-302, EMILIA). Zu den patientenrelevanten sekundären Endpunkten gehörten unerwünschte Ereignisse (in allen Studien), die gesundheitsbezogene Lebensqualität (7 Studien) sowie das Gesamtüberleben, sofern es nicht ein primärer Endpunkt der Studie war. In einzelnen Studien wurden darüber hinaus verschiedene Endpunkte der Symptomatik der Erkrankung einschließlich Opiatgebrauch (BRIM-3, RECORD-1, COU-AA-302, PROFILE1007, DC200C00058, AFFIRM, BREAK-3) und Folgekomplikationen der Erkrankung, etwa skelettale Ereignisse (AFFIRM, ALSYMPCA), erhoben.

In den meisten Studien wurde die Behandlung mit der Interventionstherapie bis zum Eintreten einer Progression, des Todes, einer unzumutbaren Toxizität oder bis zum Abbruch der Studie fortgeführt, d. h., keine der untersuchten Therapien hatte eine fixe Behandlungsdauer.

Nahezu alle Studien gleichen sich darin, dass die Beobachtungsdauer abhängig vom beobachteten Endpunkt war. Das Gesamtüberleben wurde in der Regel bis zum Tod der Patientin oder des Patienten oder bis zum Abbruch der Studie (durch die Patientin oder den Patienten oder insgesamt) beobachtet, unabhängig davon, ob zu diesem Zeitpunkt noch die Prüf- bzw. die Kontrollintervention verabreicht wurde (in der Studie ALSYMPCA wurde das Gesamtüberleben nur 3 Jahre nachbeobachtet). In einigen Fällen galt dies auch für die Endpunkte gesundheitsbezogene Lebensqualität (EMILIA), skelettale Ereignisse (AFFIRM, ALSYMPCA), schwerer Schmerz, gemessen am Beginn einer Opiattherapie (COU-AA-302), und unerwünschte Ereignisse (BRIM-3). Andere patientenrelevante Endpunkte, insbesondere unerwünschte Ereignisse, wurden, soweit dies aus den öffentlich verfügbaren Modulen der Dossiers zu entnehmen war, nur bis zum Ende der Behandlung beobachtet. Dies bedeutet, dass nicht alle Endpunkte gleichermaßen von einem Treatment Switching betroffen sind, sondern nur jene, die über den Zeitpunkt des Behandlungswechsels hinaus beobachtet werden.

5.2.3 Methoden zur Berücksichtigung des Treatment Switchings

Angaben zum Treatment Switching in den einzelnen Studien sowie zu den Methoden, die zur Berücksichtigung desselben zur Anwendung kamen, sind in Tabelle 3 zusammengefasst.

Tabelle 3: Angaben zu Methodik und zugrunde liegenden Annahmen zur Berücksichtigung von Treatment Switching (Wechsel der Behandlung von Kontrolle zu Interventionsbehandlung)

Studie (Intervention vs. Kontrolle)	Anzahl der Patientinnen und Patienten aus dem Kontrollarm, die zur Interventionsbehandlung gewechselt haben n (%)	Grund und Zeitpunkt des Treatment Switchings	Von Treatment Switching betroffene patientenrelevante Endpunkte	Endpunkt: Methode zur Berücksichtigung von Treatment Switching	Methode a priori vorgesehen	Berücksichtigung der der Methode zugrunde liegenden Annahmen (Angaben des pU)
BRIM-3 (Vemurafenib vs. Dacarbazin)	1. Datenschnitt: 0 (0) 2. Datenschnitt: 50 (14,8) 3. Datenschnitt: 81 (24,0) 4. Datenschnitt: 83 (24,6) 5. Datenschnitt: 84 (24,9)	nach dem 1. Datenschnitt zum Gesamtüberleben auf Empfehlung des Data Safety Monitoring Boards; aufgrund der Ergebnisse dieses Datenschnitts vorzeitige Beendigung der Studie	<ul style="list-style-type: none"> ▪ Gesamtüberleben ▪ unerwünschte Ereignisse 	Gesamtüberleben: für Nutzenbewertung nur Betrachtung der Hazard Ratios für den 1. Datenschnitt (kein Behandlungswechsel); Zensierung der Patientinnen und Patienten zum Zeitpunkt des Behandlungswechsels für die weiteren Datenschnitte Unerwünschte Ereignisse: nein	ja	ja, für die Zensierung der Patientinnen und Patienten mit Behandlungswechsel
TARGET (Sorafenib vs. Placebo)	216 (47,8)	nach Zwischenauswertung des PFS (ca. 1,5 Jahre nach Studienbeginn) auf Beschluss des Data Monitoring Committee (mit Entblindung)	<ul style="list-style-type: none"> ▪ Gesamtüberleben 	Gesamtüberleben: zusätzliche Analyse mit Zensierung aller Patientinnen und Patienten zum Zeitpunkt des Behandlungswechsels	k. A.	k. A.

(Fortsetzung)

Tabelle 3: Angaben zu Methodik und zugrunde liegenden Annahmen zur Berücksichtigung von Treatment Switching (Wechsel der Behandlung von Kontrolle zu Interventionsbehandlung) (Fortsetzung)

Studie (Intervention vs. Kontrolle)	Anzahl der Patientinnen und Patienten aus dem Kontrollarm, die zur Interventionsbehandlung gewechselt haben n (%)	Grund und Zeitpunkt des Treatment Switchings	Von Treatment Switching betroffene patientenrelevante Endpunkte	Endpunkt: Methode zur Berücksichtigung von Treatment Switching	Methode a priori vorgesehen	Berücksichtigung der der Methode zugrunde liegenden Annahmen (Angaben des pU)
RECORD-1 (Everolimus vs. Placebo)	vermutlich 6 (4,3)	Studienbehandlung bis Progression, danach Entblindung und Wechsel zu Everolimus für Patientinnen und Patienten der Placebogruppe möglich	<ul style="list-style-type: none"> ▪ Gesamtüberleben 	k. A. ^a	k. A.	k. A.
PROFILE 1007 (Crizotinib vs. Chemotherapie)	Im Dossier präsentierter Datenschnitt: 108 (62,1)	von Beginn der Studie an nach Eintreten einer Progression möglich (Wechsel in 1-armige Studie PROFILE 1005); keine explizite Begründung für diese Studie, aber Diskussion ethischer Gesichtspunkte	<ul style="list-style-type: none"> ▪ Gesamtüberleben 	Gesamtüberleben: Methoden nach [32] (RPSFT) für finale Analyse geplant, im Dossier jedoch Ergebnisse aus Interimsanalysen berichtet	ja	n. z., da die geplanten Auswertungen im Dossier nicht berichtet werden

(Fortsetzung)

Tabelle 3: Angaben zu Methodik und zugrunde liegenden Annahmen zur Berücksichtigung von Treatment Switching (Wechsel der Behandlung von Kontrolle zu Interventionsbehandlung) (Fortsetzung)

Studie (Intervention vs. Kontrolle)	Anzahl der Patientinnen und Patienten aus dem Kontrollarm, die zur Interventionsbehandlung gewechselt haben n (%)	Grund und Zeitpunkt des Treatment Switchings	Von Treatment Switching betroffene patientenrelevante Endpunkte	Endpunkt: Methode zur Berücksichtigung von Treatment Switching	Methode a priori vorgesehen	Berücksichtigung der der Methode zugrunde liegenden Annahmen (Angaben des pU)
COU-AA-302 (Abirateronacetat + Prednison/ Prednisolon vs. Placebo + Prednison/ Prednisolon)	1. Datenschnitt: 0 (0) 2. Datenschnitt: 54 (10,0) 3. Datenschnitt: 78 (14,4)	nach Auswertung des 2. Datenschnitts auf Empfehlung des IDMC (32 Monate nach Studienbeginn)	<ul style="list-style-type: none"> ▪ Gesamtüberleben ▪ Gebrauch von Opiattherapien 	Gesamtüberleben: Sensitivitätsanalysen für Datenschnitt 2 und 3 ^b : Rank-preserving-Failure-Time-Modell ([32]); Iterative Parameter Estimate ([36]); nicht berücksichtigt für Nutzenbewertung Gebrauch von Opiattherapien: nein	ja	k. A.
D4200C00058 (Vandetanib + BSC vs. Placebo + BSC)	38 (63,3) (bezogen auf die für die Bewertung relevante Teilpopulation)	nach Progression für Patientinnen und Patienten beider Gruppen Wechsel von der Kontrolle zu Vandetanib bzw. offener Weiterbehandlung mit Vandetanib möglich Grund: Wechsel von Kontrolle zu Intervention bei fehlenden Behandlungsalternativen ethisch notwendig (DGHO-Empfehlung)	<ul style="list-style-type: none"> ▪ Gesamtüberleben 	Gesamtüberleben: Sensitivitätsanalysen: Per-Protocol- statt ITT-Population ausgewertet, Adjustierung für Kovariaten; allerdings nur die Ergebnisse für ITT dargestellt ^c	n. z.	n. z.

(Fortsetzung)

Tabelle 3: Angaben zu Methodik und zugrunde liegenden Annahmen zur Berücksichtigung von Treatment Switching (Wechsel der Behandlung von Kontrolle zu Interventionsbehandlung) (Fortsetzung)

Studie (Intervention vs. Kontrolle)	Anzahl der Patientinnen und Patienten aus dem Kontrollarm, die zur Interventions- behandlung gewechselt haben n (%)	Grund und Zeitpunkt des Treatment Switchings	Von Treatment Switching betroffene patientenrelevante Endpunkte	Endpunkt: Methode zur Berücksichtigung von Treatment Switching	Methode a priori vorgesehen	Berücksichtigung der der Methode zugrunde liegenden Annahmen (Angaben des pU)
CLEOPATRA WO20698 (Pertuzumab + Trastuzumab + Docetaxel vs. Placebo + Trastuzumab + Docetaxel)	k. A.	nach 2. Datenschnitt (final konfirmatorische Analyse des Gesamtüberlebens; 4 Jahre nach Studienbeginn) Entblindung und Möglichkeit des Treatment Switchings aufgrund eines statistisch signifikanten Unterschieds im Gesamtüberleben	<ul style="list-style-type: none"> ▪ Gesamtüberleben 	n. z., da nach dem 2. Datenschnitt keine weiteren Analysen zum Gesamtüberleben vorlagen	n. z.	n. z.
AFFIRM (Enzalutamid + BSC vs. Placebo + BSC)	1. Datenschnitt: 0 (0) Weitere Datenschnitte: keine Angaben dazu, wie viele Patientinnen und Patienten pro Datenschnitt bereits ein Treatment Switching hatten ^d	nach 1. Datenschnitt Wechsel zu offener Behandlung mit Enzalutamid möglich; auf Empfehlung des Data Monitoring Committee	<ul style="list-style-type: none"> ▪ Gesamtüberleben ▪ Zeit bis zur 1. skelettbezogenen Komplikation 	nur Daten des 1. Datenschnitts zur Nutzenbewertung herangezogen; für weitere Datenschnitte keine Angaben	n. z.	n. z.

(Fortsetzung)

Tabelle 3: Angaben zu Methodik und zugrunde liegenden Annahmen zur Berücksichtigung von Treatment Switching (Wechsel der Behandlung von Kontrolle zu Interventionsbehandlung) (Fortsetzung)

Studie (Intervention vs. Kontrolle)	Anzahl der Patientinnen und Patienten aus dem Kontrollarm, die zur Interventionsbehandlung gewechselt haben n (%)	Grund und Zeitpunkt des Treatment Switchings	Von Treatment Switching betroffene patientenrelevante Endpunkte	Endpunkt: Methode zur Berücksichtigung von Treatment Switching	Methode a priori vorgesehen	Berücksichtigung der der Methode zugrunde liegenden Annahmen (Angaben des pU)
BRF113683 BREAK-3 (Dabrafenib vs. Dacarbazin)	1. Datenschnitt: 28 (44,4) 2. Datenschnitt: 35 (55,6) 3. Datenschnitt: 37 (58,7)	von Beginn der Studie an nach Progression Wechsel von Dacarbazin zu Dabrafenib möglich; nach Amendment (vor 2. Datenschnitt) auch vor Progression nach Ermessen des Prüfarztes aus ethischen Gründen (mutmaßlich Benachteiligung von Patientinnen und Patienten mit Standardtherapie)	<ul style="list-style-type: none"> ▪ Gesamtüberleben 	Gesamtüberleben: Rank-preserving structural Failure Time Model (RPSFT) [32]; Iterative Parameter Estimation (IPE) [36]; Inverse Probability of Censoring Weights (IPCW) [25]	ja ^e	IPCW-Methode aufgrund der Datenlage nicht angewandt; k. A. zu den anderen Methoden
EMILIA (Trastuzumab Emtansin vs. Lapatinib + Capecitabin)	k. A.	nach 2. Datenschnitt (ca. 3 Jahre nach Studienbeginn), aufgrund eines signifikanten Vorteils von Trastuzumab Emtansin für das Gesamtüberleben	<ul style="list-style-type: none"> ▪ Gesamtüberleben ▪ gesundheitsbezogene Lebensqualität 	keine, da Behandlungswechsel erst nach final confirmatorischer Analyse zum Gesamtüberleben vorgesehen	n. z.	n. z.

(Fortsetzung)

Tabelle 3: Angaben zu Methodik und zugrunde liegenden Annahmen zur Berücksichtigung von Treatment Switching (Wechsel der Behandlung von Kontrolle zu Interventionsbehandlung) (Fortsetzung)

Studie (Intervention vs. Kontrolle)	Anzahl der Patientinnen und Patienten aus dem Kontrollarm, die zur Interventionsbehandlung gewechselt haben n (%)	Grund und Zeitpunkt des Treatment Switchings	Von Treatment Switching betroffene patientenrelevante Endpunkte	Endpunkt: Methode zur Berücksichtigung von Treatment Switching	Methode a priori vorgesehen	Berücksichtigung der der Methode zugrunde liegenden Annahmen (Angaben des pU)
BC1-06 ALSYMPCA (Radium-223-Dichlorid + BSC vs. Placebo + BSC)	bis zur finalen Analyse (2. Datenschnitt) keine Patientin und kein Patient mit Therapiewechsel	Entblindung nach Datenschnitt auf Empfehlung des IDMC zur Beendigung der Studie hin aufgrund eines statistisch signifikanten Vorteils von Radium-223-Dichlorid für das Gesamtüberleben, danach Wechsel für Patientinnen und Patienten des Placeboarms zu Radium-223-Dichlorid möglich	<ul style="list-style-type: none"> ▪ Gesamtüberleben ▪ skelettale Ereignisse 	keine, da bis zur finalen Analyse kein Therapiewechsel stattgefunden hat	n. z.	n. z.
<p>a: Studie wurde nur in Studienpool für indirekten Vergleich aufgenommen, den der pU dann aber nicht durchführt</p> <p>b: jeweils unter Ausschluss von Patientinnen und Patienten der Placebogruppe, die Abirateronacetat erhalten hatten</p> <p>c: pU zieht den Endpunkt Gesamtüberleben nicht für Bewertung des Zusatznutzens heran</p> <p>d: Insgesamt waren 22 Patientinnen und Patienten (5,5 %) betroffen.</p> <p>e: Angabe aus Modul 4 wörtlich: „Die ‚On treatment‘ Analysen schätzen den Behandlungseffekt der kompletten Behandlung [...] im Vergleich zur Kontrolle unter der Annahme, dass der Behandlungseffekt nur solange vorliegt, wie die Behandlung erfolgt. Die ‚On treatment – observed‘ Analysen schätzen den Behandlungseffekt der beobachteten Behandlungsgruppe im Vergleich zur Kontrolle unter der Annahme, dass der Behandlungseffekt verschwindet, sobald die Behandlung abgesetzt wird.“ [82]</p> <p>DGHO: Deutsche Gesellschaft für Hämatologie und Onkologie; IDMC: Independent Data Monitoring Committee; ITT: Intention to treat; k. A.: keine Angabe; n. z.: nicht zutreffend; PFS: progressionsfreies Überleben</p>						

Allen Studien ist gemein, dass die pUs bezüglich eines Behandlungswechsels grundsätzlich nur Überlegungen zum Endpunkt Gesamtüberleben anstellen, obwohl mitunter weitere Endpunkte davon betroffen sind. Aufgrund der Beobachtungszeit über den Behandlungswechsel hinaus hätten in einigen Studien außerdem die Endpunkte unerwünschte Ereignisse (Studie BRIM-3), Symptomatik der Erkrankung (Opiattherapie, COU-AA-302), gesundheitsbezogene Lebensqualität (EMILIA) und skelettale Ereignisse (AFFIRM, ALSYMPCA) betrachtet werden müssen.

Wenn den Teilnehmern einer Studie die Möglichkeit des Wechsels zur Prüflintervention gegeben wurde, musste die betreffende Studie hierzu notwendigerweise entblindet werden. Die Studien unterscheiden sich bezüglich des Zeitpunkts und der Begründung für diesen Schritt. In 4 der 11 Studien (36 %) (RECORD-1, PROFILE 1007, D4200C00058 und BREAK-3) war die Möglichkeit zum Wechsel der Therapie von Studienbeginn an möglich, wenn die Patientinnen und Patienten eine Progression der Erkrankung zeigten. Begründet wurde diese Maßnahme jeweils mit der ethischen Erwägung, dass es nicht vertretbar sei, den Patientinnen und Patienten bei Versagen der Standardtherapie und Fehlen wirksamer Behandlungsalternativen die wirksame experimentelle Therapie vorzuenthalten. Eine studienspezifische Diskussion erfolgte dabei nicht. Betroffen waren im Falle der Studie RECORD-1 vermutlich 4,3 % der Patientinnen und Patienten, bei der Studie PROFILE 1007 62,1 % (Daten zum im Dossier gezeigten Datenschnitt), bei der Studie D4200C00058 63,3 % und bei der Studie BREAK-3 44,4 % bis 58,7 % vom 1. bis zum 3. Datenschnitt.

Bei den verbleibenden 7 Studien (64 %) erfolgten die Entblindung und der darauf folgende Therapiewechsel nicht zu einem präspezifizierten Zeitpunkt, sondern aufgrund der signifikant positiven Studienergebnisse an einem der geplanten Datenschnitte. In den 6 Studien (55 %) BRIM-3, COU-AA-302, CLEOPATRA, AFFIRM, EMILIA und ALSYMPCA lagen dabei verwertbare Daten zum patientenrelevanten Endpunkt Gesamtüberleben vor, nicht aber in der Studie TARGET. Die Entblindung erfolgte mitunter auch auf Empfehlung des zuständigen Data Safety Monitoring Boards bzw. Data Monitoring Committees, so in den 4 Studien (36 %) BRIM-3, TARGET, COU-AA-302 und AFFIRM.

Die Anzahl der Patientinnen und Patienten, die von der Vergleichstherapie zur experimentellen Therapie wechselten, schwankte zwischen den Studien teils erheblich. Ein Vergleich der jeweils letzten für die Nutzenbewertung verfügbaren Datenschnitte zeigt Werte von 0 % der Studienteilnehmer (AFFIRM, ALSYMPCA) bis 63,3 % (D4200C00058); im Median betrug die Zahl der Patientinnen und Patienten mit Behandlungswechsel 19,7 %. Bei Vorliegen mehrerer Datenschnitte nach der Entblindung innerhalb derselben Studie nahm der Anteil der betroffenen Patientinnen und Patienten mit fortschreitender Beobachtungsdauer erwartungsgemäß zu. So stieg der Anteil bei der Studie BRIM-3 von 14,8 % auf 24,9 %, bei der Studie COU-AA-302 von 10,0 % auf 14,4 % und bei der Studie BREAK-3 von 44,4 % auf 58,7 %.

Die Hersteller gehen mit dem Problem des Treatment Switchings unterschiedlich um. Methoden zur Berücksichtigung von Treatment Switching nach Robins und Finkelstein [10] (IPCW-Methode, siehe Abschnitt 5.1.4.1), Robins und Tsiatis [32] (RPSFT-Methode, siehe Abschnitt 5.1.4.2) oder Branson und Whitehead [36] (IPE-Methode, siehe Abschnitt 5.1.4.2) wurden insgesamt nur in 3 (27 %) der Studien a priori geplant und angewendet (PROFILE 1007, COU-AA-302, BREAK-3). In den Dossiers selbst werden die Ergebnisse dieser Analysen allerdings nur für die Studie BREAK-3 präsentiert. In der Studie PROFILE 1007 war die Adjustierung der Daten jedoch erst für die finale Analyse geplant, im Dossier werden nur Ergebnisse aus einer Interimsanalyse ohne Adjustierung berichtet. In der Studie COU-AA-302 wurden die adjustierten Daten im Rahmen von Sensitivitätsanalysen im Dossier dargestellt, wobei die Methoden IPE und RPSFT angewendet wurden. Sie wurden jedoch nicht für die Nutzenbewertung berücksichtigt, da sie aufgrund der Annahmen, die den angewandten Methoden zugrunde liegen, als zu unsicher angesehen wurden. Zudem wurde in der Nutzenbewertung der verzerrende Einfluss des Therapiewechsels auf die ITT-Analyse als eher niedrig eingestuft.

In den Studien BRIM-3 und TARGET wurden zusätzlich zur ITT-Analyse des Gesamtüberlebens Behandlungswechsler zum Zeitpunkt des Wechsels zensiert (siehe Abschnitt 5.1.3.3). Die Ergebnisse beider Auswertungsarten gingen für die Studie BRIM-3 in die Nutzenbewertung ein, wobei sich bezüglich des Ausmaßes des abgeleiteten Zusatznutzens kein Unterschied zwischen den Auswertungen zeigte. In der Studie D4200C00058 wurde im Rahmen einer Sensitivitätsanalyse statt der ITT- die Per-Protocol(PP)-Population für die Analyse des Gesamtüberlebens ausgewertet, die nur Patientinnen und Patienten ohne Behandlungswechsel umfasste, d. h., Patientinnen und Patienten, die die Behandlung gewechselt hatten, wurden von der Analyse ausgeschlossen (siehe auch Abschnitt 5.1.3.2). Allerdings werden im Dossier des Herstellers nur die Ergebnisse der ITT-Auswertung gezeigt.

Schließlich gibt es eine Reihe von Studien, in denen nach der Entblindung der Studie und einem möglichen Treatment Switching keine weiteren Analysen zum Gesamtüberleben mehr durchgeführt wurden oder bei denen im Dossier des pU nur solche Datenschnitte vorlagen, die noch nicht vom Treatment Switching betroffen waren, sodass sich eine besondere Berücksichtigung der Patientinnen und Patienten mit Behandlungswechsel erübrigte. Dies war in 4 von 11 Studien (36 %; CLEOPATRA, AFFIRM, EMILIA und ALSYMPCA) der Fall.

Sofern besondere Methoden für die Berücksichtigung von Treatment Switching zum Einsatz kamen, waren diese in allen Studien a priori geplant oder es fehlen entsprechende Angaben dazu (Studien RECORD-1 und TARGET). Wie in Abschnitt 5.1.4 beschrieben, liegen Methoden wie RPSFT und IPCW Annahmen zugrunde, deren Nichtbeachtung mit hoher Wahrscheinlichkeit zu nicht validen Ergebnissen führt [10,32]. Welche Annahmen im jeweiligen Fall zu treffen wären und ob diese erfüllt sind, wird in keinem der vorgelegten Dossiers diskutiert. Damit ist die Validität der dadurch gewonnenen Ergebnisse fraglich. Es zeigt sich somit, dass die hier betrachteten Dossiers, in denen diese Methoden zur Anwendung kamen, bezogen auf das Treatment Switching methodisch unvollständig sind.

Tabelle 4 zeigt, in wie vielen Fällen bei den hier betrachteten frühen Nutzenbewertungen zumindest ein Anhaltspunkt für einen Zusatznutzen abgeleitet werden konnte und in welchem Zusammenhang Treatment Switching damit stand.

Tabelle 4: Ableitung eines Zusatznutzens bei Dossierbewertungen, in die Studien mit Treatment Switching eingeschlossen wurden

Dossierbewertungen	N (%)
Gesamtzahl der betrachteten Arzneimittel	9 (100) ^a
▪ Arzneimittel, für die wenigstens ein Anhaltspunkt für einen Zusatznutzen abgeleitet werden konnte	6 (67)
▫ davon basierend auf Datenschnitten ohne Treatment Switching	5 (56)
▫ davon trotz möglicher Verzerrung durch Treatment Switching	1 (11)
▪ Arzneimittel, für die kein Zusatznutzen belegt war	3 (33)
▫ davon aufgrund von Treatment Switching	1 (11)
a: Nicht eingeschlossen sind hier die Studien TARGET und RECORD-1, die im Dossier des pU zu Axitinib (A12-14) aufseiten der Vergleichstherapie für einen indirekten Vergleich eingeschlossen sind, aber nicht zur Ableitung des Zusatznutzens herangezogen wurden.	

Von den 9 hier betrachteten onkologischen Arzneimitteln wurde in 6 Fällen (67 %) mindestens ein Anhaltspunkt für einen Zusatznutzen des zu bewertenden Arzneimittels abgeleitet. Charakteristisch für alle diese Bewertungen ist, dass ein Zusatznutzen nur dann abgeleitet werden konnte, wenn Auswertungen vorlagen, für die ein Behandlungswechsel zur Prüflintervention nicht berücksichtigt werden musste oder als nicht entscheidend für die Aussagekraft angesehen wurde.

Bei den Bewertungen zu Vemurafenib, Pertuzumab, Enzalutamid, Trastuzumab Emtansin und Radium-223-Dichlorid standen jeweils Datenschnitte zur Verfügung, an denen noch kein Behandlungswechsel zur Prüflintervention stattgefunden hatte. Im Falle von Abirateronacetat war das zwar der Fall, allerdings waren weniger als 15 % der Patientinnen und Patienten des Kontrollarms betroffen. Außerdem hatten zum Zeitpunkt des für die Bewertung entscheidenden Datenschnitts bereits sehr viele Patientinnen und Patienten die Studie ganz abgebrochen oder wurden bereits mit Folgetherapien behandelt, sodass der verzerrende Einfluss des Behandlungswechsels als insgesamt gering angesehen wurde. Bei der Nutzenbewertung von Axitinib wurden die Studien, in denen Treatment Switching erlaubt war (TARGET, RECORD-1), aus anderen Gründen nicht in die Bewertung eingeschlossen.

Im Falle der Nutzenbewertung von Dabrafenib, in der das Herstellerdossier umfangreiche Analysen zur Berücksichtigung des Behandlungswechsels der Patientinnen und Patienten eingereicht hatte, konnte kein Zusatznutzen abgeleitet werden, da die Annahmen, die den verwendeten Methoden zugrunde liegen, im Dossier nicht in ausreichendem Maß diskutiert wurden. Sie wurden daher als zu unsicher angesehen, um die Ergebnisse zum Gesamtüberleben in die Nutzenbewertung einzubeziehen.

Somit konnte bisher in keiner frühen Nutzenbewertung ein Zusatznutzen abgeleitet werden, für die Studien mit Auswertungszeitpunkten vorgelegt wurden, zu denen in wesentlichem Umfang Treatment Switching stattgefunden hatte.

6 Diskussion

Die in Abschnitt 5.1 dargestellten Überlegungen, insbesondere jene zur Darstellung und zum Nutzen von ITT-Analysen in Studiendesigns mit komplexeren Situationen (siehe Bild 1b in Abbildung 2), zeigen, dass Verzerrungen der Ergebnisse aufgrund von Behandlungswechseln i. d. R. nicht methodisch aufgefangen werden können. Für die frühe Nutzenbewertung von Onkologika ist dies insbesondere von Bedeutung, weil hier nicht der reine Therapieeffekt der Prüfintervention betrachtet wird, sondern die Effektschätzung der Therapiestrategie, die mit dem zu prüfenden Arzneimittel beginnt (im Vergleich zu einer Therapiestrategie ohne das Arzneimittel). Die vorgestellten komplexen Methoden zielen jedoch auf die Schätzung reiner Therapieeffekte ab.

Ein Beispiel für eine komplexe Therapiesituation ist die Nutzenbewertung zu Sipuleucel-T [87]. In den 3 Studien, die in die Bewertung eingeschlossen wurden, wurde zunächst Sipuleucel-T mit einer Scheinbehandlung verglichen. Nach bestätigter Krankheitsprogression wurden die Patientinnen und Patienten entblindet. Patientinnen und Patienten des Sipuleucel-T-Arms wechselten nach der Progression zu einer weiteren Therapie nach Ermessen des behandelnden Arztes, i. d. R. Docetaxel. Patientinnen und Patienten der Kontrollgruppe wechselten jedoch zunächst mehrheitlich zu einer Sipuleucel-T-Behandlung; nach weiterer Progression wechselten auch diese Patientinnen und Patienten zu einer Folgetherapie, häufig wiederum Docetaxel. Somit erhielten beide Patientengruppen nach (wiederholter) Progression eine Therapie mit bekanntem positivem Einfluss auf das Gesamtüberleben, jedoch die Patientinnen und Patienten des Vergleichsarms zeitlich später als jene des Sipuleucel-T-Arms. Den Patientinnen und Patienten des Vergleichsarms wurde also eine in der entsprechenden Therapielinie wirksame Therapie bis zu einem weiteren Progress vorenthalten. Somit kann nicht ausgeschlossen werden, dass unter der Annahme, dass Sipuleucel-T keinen Effekt auf das Gesamtüberleben hat, der beobachtete Effekt auf das Gesamtüberleben allein auf dem zeitlichen Versatz der Folgetherapien beruht. Damit läge eine Benachteiligung des Vergleichsarms vor, sodass eine ITT-Analyse des Gesamtüberlebens in diesem Fall mit Bezug auf die Prüfintervention nicht konservativ wäre, sondern zu einer Überschätzung des Effekts von Sipuleucel-T führen würde. Im vorliegenden Fall konnte der pU jedoch durch sinnvolle Sensitivitätsanalysen zeigen, dass der beobachtete Effekt zugunsten von Sipuleucel-T nicht vollständig auf die verzögerte Gabe von Docetaxel zurückzuführen war. Hierzu wurden u. a. das Gesamtüberleben unter Annahme eines positiven Effekts von Docetaxel sowie das Gesamtüberleben von Patientinnen und Patienten mit und ohne Docetaxel-Folgetherapie getrennt ausgewertet. Details hierzu finden sich im Addendum zur Nutzenbewertung von Sipuleucel-T [88].

Da die Verabreichung wirksamer Folgetherapien, sofern solche zur Verfügung stehen, nach Versagen der ursprünglichen Prüfintervention bzw. Vergleichstherapie ethisch und klinisch geboten erscheint, sollte in solchen Konstellationen auf Studiendesigns mit vorzeitigem Wechsel von der Vergleichs- zur Prüfintervention verzichtet werden. Insbesondere das Argument, ein Wechsel von der Vergleichs- zur Prüfintervention sei in einer solchen

Therapiesituation ethisch geboten, um Patientinnen und Patienten mit wenigen Behandlungsalternativen eine vermeintlich wirksame Therapie nicht vorzuenthalten, kann hier nicht aufrechterhalten werden. Wenn in Studien mit einem Design, welches Treatment Switching erlaubt, nicht gewährleistet werden kann, dass der (Zusatz-)Nutzen einer Therapie mit ausreichender Ergebnissicherheit nachgewiesen werden kann, kann der Einschluss von Patientinnen und Patienten in solche Studien ethisch als nicht vertretbar angesehen werden (siehe auch Prasad 2013 und 2014 [89,90]). Zudem ist der primäre Endpunkt in onkologischen Studien, anhand dessen die Wirksamkeit einer Therapie beurteilt wird, in der Regel das PFS, aber nicht der patientenrelevante Endpunkt Gesamtüberleben, sodass ein Behandlungswechsel oft stattfindet, bevor ein Effekt der Therapie auf das Gesamtüberleben nachgewiesen ist.

Insbesondere die in Abschnitt 5.1.4 vorgestellten komplexeren Methoden sind in ihrer Anwendung an eine Reihe von Voraussetzungen und Annahmen gebunden, die in den meisten klinischen Studien in der Regel nicht vollständig überprüfbar sind. So liegt der IPCW-Methode nach Robins und Finkelstein [10] u. a. die Annahme der *No unmeasured Confounders* zugrunde, d. h., ihre Validität ist an die – methodisch korrekte – Berücksichtigung aller möglichen Störgrößen gebunden, die sowohl den Effekt auf das Gesamtüberleben als auch die Wahrscheinlichkeit einer Zensierung beeinflussen. Dies ist im Rahmen einer onkologischen klinischen Studie in der Regel nicht überprüfbar, sodass eine verzerrungsfreie Anwendung der IPCW-Methode nicht möglich ist. Gleiches gilt für die Annahme des *Common Treatment Effects* im Rahmen der RPSFT-Methode. Diese setzt zum einen voraus, dass die Wirkung einer Therapie unabhängig davon ist, zu welchem Zeitpunkt im Studienverlauf eine Patientin oder ein Patient die entsprechende Therapie bekommt. Des Weiteren ist für 2 beliebige Patientinnen oder Patienten die Reihenfolge bis zum Eintreten eines untersuchten Ereignisses (Progression, Versterben) unter beiden zu vergleichenden Therapien gleich. Somit können die Annahmen, die für eine valide Anwendung dieser verbreiteten Methoden grundlegend sind, in der Regel nicht auf ihre Richtigkeit überprüft werden, was Analysen, die allein auf der Auswertung nach dem IPCW oder der RPSFT beruhen, nicht geeignet für eine Bewertung erscheinen lässt.

An klinische Studien, in denen Treatment Switching vorgesehen ist, sollten daher folgende Anforderungen gestellt werden, um eine Bewertung mit ausreichender Aussagesicherheit zu gewährleisten:

Neben der Darstellung der ITT-Daten sollte, wenn möglich, wenigstens 1 Auswertung dargestellt werden, die nicht vom Behandlungswechsel betroffen ist. In der Regel ist dies ein Datenschnitt, der stattfindet, bevor eine relevante Anzahl von Patientinnen und Patienten die Behandlung (i. d. R. nach einem Progress, aber auch vor einem a priori geplanten Wechsel) gewechselt hat. Damit solche Auswertungen aussagekräftig sind, sind wiederum die Fallzahl der Studie und das Signifikanzniveau dieser Interimsanalyse entsprechend anzupassen.

Gemäß den aktuellen Methoden des IQWiG wird das Ausmaß des Zusatznutzens eines Arzneimittels anhand der oberen Grenze des 95 %-Konfidenzintervalls des Effektschätzers bestimmt (bei diskreten Variablen, etwa dem Gesamtüberleben) [64]. Für alle im vorliegenden

Arbeitspapier genannten Methoden gilt jedoch, dass i. d. R. weder die Richtung noch das Ausmaß einer möglichen Verzerrung mit ausreichender Sicherheit abgeschätzt werden kann. Damit ist die Anwendung der oben genannten Grenze zur Abschätzung des Ausmaßes des Zusatznutzens nicht möglich. Sofern überhaupt ein Zusatznutzen der Prüfintervention gegenüber der zweckmäßigen Vergleichstherapie festgestellt werden kann, wird i. d. R. das Ausmaß des Zusatznutzens nicht quantifizierbar sein.

Weiter ist zu beachten, dass die obigen Betrachtungen nur einen einzigen Fall betreffen, und zwar die Auswirkungen eines Wechsels von der Kontrolle zur Prüfintervention auf den Endpunkt Gesamtüberleben. Grundsätzlich stellt sich das Problem aber für alle patientenrelevanten Endpunkte, die in klinischen Studien mit Treatment Switching erhoben werden. Für das Gesamtüberleben kann in bestimmten Situationen bei einem tatsächlichen Vorteil der Prüfintervention erwartet werden, dass ein derartiger Behandlungswechsel zu einer Unterschätzung des Effekts führt. Dies kann sich für andere Endpunkte jedoch völlig anders darstellen. So könnte das Nebenwirkungsprofil bei einem Wechsel von der Kontrolle zur Prüfintervention zugunsten von Letzterer verzerrt sein, falls diese im Behandlungsverlauf mehr Nebenwirkungen hervorruft. Gleiches gilt für Endpunkte zur Morbidität (Symptomskalen u. a.) und gesundheitsbezogenen Lebensqualität. Gerade bei frühen Nutzenbewertungen sind diese Endpunkte, so sie in den untersuchten Studien erhoben wurden, wesentlich für die Bestimmung des Ausmaßes eines Zusatznutzens. Es ist nicht vertretbar, hier dem Gesamtüberleben einen höheren Stellenwert gegenüber anderen patientenrelevanten Endpunkten einzuräumen, indem auf eine Berücksichtigung des Behandlungswechsels bei diesen verzichtet wird.

In den hier betrachteten Studien wurde mit wenigen Ausnahmen nur das Gesamtüberleben weiterbeobachtet, nachdem die Patientinnen und Patienten die Behandlung gewechselt hatten. Für zukünftige klinische Forschung ist daher zu fordern, alle patientenrelevanten Endpunkte in die Betrachtungen zum Treatment Switching einzubeziehen. Dazu müssen sie nicht nur generell Teil der statistischen Betrachtungen sein, sondern vor allem auch über den Behandlungswechsel hinaus weiterbeobachtet werden.

Schlussfolgerungen

Treatment Switching ist in klinischen Studien zu onkologischen Arzneimitteln weitverbreitet, wobei insbesondere ethische Gesichtspunkte als Begründung für entsprechende Studiendesigns angeführt werden. Zur Berücksichtigung der hierdurch entstehenden Verzerrung in der Bewertung des Nutzens dieser Wirkstoffe steht eine Reihe von statistischen Verfahren zur Verfügung. Ihre Anwendung birgt jedoch teils selbst das Risiko einer verzerrten Auswertung oder ist an die Berücksichtigung von Annahmen gebunden, die nicht überprüfbar sind. Aktuelle Beispiele klinischer Studien zeigen, dass es komplexe Studiendesigns mit einem Behandlungswechsel gibt, in denen isoliert betrachtet kein statistisches Verfahren eine unverzerrte Auswertung der Studiendaten gewährleistet.

Im Hinblick auf Bewertungen des Zusatznutzens von Onkologika, denen Studien mit Treatment Switching zugrunde liegen, sollte nach Möglichkeit ein Datenschnitt zur Verfügung stehen, an

dem ein Treatment Switching noch nicht oder nur in vernachlässigbarem Umfang stattgefunden hat. Nur solche Daten ermöglichen sicher durch Treatment Switching unverzerrte Aussagen zu patientenrelevanten Endpunkten.

Bei der Darstellung von Daten, die unter dem Einfluss von Treatment Switching stehen, sollten mindestens die folgenden Anforderungen an die Darstellung der Studiendaten erfüllt sein:

Grundsätzlich sollte immer das Ergebnis der ITT-Analyse gezeigt werden. Wenn zusätzliche Analysen zur Berücksichtigung von Treatment Switching durchgeführt wurden, sollten folgende Angaben vorhanden sein:

- Anzahl der Patientinnen und Patienten, die die Behandlung wechseln
- Zeit bis zum Behandlungswechsel (Kaplan-Meier-Kurven)
- Gründe für den Behandlungswechsel
- Art, Anzahl und Dauer der Folgetherapien
- Beschreibung des Mechanismus, der zum Behandlungswechsel führt (z. B. Progression der Erkrankung), unter Berücksichtigung möglichst aller zeitabhängigen Variablen, die sowohl den Behandlungswechsel als auch das Gesamtüberleben beeinflussen
- Diskussion von Faktoren, die das Verzerrungspotenzial und die Präzision der beobachteten Effekte beeinflussen können
- ausführliche, nachvollziehbare Beschreibung der statistischen Analysen, soweit es sich nicht um Standardverfahren handelt; idealerweise werden die zur Analyse verwendeten Quellcodes und Programme mit den Daten zur Verfügung gestellt

Auch unter Berücksichtigung dieser Kriterien bleibt bei der Dateninterpretation zu beachten, dass derzeit keine statistische Methodik existiert, die die Analyse des Endpunkts Gesamtüberleben unter dem Einfluss von Treatment Switching mit ausreichender Sicherheit ermöglicht. Das Ausmaß und die Richtung daraus entstehender Verzerrungen der Effekte bleiben nicht abschätzbar.

7 Fazit

Die Anwendung statistischer Verfahren zur Berücksichtigung von Treatment Switching birgt teils selbst das Risiko einer verzerrten Auswertung oder ist an die Berücksichtigung von Annahmen zur Wirksamkeit des Arzneimittels gebunden, die in der Praxis nicht überprüfbar sind. Aktuelle Beispiele klinischer Studien zeigen, dass komplexe Studiendesigns mit einem Behandlungswechsel angewendet werden, in denen kein statistisches Verfahren mehr eine unverzerrte Auswertung der Studiendaten gewährleistet. Derzeit existiert keine statistische Methodik, die die Analyse patientenrelevanter Endpunkte unter dem Einfluss von Treatment Switching mit ausreichender Sicherheit ermöglicht.

8 Literatur

1. Morden JP, Lambert PC, Latimer N, Abrams KR, Wailoo AJ. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. *BMC Med Res Methodol* 2011; 11: 4.
2. Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP et al. Adjusting survival time estimates to account for treatment switching in randomized controlled trials: an economic evaluation context; methods, limitations, and recommendations. *Med Decis Making* 2014; 34(3): 387-402.
3. Watkins C, Huang X, Latimer N, Tang Y, Wright EJ. Adjusting overall survival for treatment switches: commonly used methods and practical application. *Pharm Stat* 2013; 12(6): 348-357.
4. Latimer NR, Abrams K, Lambert P, Crowther M, Wailoo A, Morden J et al. Adjusting for treatment switching in randomised controlled trials: a simulation study and a simplified two-stage method. *Stat Methods Med Res* 2017; 26(2): 724-751.
5. Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP et al. Adjusting survival time estimates to account for treatment switching in randomised controlled trials: a simulation study [online]. 03.2013 [Zugriff: 02.07.2014]. (HEDS Discussion Papers; Band 13.06). URL: <http://eprints.whiterose.ac.uk/75249/1/HEDSDP1306.pdf>.
6. Latimer NR, Abrams KR, Lambert PC, Morden JP, Crowther MJ. Assessing methods for dealing with treatment switching in clinical trials: a follow-up simulation study. *Stat Methods Med Res* 2018; 27(3): 765-784.
7. Tappenden P, Harnan S, Ren S, Thokala P, Wong R, Mukuria C et al. Olaparib for maintenance treatment of BRCA 1 or 2 mutated, relapsed, platinum-sensitive ovarian, fallopian tube and peritoneal cancer in people whose relapsed disease has responded to platinum-based chemotherapy: an Evidence Review Group perspective of a NICE Single Technology Appraisal. *Pharmacoeconomics* 2017; 35(1): 97-109.
8. White IR, Babiker AG, Walker S, Darbyshire JH. Randomization-based methods for correcting for treatment changes: examples from the Concorde trial. *Stat Med* 1999; 18(19): 2617-2634.
9. White IR. Uses and limitations of randomization-based efficacy estimators. *Stat Methods Med Res* 2005; 14(4): 327-347.
10. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; 56(3): 779-788.
11. Korhonen PA, Laird NM, Palmgren J. Correcting for non-compliance in randomized trials: an application to the ATBC Study. *Stat Med* 1999; 18(21): 2879-2897.

12. Robins JM. The analysis of randomized and nonrandomized AIDS treatment using a new approach to causal interference in longitudinal studies. In: Sechrest L, Freeman H, Mulley A (Ed). *Health Service Research Methodology: a focus on AIDS*. Washington, D.C.: U. S. Public Health Service, National Center for Health Services Research; 1989. S. 113-159.
13. White IR. Survival analysis of randomized clinical trials adjusted for patients who switch treatments. *Stat Med* 1997; 16(22): 2619-2620.
14. Ishak KJ, Proskorovsky I, Korytowsky B, Sandin R, Faivre S, Valle J. Methods for adjusting for bias due to crossover in oncology trials. *Pharmacoeconomics* 2014; 32(6): 533-546.
15. Bellamy SL, Lin JY, Ten Have TR. An introduction to causal modeling in clinical trials. *Clin Trials* 2007; 4(1): 58-73.
16. Little RJ, Long Q, Lin X. A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics* 2009; 65(2): 640-649.
17. Yamaguchi T, Ohashi Y. Adjusting for differential proportions of second-line treatment in cancer clinical trials; part I: structural nested models and marginal structural models to test and estimate treatment arm effects. *Stat Med* 2004; 23(13): 1991-2003.
18. Yamaguchi T, Ohashi Y. Adjusting for differential proportions of second-line treatment in cancer clinical trials; part II: an application in a clinical trial of unresectable non-small-cell lung cancer. *Stat Med* 2004; 23(13): 2005-2022.
19. Bond SJ, White IR, Walker AS. Instrumental variables and interactions in the causal analysis of a complex clinical trial. *Stat Med* 2007; 26(7): 1473-1496.
20. Simes J, Voysey M, O'Connell R, Glasziou P, Best JD, Scott R et al. A novel method to adjust efficacy estimates for uptake of other active treatments in long-term clinical trials. *PLoS One* 2010; 5(1): e8580.
21. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* 2006; 60(7): 578-586.
22. Howe CJ, Cole SR, Chmiel JS, Munoz A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. *Am J Epidemiol* 2011; 173(5): 569-577.
23. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008; 168(6): 656-664.
24. Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Morden JP. Assessing methods for dealing with treatment crossover in clinical trials: a follow-up simulation study [online]. [Zugriff: 21.08.2014]. (HEDS Discussion Papers; Band 14.01). URL: http://www.shef.ac.uk/polopoly_fs/1.358847!/file/1401.pdf.

25. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; 11(5): 550-560.
26. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of Zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; 11(5): 561-570.
27. Colleoni M, Giobbie-Hurder A, Regan MM, Thurlimann B, Mouridsen H, Mauriac L et al. Analyses adjusting for selective crossover show improved overall survival with adjuvant letrozole compared with tamoxifen in the BIG 1-98 study. *J Clin Oncol* 2011; 29(9): 1117-1124.
28. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J Am Stat Assoc* 2001; 96(454): 440-448.
29. Hernán MA, Cole SR, Margolick J, Cohen M, Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiol Drug Saf* 2005; 14(7): 477-491.
30. Sternberg CN, Hawkins RE, Wagstaff J, Salman P, Mardiak J, Barrios CH et al. A randomised, double-blind phase III study of pazopanib in patients with advanced and/or metastatic renal cell carcinoma: final overall survival results and safety update. *Eur J Cancer* 2013; 49(6): 1287-1296.
31. Cox DR, Oakes D. *Analysis of survival data*. London: Chapman and Hall; 1984.
32. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Commun Stat Theory Methods* 1991; 20(8): 2609-2631.
33. Greenland S, Lanes S, Jara M. Estimating effects from randomized trials with discontinuations: the need for intent-to-treat design and G-estimation. *Clin Trials* 2008; 5(1): 5-13.
34. Mark SD, Robins JM. A method for the analysis of randomized trials with compliance information: an application to the Multiple Risk Factor Intervention Trial. *Control Clin Trials* 1993; 14(2): 79-97.
35. Robins JM, Greenland S. Adjusting for differential rates of prophylaxis therapy for PCP in high- versus low-dose AZT treatment arms in an AIDS randomized trial. *J Am Stat Assoc* 1994; 89(427): 737-749.
36. Branson M, Whitehead J. Estimating a treatment effect in survival studies in which patients switch treatment. *Stat Med* 2002; 21(17): 2449-2463.
37. Walker AS, White IR, Babiker AG. Parametric randomization-based methods for correcting for treatment changes in the assessment of the causal effect of treatment. *Stat Med* 2004; 23(4): 571-590.

38. Korhonen P, Zuber E, Branson M, Hollaender N, Yateman N, Katiskalahti T et al. Correcting overall survival for the impact of crossover via a rank-preserving structural failure time (RPSFT) model in the RECORD-1 trial of everolimus in metastatic renal-cell carcinoma. *J Biopharm Stat* 2012; 22(6): 1258-1271.
39. Odondi L, McNamee R. Performance of statistical methods for analysing survival data in the presence of non-random compliance. *Stat Med* 2010; 29(29): 2994-3003.
40. Ouwens M, Hauch O, Franzen S. A validation study of the rank-preserving structural failure time model: confidence intervals and unique, multiple, and erroneous solutions. *Med Decis Making* 2018; 38(4): 509-519.
41. Bennett I, Paracha N, Abrams K, Ray J. Accounting for uncertainty in decision analytic models using rank preserving structural failure time modeling: application to parametric survival models. *Value Health* 2018; 21(1): 105-109.
42. Bowden J, Seaman S, Huang X, White IR. Gaining power and precision by using model-based weights in the analysis of late stage cancer trials with substantial treatment switching. *Stat Med* 2016; 35(9): 1423-1440.
43. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Dabrafenib: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A13-35 [online]. 23.12.2013 [Zugriff: 02.01.2014]. (IQWiG-Berichte; Band 203). URL: https://www.iqwig.de/download/A13-35_Dabrafenib_Nutzenbewertung-35a-SGB-V.pdf.
44. White IR, Walker AS, Babiker AG. An approximate randomisation-respecting adjustment to the hazard ratio for time-dependent treatment switches in clinical trials [online]. 11.05.2004 [Zugriff: 10.11.2014]. URL: <http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/Time-dependent.pdf>.
45. Latimer NR, Abrams KR, Amonkar MM, Stapelkamp C, Swann RS. Adjusting for the confounding effects of treatment switching: the BREAK-3 trial; dabrafenib versus dacarbazine. *Oncologist* 2015; 20(7): 798-805.
46. Latimer NR, Amonkar MM, Stapelkamp C, Sun P. Adjusting for confounding effects of treatment switching in a randomized phase II study of dabrafenib plus trametinib in BRAF V600+ metastatic melanoma. *Melanoma Res* 2015; 25(6): 528-536.
47. Latimer NR, Bell H, Abrams KR, Amonkar MM, Casey M. Adjusting for treatment switching in the METRIC study shows further improved overall survival with trametinib compared with chemotherapy. *Cancer Med* 2016; 5(5): 806-815.
48. Schlumberger M, Tahara M, Wirth LJ, Robinson B, Brose MS, Elisei R et al. Lenvatinib versus placebo in radioiodine-refractory thyroid cancer. *N Engl J Med* 2015; 372(7): 621-630.
49. Demetri GD, Garrett CR, Schöffski P, Shah MH, Verweij J, Leyvraz S et al. Complete longitudinal analyses of the randomized, placebo-controlled, phase III trial of sunitinib in patients with gastrointestinal stromal tumor following imatinib failure. *Clin Cancer Res* 2012; 18(11): 3170-3179.

50. Fewell Z, Hernán MA, Wolfe F, Tilling K, Choi H, Sterne JAC. Controlling for time-dependent confounding using marginal structural models. *Stata J* 2004; 4(4): 402-420.
51. Sommer A, Zeger SL. On estimating efficacy from clinical trials. *Stat Med* 1991; 10(1): 45-52.
52. Sussman JB, Hayward RA. An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials. *BMJ* 2010; 340: c2073.
53. Matsuyama Y. A comparison of the results of intent-to-treat, per-protocol, and g-estimation in the presence of non-random treatment changes in a time-to-event non-inferiority trial. *Stat Med* 2010; 29(20): 2107-2116.
54. Parpia S, Julian JA, Thabane L, Gu C, Whelan TJ, Levine MN. Treatment crossovers in time-to-event non-inferiority randomised trials of radiotherapy in patients with breast cancer. *BMJ Open* 2014; 4(10): e006531.
55. Skaltsa K, Ivanescu C, Naidoo S, Phung, Holmstrom S, Latimer NR. Adjusting overall survival estimates after treatment switching: a case study in metastatic castration-resistant prostate cancer. *Target Oncol* 2017; 12(1): 111-121.
56. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran ME, Berry D (Ed). *Statistical models in epidemiology, the environment, and clinical trials*. New York: Springer; 2000. (IMA Volumes in Mathematics and its Applications; Band 116).
57. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JA. Methods for dealing with time-dependent confounding. *Stat Med* 2013; 32(9): 1584-1618.
58. Chanan-Khan A, Cramer P, Demirkan F, Fraser G, Silva RS, Grosicki S et al. Ibrutinib combined with bendamustine and rituximab compared with placebo, bendamustine, and rituximab for previously treated chronic lymphocytic leukaemia or small lymphocytic lymphoma (HELIOS): a randomised, double-blind, phase 3 study. *Lancet Oncol* 2016; 17(2): 200-211.
59. Robins JM, Blevins D, Ritter G, Wulfsohn M. G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology* 1992; 3(4): 319-336.
60. Fleming T, Harrington D. A class of hypothesis tests for one and two samples of censored survival data. *Commun Stat Theory Methods* 1981; 10(8): 763-794.
61. White IR. Estimating treatment effects in randomized trials with treatment switching. *Stat Med* 2006; 25(9): 1619-1622.
62. Zhang J, Chen C. Correcting treatment effect for treatment switching in randomized oncology trials with a modified iterative parametric estimation method. *Stat Med* 2016; 35(21): 3690-3703.

63. IQWiG im Dialog 2014: „Nutzenbewertung bei Studien mit erlaubtem Behandlungswechsel“. Z Evid Fortbild Qual Gesundheitswes 2014; 108(7): 449.
64. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 4.1. Köln: IQWiG; 2013. URL: https://www.iqwig.de/download/IQWiG_Methoden_Version_4-1.pdf.
65. Roche Pharma. Vemurafenib (Zelboraf): Dossier zur Nutzenbewertung gemäß § 35a SGB V; Modul 4 A; Monotherapie zur Behandlung von erwachsenen Patienten mit BRAF-V600 Mutation-positivem nicht resezierbarem oder metastasiertem Melanom; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 17.02.2012 [Zugriff: 11.06.2014]. URL: https://www.g-ba.de/downloads/92-975-112/2012-02-17_Modul4A_Vemurafenib.pdf.
66. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Vemurafenib: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A12-08 [online]. 13.06.2012 [Zugriff: 03.07.2014]. (IQWiG-Berichte; Band 133). URL: https://www.iqwig.de/download/A12-08_Vemurafenib_Nutzenbewertung_35a_SGB_V.pdf.
67. Roche Pharma. Vemurafenib (Zelboraf): Dossier zur Nutzenbewertung gemäß § 35a SGB V; Modul 4 A; Monotherapie zur Behandlung von erwachsenen Patienten mit BRAF-V600 Mutation-positivem nicht resezierbarem oder metastasiertem Melanom; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 02.09.2013 [Zugriff: 11.06.2014]. URL: https://www.g-ba.de/downloads/92-975-374/2013-09-02_Modul4A_Vemurafenib.pdf.
68. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Vemurafenib: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A13-34 [online]. 11.12.2013 [Zugriff: 16.12.2013]. (IQWiG-Berichte; Band 199). URL: https://www.iqwig.de/download/A13-34_Vemurafenib_Nutzenbewertung-35a-SGB-V.pdf.
69. Pfizer Pharma. Axitinib (Inlyta): Dossier zur Nutzenbewertung gemäß § 35a SGB V; Modul 4 A; Behandlung des fortgeschrittenen Nierenzellkarzinoms [renal cell carcinoma, RCC] bei erwachsenen Patienten nach Versagen von vorangegangener Therapie mit Sunitinib oder einem Zytokin; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 27.09.2012 [Zugriff: 20.05.2015]. URL: https://www.g-ba.de/downloads/92-975-156/2012_09_27_Modul4A_Axitinib.pdf.
70. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Axitinib: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A12-14 [online]. 21.12.2012 [Zugriff: 03.07.2014]. (IQWiG-Berichte; Band 149). URL: https://www.iqwig.de/download/A12-14_Axitinib_Nutzenbewertung_35a_SGB_V.pdf.

71. Pfizer Pharma. Crizotinib (Xalkori): Dossier zur Nutzenbewertung gemäß § 35a SGB V; Modul 4 A; Behandlung des vorbehandelten Anaplastische Lymphom-Kinase (ALK)-positiven fortgeschrittenen nicht-kleinzelligen Bronchialkarzinoms (NSCLC) bei Erwachsenen; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 07.11.2012 [Zugriff: 20.05.2015]. URL: https://www.g-ba.de/downloads/92-975-167/2012-11-07_Modul4A_Crizotinib.pdf.
72. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Crizotinib: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A12-15 [online]. 13.02.2013 [Zugriff: 03.07.2014]. (IQWiG-Berichte; Band 151). URL: https://www.iqwig.de/download/A12-15_Crizotinib_Nutzenbewertung_35a_SGB_V.pdf.
73. Janssen-Cilag. Abirateronacetat (Zytiga): Dossier zur Nutzenbewertung gemäß § 35a SGB V; Modul 4 A; Behandlung des metastasierten kastrationsresistenten Prostatakarzinoms (mCRPC) bei erwachsenen Männern mit asymptomatischem oder mild symptomatischem Verlauf der Erkrankung nach Versagen der Androgenentzugstherapie, bei denen eine Chemotherapie noch nicht klinisch indiziert ist; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 14.01.2013 [Zugriff: 20.05.2015]. URL: https://www.g-ba.de/downloads/92-975-218/2013-01-14_Modul4A_Abirateronacetat.pdf.
74. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Abirateronacetat (neues Anwendungsgebiet): Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A13-06 [online]. 11.04.2013 [Zugriff: 03.07.2014]. (IQWiG-Berichte; Band 160). URL: https://www.iqwig.de/download/A13-06_Abirateronacetat_neues_Anwendungsgebiet_Nutzenbewertung_35a_SGB_V.pdf.
75. AstraZeneca. Vandetanib (Caprelsa): Dossier zur Nutzenbewertung gemäß § 35a SGB V; Modul 4 A; medulläres Schilddrüsenkarzinom; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 14.03.2012 [Zugriff: 20.05.2015]. URL: https://www.g-ba.de/downloads/92-975-116/2012-03-14_Modul4A_Vandetanib.pdf.
76. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Vandetanib: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A12-09 [online]. 13.06.2012 [Zugriff: 03.07.2014]. (IQWiG-Berichte; Band 134). URL: https://www.iqwig.de/download/A12-09_Vandetanib_Nutzenbewertung_35a_SGB_V.pdf.
77. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Vandetanib: Erneute Nutzenbewertung gemäß § 35a Absatz 5b SGB V; Dossierbewertung; Auftrag A13-09 [online]. 13.06.2013 [Zugriff: 21.06.2013]. (IQWiG-Berichte; Band 169). URL: https://www.iqwig.de/download/A13-09_Vandetanib_Erneute-Nutzenbewertung-35a-SGB-V.pdf.

78. Roche Pharma. Pertuzumab (Perjeta): Dossier zur Nutzenbewertung gemäß § 35a SGB V; Modul 4 A; Brustkrebs; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 04.03.2013 [Zugriff: 20.05.2015]. URL: https://www.g-ba.de/downloads/92-975-256/2013-03-04_Modul4A_Pertuzumab.pdf.

79. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Pertuzumab: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A13-10 [online]. 27.06.2013 [Zugriff: 05.07.2013]. (IQWiG-Berichte; Band 177). URL: https://www.iqwig.de/download/A13-10_Pertuzumab_Nutzenbewertung-35a-SGB-V.pdf.

80. Astellas Pharma. Enzalutamid (Xtandi): Dossier zur Nutzenbewertung gemäß § 35a SGB V; Modul 4 A; Behandlung des metastasierten kastrationsresistenten Prostatakarzinoms bei erwachsenen Männern, deren Erkrankung während oder nach einer Chemotherapie mit Docetaxel fortschreitet; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 29.08.2013 [Zugriff: 20.05.2015]. URL: https://www.g-ba.de/downloads/92-975-358/2013-08-29_Modul4A_Enzalutamid.pdf.

81. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Enzalutamid: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A13-33 [online]. 28.11.2013 [Zugriff: 03.07.2014]. (IQWiG-Berichte; Band 198). URL: https://www.iqwig.de/download/A13-33_Enzalutamid_Nutzenbewertung-35a-SGB-V.pdf.

82. GlaxoSmithKline. Dabrafenib (Tafinlar): Dossier zur Nutzenbewertung gemäß § 35a SGB V; Modul 4 A; Melanom; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 20.09.2013 [Zugriff: 20.05.2015]. URL: https://www.g-ba.de/downloads/92-975-391/2013-09-20_Modul4_Dabrafenib.pdf.

83. Pharma R. Trastuzumab Emtansin (Kadcyla): Dossier zur Nutzenbewertung gemäß § 35a SGB V; Modul 4 A; Brustkrebs; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 25.11.2013 [Zugriff: 20.05.2015]. URL: https://www.g-ba.de/downloads/92-975-425/2013-11-25_Modul4A_Trastuzumab_Emtansin.pdf.

84. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Trastuzumab Emtansin: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A14-01 [online]. 28.03.2014 [Zugriff: 03.07.2014]. (IQWiG-Berichte; Band 214). URL: https://www.iqwig.de/download/A14-01_Trastuzumab-Emtansin_Nutzenbewertung-35a-SGB%20V.pdf.

85. Bayer Vital. Radium-223-Dichlorid (Xofigo): Dossier zur Nutzenbewertung gemäß § 35a SGB V; Modul 4 A; zur Behandlung von Patienten mit kastrationsresistentem Prostatakarzinom mit symptomatischen Knochenmetastasen ohne bekannte viszerale Metastasen; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. 13.12.2013 [Zugriff: 20.05.2015]. URL: https://www.g-ba.de/downloads/92-975-419/2013-12-13_Modul4_Radium-223-dichlorid_mCRPC.pdf.
86. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Radium-223-dichlorid: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A14-02 [online]. 28.03.2014 [Zugriff: 01.04.2014]. (IQWiG-Berichte; Band 215). URL: https://www.iqwig.de/download/A14-02_Radium-223-dichlorid_Nutzenbewertung-35a-SGB-V.pdf.
87. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Sipuleucel-T: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A14-38 [online]. [Zugriff: 08.01.2015]. (IQWiG-Berichte; Band 269). URL: https://www.iqwig.de/download/A14-38_Sipuleucel-T_Nutzenbewertung-35a-SGB-V.pdf.
88. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Sipuleucel-T (Addendum zum Auftrag A14-38): Auftrag A15-08 [online]. 02.03.2015 [Zugriff: 20.03.2015]. (IQWiG-Berichte; Band 283). URL: https://www.iqwig.de/download/A15-08_Addendum-zum-Auftrag-A14-38_Sipuleucel-T.pdf.
89. Prasad V. Double-crossed: why crossover in clinical trials may be distorting medical science. J Natl Compr Canc Netw 2013; 11(5): 625-627.
90. Prasad V, Grady C. The misguided ethics of crossover trials. Contemp Clin Trials 2014; 37(2): 167-169.