IQWiG Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

IQWiG-Berichte – Jahr: 2011 Nr. 80

Aussagekraft von Surrogatendpunkten in der Onkologie

Rapid Report

Auftrag: A10-05
Version: 1.1
Stand: 21.11.2011

Impressum

Herausgeber:
Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:
Aussagekraft von Surrogatendpunkten in der Onkologie

Auftraggeber:
Gemeinsamer Bundesausschuss

Datum des Auftrags:
19.08.2010

Interne Auftragsnummer:
A10-05

Anschrift des Herausgebers:
Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Dillenburger Str. 27
51105 Köln

Tel.: +49 221 35685-0
Fax: +49 221 35685-1
Berichte@iqwig.de
www.iqwig.de

ISSN: 1864-2500

Dieser Rapid Report wurde einem externen Review unterzogen. Das Review wurde durchgeführt von:

- Marcial Velasco Garrido, MPH
  Technische Universität Berlin und Universitätsklinikum Hamburg Eppendorf

Die Offenlegung der potenziellen Interessenkonflikte des externen Reviewers findet sich in Anhang C.

Kurzfassung

Hintergrund

Der Gemeinsame Bundesausschuss (G-BA) hat mit Schreiben vom 19.08.2010 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Bewertung der Aussagekraft von Surrogatendpunkten in der Onkologie beauftragt.

Fragestellung

Primäres Ziel der vorliegenden Untersuchung war es, Empfehlungen für die Validierung von Surrogatendpunkten in der Onkologie darzustellen.

Diese Empfehlungen sollten aus den Ergebnissen folgender 2 Teilziele abgeleitet werden:

1. zusammenfassende Darstellung und Bewertung methodischer Verfahren, deren Ziel die Validierung von Surrogatendpunkten ist

2. Darstellung und Bewertung von Studien zur Validierung von Parametern des Tumoransprechens als Surrogate für patientenrelevante Endpunkte, insbesondere für das Überleben von Patienten (die Bearbeitung von Teilziel 2 erfolgte beispielhaft für die Krankheitsbilder Darmkrebs und Brustkrebs)

Methoden

Zur Identifizierung relevanter Publikationen wurde für beide Teilziele jeweils eine Literaturrecherche durchgeführt.

Die Suche zu Teilziel 1 (Darstellung und Bewertung methodischer Verfahren) wurde in MEDLINE, EMBASE und dem Cochrane Methodology Register (Methods Studies) durchgeführt. Erfasst wurde der Zeitraum 2007 bis September 2010, der Zeitraum vor 2007 wurde durch andere systematische Übersichten, die in der Vorabrecherche identifiziert und die bei der Bewertung berücksichtigt wurden, erfasst. Die eingeschlossenen Publikationen wurden nach methodischen Aspekten klassifiziert und zusammenfassend beschrieben. Daraus wurden die Anforderungen an die Surrogatvalidierung abgeleitet und zusammenfassend dargestellt.

Die Suche zu Teilziel 2 (Bewertung von Validierungsstudien für Surrogatendpunkte in den Krankheitsbildern Darmkrebs und Brustkrebs) wurde ebenfalls in MEDLINE, EMBASE und dem Cochrane Methodology Register (Methods Studies) durchgeführt. Erfasst wurde der Zeitraum bis August 2010 (Darmkrebs) bzw. Oktober 2010 (Brustkrebs). Zunächst wurde die Aussagesicherheit der Validierungsstudien bewertet, und zwar anhand folgender Kriterien:

- Verwendung eines anerkannten, in der Literatur beschriebenen Verfahrens

- Durchführung von Analysen zur Prüfung der Robustheit und Generalisierbarkeit der Ergebnisse

- Systematik der für die Validierung zugrunde gelegten Daten

- hinreichende Eingrenzung der Indikationen oder Schweregrade
- hinreichende Eingrenzung der untersuchten Interventionen
- eindeutige Definitionen für die untersuchten Endpunkte

Auf Basis dieser Kriterien wurde die Aussagesicherheit der Validierungsstudien jeweils als hoch, eingeschränkt, mäßig oder niedrig eingestuft.

Im Anschluss wurden die Ergebnisse zur Validität der untersuchten Surrogatendpunkte analysiert.

Ergebnisse

Teilziel 1: Darstellung und Bewertung methodischer Verfahren

Die systematische Suche nach Publikationen zur Surrogatvalidierung zeigte, dass adäquate Vorschläge für das methodische Vorgehen bei der Validierung von Surrogatendpunkten vorliegen, die auch in der Praxis erprobt sind. Eine Validierung von Surrogaten erscheint daher grundsätzlich möglich, aus der verfügbaren Literatur können folgende grundsätzliche Aussagen abgeleitet werden:

1. Der Nachweis einer Korrelation zwischen Surrogat und klinischem Endpunkt reicht alleine zur Surrogatvalidierung nicht aus.

2. Die Übertragbarkeit von Aussagen zur Validität von Surrogaten zwischen verschiedenen Erkrankungen oder Ausprägungen einer Erkrankung bzw. zwischen verschiedenen Interventionen ist nicht ohne Weiteres möglich. Die Validität eines Surrogats kann sowohl krankheits- als auch interventionsspezifisch sein, d. h., eine abschließende Surrogatvalidierung mithilfe statistischer Methoden kann ggf. nur innerhalb einer Indikation und einer Intervention erfolgen. In welchem Ausmaß Aussagen zur Validität von Surrogaten zwischen verschiedenen Erkrankungen bzw. Interventionen übertragen werden können, muss geprüft und begründet werden. Zur Beurteilung der Übertragbarkeit sollten bei Validierungsstudien, die verschiedene Krankheitsentitäten oder Interventionen einschließen, mindestens Untersuchungen zur Heterogenität vorliegen.

3. Zur Validierung eines Surrogats werden umfangreiche Daten benötigt, vorzugsweise eine Meta-Analyse von mehreren randomisierten Studien mit ausreichender Ergebnissicherheit. In diesen Studien müssen sowohl das Surrogat als auch der klinische Endpunkt erhoben worden sein.

4. Zur Surrogatvalidierung gibt es weder ein universell anzuwendendes Maß noch eine allgemein beste Schätzmethode noch eine allgemein akzeptierte Grenze, deren Überschreitung den Nachweis der Validität bedeuten würde.

5. Zurzeit werden in der Praxis primär korrelationsbasierte Verfahren zur Surrogatvalidierung mit Schätzung von Korrelationsmaßen auf Studien- und individueller

Ebene angewendet. In der Literatur wird auf Studienebene für den Nachweis der Validität in der Regel eine „hohe“ Korrelation zwischen den Effekten für das Surrogat und den patientenrelevanten Endpunkt verlangt, verschiedene Autoren legen eine Korrelation von 0,9 als möglichen Grenzwert nahe.

Insbesondere wenn unter Berücksichtigung der zugehörigen Konfidenzintervalle keine hohe Korrelation vorliegt (untere Konfidenzgrenze der geschätzten Korrelationskoeffizienten < 0,85), die Validität des Surrogats also unklar bleibt, können durch Anwendung des STE-Konzepts nach Burzykowski & Buyse (2005, 2006) bei hinreichend großen Effekten für das Surrogat noch Aussagen bezüglich patientenrelevanter Endpunkte getroffen werden. Das STE-Konzept dient dabei zur Festlegung von Schwellenwerten für den Entscheidungsprozess, ob ein beobachteter Effekt auf das Surrogat mit ausreichender Sicherheit auch mit einem Effekt auf den interessierenden Endpunkt einhergeht. Hierzu muss die untere Konfidenzgrenze des Behandlungseffekts bezüglich des Surrogats größer sein als der STE.

Die Anwendung von Methoden, die nicht zum Bereich der korrelationsbasierten Verfahren gehören, bedarf einer ausreichenden Begründung, insbesondere dann, wenn als Datengrundlage nur eine Studie verwendet werden soll.

Teilziel 2: Bewertung von Validierungsstudien für Surrogatendpunkte in den Krankheitsbildern Darmkrebs und Brustkrebs

Die systematische Suche nach Studien zur Validierung von Surrogaten für die Krankheitsbilder Darmkrebs und Brustkrebs ergab insgesamt 15 relevante Publikationen für das Krankheitsbild Darmkrebs und 6 relevante Publikationen für das Krankheitsbild Brustkrebs.

Zusammenfassend kann festgestellt werden, dass die eingeschlossenen Validierungsstudien die Ergebnisse der Methodenreviews in großen Teilen bestätigen, jedoch keine der Studien die anerkannten Gütekriterien konsequent erfüllte. Dies mag auch dadurch bedingt sein, dass einige der Validierungsstudien nicht primär auf die Validierung eines Surrogats ausgerichtet waren, sondern methodische Arbeiten zur Anwendung der Validierungsmethodik darstellten. Entsprechend finden sich diverse der Arbeiten aus Teilziel 2 auch im Methodenreview wieder (Teilziel 1). In den Validierungsstudien wurden häufig verschiedene Interventionen oder Indikationen gemeinsam ausgewertet, oder es blieb unklar, ob sich die Datenbasis auf eine Intervention und Indikation beschränkte. Die Übertragbarkeit der Ergebnisse der Surrogatvalidierung zwischen den eingeschlossenen Interventionen oder Indikationen wurde nicht adressiert. Auch blieb zumeist unklar, wie die Datenbasis zusammengestellt war, oder die Zusammenstellung war nachweislich nicht systematisch. Die Aussagesicherheit der Validierungsstudien war insgesamt niedrig.

Auch aufgrund der statistischen Ergebnisse der Validierungsstudien blieb die Validität der untersuchten Surrogate meist unklar. Eine Ausnahme bildet das krankheitsfreie Überleben (DFS) bei Kolonkarzinom im frühen Krankheitsstadium, wo sich in beiden

Validierungsstudien eine hohe Korrelation zeigte. Die Aussagesicherheit dieser beiden Studien war jedoch ebenfalls niedrig, sodass sich aus ihnen kein Beleg für die Validiät des DFS für das Gesamtüberleben aus diesen Studien ableiten lässt.

Zusammenfassend bleibt die Validität der Parameter des Tumoransprechens als Surrogate für patientenrelevante Endpunkte (hier Gesamtüberleben) in den Krankheitsbildern Darm- und Brustkrebs aus der vorliegenden Bewertung deshalb unklar.

Vorschlag zum Umgang mit Surrogatendpunkten in der Nutzenbewertung onkologischer Behandlungen

Mit dem vorliegenden Bericht stellt das IQWiG auf Basis einer systematischen Recherche zur Methodik zur Surrogatvalidierung (Teilziel 1) sowie der beispielhaften systematischen Aufarbeitung von Validierungsstudien in den Krankheitsbildern Darm- und Brustkrebs (Teilziel 2) einen Algorithmus zum Umgang mit Surrogatendpunkten in der Nutzenbewertung von onkologischen Behandlungen zur Diskussion. Dieser Vorschlag hat das Ziel, ausgehend von den verfügbaren Daten und Informationen zu einer differenzierten Betrachtung der Ergebnisse von Surrogatendpunkten zu kommen, die die Ergebnissicherheit der verschiedenen Konstellationen berücksichtigt.

Die Ergebnisse des Projekts sollen dem Gemeinsamen Bundesausschuss als Diskussionsgrundlage für die Verwendung von Surrogatendpunkten bei der Nutzenbewertung von onkologischen Behandlungen dienen.

Bei der Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung sind 2 Ebenen zu trennen. Auf der ersten Ebene muss zunächst die Validität des Surrogats beurteilt werden. Auf der zweiten Ebene stellt sich die Frage, welche Schlussfolgerungen bezüglich patientenrelevanter Endpunkte aus den Surrogateffekten abhängig von der Ausprägung der Validität des Surrogats gezogen werden können.

Ebene 1: Beurteilung der Validität des Surrogats

Abbildung A fasst die Überlegungen zum Vorgehen bei der Surrogatvalidierung und zur Verbindung der Validierungsergebnisse zur Nutzenbewertung zusammen.

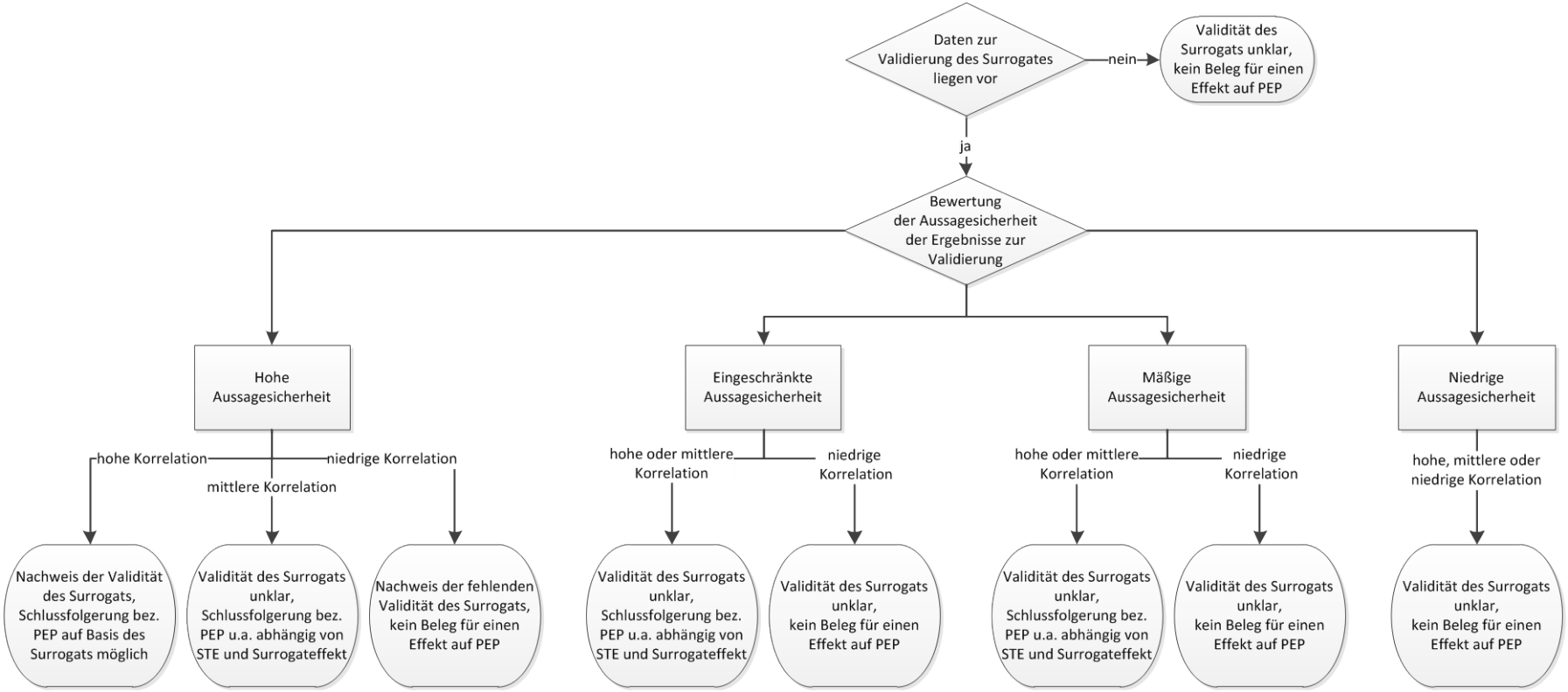

Abbildung A: Einfluss der Aussagesicherheit der Validierungsstudien und der Korrelation der Effekte von Surrogat und patientenrelevantem Endpunkt auf die Validität des Surrogats

PEP: patientenrelevanter Endpunkt; STE: Surrogate Threshold Effect

Bei hoher Aussagesicherheit der Validierungsstudie kann aus einer hohen Korrelation (untere Grenze des 95 %-Konfidenzintervalls) $R \geq 0,85$ der Effekte von Surrogat und patientenrelevantem Endpunkt ein Nachweis für die Validität des Surrogats abgeleitet werden. Aus einer niedrigen Korrelation (obere Grenze des 95 %-Konfidenzintervalls) $R \leq 0,7$ ergibt sich ein Nachweis für das Fehlen der Validität des Surrogats. Bei mittlerer Korrelation ($R < 0,85$ bis $> 0,7$) bleibt die Validität des Surrogats unklar. Ebenfalls unklar, d. h. nicht abschließend zu beurteilen, ist die Validität des Surrogats bei eingeschränkter oder mäßiger Aussagesicherheit der Validierungsstudie. Liegen keine Daten zur Validierung des Surrogats vor oder sind die vorliegenden Daten nur von geringer Aussagesicherheit, kann auf Basis der Ergebnisse zum Surrogat keine Schlussfolgerung bezüglich des patientenrelvanten Endpunkts getroffen und damit auch kein Beleg für einen Effekt auf den patientenrelevanten Endpunkt abgeleitet werden.

Das eigentliche Ziel der Surrogatvalidierung ist die Klärung der Frage, welche Schlussfolgerungen bezüglich patientenrelevanter Endpunkte aus den Surrogatendpunkten gezogen werden können (Ebene 2). Aus Effekten, die für ein valides Surrogat gemessen werden, kann direkt auf die Effekte bezüglich des patientenrelevanten Endpunkts geschlossen werden. Effekte von Surrogaten, die nachgewiesen nicht valide sind, erlauben keine Aussage bezüglich patientenrelevanter Endpunkte. Ist die Validität des Surrogats unklar, können unter Anwendung des STE-Konzepts Aussagen zu patientenrelevanten Endpunkten gemacht werden, die in ihrer Ergebnissicherheit abgestuft sind und so der Datenlage differenziert Rechnung tragen. Ein Vorschlag zur Interpretation von Surrogateffekten bei unklarer Validität wird nachfolgend vorgestellt.

Ebene 2: Schlussfolgerungen zu patientenrelevanten Endpunkten aus Surrogateffekten abhängig von der Ausprägung der Validität des Surrogats

Abbildung B zeigt die ersten Schritte des Algorithmus zur Berücksichtigung von Ergebnissen zu Surrogatendpunkten in der Nutzenbewertung.

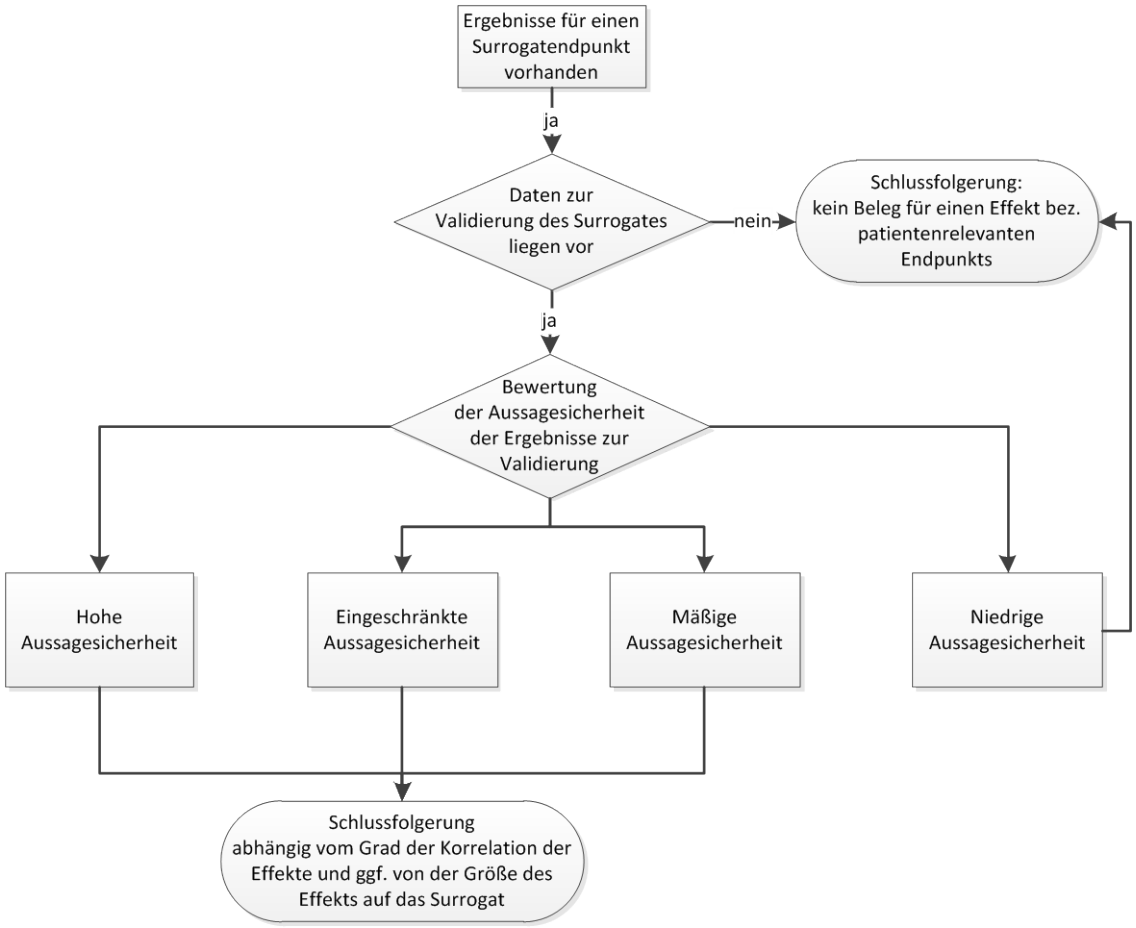

Abbildung B: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: Stellenwert von Validierungsstudien und deren Aussagesicherheit

PEP: patientenrelevanter Endpunkt

Werden für eine Nutzenbewertung Ergebnisse für einen Surrogatendpunkt vorgelegt, stellt sich die Frage, ob Daten zur Validierung des Surrogats zur Verfügung stehen. Ist das nicht der Fall, liegt aus den Ergebnissen zu dem Surrogatendpunkt kein Beleg (oder Hinweis oder Anhaltspunkt) für einen Effekt bezüglich eines patientenrelevanten Endpunkts vor. Wenn Daten zur Validierung des Surrogats verfügbar sind, werden diese in einem ersten Schritt bezüglich ihrer Aussagesicherheit bewertet. Bei niedriger Aussagesicherheit der Validierungsergebnisse gibt es aus den Ergebnissen zu dem Surrogatendpunkt grundsätzlich keinen Beleg für einen Effekt bezüglich eines patientenrelevanten Endpunkts. Bei hoher, eingeschränkter oder mäßiger Aussagesicherheit der Validierungsergebnisse ist die Schlussfolgerung zu patientenrelevanten Endpunkten abhängig vom Grad der Korrelation der Effekte von Surrogat und patientenrelevantem Endpunkt in den Validierungsstudien, sowie ggf. von der Größe des Effekts für das Surrogat in den Studien zur Nutzenbewertung. Das Grundprinzip ist dabei, dass die Sicherheit der Aussage in der Nutzenbewertung (z. B. Beleg für oder Hinweis auf einen Effekt auf einen patientenrelevanten Endpunkt) umso höher ist, je höher die Sicherheit der Aussagen zum Surrogat ist (siehe Abbildungen C bis E).

Schlussfolgerungen bei hoher Aussagesicherheit der Validierungsstudien

Abbildung C zeigt die möglichen Schlussfolgerungen für die Nutzenbewertung bei hoher Aussagesicherheit der Validierungsstudien.

Abbildung C: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: mögliche Schlussfolgerungen bei hoher Aussagesicherheit der Validierungsstudien

KI: Konfidenzintervall; STE: Surrogate Threshold Effect

Abhängig von der Korrelation zwischen Surrogat und patientenrelevantem Endpunkt aus Validierungsstudien sind unterschiedliche Schlussfolgerungen aus den Ergebnissen des Surrogatendpunkts in den Studien zur Nutzenbewertung ableitbar.

- Aus Validierungsstudien mit hoher Aussagesicherheit kann bei hoher Korrelation und einem Effekt für das Surrogat maximal auf einen Beleg für einen Effekt bezüglich des patientenrelevanten Endpunkts geschlossen werden. Die Aussage, dass maximal auf einen Beleg geschlossen werden kann, bedeutet, dass zusätzlich zur Prüfung der Validität des Surrogats in Validierungsstudien die Ergebnissicherheit der Daten zum Surrogat in den Studien zur Nutzenbewertung berücksichtigt wird. Abhängig von dieser

Ergebnissicherheit würde die Aussage bezüglich des patientenrelevanten Endpunkts ggf. von einem Beleg auf einen Hinweis herabgestuft.

Zeigen die Studien zur Nutzenbewertung keinen Effekt für das Surrogat, d. h., liegt in diesen Studien bzw. in einer Meta-Analyse der Studien kein statistisch signifikanter Gruppenunterschied vor, ist generell ein Effekt bezüglich des patientenrelevanten Endpunkts nicht belegt (gilt auch für den Pfad mit mittlerer Korrelation).

- Bei hoher Aussagesicherheit der Validierungsstudien und mittlerer Korrelation zwischen Surrogat und patientenrelevantem Endpunkt in diesen Studien soll der Effekt für das Surrogat aus den Studien zur Nutzenbewertung einem Surrogate Threshold Effect (STE) gegenübergestellt werden. Ist ein solcher STE nicht verfügbar, gilt ein Effekt bezüglich des patientenrelevanten Endpunkts als nicht belegt. Ist ein STE verfügbar, wird zunächst das 95 %-KI des Effekts für das Surrogat aus den Studien zur Nutzenbewertung zu diesem STE in Beziehung gesetzt. Liegt das 95 %-KI vollständig oberhalb des STE, liegt maximal ein Beleg für einen Effekt bezüglich des patientenrelevanten Endpunkts vor (bei erhöhter Ergebnisunsicherheit in den Studien zur Nutzenbewertung, z. B. durch ein hohes Verzerrungspotenzial, kann dieses Ergebnis wiederum zu einem Hinweis herabgestuft werden). Liegt das 95 %-KI nicht vollständig oberhalb des STE, wird das 80 %-KI des Effekts für das Surrogat zu dem STE in Beziehung gesetzt. Liegt das 80 %-KI vollständig oberhalb des STE, kann maximal auf einen Hinweis für einen Effekt bezüglich des patientenrelevanten Endpunkts geschlossen werden. Überschreitet das 80 %-KI des Surrogateffekts den STE nicht vollständig, ist bei hoher Aussagesicherheit der Validierungsstudien und mittlerer Korrelation ein Effekt bezüglich des patientenrelevanten Endpunkts nicht belegt, und es liegt auch kein Hinweis und kein Anhaltspunkt dafür vor.

- Bei hoher Aussagesicherheit der Validierungsstudien und niedriger Korrelation zwischen Surrogat und patientenrelevantem Endpunkt gilt das Surrogat als nicht valide. Aus einem Effekt auf das Surrogat ergibt sich damit grundsätzlich kein Beleg (sowie kein Hinweis und kein Anhaltspunkt) für einen Effekt bezüglich des patientenrelevanten Endpunkts.

Schlussfolgerungen bei eingeschränkter Aussagesicherheit der Validierungsstudien

Abbildung D zeigt die möglichen Schlussfolgerungen für die Nutzenbewertung bei eingeschränkter Aussagesicherheit der Validierungsstudien.

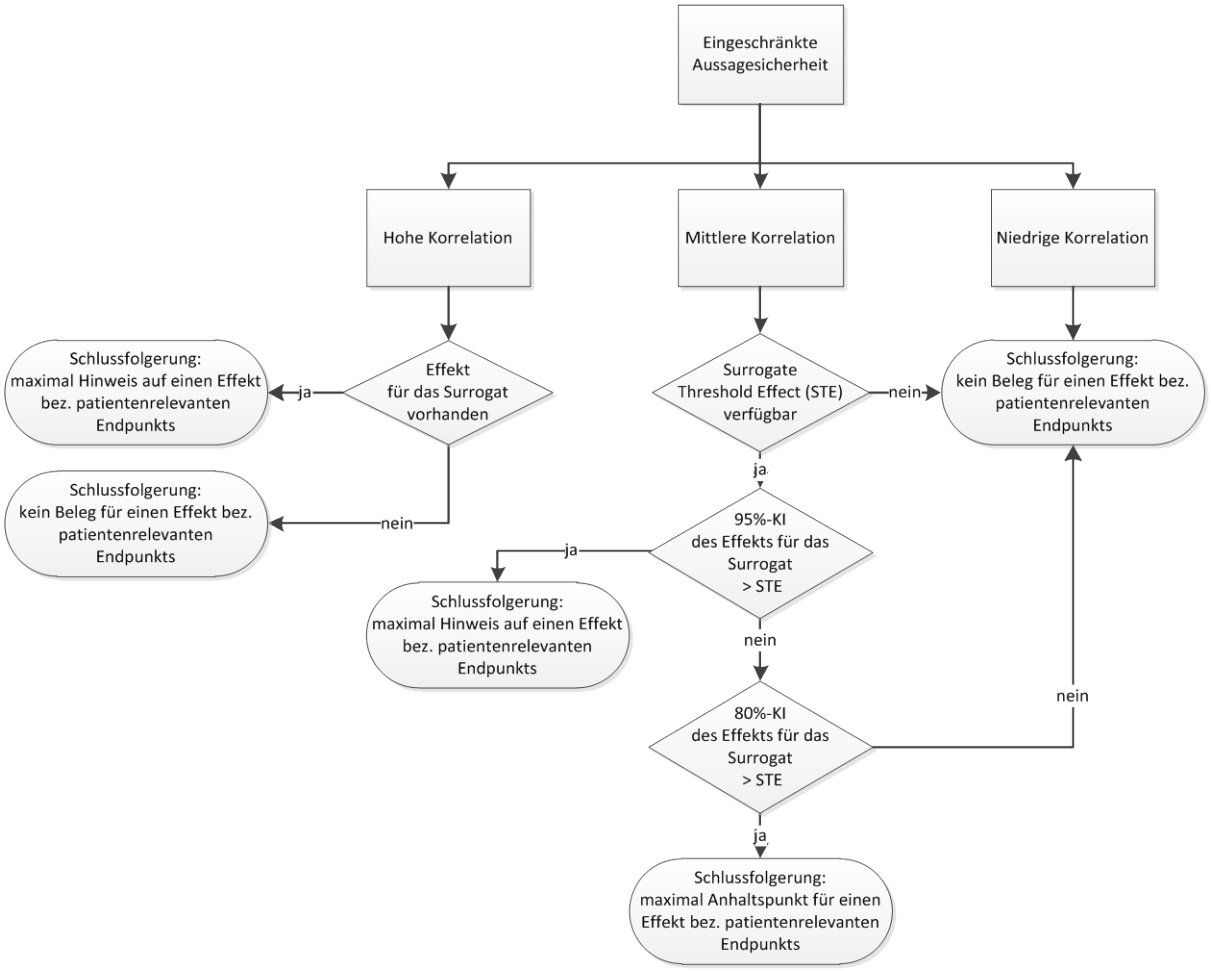

Abbildung D: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: mögliche Schlussfolgerungen bei eingeschränkter Aussagesicherheit der Validierungsstudien

KI: Konfidenzintervall; STE: Surrogate Threshold Effect

Im Vergleich zu den möglichen Schlussfolgerungen bei hoher Aussagesicherheit der Validierungsstudien haben in der Situation, in der die Aussagesicherheit dieser Studien eingeschränkt ist, alle Schlussfolgerungen zu Effekten bezüglich patientenrelevanter Endpunkte eine geringere Ergebnissicherheit.

- Bei eingeschränkter Aussagesicherheit und hoher Korrelation von Surrogat und patientenrelevantem Endpunkt kann aus Surrogateffekten in den Studien zur Nutzenbewertung maximal auf einen Hinweis auf Effekte bezüglich patientenrelevanter Endpunkte geschlossen werden.

- Bei eingeschränkter Aussagesicherheit und mittlerer Korrelation kann bei Überschreiten des STE durch das 95 %-KI des Surrogateffekts maximal ein Hinweis, bei Überschreiten des STE durch das 80 %-KI maximal ein Anhaltspunkt für einen Effekt bezüglich des patientenrelevanten Endpunkts abgeleitet werden.

- Bei eingeschränkter Aussagesicherheit und niedriger Korrelation liegt grundsätzlich kein Beleg für einen Effekt bezüglich des patientenrelevanten Endpunkts vor.

Für alle abschließenden Schlussfolgerungen wird wie oben beschrieben zusätzlich die Ergebnissicherheit der Studien zur Nutzenbewertung berücksichtigt, d. h., eine Einschränkung der Ergebnissicherheit führt zu einer Herabstufung der Aussage.

Schlussfolgerungen bei mäßiger Aussagesicherheit der Validierungsstudien

Abbildung E zeigt die möglichen Schlussfolgerungen für die Nutzenbewertung bei mäßiger Aussagesicherheit der Validierungsstudien.

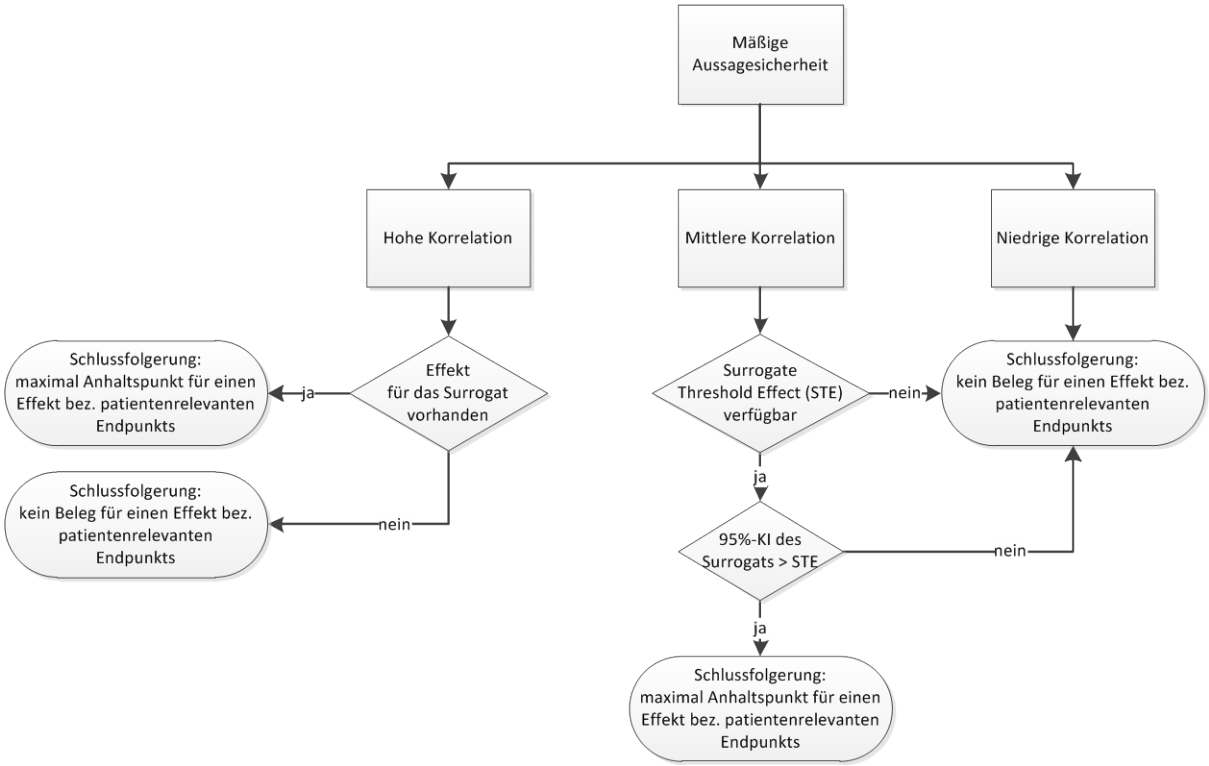

Abbildung E: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: mögliche Schlussfolgerungen bei mäßiger Aussagesicherheit der Validierungsstudien

KI: Konfidenzintervall; STE: Surrogate Threshold Effect

Bei mäßiger Aussagesicherheit der Validierungsstudien können aus den Ergebnissen zu Surrogaten in den Studien zur Nutzenbewertung maximal Anhaltspunkte für Effekte bezüglich patientenrelevanter Endpunkte abgeleitet werden. Bei mittlerer Korrelation wird hierfür ausschließlich das 95 %-KI zu dem STE in Beziehung gesetzt. Wie zuvor wird auch bei mäßiger Aussagesicherheit nach der Beurteilung der maximal abzuleitenden Aussage zum Effekt bezüglich eines patientenrelevanten Endpunkts die Ergebnissicherheit der Ergebnisse zu dem Surrogat aus den Studien zur Nutzenbewertung berücksichtigt. Bei einer Einschränkung der Ergebnissicherheit wird die Bewertung „Anhaltspunkt“ zu „kein Beleg (sowie kein Hinweis und kein Anhaltspunkt)“ herabgestuft.

Fazit

Methoden zur Surrogatvalidierung

Es liegen adäquate Vorschläge für das methodische Vorgehen bei der Validierung von Surrogatendpunkten vor, die auch in der Praxis erprobt sind. Eine Validierung von Surrogaten erscheint daher grundsätzlich möglich. Dabei besteht zu folgenden grundsätzlichen Aussagen international ein allgemeiner Konsens:

- Der Nachweis einer Korrelation zwischen Surrogat und klinischem Endpunkt reicht alleine zur Surrogatvalidierung nicht aus.

- Die Übertragbarkeit von Aussagen zur Validität von Surrogaten zwischen verschiedenen Erkrankungen oder Ausprägungen einer Erkrankung bzw. zwischen verschiedenen Interventionen ist nicht ohne Weiteres möglich. Zur Beurteilung der Übertragbarkeit sollten bei Validierungsstudien, die verschiedene Krankheitsentitäten oder Interventionen einschließen, mindestens Untersuchungen zur Heterogenität vorliegen.

- Zur Validierung eines Surrogats werden umfangreiche Daten benötigt, vorzugsweise eine Meta-Analyse von mehreren randomisierten Studien mit ausreichender Ergebnissicherheit. In diesen Studien müssen sowohl das Surrogat als auch der klinische Endpunkt erhoben worden sein.

- Zur Surrogatvalidierung gibt es weder ein universell anzuwendendes Maß noch eine allgemein beste Schätzmethode noch eine allgemein akzeptierte Grenze, deren Überschreitung den Nachweis der Validität bedeuten würde.

- Zurzeit werden in der Praxis primär korrelationsbasierte Verfahren zur Surrogatvalidierung mit Schätzung von Korrelationsmaßen auf Studien- und individueller Ebene angewendet. Das Konzept eines Surrogate Threshold Effects (STE) ist dabei hilfreich zur Festlegung von Schwellenwerten für den Entscheidungsprozess.

Validierungsstudien für die Krankheitsbilder Darmkrebs und Brustkrebs

Die zu den Krankheitsbildern Darmkrebs und Brustkrebs vorliegenden Validierungsstudien erlauben keine abschließende Aussage zur Validität von Parametern des Tumoransprechens für das Gesamtüberleben bei diesen Erkrankungen. In den meisten Fällen waren die Korrelationen der jeweiligen Effekte nicht ausreichend hoch, sodass damit die notwendige Bedingung für den Nachweis der Validität nicht erfüllt war. In allen Studien wurde außerdem die Aussagesicherheit als niedrig bewertet. Daher blieb die Validität bei allen untersuchten Surrogatendpunkten unabhängig von den statistischen Ergebnissen letztlich unklar. Für die niedrige Aussagesicherheit der Validierungsstudien waren insbesondere folgende Faktoren maßgeblich:

- Die Surrogatvalidierung wurde zumeist nicht interventionsspezifisch durchgeführt, und der Einfluss der verschiedenen Interventionen auf das Gesamtergebnis wurde nicht untersucht.

- Häufig wurden verschiedene Indikationen oder verschiedene Schweregrade der Erkrankung zusammengefasst, und der Einfluss dieser verschiedenen Behandlungssituationen auf das Gesamtergebnis wurde nicht untersucht.

- Im Krankheitsbild Darmkrebs erfolgte die Zusammenstellung der Datengrundlage häufig unsystematisch oder das Vorgehen blieb unklar.

Die Gesamtschau der Validierungsstudien zeigt, dass diese Mängel grundsätzlich behebbar sind. Es wäre im Interesse einer möglichst informierten Therapie von Patienten mit onkologischen Erkrankungen wünschenswert, wenn systematisch Validierungsstudien unter Beachtung der genannten methodischen Standards durchgeführt und transparent berichtet würden. Dies erfordert eine gemeinsame Anstrengung aller Beteiligten. Die damit verbundene Verbesserung der Erkenntnisse zum Stellenwert von Surrogatendpunkten wäre auch geeignet, Entscheidungswege für die Bewertung von neuen Arzneimitteln für onkologische Erkrankungen trotz begrenzter Datenlage zu diesen Arzneimitteln aufzuzeigen.

Vorschlag des IQWiG für den Umgang mit Surrogatendpunkten in der Nutzenbewertung onkologischer Behandlungen

Bei der Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung sind 2 Ebenen zu trennen. Auf der ersten Ebene muss zunächst die Validität des Surrogats beurteilt werden. Auf der zweiten Ebene stellt sich die Frage, welche Schlussfolgerungen bezüglich patientenrelevanter Endpunkte aus den Surrogateffekten abhängig von der Ausprägung der Validität des Surrogats gezogen werden können.

Werden für eine Nutzenbewertung Ergebnisse für einen Surrogatendpunkt vorgelegt, ist zunächst zu prüfen, ob Daten zur Validierung des Surrogats zur Verfügung stehen. Ist das nicht der Fall, liegt aus den Ergebnissen zu dem Surrogatendpunkt kein Beleg (oder Hinweis oder Anhaltspunkt) für einen Effekt bezüglich eines patientenrelevanten Endpunkts vor. Wenn Daten zur Validierung des Surrogats verfügbar sind, werden diese in einem ersten Schritt bezüglich ihrer Aussagesicherheit bewertet. Bei niedriger Aussagesicherheit der Validierungsergebnisse gibt es aus den Ergebnissen zu dem Surrogatendpunkt grundsätzlich keinen Beleg für einen Effekt bezüglich eines patientenrelevanten Endpunkts. Bei hoher, eingeschränkter oder mäßiger Aussagesicherheit der Validierungsergebnisse ist die Schlussfolgerung zu patientenrelevanten Endpunkten abhängig vom Grad der Korrelation der Effekte von Surrogat und patientenrelevantem Endpunkt in den Validierungsstudien sowie ggf. von der Größe des Effekts für das Surrogat in den Studien zur Nutzenbewertung. Hierzu wird der Effekt auf das Surrogat, der sich aus den Studien zur Nutzenbewertung ergibt, einem Surrogate Threshold Effect gegenübergestellt. Das Grundprinzip ist dabei, dass die Sicherheit der Aussage in der Nutzenbewertung (z. B. Beleg für oder Hinweis auf einen Effekt auf einen patientenrelevanten Endpunkt) umso höher ist, je sicherer ein Effekt auf das Surrogat auch einen Effekt auf den patientenrelevanten Endpunkt abbildet.

Schlagwörter: Biologische Marker; Dateninterpretation, statistische; Vorhersagewert von Tests; medizinische Onkologie; Mammatumoren; Kolontumoren; systematische Übersicht

Inhaltsverzeichnis

Seite

Tabellenverzeichnis ..... **xx**
Abbildungsverzeichnis ..... **xxi**
Abkürzungsverzeichnis ..... **xxii**
1 Hintergrund **1**
2 Ziele der Untersuchung **4**
3 Projektbearbeitung **5**
3.1 Zeitlicher Verlauf des Projekts **5**
3.2 Dokumentation der Änderungen nach Projektabschluss (Errata) **5**
4 Methoden **6**
4.1 Teilziel 1: Methodische Verfahren zur Validierung von Surrogatendpunkten **6**
4.1.1 Kriterien für den Einschluss von Studien in die Untersuchung 6
4.1.1.1 Inhalt 6
4.1.1.2 Jahr der Veröffentlichung 6
4.1.1.3 Sprache 6
4.1.1.4 Tabellarische Übersicht der Kriterien für den Studieneinschluss 7
4.1.2 Informationsbeschaffung 7
4.1.2.1 Bibliografische Literaturrecherche 7
4.1.2.2 Selektion relevanter Publikationen 7
4.1.3 Informationssynthese und -analyse 8
4.2 Teilziel 2: Studien zur Validierung von Parametern des Tumoransprechens am Beispiel Darmkrebs und Brustkrebs **8**
4.2.1 Kriterien für den Einschluss von Studien in die Untersuchung 8
4.2.1.1 Inhalt 8
4.2.1.2 Krankheitsbilder in den Studien zur Validierung 8
4.2.1.3 Therapeutische Interventionen in den Studien zur Validierung 8
4.2.1.4 Sprache 9
4.2.1.5 Tabellarische Übersicht der Kriterien für den Studieneinschluss 9
4.2.2 Informationsbeschaffung 9
4.2.2.1 Bibliografische Literaturrecherche 9
4.2.2.2 Selektion relevanter Studien 10
4.2.3 Informationsbewertung 10
4.2.4 Informationssynthese und -analyse 14
4.2.4.1 Gegenüberstellung der Ergebnisse der Einzelstudien 14

4.2.4.2 Meta-Analysen ... 14
4.2.4.3 Sensitivitätsanalyse ... 14
4.2.4.4 Faktoren, die einen Einfluss auf die Validierung haben können ... 14
4.3 Änderungen der Methodik im Vergleich zur Projektskizze ... 15
5 Ergebnisse ... 16
5.1 Teilziel 1: Methodische Verfahren zur Validierung von Surrogatendpunkten ... 16
5.1.1 Ergebnisse der Informationsbeschaffung ... 16
5.1.1.1 Ergebnis der bibliografischen Literaturrecherche ... 16
5.1.1.2 Resultierender Studienpool ... 17
5.1.2 Beschreibung der methodischen Verfahren zur Validierung von Surrogatendpunkten ... 19
5.1.2.1 Zusammenfassung der vorhandenen Übersichtsarbeiten ... 19
5.1.2.2 Klassifizierung von Methoden zur Surrogatvalidierung ... 23
5.1.2.3 Beschreibung der Arbeiten aus der Literaturrecherche ... 25
5.1.3 Zusammenfassung ... 28
5.2 Teilziel 2: Studien zur Validierung von Parametern des Tumoransprechens am Beispiel Darmkrebs und Brustkrebs ... 29
5.2.1 Ergebnisse der Informationsbeschaffung ... 29
5.2.1.1 Ergebnis der bibliografischen Literaturrecherche ... 29
5.2.1.2 Nicht weiter berücksichtigte Korrelationsstudien ... 32
5.2.1.3 Resultierender Studienpool ... 32
5.2.2 Charakteristika der in die Bewertung eingeschlossenen Studien ... 33
5.2.2.1 Indikationen, Interventionen und untersuchte Endpunkte ... 33
5.2.2.2 Validierungsverfahren in den eingeschlossenen Studien ... 47
5.2.2.3 Einschätzung der Aussagesicherheit der eingeschlossenen Studien ... 63
5.2.3 Ergebnisse zur Validierung von Surrogatendpunkten für patientenrelevante Endpunkte ... 71
5.2.3.1 Darmkrebs ... 72
5.2.3.2 Brustkrebs ... 90
5.2.4 Sensitivitätsanalyse ... 99
5.2.5 Faktoren, die einen Einfluss auf die Validierung haben können ... 99
5.2.6 Zusammenfassung der Beleglage ... 99
6 Diskussion ... 107
6.1 Ergebnis des Methodenreviews: mögliches Vorgehen zur Validierung von Surrogatendpunkten ... 107
6.2 Methodik und Ergebnisse der Untersuchungen zur Validierung von Parametern des Tumoransprechens bei den Krankheitsbildern Darmkrebs und Brustkrebs ... 112
6.3 Vorschlag für einen Algorithmus zum Umgang mit Surrogatendpunkten in onkologischen Studien in der Nutzenbewertung ... 115

6.4 Abschließende Anmerkungen 121
7 Fazit 124
8 Liste der eingeschlossenen Studien 126
8.1 Teilziel 1 126
8.2 Teilziel 2 129
8.2.1 Krankheitsbild Darmkrebs 129
8.2.2 Krankheitsbild Brustkrebs 130
9 Literatur 134
Anhang A – Suchstrategien 142
Anhang B – Liste der ausgeschlossenen Dokumente zum Thema mit Ausschlussgründen 153
B.1 – Liste der ausgeschlossenen Dokumente zu Teilziel 1 (methodische Verfahren zur Validierung von Surrogatendpunkten) 153
B.2 – Liste der ausgeschlossenen Dokumente zu Teilziel 2 - Krankheitsbild Darmkrebs 154
B.3 – Liste der ausgeschlossenen Dokumente zu Teilziel 2 - Krankheitsbild Brustkrebs 155
Anhang C – Offenlegung potenzieller Interessenkonflikte des externen Reviewers 159

Tabellenverzeichnis

Seite

Tabelle 1: Definition von Endpunkten für onkologische Studien in Leitlinien der Zulassungsbehörden 2

Tabelle 2: Teilziel 1: Übersicht der Kriterien für den Studieneinschluss (Methodenbewertung) 7

Tabelle 3: Teilziel 2: Übersicht der Kriterien für den Studieneinschluss (Validierungsstudien) 9

Tabelle 4: Studienpool für Teilziel 1 (Methodische Verfahren) 17

Tabelle 5: Studienpool für Teilziel 2 – Krankheitsbild Darmkrebs 32

Tabelle 6: Studienpool für Teilziel 2 – Krankheitsbild Brustkrebs 33

Tabelle 7: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Darmkrebs 35

Tabelle 8: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Brustkrebs 43

Tabelle 9: Methodenkategorien in den eingeschlossenen Validierungsstudien 47

Tabelle 10: Methodik zur Validierung der Surrogatendpunkte – Darmkrebs 50

Tabelle 11: Methodik zur Validierung der Surrogatendpunkte – Brustkrebs 60

Tabelle 12: Bewertung der Aussagesicherheit der eingeschlossenen Validierungsstudien – Darmkrebs 66

Tabelle 13: Bewertung der Aussagesicherheit der eingeschlossenen Validierungsstudien – Brustkrebs 69

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs 75

Tabelle 15: Ergebnisse der Validierungsstudien – Brustkrebs 92

Tabelle 16: Landkarte der Beleglage – Darmkrebs 100

Tabelle 17: Landkarte der Beleglage – Brustkrebs 105

Abbildungsverzeichnis

Abbildung 1: Literaturrecherche und Literaturscreening für Teilziel 1 (Methodische Verfahren) 16

Abbildung 2: Literaturrecherche und Literaturscreening für Teilziel 2 – Krankheitsbild Darmkrebs 30

Abbildung 3: Literaturrecherche und Literaturscreening für Teilziel 2 – Krankheitsbild Brustkrebs 31

Abbildung 4: Einfluss der Aussagesicherheit der Validierungsstudien und der Korrelation der Effekte von Surrogat und patientenrelevantem Endpunkt auf die Validität des Surrogats 111

Abbildung 5: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: Stellenwert von Validierungsstudien und deren Aussagesicherheit 116

Abbildung 6: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: mögliche Schlussfolgerungen bei hoher Aussagesicherheit der Validierungsstudien 117

Abbildung 7: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: mögliche Schlussfolgerungen bei eingeschränkter Aussagesicherheit der Validierungsstudien 119

Abbildung 8: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: mögliche Schlussfolgerungen bei mäßiger Aussagesicherheit der Validierungsstudien 120

Abkürzungsverzeichnis

Abkürzung	Bedeutung
A	Adriamycin (= Doxorubicin)
AA	Adjusted Association
AC	Adriamycin, Cyclophosphamid
ACCENT	Adjuvant Colon Cancer End Points Group
ACCMAP	Advanced Colorectal Cancer Meta-Analyis Project
ACMD	average clinically meaningful difference
AD	Adriamycin, Docetaxel
AP	Adriamycin, Paclitaxel
APEO	average prediction error of other true endpoints
APEP	average prediction error of the predicted effect
ASCO	American Society of Clinical Oncology
BCG	Bacillus Calmette-Guérin
BV	Bevacizumab
CR	Complete response (komplettes Ansprechen)
D	Docetaxel
DAC	Docetaxel, Adriamycin, Cyclophosphamid
DC	Disease control (Krankheitskontrolle)
DFS	Disease-free survival (krankheitsfreies Überleben)
E	Epirubicin
EC	Epirubicin, Cyclophosphamid
ECCO	European Cancer Organisation
ECOG	Eastern Cooperative Oncology Group
ED	Epirubicin, Docetaxel
EFS	Event-free survival (eventfreies Überleben)
EMA	European Medicines Agency
EP	Epirubicin, Paclitaxel
ERO	Excess Relative Odds
ESMO	European Society of Medical Oncology
FAC	Fluorouracil, Adriamycin, Cyclophosphamid
FDA	Food and Drug Administration

Abkürzung	Bedeutung
FEC	Fluorouracil, Epirubicin, Cyclophosphamid
FECS	Federation of European Cancer Societies, (heute ECCO: European Cancer Organisation)
5-FU	Fluorouracil
FUDR	Floxuridin
G-BA	Gemeinsamer Bundesausschuss
HER 2	Human epidermal growth factor receptor 2 (humaner epidermaler Wachstumsfaktorrezeptor 2)
HR	Hazard ratio
HTA	Health technology assessment
INF-α	Interferon-alpha
i. v.	intravenös
IPD	individuelle Patientendaten
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
k. A.	keine Angabe
KI	Konfidenzintervall
LEV	Levamisol
LRF	Likelihood Reduction Factor
LV	Leucovorin
MAGIC	Meta-Analysis Group in Cancer
MBC	metastatic breast cancer
MCRC	metastatic colorectal cancer
MOF	Semustin, Vincristin, Fluorouracil
MTX	Methotrexat
NCI	National Cancer Institute
NSABP	National Surgical Adjuvant Breast and Bowel Project
OP	Operation
OR	Odds ratio
ORR	Objective response rate (objektive Ansprechrate)
OS	Overall survival (Gesamtüberleben)
P	Paclitaxel
PD	progressive disease (progressive Erkrankung)

Abkürzung	Bedeutung
PEP	patientenrelevanter Endpunkt
PFS	Progression-free survival (progressionsfreies Überleben)
PH-Modell	Proportional-Hazards-Modell
PIG	Proportion of Information Gain
PR	partial response (partielles Ansprechen)
PTE	Proportion of treatment effect explained
RCT	randomised controlled trial (randomisierte kontrollierte Studie)
RE	relativer Effekt
RFS	Recurrence-free survival (Überleben ohne Wiederauftreten)
RR	Response rate (Ansprechrate)
SABCS	San Antonio Breast Cancer Symposium
SD	stable disease (stabile Erkrankung)
SEP	Surrogatendpunkt
STE	Surrogat Threshold Effect
TR	Tumor response (Tumoransprechen)
TTF	Time to treatment failure (Zeit bis zum Therapieversagen)
TTP	Time to tumor progression (Zeit bis zur Tumorprogression)
TTR	Time to tumor recurrence (Zeit bis zum Wiederauftreten des Tumors)
VM	Vincristin, Methotrexat
VP	Vincristin, Prednison
VRF	Variance Reduction Factor
vs.	versus
WHO	World Health Organization

1 Hintergrund

Bei der Behandlung onkologischer Erkrankungen stellt die Verlängerung des Überlebens des Patienten i. d. R. ein herausragendes Therapieziel dar. Das Überleben des Patienten (Gesamtüberleben, overall survival) wird deshalb zumeist auch als Endpunkt in klinischen Studien untersucht. Weitere patientenrelevante Endpunkte sind u. a. die Symptomatik der Erkrankung, die unerwünschten Wirkungen der Therapie und die Lebensqualität der Patienten.

Da Aussagen zum Einfluss einer Therapie auf das Überleben von Patienten häufig erst nach längerer Beobachtungsdauer möglich sind und entsprechende Studien häufig große Fallzahlen benötigen, sind eine Reihe von intermediären Endpunkten (Surrogatendpunkte) entwickelt worden, mit denen versucht wird, den Nutzen einer Therapie abzuschätzen. Surrogatendpunkte sind dabei solche Endpunkte, die anstelle von den eigentlichen interessierenden (patientenrelevanten) Endpunkten gemessen werden [1].

In der Onkologie werden dazu insbesondere Parameter des Tumoransprechens verwendet. Beispiele für Parameter des Tumoransprechens sind ORR (objective response rate, objektive Ansprechrate), TTP (time to tumor progression, Zeit bis zur Progression des Tumors), TTF (time to treatment failure, Zeit bis zum Therapieabbruch) oder PFS (progression-free survival, progressionsfreies Überleben) [2,3].

Die folgende Tabelle zeigt Definitionen des Endpunkts Gesamtüberleben bzw. von Endpunkten des Tumoransprechens aus Leitlinien der europäischen und US-amerikanischen Zulassungsbehörden.

Tabelle 1: Definition von Endpunkten für onkologische Studien in Leitlinien der Zulassungsbehörden

Endpunkt	**Definition EMA**	**Definition FDA**
Gesamtüberleben		
OS: Overall survival	time from randomisation to death from any cause	Overall survival is defined as the time from randomization until death from any cause
Endpunkte auf Basis von Parametern des Tumoransprechens		
DFS: Disease-free survival	time from randomisation to (objective) recurrence or death from any cause	time from randomization until recurrence of tumor or death from any cause
CR: Complete response		no detectable evidence of tumor
PR: Partial response		
ORR: Objective response rate	the proportion of patients in whom a CR or PR was observed	proportion of patients with tumor size reduction of a predefined amount and for a minimum time period Response duration usually is measured from the time of initial response until documented tumor progression. Generally, the FDA has defined ORR as the sum of partial responses plus complete responses. Stable disease should not be a component of ORR.
TTP: Time to tumour progression	time from randomisation to observed tumour progression, censoring for death without progression	TTP is defined as the time from randomization until objective tumor progression. TTP does not include deaths.
PFS: Progression-free survival	time from randomisation to objective tumour progression or death from any cause	PFS is defined as the time from randomization until objective tumor progression or death. The precise definition of tumor progression is important and should be carefully detailed in the protocol.
EFS: Event-free survival	variable protocol specific definitions, e.g. time from randomisation to objective tumour progression, secondary malignancy, or cancer-related death	
TTF: Time to treatment failure	time from randomisation to discontinuation of therapy or add-on of new anti-cancer therapy for any reason including death, progression, toxicity	TTF is defined as a composite endpoint measuring time from randomization to discontinuation of treatment for any reason, including disease progression, treatment toxicity, and death.

Für die Messung der genannten Parameter des Tumoransprechens werden in der Regel bildgebende Verfahren eingesetzt, die darauf abzielen, Veränderungen der Größe des Tumors darzustellen.

Tumorwachstum bzw. die Progression einer onkologischen Erkrankung kann sich auch in einer Symptomatik oder in einer Veränderung der gesundheitsbezogenen Lebensqualität der Patienten niederschlagen. Die Symptomatik bzw. die gesundheitsbezogene Lebensqualität ist ein direkt patientenrelevantes Maß für die Progression der Erkrankung. So grenzt auch die FDA z. B. den Endpunkt „Zeit bis zur Progression von Symptomen“ als direktes Maß des klinischen Nutzens von der „Zeit bis zur Progression des Tumors“ als potenzielles Surrogat ab [2].

Der Gemeinsame Bundesausschuss (G-BA) hat das IQWiG mit der methodischen Bewertung der Aussagekraft von Surrogatendpunkten in Arzneimittelstudien in der Onkologie beauftragt. Dabei sollte laut Auftrag untersucht werden, welche Parameter des Tumoransprechens wie zum Beispiel DFS (disease-free survival, krankheitsfreies Überleben), ORR (objective response rate, objektive Ansprechrate), TTP (time to progression, Zeit bis zur Progression), TTF (time to treatment failure, Zeit bis zum Therapieabbruch), PFS (progression-free survival, progressionsfreies Überleben) valide Surrogate für patientenrelevante Endpunkte sein können. Die Bewertung sollte sowohl allgemein als auch für bestimmte Krankheitsbilder erfolgen, zu denen ausreichende Daten vorliegen.

Der vorliegende Bericht untersuchte daher zum einen die Frage, welche methodischen Verfahren zur Validierung von Surrogatendpunkten vorliegen. Zum anderen wurde exemplarisch für 2 Krankheitsbilder (Brustkrebs und Darmkrebs) untersucht, für welche Parameter des Tumoransprechens gezeigt wurde, dass der Effekt von Interventionen mit vergleichbarem Wirkmechanismus auf den zu ersetzenden patientenrelevanten Endpunkt durch den Effekt auf den Surrogatendpunkt in einem ausreichenden Ausmaß erklärt wird. Aus diesen Analysen sollten Empfehlungen für die Validierung von Surrogatendpunkten in der Onkologie abgeleitet werden.

2 Ziele der Untersuchung

Primäres Ziel der vorliegenden Untersuchung war es, Empfehlungen für die Validierung von Surrogatendpunkten in der Onkologie darzustellen.

Diese Empfehlungen sollten aus den Ergebnissen folgender 2 Teilziele abgeleitet werden:

1. zusammenfassende Darstellung und Bewertung methodischer Verfahren, deren Ziel die Validierung von Surrogatendpunkten ist,

2. Darstellung und Bewertung von Studien zur Validierung von Parametern des Tumoransprechens als Surrogate für patientenrelevante Endpunkte, insbesondere für das Überleben von Patienten.

Die Bearbeitung von Teilziel 2 erfolgte beispielhaft für die Krankheitsbilder Darmkrebs und Brustkrebs.

3 Projektbearbeitung

3.1 Zeitlicher Verlauf des Projekts

Der Gemeinsame Bundesausschuss (G-BA) hat am 19.08.2010 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Bewertung der Aussagekraft von Surrogatendpunkten in der Onkologie beauftragt.

Die Bearbeitung des Projekts erfolgte durch eine interne Projektgruppe. Auf Basis der Projektskizze wurde ein Rapid Report erstellt. Der Rapid Report wurde einem externen Review unterzogen. Im Anschluss wurde der Bericht an den G-BA übermittelt und auf der Website des IQWiG veröffentlicht.

3.2 Dokumentation der Änderungen nach Projektabschluss (Errata)

Nach Verabschiedung des Rapid Reports Version 1.0 vom 31.01.2011 und Versendung an den G-BA wurde die unten aufgeführte Änderung vorgenommen. Diese Änderung führte nicht zu einer Änderung des Fazits.

- In der Kurzfassung auf Seite viii, 1. Absatz von oben, wurden der erste und zweite Satz „Bei hoher Aussagesicherheit der Validierungsstudie kann aus einer hohen Korrelation ($R \geq 0,85$) der Effekte von Surrogat und patientenrelevantem Endpunkt ein Nachweis für die Validität des Surrogats abgeleitet werden. Aus einer niedrigen Korrelation ($R \leq 0,7$) ergibt sich ein Nachweis für das Fehlen der Validität des Surrogats.“ gegen die folgenden Sätze ausgetauscht: „Bei hoher Aussagesicherheit der Validierungsstudie kann aus einer hohen Korrelation (untere Grenze des 95 %-Konfidenzintervalls) $R \geq 0,85$ der Effekte von Surrogat und patientenrelevantem Endpunkt ein Nachweis für die Validität des Surrogats abgeleitet werden. Aus einer niedrigen Korrelation (obere Grenze des 95 %-Konfidenzintervalls) $R \leq 0,7$ ergibt sich ein Nachweis für das Fehlen der Validität des Surrogats.“

4 Methoden

Die Methoden werden getrennt für die beiden Teilziele dargestellt. Abschnitt 4.1 beschreibt die Methoden zur Darstellung und Bewertung methodischer Verfahren, deren Ziel die Validierung von Surrogatendpunkten ist. Abschnitt 4.2 stellt die Methodik zur Bewertung von Studien zur Validierung von Parametern des Tumoransprechens als Surrogate für patientenrelevante Endpunkte für ausgewählte Krankheitsbilder dar.

4.1 Teilziel 1: Methodische Verfahren zur Validierung von Surrogatendpunkten

4.1.1 Kriterien für den Einschluss von Studien in die Untersuchung

4.1.1.1 Inhalt

Eingeschlossen wurden Publikationen, in denen ein neues Verfahren zur Validierung von Surrogatendpunkten vorgestellt oder ein bereits existierendes Verfahren methodisch erweitert oder kommentiert wurde, sowie systematische Übersichten, in denen Validierungsverfahren vorgestellt wurden. Im Fall einer systematischen Übersicht mussten Arbeiten mit mindestens einer der oben genannten Eigenschaften Teil der Suchstrategie und der Ergebnisdarstellung dieser systematischen Übersicht sein.

Publikationen, die Validierungsverfahren vorstellten, die spezifisch für eine nicht-onkologische Indikation entwickelt oder betrachtet werden, wurden nicht berücksichtigt.

4.1.1.2 Jahr der Veröffentlichung

Im Rahmen einer Vorabrecherche wurden 3 relevante systematische Übersichten zum Thema Validierung von Surrogatendpunkten identifiziert (Weir 2006 [4], Lassere 2008 [5] und Mangiapane 2009 [6]). Die aktuellste systematische Übersicht Mangiapane 2009 umfasste einen Suchzeitraum bis einschließlich März 2007. Die Recherche für Teilziel 1 beschränkt sich daher auf den Zeitraum ab 2007.

4.1.1.3 Sprache

Es wurden Publikationen in deutscher oder englischer Sprache eingeschlossen.

4.1.1.4 Tabellarische Übersicht der Kriterien für den Studieneinschluss

Die folgende Tabelle zeigt die Kriterien für den Einschluss von Studien in die Bearbeitung von Teilziel 1 (Darstellung und Bewertung der methodischen Ansätze zur Validierung von Surrogatendpunkten).

Tabelle 2: Teilziel 1: Übersicht der Kriterien für den Studieneinschluss (Methodenbewertung)

Einschlusskriterien	
ME1	Methodische Arbeiten, die ein Validierungsverfahren für Surrogate entwickeln, erweitern oder kommentieren (siehe auch Abschnitt 4.1.1.1)
ME2	Publikation 2007 oder später (siehe auch Abschnitt 4.1.1.2)
ME3	Sprache Deutsch oder Englisch
ME4	Vollpublikation verfügbar

4.1.2 Informationsbeschaffung

4.1.2.1 Bibliografische Literaturrecherche

Die systematische Literaturrecherche nach relevanten Publikationen wurde in MEDLINE, dem Cochrane Methodology Register (Methods Studies) sowie in der Health Technology Assessment Database (Technology Assessments) durchgeführt.

Die Suchstrategien für die Suche in bibliographischen Datenbanken finden sich in Anhang A.

Die Recherche nach methodischen Arbeiten fand am 24.09.2010 statt.

4.1.2.2 Selektion relevanter Publikationen

Die Selektion relevanter Publikationen erfolgte durch 2 Reviewer unabhängig voneinander. Dazu wurde das Ergebnis der Recherche in den oben genannten Quellen herangezogen.

In einem ersten Selektionsschritt wurden auf Basis des Titels und, sofern vorhanden, des Abstracts, alle gemäß den Einschlusskriterien sicher nicht themenrelevanten Dokumente aussortiert. Im zweiten Selektionsschritt wurden alle verbliebenen Arbeiten im Volltext gesichtet. Diskrepante Bewertungen durch die Gutachter wurden am Ende jedes Selektionsschrittes durch Diskussion aufgelöst.

Zusätzlich zu den Publikationen, die durch die in Abschnitt 4.1.2.1 beschriebene Suche identifiziert bzw. darauffolgend selektiert wurden, wurden auch die im Rahmen der Vorabrecherche identifizierten systematischen Übersichten Weir 2006 [4], Lassere 2008 [5] und Mangiapane 2009 [6] herangezogen. Darüber hinaus wurden auch das Standardwerk Burzykowski 2005 [7] sowie die Übersicht Molenberghs 2010 [8], die das Standardwerk Burzykowski 2005 referenziert und aktuelle Entwicklungen beschreibt, berücksichtigt.

4.1.3 Informationssynthese und -analyse

Die Informationen wurden einer Informationssynthese und -analyse unterzogen.

Die nach Abschluss der Studienselektion als relevant identifizierten Publikationen wurden zunächst tabellarisch mit Angaben zu Autoren, Jahr der Veröffentlichung und Titel aufgeführt.

Die in den Publikationen behandelten Verfahren wurden unter grundsätzlichen methodischen Gesichtspunkten klassifiziert und zusammenfassend beschrieben. Die Klassifizierung erfolgte anhand der in den Publikationen verwendeten Methoden zur Validierung. Aus den Publikationen wurden die grundsätzlichen methodischen Anforderungen an die Validierung eines Surrogats abgeleitet und beschrieben.

4.2 Teilziel 2: Studien zur Validierung von Parametern des Tumoransprechens am Beispiel Darmkrebs und Brustkrebs

4.2.1 Kriterien für den Einschluss von Studien in die Untersuchung

4.2.1.1 Inhalt

Eingeschlossen wurden Studien, in denen eine Validierung von Endpunkten des Tumoransprechens als Surrogate patientenrelevanter Endpunkte vorgenommen wurde. Eine Einschränkung bezüglich der für die Validierung verwendeten Methodik wurde für die Recherche nach Validierungsstudien nicht vorgenommen.

4.2.1.2 Krankheitsbilder in den Studien zur Validierung

Die Vorabrecherche ergab Hinweise darauf, dass die Validität von Parametern des Tumoransprechens als Surrogate für patientenrelevante Endpunkte ggf. für verschiedene Krankheitsbilder variiert [9]. Aus diesem Grund wurden Studien eingeschlossen, die eine Validierung innerhalb eines definierten Krankheitsbildes vornehmen. Für das vorliegende Projekt wurden die Krankheitsbilder Darmkrebs und Brustkrebs ausgewählt, da zum einen für diese Erkrankungen vergleichsweise häufig Arzneimittel neu zugelassen werden und zum anderen die Vorabrecherche auf eine relevante Studienmenge schließen ließ.

4.2.1.3 Therapeutische Interventionen in den Studien zur Validierung

Studien zur Validierung von Parametern des Tumoransprechens können auf Basis der Ergebnisse unterschiedlicher therapeutischer Interventionen vorgenommen werden. So kann die Behandlung mit verschiedenen medikamentösen Verfahren, aber auch mit nichtmedikamentösen Verfahren erfolgen.

Eine Einschränkung der Auswahl von Validierungsstudien auf Basis der eingeschlossenen Interventionen erfolgte nicht.

4.2.1.4 Sprache

Es wurden Publikationen in deutscher oder englischer Sprache eingeschlossen.

4.2.1.5 Tabellarische Übersicht der Kriterien für den Studieneinschluss

Die folgende Tabelle zeigt die Kriterien für den Einschluss von Studien in die Bearbeitung von Teilziel 2 (Darstellung und Bewertung von Studien zur Validierung von Parametern des Tumoransprechens als Surrogate für patientenrelevante Endpunkte).

Tabelle 3: Teilziel 2: Übersicht der Kriterien für den Studieneinschluss (Validierungsstudien)

Einschlusskriterien	
VE1	Studien zur Validierung von Parametern des Tumoransprechens als Surrogate für patientenrelevante Endpunkte (siehe auch Abschnitt 4.2.1.1)
VE2	Krankheitsbild: Darmkrebs oder Brustkrebs (siehe auch Abschnitt 4.2.1.2)
VE3	Validierung auf Basis der Ergebnisse jeder therapeutischen Intervention (siehe auch Abschnitt 4.2.1.3)
VE4	Sprache Deutsch oder Englisch
VE5	Vollpublikation verfügbar

4.2.2 Informationsbeschaffung

4.2.2.1 Bibliografische Literaturrecherche

Die systematische Literaturrecherche nach relevanten Studien wurde in folgenden Quellen durchgeführt:

- Suche nach Primärstudien in den bibliografischen Datenbanken MEDLINE, EMBASE sowie dem Cochrane Methodology Register (Methods Studies)
- Sichtung relevanter Sekundärpublikationen zur Identifikation weiterer Primärstudien; Identifizierung der Sekundärpublikationen mittels Suche in den Datenbanken MEDLINE und EMBASE

Die Suchstrategien für die Suche in bibliographischen Datenbanken finden sich in Anhang A.

Die Recherche nach Validierungsstudien zu Darmkrebs fand am 24.08.2010 statt. Die Recherche nach Validierungsstudien zu Brustkrebs fand am 07.10.2010 statt.

4.2.2.2 Selektion relevanter Studien

Die Selektion relevanter Studien erfolgte durch 2 Reviewer unabhängig voneinander. Dazu wurde das Ergebnis der Recherche in den oben genannten Quellen herangezogen. Diskrepanzen wurden durch Diskussion der beiden Reviewer konsentiert.

Im Selektionsprozess wurden die Suchergebnisse zunächst 3 Kategorien zugeordnet:

1. Dokumente zum Thema

   Dies sind Dokumente, die folgende minimale Einschlusskriterien erfüllen:

   a) Studie zur Validierung eines Surrogatendpunktes

   b) Humanstudie (Beobachtung am Menschen, Originalarbeit)

2. Sekundärpublikationen (systematische Übersichten, HTA-Berichte) zur Fragestellung der Bewertung

3. Dokumente, die nicht themenrelevant sind, da sie nicht die minimalen Einschlusskriterien erfüllen bzw. keine Sekundärpublikationen zur vorliegenden Fragestellung sind

Innerhalb der Dokumente zum Thema wurden dann die Studien selektiert, die die vollständigen Kriterien für den Studieneinschluss erfüllten (siehe 4.2.1.5).

Im ersten Selektionsschritt wurden auf Basis des Titels und, sofern vorhanden, des Abstracts alle sicher nicht themenrelevanten Dokumente (also Dokumente zu 3.) gemäß den minimalen Einschlusskriterien aussortiert. Im zweiten Selektionsschritt wurden aus den verbleibenden Zitaten ebenfalls auf Basis des Titels und ggf. Abstracts alle Zitate aussortiert, die gemäß den vollständigen Ein- und Ausschlusskriterien sicher als nicht für die Bewertung relevante Dokumente zum Thema klassifiziert werden konnten (Pool der begründet auszuschließenden Studien). Alle verbleibenden Dokumente wurden im dritten Selektionsschritt im Volltext gesichtet und abschließend einer der oben genannten 3 Kategorien zugeordnet. Innerhalb der Dokumente zum Thema (also Dokumente zu 1.) wurden die Studien dem für die Bewertung relevanten Studienpool oder dem Pool der begründet auszuschließenden Studien zugeordnet.

Ausgeschlossene Studien zu 1. wurden mit Ausschlussgrund dokumentiert (s. Anhang B); nicht themenrelevante Dokumente (zu 3.) wurden nicht dokumentiert.

4.2.3 Informationsbewertung

Die Bewertung der Informationen der eingeschlossenen Studien hing stark von den verfügbaren Angaben und der Qualität der jeweiligen Publikationen ab.

Datenextraktion

Alle für die Validierung der Parameter des Tumoransprechens notwendigen Informationen wurden aus den Publikationen zu den eingeschlossenen Studien in standardisierte Tabellen extrahiert (siehe Tabelle 7 bis Tabelle 15).

Bewertung der Aussagesicherheit der Ergebnisse zur Validierung

Bei der Bewertung der Aussagesicherheit der Ergebnisse zur Validierung wurden folgende Kriterien berücksichtigt:

Anerkanntes, beschriebenes Verfahren verwendet

Die Adäquatheit der statistischen Validierungsansätze wurde danach beurteilt, ob das jeweilige Verfahren in der Fachliteratur anerkannt bzw. angemessen beschrieben wurde. Dies galt als erfüllt, sofern ein Verfahren in einer biometrischen Fachzeitschrift (mit Peer-review-Prozess) vorgestellt und nachfolgend von anderen Autoren nicht grundsätzlich infrage gestellt wurde. Davon wurden Verfahren abgegrenzt, für die keine methodische Beschreibung in der Literatur bekannt ist und die noch nicht systematisch untersucht wurden.

Eine vollständige methodische Bewertung konnte auf derzeitigem Wissensstand für kein Verfahren erfolgen, weil die jeweiligen methodischen Ansätze nicht umfassend und abschließend erforscht waren.

Verfahren, die in der Fachliteratur anerkannt bzw. angemessen beschrieben wurden, wurde eine höhere Aussagesicherheit zugesprochen.

Analysen zur Robustheit und Generalisierbarkeit

Für die eingeschlossenen Validierungsstudien wurde bewertet, ob durch geeignete Analysen die Robustheit bzw. die Generalisierbarkeit der Ergebnisse überprüft wurde. Zur Untersuchung der Robustheit eignen sich Analysen, bei denen z. B. die Analyse für eine bestimmte Teilmenge des verfügbaren Datensatzes wiederholt wird oder mit denen der Einfluss weniger einzelner Beobachtungen auf das Gesamtergebnis beurteilt werden kann. Mit Analysen zur Generalisierbarkeit wird überprüft, ob die für einen bestimmten Datensatz gewonnenen Ergebnisse als übertragbar auf vergleichbare Datensätze angesehen werden können. Dies lässt sich z. B. durch Kreuzvalidierungsverfahren erreichen oder durch Analysen, bei denen die Ergebnisse auf Basis eines Testdatensatzes auf ihre Gültigkeit bzgl. eines neuen Datensatzes überprüft werden. Die hier genannten Analysen beziehen sich nur auf die Überprüfung der statistischen Ergebnisse. Medizinisch-inhaltliche Aspekte (siehe unten) werden hierdurch nicht berührt. Waren solche Analysen jeweils verfügbar, wurde dies mit „ja“ bewertet. Das Vorliegen von solchen Analysen führte zu einer höheren Aussagesicherheit, wobei den Analysen zur Generalisierbarkeit ein höherer Stellenwert zugesprochen wurde.

Datengrundlage

Die für die Validierung eines Surrogatendpunktes zugrunde gelegten Daten bzw. Studien wurden nach der Systematik in der Datengewinnung klassifiziert. Demnach war eine Einschätzung als „systematisch“ und „nicht systematisch“ möglich. Gab es für die Validierungsstudien, bei denen die Datengewinnung nicht systematisch erfolgte, Hinweise darauf, dass der Datensatz repräsentativ war, wurde dies mit „nicht systematisch, repräsentativer Datensatz“ bewertet. Waren die in der jeweiligen Publikation vorhandenen Informationen für eine eindeutige Einschätzung unzureichend, wurde die Datengrundlage als „unklar“ bewertet. In jedem Fall wurden randomisierte kontrollierte Studien als Voraussetzung angesehen. Wurden die Daten bzw. Studien, die zur Validierung herangezogen wurden, mittels systematischer Suche identifiziert, wurde diesen Validierungsstudien eine höhere Aussagesicherheit zugesprochen.

Eingrenzung der Indikationen und / oder Schweregrade

Unterschiedliche Indikationen (z. B. Kolonkarzinom und Rektumkarzinom) und / oder Schweregrade (z. B. nicht metastasierender oder lokal fortgeschrittener Brustkrebs sowie Kolonkarzinom Stadium I, II und III) innerhalb eines Krankheitsbildes sind zum Teil mit deutlich unterschiedlichen Prognosen behaftet, sodass die Validität eines Surrogatendpunktes, die für eine Indikation gezeigt ist, nicht ohne Weiteres auf eine andere Indikation übertragbar ist. Aus diesem Grund wurde überprüft, ob für die Validierung eines Surrogatendpunktes eine hinreichend klare Eingrenzung von Indikation und / oder Schweregrad vorgenommen wurde. Bei hinreichend klarer Eingrenzung der Indikation und / oder des Schweregrades erfolgte die Bewertung „ja“. Gab es Hinweise darauf, dass eine Zusammenfassung von Patientengruppen mit unterschiedlicher Prognose vorgenommen wurde, wurde dies als „mit Einschränkungen“ beurteilt. Wurden in den Validierungsstudien verschiedene Indikationen und / oder Schweregrade, die eindeutig mit unterschiedlichen Prognosen behaftet sind, zusammenfassend ausgewertet, erfolgte die Bewertung „nein“. Wurde die Krebserkrankung (Indikation und / oder Schweregrad) der in den Studien eingeschlossenen Patientenpopulationen nicht näher erläutert, erfolgte die Bewertung „unklar“. Je enger die für die Validierung des Surrogatendpunkts herangezogene Patientenpopulation definiert war, desto höher wurde die Aussagesicherheit bezogen auf dieses Kriterium bewertet. Bei Vorliegen geeigneter Sensitivitätsanalysen wie z. B. Analysen für bestimmte Teilpopulationen konnte eine Einschränkung der Aussagesicherheit vermieden werden.

Eingrenzung der Interventionen

Die Eingrenzung der für die Surrogatvalidierung untersuchten Interventionen wurde in den folgenden drei Abstufungen bewertet: Validierung für einen Wirkstoff, Validierung innerhalb einer Wirkstoffklasse (Wirkstoffe mit gleichem Wirkprinzip) sowie Validierung innerhalb verschiedener Wirkstoffklassen. Wurden die eingesetzten Interventionen in den Validierungsstudien nicht näher beschrieben, erfolgte die Bewertung „unklar“. Diese Bewertung wurde gleichermaßen für die in den Studien angewendete Kontrollintervention vorgenommen. Je höher die Eingrenzung der untersuchten Interventionen vorgenommen

wurde, desto höher wurde die Aussagesicherheit bezogen auf dieses Kriterium bewertet. Bei Vorliegen geeigneter Sensitivitätsanalysen, in denen z. B. die Heterogenität der Ergebnisse bei Einschränkung auf spezifische Interventionen überprüft wurde, konnte eine Einschränkung der Aussagesicherheit vermieden werden.

Endpunktdefinitionen

Zur Prüfung der Aussagesicherheit bezüglich der Endpunktdefinitionen wurden 2 Aspekte berücksichtigt. Zum einen wurde überprüft, ob in der Validierungsstudie eindeutige Definitionen für die untersuchten Endpunkte angegeben wurden (lag nur für das Gesamtüberleben keine Definition vor, führte das nicht zur Abwertung). War dies der Fall, wurde darüber hinaus anhand der Publikation der Validierungsstudie geprüft, ob diese Definitionen über die zur Validierung herangezogenen Einzelstudien einheitlich waren oder nicht. Entsprechend wurden die Endpunktdefinitionen als „einheitlich“ oder „uneinheitlich“ bewertet. Mit „einheitlich“ wurden auch solche Validierungsstudien bewertet, bei denen die Endpunktdefinitionen auch nicht relevant uneinheitlich waren. Waren die entsprechenden Informationen unzureichend, wurde die Einheitlichkeit der Endpunktdefinitionen als „unklar“ bewertet. Das Vorliegen einheitlicher Endpunktdefinitionen führte zu einer höheren Aussagesicherheit.

Gesamtbewertung

Die Bewertung der Aussagesicherheit insgesamt erfolgt unter Berücksichtigung aller zuvor genannten Kriterien. Es wird zwischen folgenden möglichen Ausprägungen unterschieden:

- Hohe Aussagesicherheit: Diese lag dann vor, wenn allen der oben genannten Kriterien die höchstmögliche Aussagesicherheit zugesprochen wurde.

- Eingeschränkte Aussagesicherheit: Diese lag dann vor, wenn in einzelnen Kriterien geringe Abweichungen von der höchstmöglichen Aussagesicherheit vorlagen (z.B. ggf. Validierung innerhalb einer Wirkstoffklasse ohne Prüfung des Einflusses einzelner Wirkstoffe auf das Ergebnis).

- Mäßige Aussagesicherheit: Eine mäßige Aussagesicherheit lag in den Fällen vor, in denen einzelne Kriterien in relevantem Maße oder mehrere Kriterien in geringem Ausmaß von der höchstmöglichen Aussagesicherheit abwichen.

- Niedrige Aussagesicherheit: Eine niedrige Aussagesicherheit lag zum einen dann vor, wenn mehrere Kriterien in relevantem Maße von der höchstmöglichen Aussagesicherheit abwichen. Andererseits konnte sie auch auf ein einzelnes Kriterium zurückzuführen sein, wenn für dieses so hohe Abweichungen von der höchstmöglichen Aussagesicherheit bestand, dass keine Aussagen über die Validität des Surrogatendpunktes und damit zur Vorhersagefähigkeit bez. des patientenrelevanten Endpunkts mehr getroffen werden konnten (zum Beispiel durch die Verwendung eines inadäquaten methodischen Verfahrens).

4.2.4 Informationssynthese und -analyse

Die Informationen wurden einer Informationssynthese und -analyse unterzogen. Wenn möglich, sollten über die Gegenüberstellung der Ergebnisse der Einzelstudien hinaus die unten beschriebenen Werkzeuge einzusetzen. Eine abschließende zusammenfassende Bewertung der Informationen erfolgte darüber hinaus in jedem Fall.

4.2.4.1 Gegenüberstellung der Ergebnisse der Einzelstudien

Die in den Studien berichteten Ergebnisse zur Validierung von Parametern des Tumoransprechens als Surrogate für patientenrelevante Endpunkte wurden im Bericht vergleichend beschrieben.

4.2.4.2 Meta-Analysen

Sofern die Studien hinsichtlich der Fragestellung und relevanter Charakteristika vergleichbar waren, sollten die Einzelergebnisse, soweit möglich, mithilfe von Meta-Analysen quantitativ zusammenzufassen.

4.2.4.3 Sensitivitätsanalyse

Zur Einschätzung der Robustheit der Ergebnisse waren, soweit möglich, Sensitivitätsanalysen hinsichtlich methodischer Faktoren geplant. Das Ergebnis der Sensitivitätsanalysen hätte die Sicherheit der aus den beobachteten Effekten abgeleiteten Aussagen beeinflussen können.

4.2.4.4 Faktoren, die einen Einfluss auf die Validierung haben können

Die Ergebnisse zur Validität von Parametern des Tumoransprechens als Surrogate patientenrelevanter Endpunkte sollten, soweit möglich, hinsichtlich potenzieller Faktoren, die einen Einfluss auf die Validierung haben können, untersucht werden. Dies konnten direkte Patientencharakteristika (Subgruppenmerkmale) sowie Spezifika der Behandlungen sein. Im Gegensatz zu den in Abschnitt 4.2.4.3 beschriebenen methodischen Faktoren für Sensitivitätsanalysen bestand hier das Ziel, mögliche Unterschiede zwischen Patientengruppen und Behandlungsspezifika aufzudecken.

Für einen Nachweis unterschiedlicher Ergebnisse zur Validität eines Surrogats (abhängig von Patientencharakteristika oder Behandlungsspezifika) war geplant, sofern möglich, einen Homogenitäts- bzw. Interaktionstest heranzuziehen. In die Untersuchung der Faktoren sollten die vorliegenden Ergebnisse aus Regressionsanalysen, die Interaktionsterme beinhalteten, und aus Subgruppenanalysen einbezogen werden. Außerdem hätten eigene Analysen unter Kategorisierung der Studien bezüglich der möglichen Faktoren erfolgen können.

Es war vorgesehen, folgende Faktoren in die Analysen einzubeziehen:

- Charakteristika der Tumorerkrankung (z. B. nicht-metastasierend vs. metastasierend)
- Therapeutische Intervention

Anhaltspunkte für weitere mögliche Faktoren, die einen Einfluss auf die Validierung haben können, sollten ebenfalls dazu führen, dass diese Faktoren begründet einzubeziehen.

Bei Identifizierung von Faktoren, die einen Einfluss auf die Validierung haben können, sollte gegebenenfalls eine Präzisierung der aus dem Gesamtergebnis abgeleiteten Aussagen erfolgen.

4.3 Änderungen der Methodik im Vergleich zur Projektskizze

Im Vergleich zu der in der Projektskizze beschriebenen Methodik wurden folgende Änderungen vorgenommen:

- Für Teilziel 1 wurden neben den 3 systematischen Übersichten Weir 2006, Lassere 2008 und Mangiapane 2009 auch das Standardwerk Burzykowski 2005 sowie die Übersicht Molenberghs 2010 berücksichtigt.

5 Ergebnisse

5.1 Teilziel 1: Methodische Verfahren zur Validierung von Surrogatendpunkten

5.1.1 Ergebnisse der Informationsbeschaffung

5.1.1.1 Ergebnis der bibliografischen Literaturrecherche

Abbildung 1 zeigt das Ergebnis der systematischen Literaturrecherche nach Publikationen über methodische Verfahren zur Validierung von Surrogatendpunkten sowie das Literaturscreening gemäß Ein- / Ausschlusskriterien.

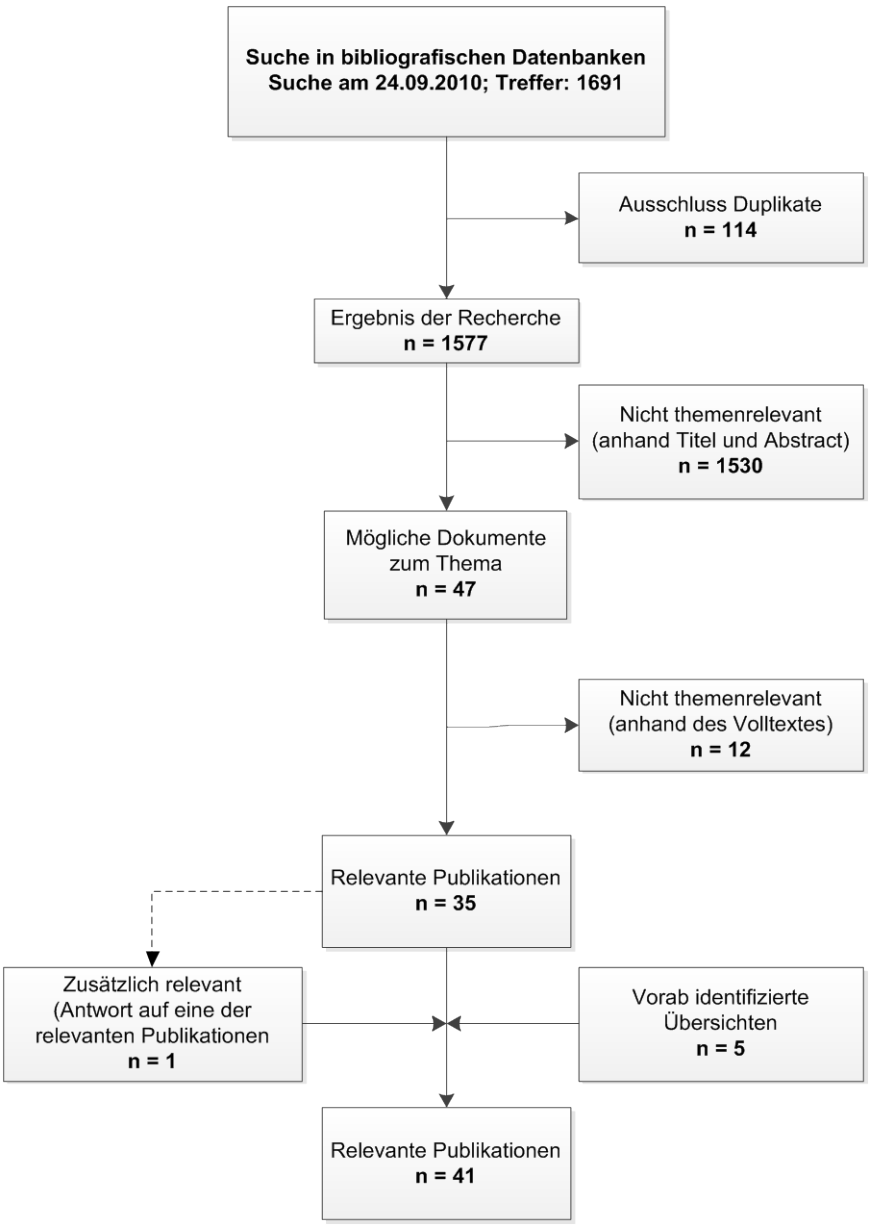

Abbildung 1: Literaturrecherche und Literaturscreening für Teilziel 1 (Methodische Verfahren)

Zunächst wurden Duplikate aus dem Rechercheergebnis entfernt. Daraus ergaben sich insgesamt 1577 zu bewertende Treffer. Von diesen wurden 1530 als sicher nicht themenrelevant aussortiert. Unter den verbleibenden 47 Dokumenten befanden sich 12 Publikationen, die aufgrund des Nichterfüllens von Einschlusskriterien für die Bewertung nicht relevant waren (siehe Anhang B.1). Damit verblieben 35 relevante Publikationen. Zusätzlich wurde die Publikation Alonso 2007 [10] als Antwort auf eine dieser 35 Publikationen (Qu 2007 [11]) identifiziert. Damit ergaben sich insgesamt 36 durch die Literaturrecherche identifizierte Publikationen.

Zusätzlich zu den 36 Publikationen wurden, wie in Abschnitt 4.1.2.2 beschrieben, die bereits vorab identifizierten systematischen Übersichten Weir 2006, Lassere 2008 und Mangiapane 2009 sowie das Standardwerk Burzykowski 2005 und die Übersicht Molenberghs 2010 berücksichtigt. Insgesamt ergaben sich damit 41 relevante Publikationen.

5.1.1.2 Resultierender Studienpool

Tabelle 4 zeigt den Pool relevanter Publikationen über methodische Verfahren zur Surrogatvalidierung. Insgesamt ergaben sich 41 relevante Publikationen, davon 5 Übersichtsarbeiten.

Tabelle 4: Studienpool für Teilziel 1 (Methodische Verfahren)

Arbeit zu methodischen Verfahren	**Zugehörige Publikationen**
Vorab identifizierte Übersichtsarbeiten	
Burzykowski 2005	[7]
Lassere 2008	[5]
Mangiapane 2009	[6]
Molenberghs 2010	[8]
Weir 2006	[4]
Originalarbeiten, Übersichten und Kommentare	
Alonso 2007	[12]
Alonso 2008a	[13]
Alonso 2008b	[14]
Baker 2008	[15]
Burzykowski 2008a	[16]
Buyse 2009	[17]
Buyse 2010	[18]
Cortiñas Abrahantes 2008	[19]

(Fortsetzung)

Tabelle 4: Studienpool für Teilziel 1 (Methodische Verfahren) (Fortsetzung)

Arbeit zu methodischen Verfahren	**Zugehörige Publikationen**
Originalarbeiten und Kommentare	
Deslandes 2007	[20]
Ditlevsen 2007	[21]
Duffy 2009	[22]
Dunning 2008	[23]
Fleischer 2009	[24]
Freedman 2007	[25]
Ghosh 2009	[26]
Ghosh 2010	[27]
Gilbert 2008	[28]
Gilbert 2009	[29]
Green 2008	[30]
Huang 2010	[31]
Joffe 2009	[32]
Lassere 2007a	[33]
Lassere 2007b	[34]
Li 2010	[35]
MacKinnon 2007	[36]
Piedbois 2008	[37]
Pryseley 2007	[38]
Qu 2007 dazu: Alonso 2007	[10,11]
Sarkar 2007	[39]
Sertdemir 2009	[40]
Shi 2009	[41]
Tilahun 2009	[42]
Tilahun 2008	[43]
van Walraven 2009	[44]
Wolfson 2010	[45]

5.1.2 Beschreibung der methodischen Verfahren zur Validierung von Surrogatendpunkten

5.1.2.1 Zusammenfassung der vorhandenen Übersichtsarbeiten

Die Arbeiten Weir 2006, Lassere 2008 sowie Mangiapane 2009 stellen systematische Übersichten auf der Basis einer systematischen Literaturrecherche in Literaturdatenbanken dar. Das Buch Burzykowski 2005 beschrieb die wesentlichen methodischen Entwicklungen auf dem Gebiet der Surrogatvalidierung bis zum Jahr 2005. Darauf aufbauend fasste die Übersichtsarbeit Molenberghs 2010 neuere Ansätze zusammen, die nach 2005 entstanden sind.

(a) Weir 2006

Die Arbeit Weir 2006 gab eine zusammenfassende Übersicht über statistisch-methodische Verfahren von Surrogatendpunkten. Als Surrogatendpunkte wurden hier ausschließlich sogenannte Biomarker betrachtet. Dabei war ein Biomarker gemäß der Biomarkers Definitions Working Group (2001) definiert als eine objektiv messbare Größe entweder eines biologischen oder pathogenen Prozesses oder einer pharmakologischen Reaktion auf eine Behandlung [46]. Ein Endpunkt war dann ein Surrogat, wenn beabsichtigt war, diesen statt eines eigentlich interessierenden klinischen Endpunktes zu betrachten, und erwartet wurde, dass der klinische Nutzen darüber vorhersagbar war. Die Autoren strebten eine ausführliche Darstellung der bis zum Zeitpunkt der Suche existierenden Methoden an sowie deren Kategorisierung aufgrund ihres methodischen Ansatzes.

Die Suche nach relevanten Arbeiten erfolgte systematisch in den Datenbanken ISI Web of Science (1981 bis 2003) sowie in MEDLINE (1966 bis 2003). Es wurde außerdem untersucht, welche Publikationen eine der relevanten Arbeiten zitieren. Des Weiteren wurden Internetseiten von Zulassungsbehörden auf relevante Vorgaben und Verfahren hin durchsucht.

Der erste methodische Bereich umfasste die von Prentice (1989) [47] vorgestellten Kriterien, die über Signifikanztests formuliert wurden, und darauf aufbauende Arbeiten und Kommentare. Trotz ihres bedeutsamen Einflusses auf die weitere Entwicklung von Methoden zur Validierung von Surrogaten hatten die Prentice-Kriterien nach Ansicht der Autoren ungünstige Eigenschaften. Sie seien zum einen schwierig zu erfüllen, auf der anderen Seite gebe es perfekte Surrogate, die die Prentice-Kriterien nicht erfüllten. Aus diesen Gründen seien alternative Verfahren zur Validierung von Surrogaten entwickelt worden.

Der zweite Bereich von Methoden zur Validierung von Surrogaten umfasste Verfahren, die versuchen, das Ausmaß der Validität des Surrogats zu schätzen. Hierfür seien eine Reihe von Korrelationsmaßen entwickelt worden, insbesondere die Maße „Proportion of Treatment Effect Explained“ (PTE), „Relative Effect“ (RE), „Adjusted Association“ (AA), R^2_{trial}, R^2_{indiv} sowie der „Likelihood Reduction Factor“ (LRF). Nach Darstellung der Autoren versuchen diese Maße die Validität des Surrogats getrennt nach individueller Ebene (AA, R^2_{indiv}, LRF) und Studienebene (RE, R^2_{trial}) oder kombiniert (PTE) darzustellen. Für einen Teil dieser Maße

sei die Durchführung einer Meta-Analyse mehrerer Studien notwendig oder zumindest wünschenswert. Je nach Datensituation (insbesondere in Abhängigkeit davon, ob es sich bei den betrachteten Surrogaten und Endpunkten um binäre, stetige, ordinale Daten oder Überlebenszeiten handelt) gebe es eine Vielzahl verschiedener Techniken und Methoden zur Parameterschätzung, die hier nicht dargestellt werden. Die Autoren der Arbeit Weir 2006 schlussfolgerten, dass sich die Verwendung solcher Assoziationsmaße insbesondere in Verbindung mit Meta-Analysen als Verfahren zur Validierung von Surrogaten etabliert hat.

Im Folgenden wurden 2 spezielle Situationen betrachtet, nämlich der Fall multipler Surrogate sowie die Situation, in der die potenzielle Surrogatvariable wiederholt über die Zeit gemessen wird und der klinische Endpunkt eine Überlebenszeitvariable darstellt. Die zweite Situation war nach Ansicht der Autoren insbesondere in der AIDS-Forschung von Bedeutung.

Den letzten methodischen Bereich bildeten Verfahren, die auf Kausalmodellen basieren. Einen grundlegenden Ansatz stelle hierbei das Prinzip der „Principal Stratification“ nach Frangakis & Rubin (2002) dar [48]. Ein grundlegender Nachteil dieser Verfahren sei, dass ein Teil der benötigten Information nicht beobachtbar ist. Aus diesem Grund seien zur Anwendung dieser Verfahren starke Annahmen notwendig. Zumindest bis zum Jahr 2005 habe es keine Anwendung dieser Verfahren in der Praxis gegeben.

Die Autoren der Arbeit Weir 2006 machten noch die folgenden grundlegenden Aussagen. Die Validierung eines Surrogats sei spezifisch für die jeweilige Behandlung oder bestenfalls für eine Klasse von Behandlungen mit gleichem Wirkmechanismus, d. h., man könne nicht aus der Validierung eines Surrogats für eine bestimmte Behandlung schließen, dass die Surrogatvariable auch für eine andere Behandlung valide ist. Die Bestimmung von Korrelationsmaßen auf Studien- und individueller Ebene mithilfe adäquater Meta-Analysen stelle die am meisten akzeptierte und angewendete Methode zur Validierung von Surrogaten dar. Allerdings gebe es weder ein universell anzuwendendes Maß noch eine allgemein beste Schätzmethode noch eine allgemein akzeptierte Grenze, deren Überschreitung den Nachweis der Validität bedeuten würde. Die Wahl der adäquaten Maße, der Schätzmethoden und der Grenzen für den Validitätsnachweis sei immer kontextabhängig.

(b) Lassere 2008

Die Publikation Lassere 2008 hatte die Zielsetzung, ein mehrstufiges Schema zur Surrogatbeurteilung als Handlungsanleitung vorzuschlagen, wobei außer methodischen Aspekten auch inhaltliche wie z. B. biologische Plausibilität eingingen. Dabei wurden wie in Weir 2006 die Definitionen der Biomarkers Definitions Working Group (2001) [46] für Surrogatendpunkte und Biomarker verwendet. Ein Teil dieser Arbeit stellte eine ausführliche Literaturübersicht über existierende statistische Verfahren dar, die im Folgenden kurz zusammengefasst wird.

Die Suche nach relevanten Publikationen erfolgte systematisch in MEDLINE (1950 bis 2007). Ergänzend gab es Recherchen über die Literaturverzeichnisse der gefundenen Arbeiten sowie in Google Scholar (Mai 2005). Als relevant wurden Arbeiten u. a. dann bewertet, wenn

es sich um methodische Entwicklungen oder Bewertungen der Validität von Methoden handelte.

Wie in Weir 2006 wurde nach kurzer Zusammenfassung der Kriterien von Prentice (1989) [47] ausführlicher auf die korrelationsbasierten Verfahren und die verschiedenen Maße auf Studien- und individueller Ebene eingegangen. Einzige aktuelle Ergänzung im Vergleich zu Weir 2006 war die Beschreibung des „Surrogate Threshold Effect“ (STE) von Burzykowski & Buyse (2005, 2006) [49,50]. Es wurde erklärt, dass der STE den minimalen Behandlungseffekt darstellt, den man bezüglich des Surrogatendpunkts beobachten muss, damit mit hoher Sicherheit auch ein Effekt bezüglich des klinischen Endpunkts vorhanden ist. Die Schätzung des STE lasse sich mithilfe meta-regressionsanalytischer Techniken durchführen. Andere Verfahren wie die spezielle Situation multipler Surrogate oder die kausalmodellbasierten Verfahren wurden nur kurz ohne weitere Erklärungen erwähnt.

(c) Mangiapane 2009

Der HTA-Bericht Mangiapane 2009 behandelte u. a. folgende Forschungsfragen:

1. Was sind notwendige Kriterien für einen validen Surrogatendpunkt?

2. Welche Methoden zur Validierung von Surrogaten gibt es?

Zur Beantwortung dieser Fragen wurde eine systematische Literaturrecherche in einer Reihe von Literaturdatenbanken (u. a. MEDLINE, EMBASE, DARE) bezüglich des Zeitraums 1990 bis 2007 durchgeführt und eine Übersicht über die relevanten Publikationen erstellt, die im Folgenden zusammengefasst wird.

Nach einer Beschreibung konzeptioneller Aspekte bezüglich der Anforderungen an Surrogatendpunkte wurden – ähnlich wie in Weir 2006 – die statistischen Methoden zur Validierung von Surrogaten beschrieben. Wie in Lassere 2008 wurde zusätzlich das Konzept des „Surrogate Threshold Effect“ (STE) von Burzykowski (2005, 2006) dargestellt.

Als zusätzliche Methoden im Vergleich zu Weir 2006 und Lassere 2008 wurden hier der „Variance Reduction Factor“ (VRF) von Alonso et al. (2003, 2006) ([51], [52]) und der „Average Prediction Error of the Predicted Effect“ (APEP) von Baker 2006 vorgestellt. VRF wurde hierbei als Verallgemeinerung des Maßes R^2_{indiv} für den Fall von Messwertwiederholungen beschrieben. APEP stelle die durchschnittliche absolute Differenz zwischen dem prognostizierten Interventionseffekt bezüglich des klinischen Endpunkts auf der Basis des Surrogats und dem beobachteten Interventionseffekt bezüglich des klinischen Endpunkts dar. Zur Berechnung des APEP sei wieder eine Meta-Analyse erforderlich. Die Validierung des Surrogats könne durch einen Vergleich des APEP mit der „Average Clinically Meaningful Difference“ (ACMD) und dem „Average Prediction Error of Other True Endpoints“ (APEO) erfolgen. Baker 2006 schlage hierzu vor, dass ein Surrogat validiert ist, wenn (1) die obere Grenze des 95%-Konfidenzintervalls von APEP kleiner als ACMD und (2) APEP kleiner als APEO ist.

Die Übersicht Mangiapane 2009 über methodische Arbeiten schloss mit der Feststellung, dass es „derzeit keinen Goldstandard für die Validierung von Surrogatendpunkten gibt". Ansonsten wurde wegen der Fehleranfälligkeit von Einzelstudienergebnissen eine Surrogatvalidierung auf der Basis von Meta-Analysen präferiert. Außerdem wurde betont, dass die Validität eines Surrogats sowohl krankheits- als auch technologiespezifisch ist, d. h., eine Übertragbarkeit sei sogar innerhalb derselben Wirkstoffklasse fraglich. Eine Surrogatvalidierung im engeren Sinn könne daher nur indikations- und behandlungs- bzw. technologiespezifisch erfolgen.

(d) Burzykowski 2005

Das Buch Burzykowski 2005 lieferte die meisten Details zu Methoden der Surrogatvalidierung. Es gab einen umfassenden Überblick über die von der entsprechenden Arbeitsgruppe entwickelten Validierungsverfahren, in denen Korrelationen der Behandlungseffekte auf Surrogate und wahre (klinische) Endpunkte untersucht werden. Die Definition von Surrogatendpunkten erfolgte mit Verweis auf die nach Ellenberg und Hamilton (1989) [53].

Auch in Burzykowski 2005 wurden zunächst die Prentice-Kriterien [47] erläutert und anschließend die korrelationsbasierten Verfahren mit den bereits erwähnten Maßen PTE (hier PE genannt), RE, AA, R^2_{indiv}, R^2_{trial}, VRF und LRF vorgestellt. Für die statistische Validierung eines Surrogats wurde verlangt, dass die damit ermittelten Korrelationsmaße auf Studienebene und auf individueller Patientenebene hoch sind. Klare Regeln und Grenzwerte wurden nicht genannt, jedoch wurden R^2-Maße von ungefähr 0,9 oder höher als ausreichend valide und solche unter 0,6 als unzureichend valide bezeichnet.

Da die statistischen Modelle, die diesen Korrelationsmaßen zugrunde liegen, zum Teil kompliziert sind und es möglich ist, dass die bei der Berechnung verwendeten Schätzverfahren für gegebene Datensituationen nicht konvergieren, wurden zahlreiche Modellvarianten vorgestellt und mithilfe von Simulationsverfahren untersucht und diskutiert. Da verschiedene Messniveaus von Surrogaten und klinischen Endpunkten eine Anpassung der statistischen Methodik erfordern, wurden Lösungsmöglichkeiten neben der Standardsituation normalverteilter Daten auch für binäre, ordinale, longitudinale Endpunkte sowie Überlebenszeiten präsentiert.

Zusätzlich zu den bisher erwähnten Maßen wurde auf individueller Ebene für den Fall von Messwertwiederholungen des Surrogats und des klinischen Endpunkts das Maß R^2_Λ vorgeschlagen, das auf Wilks' Lambda-Statistik basiere und nicht so stark von der Normalverteilungsannahme abhänge wie der VRF ([54]). Im einfachen Fall ohne Messwertwiederholungen sowie Normalverteilung gelte: $VRF_{indiv} = R^2_\Lambda = R^2_{indiv} = LRF$.

Nach einer Beschreibung Bayesianischer Methoden zur Schätzung der Korrelationsmaße auf Studienebene und auf individueller Patientenebene wurde das STE-Konzept von Burzykowski & Buyse (2005, 2006) [49,50] vorgestellt. Der STE stelle den minimalen Behandlungseffekt dar, den man bezüglich des Surrogatendpunkts beobachten muss, damit mit hoher Sicherheit auch ein Effekt bezüglich des klinischen Endpunkts vorhanden ist. Als

Entscheidungskriterium könne der Vergleich der unteren Konfidenzgrenze des Behandlungseffekts bezüglich des Surrogats mit dem STE dienen [49,50].

Das Buch Burzykowski 2005 schloss mit einer Diskussion, in der betont wurde, dass die Validierung eines Surrogats für jede Intervention extra durchgeführt werden muss. Allerdings wurde die Möglichkeit der Übertragung innerhalb einer Klasse biologisch vergleichbarer Interventionen gesehen. Eine nützliche Rolle wurde Surrogatendpunkten hauptsächlich in Phase-2-Studien zugeschrieben.

(e) Molenberghs 2010

Die aktuelle Übersicht Molenberghs 2010 verwendete ebenfalls die Definitionen der Biomarkers Definitions Working Group (2001) für Surrogatendpunkte [46]. Es wurden kurz die Prentice-Kriterien [47] dargestellt und anschließend ausführlicher die korrelationsbasierten Verfahren mit den Maßen PTE (hier PE genannt), RE, AA, R^2_{indiv} und R^2_{trial} in verschiedenen Datensituationen vorgestellt. Ähnlich wie in Burzykowski 2005 wurden die neueren Maße VRF, R^2_{Λ} und LRF beschrieben [51,52,55].

Zusätzlich zu den bisher angeführten Verfahren wurde in Molenberghs 2010 ein weiteres Maß R^2_h vorgestellt. Dieses Maß gehe auf eine neuere Arbeit (Alonso 2007) zurück und stelle durch seine informationstheoretische Grundlage einen einheitlichen Rahmen dar, die Surrogatvalidität auf individueller Ebene in ganz verschiedenen Datensituationen zu beschreiben. Es wurde erwähnt, dass unter bestimmten Voraussetzungen $R^2_h = R^2_{indiv}$ gilt. Die Anwendung des informationstheoretischen Ansatzes auf Studienebene sei auch möglich und liefere das Maß R^2_{ht}, das ebenfalls unter bestimmten Voraussetzungen identisch sei mit R^2_{trial}.

Anschließend wurde noch das STE-Konzept von Burzykowski & Buyse (2005, 2006) [49,50] vorgestellt. Andere Methoden wie die kausalmodellbasierten Verfahren (z. B. [48]) wurden nur kurz ohne weitere Erklärungen erwähnt.

Der neue informationstheoretische Ansatz wurde als vielversprechende Methode für die Validierung von Surrogaten in der Praxis angesehen. Es wurde darauf hingewiesen, dass gewöhnlich die Validität eines Surrogates als übertragbar auf andere Medikamente derselben Wirkstoffklasse angesehen wird. Hierbei wurde jedoch betont, dass diese Einschätzung subjektiv ist und nur unter der Einbeziehung klinischer Expertise getroffen werden kann.

5.1.2.2 Klassifizierung von Methoden zur Surrogatvalidierung

Es wird davon ausgegangen, dass der methodische Wissensstand zu Validierungsverfahren von Surrogatendpunkten bis zum Jahr 2007 durch die 4 Übersichtsarbeiten (Weir 2006; Lassere 2008; Mangiapane 2009; Molenberghs 2010) und das Buch Burzykowski 2005 ausreichend wiedergegeben ist.

Die Gliederung bzw. Klassifizierung der methodischen Ansätze war in keiner der Arbeiten systematisch. Beispielsweise gilt sowohl für Weir 2006 als auch für Mangiapane 2009, dass Bereiche nicht nur wegen grundsätzlicher methodischer Verfahrenskonzepte, sondern vor

allem wegen bestimmter Dateneigenschaften voneinander abgegrenzt wurden. Hierbei ist z. B. problematisch, dass Methoden zur Surrogatvalidierung mithilfe von Meta-Analysen eine eigene Klassifikationskategorie darstellen, obwohl Meta-Analysen auch in anderen methodischen Kategorien angewendet werden.

Zur Klassifizierung der in den Übersichten dargestellten methodischen Verfahren zur Validierung von Surrogatvariablen wird daher die nachfolgende grobe Einteilung vorgenommen. Diese Einteilung wird auch für die Zuordnung der durch die Literaturrecherche identifizierten Einzelarbeiten herangezogen.

Kategorie 1: Testbasierte Verfahren

Der Bereich der testbasierten Verfahren geht auf Prentice (1989) [47] zurück, der forderte, dass der Signifikanztest auf Therapieeffekt bezüglich des Surrogatendpunkts auch ein valider Signifikanztest bezüglich des klinischen Endpunkts sein muss.

Kategorie 2: Korrelationsbasierte Verfahren

Im Unterschied zu den testbasierten Verfahren wird hier versucht, das Ausmaß des Zusammenhangs zwischen dem Surrogatendpunkt und dem klinischen Endpunkt mithilfe von Korrelationsmaßen abzuschätzen und bei ausreichend hohem Zusammenhang von einem validierten Surrogatendpunkt zu sprechen. Hierbei gibt es – abhängig von den statistischen Eigenschaften des betrachteten Surrogats und des klinischen Endpunkts wie z. B. dem Messniveau – eine Vielzahl von Maßen und Techniken (zum Beispiel R^2_{trial}, R^2_{indiv}). Dieser Zusammenhang ist sowohl auf Einzelstudienebene wie auch für mehrere Studien über Meta-Analysen zu untersuchen. Die Größe des STE, die den minimalen Behandlungseffekt bezüglich des Surrogatendpunkts bezeichnet, aus der auf einen signifikanten Effekt bezüglich des klinischen Endpunktes gefolgert werden kann, kann dabei als Entscheidungskriterium genutzt werden.

Kategorie 3: Kausalmodellbasierte Verfahren

Im dritten Bereich werden Verfahren eingeschlossen, die auf Kausalmodellen basieren. Zu dieser Gruppe gehören z. B. das Kausalmodell von Robins & Greenland (1992) [56] sowie der Ansatz der „Principal Stratification“ von Frangakis & Rubin (2002) [48]. Diese Verfahren benötigen sehr strenge Annahmen, da nicht alle benötigten Größen beobachtbar sind, und werden bislang in der Praxis zur Validierung von Surrogatendpunkten kaum eingesetzt, vgl. Weir 2006.

Kategorie 4: Sonstige Verfahren

Verfahren, die keinem der ersten drei Bereiche zugeordnet werden können, werden in einem vierten Bereich zusammengefasst. Hierzu zählen z. B. spezielle Verfahren bei Verwendung multipler Endpunkte wie das von [57].

5.1.2.3 Beschreibung der Arbeiten aus der Literaturrecherche

Im Folgenden wird die durch die systematische Recherche identifizierte Literatur beschrieben. Übersichtsarbeiten, die nicht vorab identifiziert und getrennt dargestellt sind, werden hierbei wie Einzelarbeiten behandelt. Die Arbeiten werden dabei nach ihrem methodischen Schwerpunkt in die oben genannten Kategorien eingeordnet. Falls eine Publikation mehrere methodische Schwerpunkte gleichrangig behandelt, die zu mehr als einer der Kategorien gehören, wird diese mehrfach dargestellt.

Kategorie 1: Testbasierte Verfahren

Freedman 2007 gab einen kurzen Überblick über die Prentice-Kriterien [47] sowie die testbasierten Verfahren zur Validierung von Surrogaten.

Baker 2008 stellte ein neues testbasiertes Verfahren vor, das den für den Surrogatendpunkt angemessenen kritischen Wert unter Berücksichtigung der Unsicherheit bezüglich des Zusammenhangs zwischen Surrogat und klinischem Endpunkt liefert. Die neue Methode sei nur für den Fall binärer Surrogatparameter und zweiarmige Studien gültig.

Kategorie 2: Korrelationsbasierte Verfahren

Freedman 2007, Green 2008, Piedbois 2008, Buyse 2009, Shi 2009, Buyse 2010 sowie Duffy 2010 gaben einen Überblick bzw. Kommentare über die schätzbasierten Verfahren zur Validierung von Surrogaten. Es wurde betont, dass die ideale Datengrundlage eine Meta-Analyse auf der Basis individueller Patientendaten (IPD) darstellt (Piedbois 2008, Buyse 2009, Shi 2009). Als unterer Wert für die Korrelationskoeffizienten zwischen Surrogat und Endpunkt sowie zwischen den Effekten bezüglich Surrogat und Endpunkt, ab dem ein Surrogat akzeptiert werden kann, wurde 0,7 genannt (Piedbois 2008). Shi 2009 sowie Buyse 2010 betonten, dass statistische Kriterien eine wichtige Rolle spielen, aber in Kombination mit klinischen und biologischen Betrachtungen eingesetzt werden müssen.

Ditlevsen 2007 kommentierte die Übersicht Weir 2006 und wies darauf hin, dass es einen weiteren Artikel [58] gibt, der Methoden zur Schätzung des PTE beschreibt.

Deslandes 2007 stellte auf der Basis eines Multi-State-Modells sowie eines gemeinsamen longitudinalen Modells neue Verfahren zur Schätzung der PTE vor, wenn longitudinale Daten für das Surrogat vorliegen und der klinische Endpunkt eine Überlebenszeitvariable darstellt. McKinnon 2007 beschrieb im Rahmen der logistischen sowie der Probit-Regression Limitierungen des Maßes PTE und gab Empfehlungen zur Punkt- und Intervallschätzung. Für die gleichen Regressionsmodelle schlug Huang 2010 ein zur PTE alternatives Maß F mit verbesserten statistischen Eigenschaften vor, um den Anteil des Behandlungseffekts zu beschreiben, der durch das Surrogat erklärt wird. Im Fall normalverteilter Daten gelte hierbei F = PTE. Sarkar 2007 verwendete als alternatives Maß das Excess Relative Odds (ERO) und entwickelte Schätzmethoden, mit denen für Messfehler adjustiert werden kann.

Gosh 2009 entwickelte für den Fall von Überlebenszeiten neue Schätzverfahren für die Maße RE und AA. Dieser Ansatz erfordere Daten aus nur einer Studie und solle erweitert werden, um in Meta-Analysen anwendbar zu sein.

Sertdemir 2009 untersuchte den Einfluss der Signifikanz des Tests auf Behandlungseffekt auf das Maß R^2_{trial}. In einer Fallstudie zeigten sich unterschiedliche Werte für R^2_{trial} zwischen Studien mit signifikantem und nicht signifikantem Effekt, sodass dies in einem Validierungsprozess berücksichtigt werden sollte.

Tilahun 2009 erweiterte die Methoden zur Schätzung der Maße VRF [51] und R^2_Λ [52] auf den Fall gemischter einfacher und wiederholter Messungen von Surrogat und Endpunkt.

Qu 2007 kritisierte die Eigenschaften von PTE und LRF und schlug das alternative Maß „Proportion of Information Gain“ (PIG) vor. Alonso 2007 wies die Kritik zurück und erläuterte, dass die Arbeit Qu 2007 eher wünschenswerte Eigenschaften des LRF aufzeigt.

Baker 2008 beschrieb eine Erweiterung der Methode zur Schätzung des APEP nach Baker 2006, die nach eigenen Angaben eine bessere Möglichkeit zur Identifizierung der Kontroll- und Interventionsgruppen der Einzelstudien ermöglicht.

Wie bereits bei der Beschreibung der Übersicht Molenberghs 2010 erwähnt, schlug Alonso 2007 mithilfe der Informationstheorie einen einheitlichen Rahmen vor, die Surrogatvalidität in ganz verschiedenen Datensituationen zu beschreiben. Das grundlegende Maß R^2_h sei auf Studien- und individueller Ebene sowie für normalverteilte, binäre, kategorielle und longitudinale Endpunkte anwendbar. Eine Reihe der bisher vorgestellten Maße seien in gewissen Datensituationen Spezialfälle von R^2_h. Pryseley 2007 diskutierte den informationstheoretischen Ansatz für den Fall gemischter Messniveaus (stetig / diskret) von Surrogat und Endpunkt, Tilahun 2008 für den Fall binärer Daten und Alonso 2008a für den Fall von Überlebenszeiten. Cortiñas Abrahantes 2008 schlug im Rahmen des informationstheoretischen Ansatzes eine Reihe von neuen Prozeduren zur Schätzung der Assoziation zwischen den Effekten bezüglich des Surrogats und des klinischen Endpunkts auf Studienebene vor und verglich diese mit der Methode nach Alonso 2007. Die neuen Prozeduren seien flexibler, aber auch sehr rechenintensiv. Zur Berechnung von Konfidenzintervallen wurden Bootstrap-Techniken eingesetzt. Cortiñas Abrahantes 2008 schlussfolgerte, dass die neuen Verfahren in Kombination mit Kreuzvalidierung eingesetzt werden sollen. Alonso 2008b gab eine kurze Übersicht über die bisherigen Verfahren zur Validierung von Surrogaten, insbesondere der Korrelationsmaße auf Studien- und individueller Ebene inklusive des neuen informationstheoretischen Ansatzes von Alonso 2007. Bei der Kommentierung der Verfahren wurde hervorgehoben, dass es sich um einen Bereich aktiver Forschung handelt und in den folgenden Jahren weitere methodische Verfahren zu erwarten sind. Das Editorial Burzykowski 2008a kommentierte die neuen Validierungsverfahren.

Fleischer 2009 entwickelte ein statistisches Modell für die Abhängigkeit zwischen progressionsfreiem Überleben (PFS) und dem Gesamtüberleben (OS), aus dem sich die Korrelation zwischen PFS und OS auf individueller Ebene ableiten lasse. Dieses Modell kann somit einen Baustein zur Surrogatvalidierung liefern, ist aber alleine nicht ausreichend.

Ähnlich wie in Lassere 2008 schlug Lassere 2007a ein Evaluationsschema zur Einschätzung von Surrogaten vor. Hierbei wurde auch kurz auf statistische Methoden zur Validierung von Surrogaten eingegangen. Es wurde betont, dass es keinen Konsens darüber gibt, welches Maß am besten geeignet ist und welche Schwellenwerte zur Validierung verwendet werden sollten. Lassere 2007b ging etwas ausführlicher auf den Teil der statistischen Validierung von Surrogaten ein. Es wurde das Ergebnis eines Workshops berichtet, in dem die Teilnehmer darin übereinstimmten, dass die Verwendung von Daten einer einzigen Studie unzureichend für eine Surrogatvalidierung ist. Als Methode wurde das STE-Konzept von Burzykowski & Buyse (2005, 2006) [49,50] auf der Basis von Meta-Analysen favorisiert.

Kategorie 3: Kausalmodellbasierte Verfahren

Gosh 2010 schlug ein Strukturgleichungsmodell zur Schätzung des kausalen Effekts des Surrogats auf den klinischen Endpunkt vor. Hierbei wurde die Randomisierung als Instrumentenvariable eingesetzt. Zur Anwendung dieser Verfahren sei es notwendig, die korrekten kausalen Pfade des Krankheitsprozesses zu kennen.

Gilbert 2008 stellte ein dreistufiges Validierungsmodell zur Anwendung in der Impfstoffentwicklung vor, in dem zuerst die Korrelation von Surrogat und klinischem Endpunkt untersucht wird, anschließend die Validierung auf Einzelstudienebene und schließlich die Validierung mithilfe von Meta-Analysen erfolgt. Methodische Grundlage bildet hierbei der Ansatz von Frangakis & Rubin (2002) [48]. Dunning 2008 kommentierte, dass im betrachteten Kontext von Impfstoffen ein weiteres wichtiges Kriterium die Vorhersage der Wirksamkeit der Impfung durch das Surrogat ist, und wies auf weitere spezielle statistische Methoden hin. Gilbert 2009 antwortete, dass ein solches Kriterium in dem Ansatz enthalten ist, und wies auf die Limitierungen der angesprochenen statistischen Methoden hin.

Joffe 2009 verglich aus kausalinferenzstatistischer Sichtweise die wesentlichen bisherigen Validierungsverfahren der Kategorien 1 bis 3 und leitete aus verallgemeinerten Modellannahmen paradigmenspezifische Eigenschaften ab.

Li 2010 entwickelte im Rahmen des Ansatzes von Frangakis & Rubin (2002) [48] Bayesianische Schätzverfahren für den Fall binärer Surrogate und klinischer Endpunkte. Die Autoren wiesen darauf hin, dass die Entwicklung der Methode für Anwendungen in der Onkologie mit Überlebenszeiten noch aussteht.

Wolfson 2010 untersuchte, wie sich verschiedene Annahmen auf die Schätzungen im Ansatz von Frangakis & Rubin (2002) [48] auswirken, und leitete daraus Möglichkeiten der Sensitivitätsanalyse ab. Es wurde betont, dass die Validierung von Surrogaten mit erheblichen statistischen Schwierigkeiten verbunden ist.

Kategorie 4: Sonstige Verfahren

Van Walraven 2009 präsentierte einen Ansatz, in dem mithilfe von Monte-Carlo-Simulationen unter Verwendung von bereits vorliegenden Ergebnissen aus Meta-Analysen aus den Effekten bezüglich des Surrogats die Effekte bezüglich des klinischen Endpunkts vorhergesagt werden können. Die Autoren betonten, dass sie ihre Methode für valide halten. Es muss jedoch klar gesagt werden, dass in diesem Ansatz eine Surrogatvalidierung nicht auf der Basis von Evidenz, sondern auf der Basis von simulierten Daten vorgenommen wird.

5.1.3 Zusammenfassung

Aus den Übersichten sowie den in der Literaturrecherche gefundenen methodischen Arbeiten ab 2007 geht hervor, dass die dominierende Rolle bei der Validierung von Surrogaten den korrelationsbasierten Verfahren zukommt. Die neusten Entwicklungen hierbei stellen das STE-Konzept nach Burzykowski & Buyse (2005, 2006) [49,50] sowie der informations-theoretische Ansatz nach Alonso 2007 dar. Testbasierte Verfahren spielen praktisch keine Rolle mehr. Die kausalmodellbasierten Verfahren finden in der Praxis zurzeit keine Anwendung, könnten aber zukünftig an Bedeutung gewinnen.

5.2 Teilziel 2: Studien zur Validierung von Parametern des Tumoransprechens am Beispiel Darmkrebs und Brustkrebs

5.2.1 Ergebnisse der Informationsbeschaffung

5.2.1.1 Ergebnis der bibliografischen Literaturrecherche

Die Recherchen zu den beiden Krankheitsbildern Darmkrebs und Brustkrebs wurden getrennt voneinander durchgeführt.

Recherche nach Validierungsstudien für das Krankheitsbild Darmkrebs

Abbildung 2 zeigt für das Krankheitsbild Darmkrebs das Ergebnis der systematischen Literaturrecherche nach Studien in den bibliografischen Datenbanken und das Literaturscreening gemäß Ein- / Ausschlusskriterien.

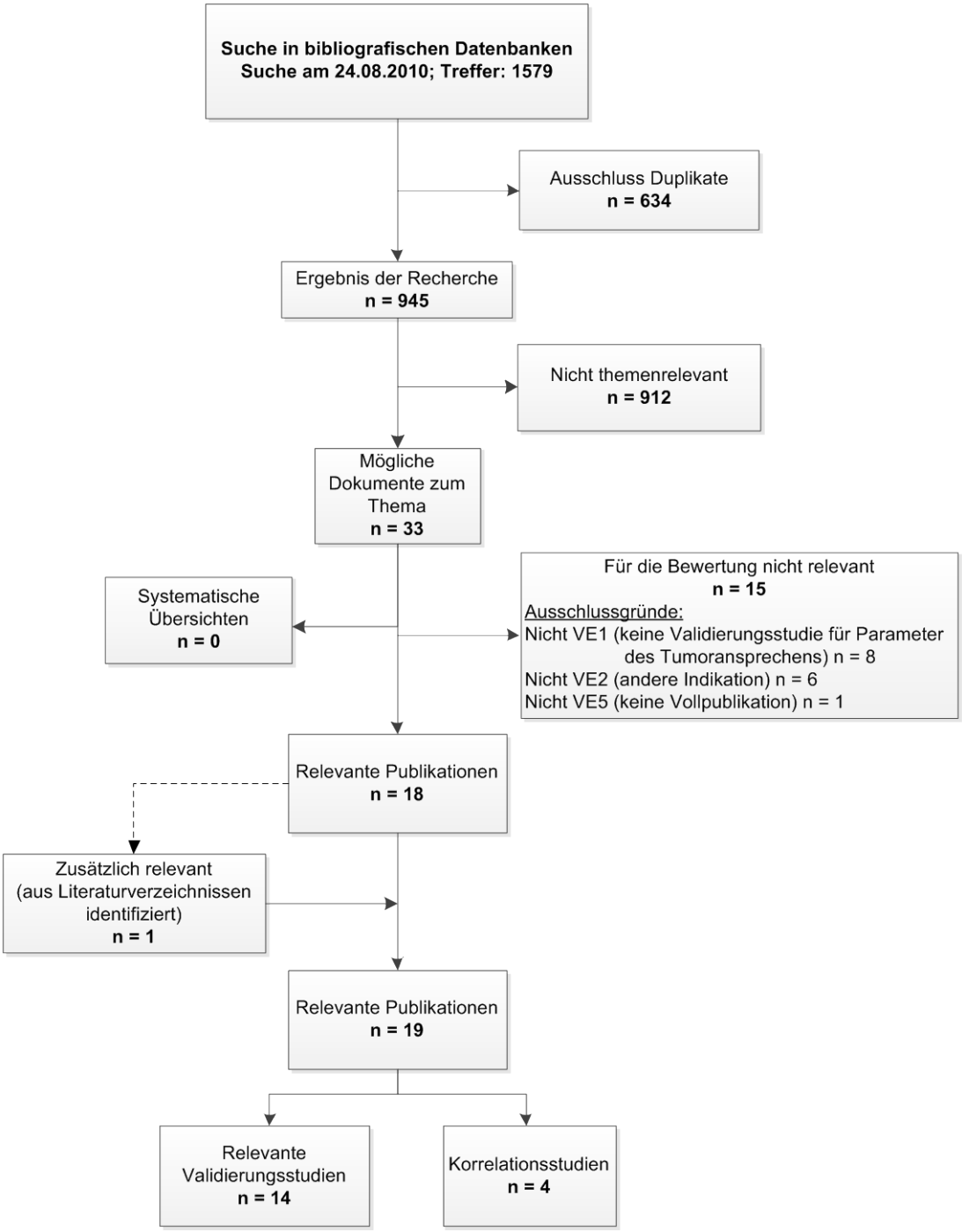

Abbildung 2: Literaturrecherche und Literaturscreening für Teilziel 2 – Krankheitsbild Darmkrebs

Zunächst wurden Duplikate aus dem Rechercheergebnis entfernt. Daraus ergaben sich insgesamt 945 zu bewertende Treffer. Von diesen wurden 912 als sicher nicht themenrelevant aussortiert. Unter den verbleibenden 33 Dokumenten befanden sich 15 Publikationen, die aufgrund des Nichterfüllens von Einschlusskriterien für die Bewertung nicht relevant waren (siehe Anhang B.2). Damit verblieben 18 relevante Publikationen, darunter keine systematische Übersicht zum Thema. Eine weitere relevante Publikation (Burzykowski 2004

[59]) wurde identifiziert, indem in einer der 18 Publikationen (Baker 2006 [60]) auf diese verwiesen wurde. Insgesamt ergaben sich damit 19 relevante Publikationen.

Recherche nach Validierungsstudien für das Krankheitsbild Brustkrebs

Abbildung 3 zeigt für das Krankheitsbild Brustkrebs das Ergebnis der systematischen Literaturrecherche nach Studien in den bibliografischen Datenbanken und das Literaturscreening gemäß Ein- / Ausschlusskriterien

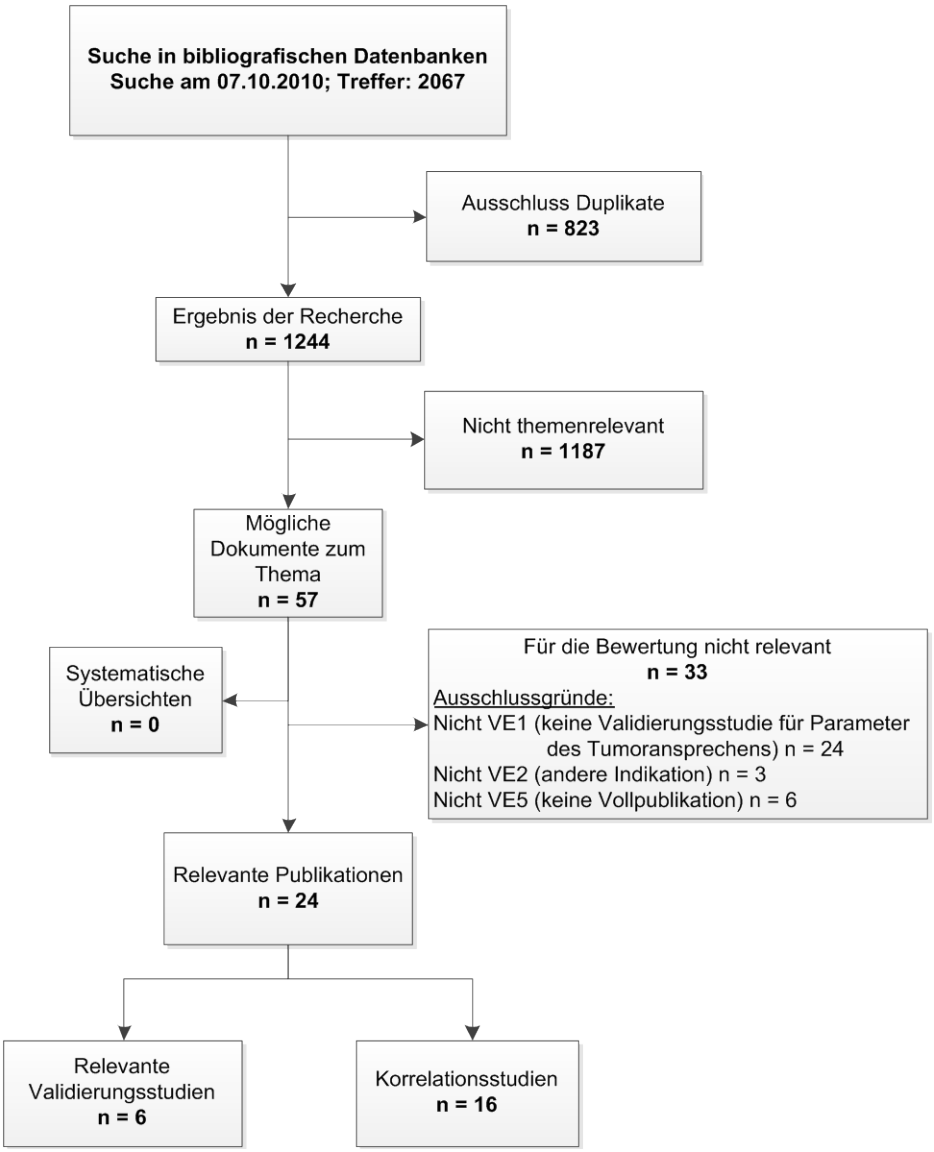

Abbildung 3: Literaturrecherche und Literaturscreening für Teilziel 2 – Krankheitsbild Brustkrebs

Zunächst wurden Duplikate aus dem Rechercheergebnis entfernt. Daraus ergaben sich insgesamt 1244 zu bewertende Treffer. Von diesen wurden 1187 als sicher nicht themenrelevant aussortiert. Unter den verbleibenden 57 Dokumenten befanden sich 33 Publikationen, die aufgrund des Nichterfüllens von Einschlusskriterien für die Bewertung

nicht relevant waren (siehe Anhang B.3). Damit verblieben 24 relevante Publikationen, darunter keine systematische Übersicht zum Thema.

5.2.1.2 Nicht weiter berücksichtigte Korrelationsstudien

Unter den 19 für das Krankheitsbild Darmkrebs und den 24 für das Krankheitsbild Brustkrebs identifizierten Studien waren 4 (Darmkrebs) bzw. 18 (Brustkrebs) Studien, die lediglich einen Zusammenhang zwischen Endpunkten untersucht haben, ohne dass dabei die Behandlungseffekte berücksichtigt wurden. Solche Ansätze sind grundsätzlich ungeeignet, um eine abschließende Aussage über die Validität eines Surrogats treffen zu können. Diese Studien wurden daher nicht weiter berücksichtigt. Eine Liste dieser Studien findet sich in Kapitel 8.

5.2.1.3 Resultierender Studienpool

Darmkrebs

Tabelle 5 zeigt den Pool relevanter Validierungsstudien für das Krankheitsbild Darmkrebs. Insgesamt ergaben sich aus der Recherche 15 relevante Validierungsstudien.

Tabelle 5: Studienpool für Teilziel 2 – Krankheitsbild Darmkrebs

Studie	**Zugehörige Publikationen**
Alonso 2008a	[13]
Baker 2006	[60]
Burzykowski 2004	[59]
Burzykowski 2008b	[61]
Buyse 2000b	[62]
Buyse 2007	[63]
Buyse 2008	[64]
Cortiñas Abrahantes 2008	[19]
Green 2008	[30]
Johnson 2006	[65]
Montagnani 2009	[66]
Sargent 2005	[67]
Sargent 2007	[68]
Sertdemir 2009	[40]
Tang 2007	[69]

Brustkrebs

Tabelle 6 zeigt den Pool relevanter Validierungsstudien für das Krankeitsbild Brustkrebs. Insgesamt ergaben sich aus der Recherche 6 relevante Validierungsstudien.

Tabelle 6: Studienpool für Teilziel 2 – Krankheitsbild Brustkrebs

Studie	Zugehörige Publikationen
Bruzzi 2005	[70]
Burzykowski 2008c	[71]
Hackshaw 2005	[72]
Miksad 2008	[73]
Ng 2007	[74]
Sherrill 2008	[75]

5.2.2 Charakteristika der in die Bewertung eingeschlossenen Studien

5.2.2.1 Indikationen, Interventionen und untersuchte Endpunkte

Darmkrebs

In Tabelle 7 sind Angaben zur Indikation, zu den eingesetzten Interventionen und den untersuchten Surrogatendpunkten sowie patientenrelevanten Endpunkten für die insgesamt 15 Validierungsstudien des Krankheitsbildes Darmkrebs aufgeführt.

Fünf Validierungsstudien schlossen ausschließlich Studien mit Patienten mit fortgeschrittenem Kolorektalkarzinom (inklusive metastasierenden Kolorektalkarzinoms) ein, 3 schlossen Studien mit Patienten mit metastasierendem Kolorektalkarzinom, 4 schlossen Studien mit Patienten mit nicht metastasierendem Kolonkarzinom der Stadien I bis III bzw. Dukes‘ A bis C ein. Drei weitere Validierungsstudien (Alonso 2008a, Buyse 2008, Green 2008) unterteilten die zugrunde liegenden Studien je nach Indikation in Patienten mit Kolonkarzinom im frühen Krankheitsstadium sowie Patienten mit fortgeschrittenem Kolorektalkarzinom (inklusive metastasierenden Kolorektalkarzinoms). Diese 3 Validierungsstudien sind in Tabelle 7 jeweils zweimal aufgeführt.

Die eingesetzten Interventionen in den, den Validierungsstudien zugrunde liegenden Studien waren sehr heterogen. Sie reichten von einer alleinigen OP in den Kontrollarmen bis hin zu komplexen Kombinationschemotherapieregimen wie z. B. 5-FU/LV + Oxaliplatin + Bevacizumab. In einigen Validierungsstudien wurden die angewendeten Interventionen nicht genau beschrieben. Weiterhin war teilweise nicht ersichtlich, wie die Zuordnung einer Intervention zum Kontrollarm bzw. zum Interventionsarm erfolgte. Die unterschiedlichen

Interventionen verfolgten teilweise ein kurativ es Ziel oder sie wurden in palliativen Situationen eingesetzt.

Als Surrogatendpunkte wurden am häufigsten das Tumoransprechen (TR), das krankheitsfreie Überleben (DFS), das progressionsfreie Überleben (PFS) und die Zeit bis zur Progression (TTP) untersucht. In 13 von 15 Validierungsstudien wurde das Gesamtüberleben (OS) als patientenrelevanter Endpunkt untersucht. Zwei Validierungsstudien analysierten das Überleben, ohne diesen Endpunkt näher zu definieren. Insgesamt fehlten in 6 Validierungsstudien Definitionsangaben zu Surrogatendpunkten und / oder patientenrelevanten Endpunkten.

Brustkrebs

In Tabelle 8 sind Angaben zur Indikation, zu den eingesetzten Interventionen und den untersuchten Surrogatendpunkten sowie patientenrelevanten Endpunkten für die insgesamt 6 Validierungsstudien des Krankheitsbildes Brustkrebs aufgeführt.

Drei Validierungsstudien schlossen Studien mit Patientinnen mit fortgeschrittenem Brustkrebs (inklusive metastasierenden Brustkrebses) ein, 2 schlossen Studien mit Patientinnen mit metastasierendem Brustkrebs und 1 schloss Studien mit Patientinnen mit Brustkrebs (ohne lokal fortgeschrittenen oder metastasierenden Brustkrebs) ein.

Die eingesetzten Interventionen in den, den Validierungsstudien zugrunde liegenden Studien waren sehr heterogen. Es wurden Anthrazykline, Taxane, sonstige Chemotherapien, Hormontherapien, Biologika oder sonstige Therapien, jeweils als Mono- oder Kombinationstherapie, sowohl im Kontroll- als auch im Interventionsarm angewendet. In einigen Validierungsstudien wurden die angewendeten Interventionen nicht genau beschrieben. Weiterhin war teilweise nicht ersichtlich, wie die Zuordnung einer Intervention zum Kontrollarm bzw. zum Interventionsarm erfolgte. Die unterschiedlichen Interventionen verfolgten teilweise ein kurativ es Ziel oder sie wurden in palliativen Situationen eingesetzt.

Als Surrogatendpunkte wurden am häufigsten das Tumoransprechen (TR), das progressionsfreie Überleben (PFS) und die Zeit bis zur Progression (TTP) untersucht. In allen 6 Validierungsstudien wurde das Gesamtüberleben (OS) als patientenrelevanter Endpunkt untersucht. Insgesamt fehlten in 5 Validierungsstudien Definitionsangaben zu Surrogatendpunkten und / oder dem patientenrelevanten Endpunkt.

Tabelle 7: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Darmkrebs

Studie	Indikation	Intervention	untersuchte Endpunkte (Name, Definition)	
			Surrogatendpunkte (SEP)	patientenrelevanter Endpunkt (PEP)
Kolonkarzinom im frühen Krankheitsstadium				
Alonso 2008a	Teilpopulation: Kolonkarzinom in frühem Krankheitsstadium (Verweis auf den in Sargent 2005 beschriebenen Datensatz mit 18 Studien. Von diesen wurden aber nur 10 Studien ohne weitere Angaben berücksichtigt)	k. A. (Verweis auf den in Sargent 2005 beschriebenen Datensatz mit 18 Studien. Von diesen wurden aber nur 10 Studien ohne weitere Angaben berücksichtigt)	▪ Zeit bis zum Wiederauftreten des Tumors (TTR); Definition: k. A.	▪ Gesamtüberleben (OS); Definition: k. A.
Buyse 2008	Teilpopulation: Kolonkarzinom in frühem Krankheitsstadium (Verweis auf den in Sargent 2005 beschriebenen Datensatz mit 18 Studien. Von diesen wurden aber nur 10 Studien ohne weitere Angaben berücksichtigt)	k. A. (Verweis auf den in Sargent 2005 beschriebenen Datensatz mit 18 Studien. Von diesen wurden aber nur 10 Studien ohne weitere Angaben berücksichtigt)	▪ Krankheitsfreies Überleben (DFS); Definition: Zeit von der Randomisierung bis zum Wiederauftreten der Erkrankung oder bis zum Tod, unabhängig von der Todesursache	▪ Gesamtüberleben (OS); Definition: Zeit von der Randomisierung bis zum Tod, unabhängig von der Todesursache
Green 2008	Teilpopulation: Kolonkarzinom in frühem Krankheitsstadium[1]	k. A.[1]	▪ Krankheitsfreies Überleben (DFS); Definition: Zeit von der Randomisierung bis zum Wiederauftreten der Erkrankung oder bis zum Tod	▪ Gesamtüberleben (OS); Definition: Zeit von der Randomisierung bis zum Tod, unabhängig von der Todesursache

(Fortsetzung)

Tabelle 7: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Darmkrebs (Fortsetzung)

Studie	Indikation	Intervention	untersuchte Endpunkte (Name, Definition)	
			Surrogatendpunkte (SEP)	patientenrelevanter Endpunkt (PEP)
Nicht metastasierendes Kolonkarzinom				
Burzykowski 2008b	Kolonkarzinom (nicht metastasierend) ▪ Stadium I: 1%, II: 33%, III: 66% der Patienten, 2 Studien beinhalteten auch Patienten mit Rektalkarzinom (Verweis auf Sargent 2005)	Unterschiedliche adjuvante Therapien in Interventions- und Kontrollarmen[2] In mindestens einem Studienarm wurden die Patienten mit einem Therapieregime behandelt, das 5-FU enthielt. (Verweis auf Sargent 2005)	▪ Krankheitsfreies Überleben (DFS)[3]; Definition: Zeit von der Randomisierung bis zum Wiederauftreten der Erkrankung oder bis zum Tod, unabhängig von der Todesursache	▪ Gesamtüberleben (OS)[4]; Definition: Zeit von der Randomisierung bis zum Tod, unabhängig von der Todesursache
Sargent 2005	Kolonkarzinom (nicht metastasierend) ▪ Stadium I: 1%, II: 33%, III: 66% der Patienten, 2 Studien beinhalteten auch Patienten mit Rektalkarzinom	Unterschiedliche adjuvante Therapien in Interventions- und Kontrollarmen[2] In mindestens einem Studienarm wurden die Patienten mit einem Therapieregime behandelt, das 5-FU enthielt.	▪ krankheitsfreies Überleben (DFS)[5]; Definition DFS: Zeit von der Randomisierung bis zum Wiederauftreten der Erkrankung oder bis zum Tod (ohne Sekundärtumore)	▪ Gesamtüberleben (OS)[6]; Definition OS: Zeit von der Randomisierung bis zum Tod, unabhängig von der Todesursache

(Fortsetzung)

Tabelle 7: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Darmkrebs (Fortsetzung)

Studie	Indikation	Intervention	untersuchte Endpunkte (Name, Definition)	
			Surrogatendpunkte (SEP)	patientenrelevanter Endpunkt (PEP)
Nicht metastasierendes Kolonkarzinom				
Sargent 2007	Kolonkarzinom (nicht metastasierend) ▪ Stadium I: 1%, II: 33%, III: 66% der Patienten, 2 Studien beinhalteten auch Patienten mit Rektalkarzinom (Verweis auf Sargent 2005)	Unterschiedliche adjuvante Therapien in Interventions- und Kontrollarmen [2] In mindestens einem Studienarm wurden die Patienten mit einem Therapieregime behandelt, das 5-FU enthielt. (Verweis auf Sargent 2005)	▪ krankheitsfreies Überleben (DFS)[5]; Definition: Zeit von der Randomisierung bis zum Wiederauftreten der Erkrankung oder bis zum Tod (ohne Sekundärtumore)	▪ Gesamtüberleben (OS)[6]; Definition: Zeit von der Randomisierung bis zum Tod, unabhängig von der Todesursache
Sertdemir 2009	Kolonkarzinom (nicht metastasierend) ▪ Stadien Dukes‘ A bis C	C-01: OP vs. OP + MOF vs. OP + BCG C-02: OP vs. OP +5-FU/Heparin (Portalvenenapplikation, 1. Applikation am Tag der OP) C-03: MOF vs. 5-FU/LV C-04: 5-FU + LV vs. 5-FU + LEV vs. 5-FU + LV+LEV	▪ krankheitsfreies Überleben (DFS)[7]; Definition: Wiederkehren der Erkrankung, Auftreten eines zweiten primären Tumors oder Tod	▪ Gesamtüberleben (OS); Definition: Tod, unabhängig von der Todesursache.

(Fortsetzung)

Tabelle 7: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Darmkrebs (Fortsetzung)

Studie	Indikation	Intervention	untersuchte Endpunkte (Name, Definition)	
			Surrogatendpunkte (SEP)	patientenrelevanter Endpunkt (PEP)
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)				
Alonso 2008a	Teilpopulation: Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms) (Verweis auf MAGIC 2004)	5-FU (± INF-α) vs. 5-FU/LV (± INF-α)	▪ Zeit bis zum Wiederauftreten des Tumors (TTR); Definition: k. A.	▪ Gesamtüberleben(OS); Definition: k. A.
Baker 2006	Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)[8] (Verweis auf Buyse 2000b und Burzykowski 2004)	First-Line-Chemotherapieregime: Kontrollarm: 5-FU oder FUDR (jeweils i.v. Bolus-Applikation) Interventionsarm: 5-FU/LV oder 5-FU + MTX oder 5-FU (Dauerinfusion) oder FUDR (hepatisch-arterielle Infusion) (Verweis auf Burzykowski 2004 und Buyse 2000b)	▪ Tumoransprechen (TR) Definition: CR + PR (Responder), SD + PD (Non-Responder), keine weiteren Angaben	▪ Überleben; Definition: k. A.
Burzykowski 2004	Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)[9] (Verweis auf ACCMAP 1992, 1994 und MAGIC 1996, 1998)	First-Line-Chemotherapieregime: Kontrollarm: 5-FU oder FUDR (jeweils i.v. Bolus-Applikation) Interventionsarm: 5-FU/LV oder 5-FU + MTX oder 5-FU (Dauerinfusion) oder FUDR (hepato-arterielle Infusion) (Verweis auf ACCMAP 1992, 1994 und MAGIC 1996, 1998)	▪ Tumoransprechen (TR): <u>TR in 4 Kategorien</u>; Definition: CR, PR, SD, PD nach WHO 1979 <u>binäres TR</u>: Definition: CR + PR (Responder), SD + PD (Non-Responder)	▪ Gesamtüberleben (OS); Definition: Zeit von der Randomisierung bis zum Tod, unabhängig von der Todesursache

(Fortsetzung)

Tabelle 7: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Darmkrebs (Fortsetzung)

Studie	Indikation	Intervention	untersuchte Endpunkte (Name, Definition)	
			Surrogatendpunkte (SEP)	patientenrelevanter Endpunkt (PEP)
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)				
Buyse 2000b	Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)[9] (Verweis auf ACCMAP 1992, 1994 und MAGIC 1996, 1998)	First-Line-Chemotherapieregime: Kontrollarm: 5-FU oder FUDR (jeweils i.v. Bolus-Applikation) Interventionsarm: 5-FU/LV oder 5-FU + MTX oder 5-FU (Dauerinfusion) oder FUDR (hepatisch-arterielle Infusion)	▪ Tumoransprechen (TR); Definition: CR + PR (Responder), SD + PD (Non-Responder) CR: Tumor nicht mehr nachweisbar PR: Abnahme der Tumoroberfläche um ≥ 50% ohne Auftreten neuer Herde	▪ Gesamtüberleben (OS); Definition: Zeit von der Randomisierung bis zum Tod, unabhängig von der Todesursache
Buyse 2007	Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)[10]	Kontrollarm: 5-FU/LV Interventionsarm: 5-FU oder Raltitrexed	▪ Progressionsfreis Überleben (PFS); Definition: Zeit von der Randomisierung bis zur Progression der Erkrankung oder bis zum Tod, unabhängig von der Todesursache	▪ Gesamtüberleben (OS); Definition: Zeit von der Randomisierung bis zum Tod, unabhängig von der Todesursache
Buyse 2008	Teilpopulation: Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)[11] (Verweis auf Buyse 2007 und Burzykowski 2008b[12])	Kontrollarm: 5-FU/LV Interventionsarm: 5-FU oder Raltitrexed (Verweis auf Buyse 2007 und Burzykowski 2008b[12])	▪ Progressionsfreies Überleben (PFS); Definition: Zeit von der Randomisierung bis zur Progression der Erkrankung oder bis zum Tod, unabhängig von der Todesursache	▪ Gesamtüberleben (OS); Definition: Zeit von der Randomisierung bis zum Tod, unabhängig von der Todesursache

(Fortsetzung)

Tabelle 7: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Darmkrebs (Fortsetzung)

Studie	Indikation	Intervention	untersuchte Endpunkte (Name, Definition)	
			Surrogatendpunkte (SEP)	patientenrelevanter Endpunkt (PEP)
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)				
Cortiñas Abrahantes 2008	Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)[10] (Verweis auf Burzykowski 2004 und 2005)	k. A. (Daten beruhen laut Autor auf dem MAGIC-Datensatz, zitiert werden aber 2 in diesem nicht enthaltene RCTs. Der genannte Interventionsvergleich 5-FU vs. 5-FU + INF-α findet sich in den RCTs, nicht aber in den MAGIC-Daten.)	▪ Tumoransprechen (TR); Definition: komplettes Ansprechen (CR), partielles Ansprechen (PR), stabile Erkrankung (SD), progressive Erkrankung (PD) nach WHO 1979	▪ Überleben; Definition: k. A.
Green 2008	Teilpopulation: Fortgeschrittenes Kolorektalkarzinom (vermutlich inkl. metastasierenden Kolorektalkarzinoms)[13]	k. A.[13]	▪ Progressionsfreies Überleben (PFS); Definition: Zeit von Randomisierung bis zur Progression der Erkrankung oder bis zum Tod	▪ Gesamtüberleben (OS); Definition: Zeit von der Randomisierung bis zum Tod, unabhängig von der Todesursache
Metastasierendes Kolorektalkarzinom				
Johnson 2006	Metastasierendes Kolorektalkarzinom	First-Line-Chemotherapieregime mit u. a. 5-FU, Capecitabin, Irinotecan, Oxaliplatin, Bevacizumab, Tirapazamin, Gefitinib (keine Differenzierung zwischen Kontroll- und Interventionsarmen angegeben)	▪ Tumoransprechen (TR); Definition: komplettes oder partielles Ansprechen ohne weitere Spezifizierung ▪ Zeit bis zur Progression (TTP); Definition: k. A.	▪ Gesamtüberleben; Definition: k. A.

(Fortsetzung)

Tabelle 7: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Darmkrebs (Fortsetzung)

Studie	Indikation	Intervention	untersuchte Endpunkte (Name, Definition)	
			Surrogatendpunkte (SEP)	patientenrelevanter Endpunkt (PEP)
Metastasierendes Kolorektalkarzinom				
Montagnani 2009	Metastasierendes Kolorektalkarzinom	Unterschiedliche First-Line-Therapien (5-FU, 5-FU oder Capecitabin + Irinotecan oder Oxaliplatin), die alle Bevacizumab enthielten im Vergleich zu unterschiedlichen Kontrolltherapien jeweils ohne Bevacizumab	▪ Progressionsfreies Überleben (PFS); Definition: k. A.	▪ Gesamtüberleben (OS); Definition: k. A.
Tang 2007	Metastasierendes Kolorektalkarzinom (ausgenommen lokal fortgeschrittene, nicht resezierbare Erkrankungen)	Unterschiedliche Chemotherapien in Interventions- und Kontrollarmen[14]	▪ Tumoransprechen (TR); Definition: k. A. ▪ Zeit bis zur Progression (TTP); Definition: k. A. ▪ progressionsfreies Überleben (PFS); Definition: k. A.	▪ Gesamtüberleben (OS); Definition: Zeit von der Randomisierung oder der ersten Behandlung bis zum Tod.

Abkürzungen: ACCMAP: Adjuvant Colorectal Cancer Meta-Analysis Project, BCG: Bacillus Calmette-Guérin, CR: Komplettes Ansprechen, 5-FU: Fluorouracil, FUDR: Floxuridin, INF: Interferon, INF-α: Interferon-alpha, k. A.: keine Angabe, LEV: Levamisol, LV: Leucovorin, MAGIC: Meta-Analysis Group in Cancer, MOF: Semustin/Vincristin/Fluorouracil, MTX: Methotrexat, OP: Operation, OS: Overall survival, PD: progressive disease, PFS: Progression-free survival, PR: Partial response, RCT: randomisierte kontrollierte Studie, RR: Response rate, SD: Stable disease, TR: Tumor response, TTP: Time to tumor progression, TTR: Time to tumor recurrence, WHO: World Health Organization.

[1] In Green 2008 werden Sargent 2005 sowie Burzykowski 2001 zitiert, es ist unklar, welche Daten in die Analyse eingeflossen sind.
[2] Gemäß dem ursprünglichen Studienprotokoll wurde ein Studienarm von den Autoren als Kontrollarm definiert. In mindestens einem Studienarm wurden die Patienten mit einem Therapieregime, das 5-FU enthielt, behandelt. In 9 Studien erhielt die Kontrollgruppe nach erfolgter Tumorresektion keine weitere Therapie.
[3] Beobachtungen werden nach 1, 2 bzw. 3 Jahren zensiert.
[4] Beobachtungen werden nach 5 Jahren zensiert.
[5] Beobachtungen werden so zensiert, dass die mediane (bzw. minimale) Nachbeobachtungszeit 1, 2 bzw. 3 Jahre beträgt.
[6] Beobachtungen werden so zensiert, dass die mediane (bzw. minimale) Nachbeobachtungszeit 5 Jahre beträgt.
[7] Entspricht den Informationen aus den Originalarbeiten C01bis C04. Surrogatendpunkt wird in Sertdemir 2009 „recurrence free survival“ genannt und nicht näher definiert.

(Fortsetzung)

Tabelle 7: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Darmkrebs (Fortsetzung)

[8] Die Begriffe „metastatic“ und „advanced“ wurden in der Publikation Buyse 2000 parallel verwendet.
[9] Die Begriffe „metastatic“ und „advanced“ wurden in den Publikationen ACCMAP 1992, 1994 und MAGIC 1996, 1998 parallel verwendet.
[10] Die Begriffe „metastatic“ und „advanced“ wurden parallel in der Publikation verwendet.
[11] Die Begriffe „metastatic“ und „advanced“ wurden parallel in Buyse 2007 verwendet.
[12] Extrahiert wurden die Angaben aus Buyse 2007. Burzykowski 2008b bezieht sich auf die ACCENT-Daten.
[13] In Green 2008 werden Sargent 2005 (ACCENT-Daten, 18 RCTs) sowie Burzykowski 2001 (Auswertung von 2 RCTs mit fortgeschrittenem Kolorektalkarzinom) zitiert. Nähere Angaben zur Auswahl der Studien und Zuordnung zu den Indikationen ist nicht erfolgt.
[14] Zur Validierung wurde eine systematische Suche nach randomisierten kontrollierten Studien durchgeführt. Für jede Studie wurde von den Autoren ein Arm als Kontrollgruppe festgelegt, alle anderen Studienarme galten als Interventionsgruppe. In den unterschiedlichen Kontrollgruppen wurden hauptsächlich 5-FU oder 5-FU/LV appliziert, darüber hinaus gab es jedoch auch Kombinationen aus 5-FU/LV mit Oxaliplatin oder Irinotecan. In den Interventionsgruppen kamen u. a. folgende Therapieregime zum Einsatz: 5-FU, 5-FU in Kombination mit Interferon, 5-FU/LV, 5-FU/LV in Kombination mit Oxaliplatin oder Irinotecan mit und ohne Bevacizumab, Irinotecan in Kombination mit Oxaliplatin, Capecitabin sowie Uracil/Tegafur. In beiden Studienarmen wurde 5-FU sowohl als Bolus als auch als Dauerinfusion appliziert.

Tabelle 8: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Brustkrebs

Studie	Indikation	Intervention	untersuchte Endpunkte (Name, Definition)	
			Surrogatendpunkte (SEP)	patientenrelevanter Endpunkt (PEP)
Nicht metastasierender oder lokal fortgeschrittener Brustkrebs				
Ng 2007	Brustkrebs (ohne lokal fortgeschrittenen oder metastasierenden Brustkrebs)	Adjuvante systemische Behandlungen (z. B. Chemotherapie, hormonelle Therapie und/oder Immuntherapie), unterschiedliche Therapien nicht näher spezifiziert. Definition Kontrollarm: Diejenige Behandlung, die am besten die Standardtherapie der jeweiligen Zeitperiode widerspiegelte	▪ krankheitsfreies Überleben (DFS)[1]; Definition: k. A. [2]	▪ Gesamtüberleben (OS)[3]; Definition: k. A.
Fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses)				
Bruzzi 2005	Fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses)[4]	Interventionsarm: Dosisintensivierte bzw. dosisdichte Epirubicin-basierte Chemotherapieregime (FEC, EC oder E) Kontrollarm: unterschiedliche Epirubicin-basierte Standardchemotherapieregime (FEC, EC, oder E)	▪ objektives Tumoransprechen (TR); Definition: CR + PR (Responder) hauptsächlich gemäß WHO- oder ECOG-Kriterien[5]	▪ Gesamtüberleben (OS); Definition: Zeit von der Randomisierung bis zum Tod, unabhängig von der Todesursache

(Fortsetzung)

Tabelle 8: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Brustkrebs (Fortsetzung)

Studie	Indikation	Intervention	untersuchte Endpunkte (Name, Definition)	
			Surrogatendpunkte (SEP)	patientenrelevanter Endpunkt (PEP)
Fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses)				
Burzykowski 2008c	Fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses)[4]	First-Line-Chemotherapieregime: Kontrollarm: Anthrazyklin (A) vs. Interventionsarm: Taxan (P oder D) oder Kontrollarm: Anthrazyklin-Kombinationstherapie (EC, AC, FAC oder FEC) vs. Interventionsarm: Anthrazyklin-Taxan-Kombinationstherapie (EP, AP, DAC, AD oder ED)	▪ Tumoransprechen (TR); Definition: k.A ▪ Krankheitskontrolle (DC); Definition: k. A. ▪ Progressionsfreies Überleben (PFS); Definition: Zeit von der Randomisierung bis zur Progression der Erkrankung oder bis zum Tod, unabhängig von der Todesursache ▪ Zeit bis zur Progression (TTP) Definition: Zeit von der Randomisierung bis zur Progression der Erkrankung oder bis zum Tod aufgrund der Krebserkrankung	▪ Gesamtüberleben (OS); Definition: Zeit von der Randomisierung bis zum Tod, unabhängig von der Todesursache
Miksad 2008	Fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses)[4, 6]	▪ First-Line-Chemotherapieregime: Anthrazyklin-basierte Kombinationstherapie (FAC oder FEC) ▪ First-/ Second-Line oder nachfolgende Chemotherapieregime: Taxan-basierte Kombinationstherapien[7] Vergleichstherapien jeweils nicht näher definiert.	▪ Progressionsfreies Überleben (PFS); Definition: Zeit von Randomisierung bis zur Progression der Erkrankung oder bis zum Tod, unabhängig von der Todesursache (FDA-Definition)	▪ Gesamtüberleben (OS); Definition: k. A.

(Fortsetzung)

Tabelle 8: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Brustkrebs (Fortsetzung)

Studie	Indikation	Intervention	untersuchte Endpunkte (Name, Definition)	
			Surrogatendpunkte (SEP)	patientenrelevanter Endpunkt (PEP)
Metastasierender Brustkrebs				
Hackshaw 2005	Metastasierender Brustkrebs[4, 8]	First-Line-Chemotherapieregime: Anthrazyklin-Kombinationstherapie (FAC oder FEC, teilweise + VP, VM, MTX, Tamoxifen, Lonidamin oder Progesteron) vs. unterschiedliche Kombinationsregime mit und ohne Anthrazyklin	▪ Zeit bis zur Progression (TTP); Definition: Zeit von der Randomisierung (oder Beginn der Behandlung) bis zur Progression der Erkrankung oder bis zum Auftreten eines Rezidivs[9] ▪ Tumoransprechen (TR); Definition: CR + PR (Responder), keine weiteren Angaben ▪ Progression der Erkrankung (PD): Definition: k.A	▪ Gesamtüberleben (OS) Definition: Zeit von der Randomisierung (oder Beginn der Behandlung) bis zum Tod, unabhängig von der Todesursache
Sherrill 2008	Metastasierender Brustkrebs	Unterschiedliche First-Line- oder nachfolgende Behandlungen: Anthrazykline, Taxane, sonstige Chemotherapien, Hormontherapien, Biologika oder sonstige Therapien, jeweils als Mono- oder Kombinationstherapie, sowohl im Kontroll- als auch im Interventionsarm; Festlegung von Kontroll- und Interventionsarm unklar	▪ Zeit bis zur Progression (TTP) bzw. Progressionsfreies Überleben (PFS); Definition. k. A. Beide Endpunkte wurden in der Auswertung kombiniert, da inkonsistente Definitionen in den eingeschlossenen Studien vorlagen	▪ Gesamtüberleben (OS); Definition: k. A.

(Fortsetzung)

Tabelle 8: Charakterisierung der Validierungsstudien (Indikation, Intervention, untersuchte Endpunkte) – Brustkrebs (Fortsetzung)

Abkürzungen: A: Doxorubicin (= Adriamycin), AC: Adriamycin, Cyclophosphamid, AD: Adriamycin, Docetaxel, AP: Adriamycin, Paclitaxel, CR: Complete Response, D: Docetaxel, DAC: Docetaxel, Adriamycin, Cyclophosphamid, DC: Disease control, DFS: Disease-free survival, E: Epirubicin, EC: Epirubicin, Cyclophosphamid, ED: Epirubicin, Docetaxel, EP: Epirubicin, Paclitaxel, FAC: Fluorouracil + Adriamycin + Cyclophosphamid, FDA: Food and Drug Administration, FEC: Fluorouracil, Epirubicin, Cyclophosphamid, ECOG: Eastern Cooperative Oncology Group, k. A.: keine Angabe, MTX: Methotrexat, OS: Overall survival, P: Paclitaxel, PD: Progressive disease, PFS: Progression-free survival, PR: Partial response, VM: Vincristin, Methotrexat, VP: Vincristin, Prednison, TTP: Time to tumor progression, WHO: World Health Organization.

[1] Untersucht wurden 2- und 3-Jahres-DFS. Da die Ergebnisse ähnlich waren, werden in der Publikation nur die Ergebnisse zu 2-Jahres-DFS präsentiert.
[2] Unterschiedliche DFS-Definitionen in den eingeschlossenen Studien
[3] Geplant waren 5- und 10-Jahres-OS. Aufgrund unzureichender Daten wurde keine Berechnung für 10-Jahres-OS durchgeführt.
[4] Die Begriffe „metastatic“ und „advanced“ wurden parallel in der Publikation verwendet.
[5] Die WHO-Kriterien wurden in 8 der 10 eingeschlossenen Studien angewendet, die Kriterien der Eastern Cooperative Oncology Group in einer Studie, in einer Studie wurde keine Definition genannt.
[6] Im Flow-Chart der Literaturrecherche wurde „not a study of metastatic breast cancer“ ausgeschlossen.
[7] Die Ergebnisse zu anthrazyklin- und thaxanbasierten Therapien wurden jeweils getrennt ausgewertet.
[8] Einige Studien schlossen Frauen mit Brustkrebsrezidiv ein. Es ist unklar, ob diese Patientinnen auch Metastasen aufwiesen.
[9] In 9 Studien wurde auch Tod als Ereignis mitgezählt, diese wurden von der Hauptanalyse für TTP ausgeschlossen und separat ausgewertet.

5.2.2.2 Validierungsverfahren in den eingeschlossenen Studien

Tabelle 9 gibt einen Überblick über die in den eingeschlossenen Validierungsstudien eingesetzten Verfahren klassifiziert nach den in Abschnitt 5.1.2.2 beschriebenen Methodenkategorien.

Tabelle 9: Methodenkategorien in den eingeschlossenen Validierungsstudien

Publikation	Methodenkategorie			
	Testbasierte Verfahren	Korrelations-basierte Verfahren	Kausalmodell-basierte Verfahren	Sonstige Verfahren
Darmkrebs				
Alonso 2008		●		
Baker 2006	●	●		
Burzykowski 2004		●		
Burzykowski 2008b		●		
Buyse 2000b		●		
Buyse 2007		●		
Buyse 2008		●		
Cortiñas Abrahantes 2008		●		
Green 2008		●		
Johnson 2006		●		
Montagnani 2009		●		
Sargent 2005		●		
Sargent 2007		●		
Sertdemir 2009		●		
Tang 2007		●		
Brustkrebs				
Bruzzi 2005	●			
Burzykowski 2008c		●		
Hackshaw 2005		●		
Miksad 2008		●		
Ng 2007		●		
Sherrill 2008		●		

Mit einer Ausnahme wurden in allen Studien korrelationsbasierte Validierungsverfahren eingesetzt. In einer Studie wurden ausschließlich testbasierte Verfahren verwendet, in einer weiteren Studie kamen sowohl testbasierte als auch korrelationsbasierte Verfahren zum

Einsatz. Es wurden keine Studien identifiziert, in denen kausalmodellbasierte oder sonstige Verfahren zur Validierung von Surrogatendpunkten verwendet wurden.

Im Folgenden werden die eingesetzten Validierungsverfahren getrennt für die Studien zur Validierung von Surrogatendpunkten in den Krankheitsbildern Darm- und Brustkrebs im Detail beschrieben.

Darmkrebs

In Tabelle 10 sind Angaben zur Datengewinnung, zur Datengrundlage und zur Methodik der Gegenüberstellung von Surrogatendpunkt und patientenrelevantem Endpunkt für die Validierungsstudien des Krankheitsbildes Darmkrebs aufgeführt.

In 3 der 15 Validierungsstudien zu Darmkrebs erfolgte die Datengewinnung auf Basis einer systematischen Suche in bibliografischen Datenbanken. In 5 Fällen erfolgte der Einschluss von Studien auf nicht systematische Weise. Teilweise wurden ausgewählte Datensätze verwendet (z. B. Sertdemir 2009). Anderen Validierungsstudien lag der Datensatz bzw. ein Teil des Datensatzes der ACCENT-Gruppe zugrunde. In diesem Datensatz sind RCTs zusammengefasst, die von leitenden Prüfärzten zur Verfügung gestellt wurden. Für 7 der Validierungsstudien blieb die Art der Datengewinnung aufgrund fehlender oder diskrepanter Informationen unklar.

Die Datengrundlage für die Validierung beruhte in allen Fällen auf randomisierten kontrollierten Studien. Die Anzahl der eingeschlossenen RCTs reichte von 4 (Sertdemir 2009) bis 146 (Johnson 2006), die Gesamtzahl der Patienten reichte (soweit bekannt) von knapp 3000 (Montagnani 2009) bis über 35 000 (Johnson 2006). In mehreren Fällen war auch die Anzahl der Vergleiche angegeben, die in die Untersuchung der Validität des entsprechenden Endpunkts eingeschlossen wurden. Diese konnte die Anzahl der eingeschlossenen Einzelstudien überschreiten. Dies war dann der Fall, wenn auch Studien mit mehr als 2 Studienarmen eingeschlossen wurden. Hier wurde die jeweilige Kontrollgruppe zumeist mehrfach verwendet. Für 9 der 15 Validierungsstudien standen den Autoren individuelle Patientendaten (IPD) zur Verfügung, für 4 Studien lediglich aggregierte Daten, für 2 Studien war die Art der Datengrundlage unklar.

Die in den Studien verwendeten methodischen Verfahren sind fast ausnahmslos den korrelationsbasierten Verfahren zuzuordnen. Allein die Studie Baker 2006 wählte einen Ansatz, der Kriterien sowohl von test- als auch korrelationsbasierten Verfahren erfüllt. Kausalmodellbasierte Verfahren oder sonstige Verfahren wurden gar nicht verwendet Für den häufigsten Fall, dass bei Verfügbarkeit mehrerer Studien eine meta-analytische Fragestellung vorlag, konnten in der Regel Verfahren eingesetzt werden, die individuelle Patientendaten erfordern, weil diese häufig vorlagen. In den 4 Studien jedoch, die lediglich aggregierte Daten nutzten, wurden entsprechend andere Verfahren eingesetzt. Berechnungen der Assoziation von Surrogatendpunkt und Überleben gemessen durch Bestimmtheitsmaß oder Korrelationskoeffizienten erfolgten bei allen Publikationen mindestens für die jeweiligen Effekte, d. h. auf Studienebene. Der Zusammenhang auf individueller Ebene wurde in 8 Studien ebenfalls untersucht. In wenigen Fällen berichteten Studien stattdessen von einem

Zusammenhang bezogen auf die jeweiligen Interventionen (Studienarme). Vier Studien berechneten außerdem den Surrogate Threshold Effect (STE).

Brustkrebs

In Tabelle 11 sind Angaben zur Datengewinnung, zur Datengrundlage und zur Methodik der Gegenüberstellung von Surrogatendpunkt und patientenrelevantem Endpunkt für die Validierungsstudien des Krankheitsbildes Brustkrebs aufgeführt.

Für alle 6 Validierungsstudien erfolgte die Datengewinnung mittels systematischer Suche in bibliografischen Datenbanken.

Die Datengrundlage für die Validierung waren ausschließlich randomisierte kontrollierte Studien. Die Anzahl der eingeschlossenen RCTs reichte von 10 (Bruzzi 2005) bis 126 (Ng 2007), die Gesamtzahl der Patienten reichte von über 2000 (Bruzzi 2005) bis über 17 000 (Sherrill 2008). In 2 Fällen war auch die Anzahl der Vergleiche angegeben, die in die Untersuchung der Validität des entsprechenden Endpunkts eingeschlossen wurde. Diese konnte die Anzahl der eingeschlossenen Einzelstudien überschreiten. Dies war dann der Fall, wenn auch Studien mit mehr als 2 Studienarmen eingeschlossen wurden. Hier wurde die jeweilige Kontrollgruppe zumeist mehrfach verwendet. Für 2 der 6 Validierungsstudien standen den Autoren individuelle Patientendaten (IPD) zur Verfügung, für 4 Studien lediglich aggregierte Daten.

Eine der Studien (Bruzzi 2005) verwendete ausschließlich den Validierungsansatz nach Prentice (1989) [47] als testbasiertes Verfahren. In den übrigen Studien wurden korrelationsbasierte Verfahren eingesetzt. Nur in einer Studie (Burzykowski 2008c) wurde dieses zur Untersuchung des Zusammenhangs sowohl auf individueller wie auch auf Effekt- bzw. Studienebene eingesetzt. Die anderen 3 Studien verwendeten ebenfalls meta-analytische Ansätze, die jedoch wegen der nur in aggregierter Form vorliegenden Daten nur Aussagen über den Zusammenhang von Surrogatendpunkt und Überleben bezogen auf die Effekte, also auf Studienebene, erlaubten.

Tabelle 10: Methodik zur Validierung der Surrogatendpunkte – Darmkrebs

Studie	Datengewinnung	Datengrundlage	Methodik zur Gegenüberstellung von Surrogat und patientenrelevantem Endpunkt	Methoden-kategorie gemäß Abschnitt 5.1.2.2
Kolonkarzinom im frühen Krankheitsstadium				
Alonso 2008a	Auswahl von Studien zu Kolonkarzinom im frühen Krankheitsstadium aus dem ACCENT-Datensatz mit Verweis auf Sargent 2005; Auswahlprozess unklar[1]	10 RCTs Patientenzahl: individuelle Daten zu > 10 000 Patienten Beobachtungsdauer: k. A.	Zweistufiger meta-analytischer Ansatz nach Buyse (2000a) [76]. 1. Berechnung des informationstheoretischen Assoziationsmaßes R^2_h nach Alonso (2007) auf Studien- und individueller Ebene bei univariater Modellierung von SEP und OS durch PH-Modell [12]. 2. Berechnung von R^2_{trial} bei bivariater Copula-Modellierung (Randverteilungen aus PH-Modell) auf Studienebene.	2
Buyse 2008	Auswahl von Studien zu Kolonkarzinom im frühen Krankheitsstadium aus dem ACCENT-Datensatz mit Verweis auf Sargent 2005; Auswahlprozess unklar[1]	10 RCTs Patientenzahl: individuelle Daten zu 10 255 Patienten Beobachtungsdauer: 8 Jahre (Median)	Zweistufiger meta-analytischer Ansatz nach Buyse (2000a) mit bivariater Copula-Modellierung nach Burzykowski (2001) [76,77]. ▪ Auf Studienebene Berechnung des Korrelationskoeffizienten der Behandlungseffekte (jeweils logHR aus Weibull-PH-Modell) aus linearem Regressionsmodell. Auf individueller Ebene Berechnung des Rangkorrelationskoeffizienten R von SEP und OS. ▪ STE von SEP auf OS nach Burzykowski (2006) [50]	2

(Fortsetzung)

Tabelle 10: Methodik zur Validierung der Surrogatendpunkte – Darmkrebs (Fortsetzung)

Studie	Datengewinnung	Datengrundlage	Methodik zur Gegenüberstellung von Surrogat und patientenrelevantem Endpunkt	Methoden-kategorie gemäß Abschnitt 5.1.2.2
Kolonkarzinom im frühen Krankheitsstadium				
Green 2008	k. A.[2]	13 RCTs Patientenzahl: k. A.[2] Beobachtungsdauer: k. A.	1. Zweistufiger meta-analytischer Ansatz auf Studienarm- und Studienebene: ▪ Lineare Regression der Effektschätzer (HR aus Cox PH-Modell) von OS auf SEP. ▪ Auf Studienarmebene: Effektschätzer Überlebensrate. Auf Studienebene: Effektschätzer HR aus Cox PH-Modell. 2. PTE nach Freedman (1992) [78]. Effektschätzer: Hazard Ratio nach dem Cox PH-Modell 3. Zweistufiger meta-analytischer Ansatz nach Burzykowski (2001) [77]. ▪ Auf Studienebene Berechnung von R^2_{trial}, auf individueller Ebene Berechnung von R^2_{indiv}. 4. Anteil der Übereinstimmung der Studienergebnisse anhand statistischer Tests für SEP und PEP gemäß Begg und Leung 2000 [79]	2
Nicht metastasierendes Kolonkarzinom				
Burzykowski 2008b	Die ACCENT-Gruppe fasste Studiendaten aus RCTs zusammen, die von den leitenden Prüfärzten (principal investigators) der Studien zur Verfügung gestellt wurden (Verweis auf Sargent 2005)	18 RCTs (43 Behandlungsarme, 25 Vergleiche[3]) Patientenzahl: individuelle Daten zu 20 898 Patienten Beobachtungsdauer: 8 Jahre (Median) (Verweis auf Sargent 2005)	Zweistufiger meta-analytischer Ansatz nach Buyse (2000a) mit bivariater Copula-Modellierung [76]. ▪ Auf Studienebene Berechnung der Korrelation der Behandlungseffekte (jeweils logHR aus Weibull-PH-Modell) geschätzt durch lineares Regressionsmodell mit (bzw. ohne[4]) Adjustierung für Meßfehler des Behandlungseffekts nach Burzykowski (2005)[7]. Auf individueller Ebene Rangkorrelationskoffizient von SEP und OS. ▪ Vorhersage des geschätzten Behandlungseffekts für OS in den japanischen Studien ▪ STE von SEP auf OS nach Burzykowski (2006) [50]	2

(Fortsetzung)

Tabelle 10: Methodik zur Validierung der Surrogatendpunkte – Darmkrebs (Fortsetzung)

Studie	Datengewinnung	Datengrundlage	Methodik zur Gegenüberstellung von Surrogat und patientenrelevantem Endpunkt	Methoden-kategorie gemäß Abschnitt 5.1.2.2
Nicht metastasierendes Kolonkarzinom				
Sargent 2005	Die ACCENT-Gruppe fasste Studiendaten aus RCTs zusammen, die von den leitenden Prüfärzten (principal investigators) der Studien zur Verfügung gestellt wurden[5]	18 RCTs (43 Behandlungsarme und 25 Vergleiche[3]) Patientenzahl: individuelle Daten zu insgesamt 20 898 Patienten Beobachtungsdauer: 8 Jahre (Median)	1. Meta-analytischer Ansatz auf Studienarm- und Studienebene: ▪ Lineare Regression der Effektschätzer von OS auf SEP. Berechnung von R^2 sowie Spearman-Rangkorrelationskoeffizient r. ▪ Auf Studienarmebene: Effektschätzer Überlebensrate. Auf Studienebene: Effektschätzer HR aus Cox PH-Modell, Kreuzvalidierung. 2. Zweistufiger meta-analytischer Ansatz nach Burzykowski (2001) mit bivariater Modellierung [77]. ▪ Auf Studienebene Korrelation aus linearem Regressionsmodell von Effektschätzer (jeweils HR) OS auf SEP. Auf individueller Ebene Korrelation von OS und SEP. 3. PTE nach Freedman (1992) [78], Effektschätzer jeweils aus Cox-PH-Modell für OS und SEP 4. Anteil der Übereinstimmung der Studienergebnisse nach Begg (2000) [79]	2
Sargent 2007	Die ACCENT-Gruppe fasste Studiendaten aus RCTs zusammen, die von den leitenden Prüfärzten (principal investigators) der Studien zur Verfügung gestellt wurden (Verweis auf Sargent 2005)	18 RCTs (43 Behandlungsarme, 25 Vergleiche)[3] Patientenzahl: individuelle Daten zu insgesamt 20 898 Patienten Beobachtungsdauer: 8 Jahre (Median) (Verweis auf Sargent 2005)	1. Meta-analytischer Ansatz auf Studienebene: ▪ Lineare Regression der Effektschätzer von OS (HR aus Cox-PH-Modell) auf SEP. ▪ Berechnung von R^2 sowie Spearman-Rangkorrelationskoeffizient ρ. 2. Zweistufiger meta-analytischer Ansatz nach Burzykowski (2001) mit bivariater Modellierung [77]. ▪ Auf Studienebene Assoziation[6] aus linearem Regressionsmodell von Effektschätzer (jeweils HR) OS auf SEP (keine Berechnungen auf individueller Ebene). 3. Anteil der Übereinstimmung der Studienergebnisse nach Begg (2000) [79]	2

(Fortsetzung)

Tabelle 10: Methodik zur Validierung der Surrogatendpunkte – Darmkrebs (Fortsetzung)

Studie	Datengewinnung	Datengrundlage	Methodik zur Gegenüberstellung von Surrogat und patientenrelevantem Endpunkt	Methoden-kategorie gemäß Abschnitt 5.1.2.2
Nicht metastasierendes Kolonkarzinom				
Sertdemir 2009	Ausgewählte Datensätze des National Surgical Adjuvant Breast and Bowel Project (NSABP)	4 RCTs (C-01 bis C-04) mit insgesamt 5506 Patienten [80-83] C-01: 1166 Patienten Beobachtungsdauer 77,3 Monate (Mittelwert) C-02: 1158 Patienten Beobachtungsdauer 41,8 Monate (Mittelwert) C-03: 1081 Patienten Beobachtungsdauer 47,6 Monate (Mittelwert) C-04: 2151 Patienten Beobachtungsdauer 86 Monate (Mittelwert)	1. Berechnung des Bestimmtheitsmaßes R^2_{trial} nach Buyse (2000a) mit Effektschätzern für OS und SEP aus univariater Cox-Regression [76]. 2. PTE nach Freedman (1992) [78], Effektschätzer jeweils aus Cox-Regression.	2
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)				
Alonso 2008a	Auswahl von Studien aus dem Datensatz in MAGIC 2004 (Verweis auf MAGIC 2004) Auswahlprozess unklar[7]	10 RCTs Patientenzahl: k. A. Beobachtungsdauer: k. A.	Zweistufiger meta-analytischer Ansatz nach Buyse (2000a) [76]. 3. Berechnung des informationstheoretischen Assoziationsmaßes R^2_h nach Alonso (2007) auf Studien- und individueller Ebene bei univariater Modellierung von SEP und OS durch PH-Modell [12]. 4. Berechnung von R^2_{trial} bei bivariater Copula-Modellierung (Randverteilungen aus PH-Modell) auf Studienebene.	2

(Fortsetzung)

Tabelle 10: Methodik zur Validierung der Surrogatendpunkte – Darmkrebs (Fortsetzung)

Studie	Datengewinnung	Datengrundlage	Methodik zur Gegenüberstellung von Surrogat und patientenrelevantem Endpunkt	Methoden-kategorie gemäß Abschnitt 5.1.2.2
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)				
Baker 2006	Verweis auf 4 Publikationen: ACCMAP 1992, 1994 und MAGIC 1996, 1998 [84-87] (Verweis auf Burzykowski 2004 und Buyse 2000b) Auswahlprozess unklar[8]	Anzahl an RCTs: k. A. (26 Vergleiche) Patientenzahl: k. A. Beobachtungsdauer: k. A.	Meta-analytischer Ansatz nach eigenem methodischen Vorschlag (Baker 2006) für aggregierte Daten: ▪ Berechnung und Vergleich der Kennzahlen Average prediction error for the predicted effect (APEP), average prediction error of other true endpoints (APEO), average clinically meaningful difference (ACMD)	1, 2
Burzykowski 2004	Suche durch Meta-Analysis Group in Cancer (Verweis auf 4 Publikationen: ACCMAP 1992, 1994 und MAGIC 1996, 1998 [84-87]) Auswahlprozess unklar[8]	24 RCTs (27 Vergleiche[3]) Patientenzahl: individuelle Daten zu 4010 Patienten; Beobachtungsdauer: k. A.	Zweistufiger meta-analytischer Ansatz nach Buyse (2000a) und Burzykowski (2001) mit bivariater Copula-Modellierung [76,77]. ▪ Auf Studienebene Berechnung des 1. Bestimmtheitsmaß $R^2_{trial}(\alpha)$ aus lin. Regressionsmodell 2. Bestimmtheitsmaß $R^2_{trial}(\alpha,\eta)$ aus Proportional Odds Modell ▪ Auf individueller Ebene: Assoziationsparameter θ der Plackett-Copula. ▪ Für SEP binär: Durchführung aller Verfahren für varierende Landmark-Zeiten sowie mit und ohne Adjustierung für Performance-Status	2
Buyse 2000b	Suche durch Meta-Analysis Group in Cancer (Verweis auf 4 Publikationen: ACCMAP 1992, 1994 und MAGIC 1996, 1998 [84-87]) Auswahlprozess unklar [8]	25 RCTs, Patientenzahl: individuelle Daten zu 3791 Patienten Beobachtungsdauer: k. A.	Meta-analytischer Ansatz nach Buyse (1998) auf Studienebene [88]. ▪ Berechnung des RE der Effektschätzer für OS (logHR) und SEP (logOR). ▪ Gewichtete Regression von Effektschätzer des OS auf Effektschätzer des SEP und Vorhersage des Behandlungseffekts auf OS durch SEP und Berechnung von R^2. ▪ Modellierung für verschiedene Landmark-Zeiten (1 – 12 Monate). PTE nach Freedman (1992) [78], Lin (1997) [89].	2

(Fortsetzung)

Tabelle 10: Methodik zur Validierung der Surrogatendpunkte – Darmkrebs (Fortsetzung)

Studie	Datengewinnung	Datengrundlage	Methodik zur Gegenüberstellung von Surrogat und patientenrelevantem Endpunkt	Methoden-kategorie gemäß Abschnitt 5.1.2.2
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)				
Buyse 2007	k. A.	10 RCTs Patientenzahl: 3089 Beobachtungsdauer: 30,4 Monate (Median)	Zweistufiger meta-analytischer Ansatz nach Buyse (2000a) mit bivariater Copula-Modellierung nach Burzykowski (2001) [76,77]. ▪ Auf Studienebene Berechnung der Korrelation R der Behandlungseffekte (HRs aus Kaplan-Meier-Schätzern) aus linearem Regressionsmodell. Auf individueller Ebene: Rangkorrelation von SEP und OS. ▪ Berechnungen jeweils 1. für Beobachtungszeitraum 6 Monate (SEP) und 12 Monaten (OS)) und 2. für vollständigen Beobachtungszeitraum ▪ STE von SEP auf OS nach Burzykowski (2005) [7] ▪ Vorhersage des geschätzten Behandlungseffekts für OS in den Validierungsstudien.	2
Buyse 2008	k. A. (Verweis auf Burzykowski 2008b und Buyse 2007)	10 RCTs Patientenzahl: individuelle Daten zu 3089 Patienten Beobachtungsdauer: 30,4 Monate (Median)	Zweistufiger meta-analytischer Ansatz nach Buyse (2000a) mit bivariater Copula-Modellierung nach Burzykowski (2001) [76,77]. ▪ Auf Studienebene Berechnung des Korrelationskoeffizienten der Behandlungseffekte (jeweils logHR aus Weibull-PH-Modell) aus linearem Regressionsmodell. Auf individueller Ebene Berechnung des Rangkorrelationskoeffizienten R von SEP und OS. ▪ STE von SEP auf OS nach Burzykowski (2006) [50]	2

(Fortsetzung)

Tabelle 10: Methodik zur Validierung der Surrogatendpunkte – Darmkrebs (Fortsetzung)

Studie	Datengewinnung	Datengrundlage	Methodik zur Gegenüberstellung von Surrogat und patientenrelevantem Endpunkt	Methoden-kategorie gemäß Abschnitt 5.1.2.2
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)				
Cortiñas Abrahantes 2008	k. A.[9] (Verweis auf Burzykowski 2004 und 2005, Verweis auf Corfu-Studie [90] und Greco 1996 [91])	27 Studien[9], Patientenzahl: individuelle Daten zu 4010 Patienten; Beobachtungsdauer: k. A.	Zweistufiger meta-analytischer Ansatz nach Buyse (2000a) mit bivariater Copula-Modellierung der Endpunkte [76]. ▪ Berechnung des informationstheoretischen Assoziationsmaßes R^2_h nach Alonso (2007) auf Studienebene mit folgenden Verfahren [12]: 1. Ursprünglicher Vorschlag nach Alonso (2007), R^2_h 2. Lineares Regressionsmodell (Konfidenzintervalle durch Bootstrap), R^2_h 3. Lineares Regressionsmodell (Konfidenzintervalle über Delta-Methode), R^2_h 4. Lineares Regressionsmodell (Konfidenzintervalle nach Ding (1996) [92]), R^2_h 5. Regression trees, RD_{tree} 6. Regression trees mit Bagging, RD_{tree} 7. Random forests, R^2_h 8. Support vector machines, RD_{SVMR} ▪ Durchführung der Verfahren 2. – 8. mit und ohne Kreuzvalidierung	2

(Fortsetzung)

Tabelle 10: Methodik zur Validierung der Surrogatendpunkte – Darmkrebs (Fortsetzung)

Studie	Datengewinnung	Datengrundlage	Methodik zur Gegenüberstellung von Surrogat und patientenrelevantem Endpunkt	Methoden-kategorie gemäß Abschnitt 5.1.2.2
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)				
Green 2008	k.A[2].	10 RCTs Patientenzahl: k. A.[2] Beobachtungsdauer: k. A.	1. Zweistufiger meta-analytischer Ansatz auf Studienarm- und Studienebene: ▪ Lineare Regression der Effektschätzer (HR aus Cox PH-Modell) von OS auf SEP. ▪ Auf Studienarmebene: Effektschätzer Überlebensrate. Auf Studienebene: Effektschätzer HR aus Cox PH-Modell. 2. PTE nach Freedman (1992) [78]. Effektschätzer: Hazard Ratio nach dem Cox PH-Modell 3. Zweistufiger meta-analytischer Ansatz nach Burzykowski (2001) [77]. ▪ Auf Studienebene Berechnung von R^2_{trial}, auf individueller Ebene Berechnung von R^2_{indiv}. 4. Anteil der Übereinstimmung der Studienergebnisse anhand statistischer Tests für SEP und PEP gemäß Begg und Leung 2000[79].	2
Metastasierendes Kolorektalkarzinom				
Johnson 2006	Systematische Suche in MEDLINE, EMBASE, Cochrane, PubMed, Current Contents, CINAHL, OLDMEDLINE; Handsuche in systematischen Reviews, führenden Zeitschriften, Online-Abstractbänden von Konferenzen (ASCO, FECS, NCI)	146 RCTs Patientenzahl: 35 337 Beobachtungsdauer: k. A.	Meta-analytischer Ansatz für aggregierte Daten auf Studienebene: ▪ Lineare Regression von Effektschätzer des OS auf Effektschätzer des SEP und Berechnung von R^2. Effektschätzer: Differenz der medianen Überlebenszeit zwischen den Behandlungsgruppen (TTP) bzw. Differenz des Responderanteils zwischen den Behandlungsgruppen (TR). ▪ Surrogate Threshold von SEP auf OS	2

(Fortsetzung)

Tabelle 10: Methodik zur Validierung der Surrogatendpunkte – Darmkrebs (Fortsetzung)

Studie	Datengewinnung	Datengrundlage	Methodik zur Gegenüberstellung von Surrogat und patientenrelevantem Endpunkt	Methoden-kategorie gemäß Abschnitt 5.1.2.2
Metastasierendes Kolorektalkarzinom				
Montagnani 2009	Suche in MEDLINE und Cochrane-Register sowie Handsuche in Abstractbänden der ASCO und der ESMO	5 RCTs (18 Studienarme und 9 Vergleiche) Patientenzahl: 2901 Beobachtungsdauer: k. A.	Meta-analytischer Ansatz auf Studienarm- und Studienebene für aggregierte Daten: ▪ Auf Studienarmebene Rangkorrelationskoeffizient der Endpunkte OS und SEP ▪ Auf Studienebene Rangkorrelationskoeffizient der Effektschätzer (Differenz der medianen Überlebenszeit zwischen den Behandlungsgruppen) für OS und SEP.	2
Tang 2007	Systematische Suche in MEDLINE, EMBASE und Cochrane Central Register of Controlled Trials.	39 RCTs (87 Studienarme) Patientenzahl: 18 668 Beobachtungsdauer: 12 bis 57,6 Monate (Median)	1. Meta-analytischer Ansatz auf Studienarm- und Studienebene für aggregierte Daten: ▪ Auf Studienarmebene Rangkorrelationskoeffizient r_s der Endpunkte OS und SEP. Auf Studienebene Rangkorrelationskoeffizient r_s für die Effektschätzer (Differenz der Überlebenszeit) des OS und des SEP. 2. Lineare Regression von Effektschätzer des OS auf Effektschätzer des SEP, Berechnung des Bestimmtheitsmaßes R^2	2

Abkürzungen: ACCENT: Adjuvant Colon Cancer End Points Group, ACCMAP: Advanced Colorectal Cancer Meta-Analyis Project, ACMD: average clinically meaningful difference, APEP: average prediction error for the predicted effect, APEO: average prediction error of other true endpoints, ASCO: American Society of Clinical Oncology, ESMO: European Society of Medical Oncology, FECS: Federation of European Cancer Societies, (heute ECCO: European Cancer Organisation), HR: Hazard Ratio, k. A.: keine Angabe, MAGIC: Meta-Analysis Group in Cancer, NCI: National Cancer Institute, NSABP: National Surgical Adjuvant Breast and Bowel Project, OR: Odds Ratio, OS: Overall survival, PH-Modell: Proportional Hazards Modell, PEP: patientenrelevanter Endpunkt. PFS: Progression-free survival, PTE: Proportion of treatment effect, RCT: randomisierte kontrollierte Studie, RE: relativer Effekt, RR: Response rate, SEP: Surrogatendpunkt, STE: Surrogat Threshold Effect, TTP: Time to tumor progression, TR: Tumor response.

[1] Die Daten stammen aus dem ACCENT-Datensatz. Dieser umfasst 18 RCTs mit 20 898 Patienten.
[2] In Green 2008 werden Sargent 2005 (ACCENT-Daten, 18 RCTs) sowie Burzykowski 2001 (Auswertung von 2 RCTs mit fortgeschrittenem Kolorektalkarzinom) zitiert. Nähere Angaben zur Auswahl der Studien und Zuordnung zu der Indikation ist nicht erfolgt.

(Fortsetzung)

Tabelle 10: Methodik zur Validierung der Surrogatendpunkte – Darmkrebs (Fortsetzung)

[3] Bei mehrarmigen Studien ging der Kontrollarm mehrfach in die Auswertung ein.
[4] Die Ergebnisse der Auswertung „ohne Adjustierung“ werden in der Publikation nicht dargestellt.
[5] Gemäß den Autoren wurden Studien nach Relevanz, Stand der Fertigstellung und Datenverfügbarkeit ausgewählt.
[6] Damit ist in dieser Publikation vermutlich das Bestimmtheitsmaß R^2 gemeint.
[7] Der in MAGIC 2004 beschriebene Datensatz. umfasst 19 Studien mit 3338 Patienten.
[8] In den 4 Publikationen werden insgesamt 28 Studien mit IPD zu 4289 Patienten beschrieben.
[9] Angabe in der Publikation mit Verweis auf Burzykowski 2004. Dort ist die Datengrundlage 27 Vergleiche zu 24 Studien. Die Patientenzahlen in beiden Publikationen stimmen überein. Es wird jedoch auch auf andere Studien mit anderen Patientenzahlen verwiesen.

Tabelle 11: Methodik zur Validierung der Surrogatendpunkte – Brustkrebs

Studie	Datengewinnung	Datengrundlage	Methodik zur Gegenüberstellung von Surrogat und patientenrelevantem Endpunkt	Methoden-kategorie gemäß Abschnitt 5.1.2.2
Nicht metastasierender oder lokal fortgeschrittener Brustkrebs				
Ng 2007	Systematische Suche in MEDLINE und Pubmed	126 RCTs (149 Vergleiche[1]); Patientenzahl: 533 (Median; Spannweite 216 – 3757); Beobachtungsdauer: 81 Monate (Median; Spannweite 34 – 258)	Meta-analytischer Ansatz für aggregierte Daten (auf Studienebene): ▪ Regression von Effektschätzer für SEP auf Effektschätzer des OS (jeweils Differenzen der Überlebenszeiten der Studienarme) mit generalized estimating equations: 1. Basis-Modell nur mit Prädiktor des SEP, 2. vollständiges Modell mit studienspezifischen Kovariablen als weiteren Prädiktoren. ▪ Berechnung des Korrelationskoeffizienten r und des Bestimmtheitsmaßes R^2 a) aus dem o. g. Modell, b) nach Chu (2002) [93].	2
Fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses)				
Bruzzi 2005	Systematische Suche in MEDLINE, Handsuche in Originalpublikationen und Reviews Die leitenden Prüfärzte (principal investigators) oder die Hersteller stellten anschließend individuelle Patientendaten zur Verfügung.	10 RCTs Patientenzahl: individuelle Daten von insgesamt 2126 Patienten Beobachtungsdauer: k. A.	Nachweis von[2] ▪ Behandlung verlängert Überleben ▪ Behandlung steht in Zusammenhang mit Änderung des SEP ▪ Hohe Werte des SEP stehen in Zusammenhang mit hohen Werten des OS ▪ Kein Behandlungseffekt auf OS, wenn Adjustierung des Effekts um SEP	1

(Fortsetzung)

Tabelle 11: Methodik zur Validierung der Surrogatendpunkte – Brustkrebs (Fortsetzung)

Studie	Datengewinnung	Datengrundlage	Methodik zur Gegenüberstellung von Surrogat und patientenrelevantem Endpunkt	Methoden-kategorie gemäß Abschnitt 5.1.2.2
Fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses)				
Burzykowski 2008c	Systematische Suche in MEDLINE und CANCERLIT (Informationen aus Piccart-Gebhart et al. 2008 [94])	11 RCTs Patientenzahl: individuelle Daten zu 3959 Patienten Beobachtungsdauer: k. A.	Zweistufiger meta-analytischer Ansatz nach Buyse (2000a) mit bivariater Copula-Modellierung [76]. ▪ Für SEP TR und DC auf Studienebene Berechnung von ρ der Behandlungseffekte[3] (logOR aus Proportional Odds-Modell bzw. logHR aus Weibull-PH-Modell). Auf individueller Ebene Berechnung des OR. Modellierung für verschiedene Landmark-Zeiten (1- 6 Monate). ▪ Für SEP PFS und TTP auf Studienebene Berechnung des Rangkorrelations-koffizienten (logHR aus Weibull-PH-Modell) nach Burzykowski (2001). Auf individueller Ebene Berechnung des Rangkorrelationskoeffizienten.	2
Miksad 2008	Systematische Suche in MEDLINE, Suche in Abstractbänden von Konferenzen (ASCO, SABCS)	*Anthrazyklin-basierte Kombinationstherapien:* 16 RCTs (17 Vergleiche) Patientenzahl: 4323 Beobachtungsdauer: k. A. *Taxan-basierte Kombinationstherapien:* 15 RCTs (17 Vergleiche) Patientenzahl: 5893 Beobachtungsdauer: k. A.	1. Grad der Übereinstimmung der Richtung der Effektschätzer (HR) für SEP und OS mittels Kappa 2. Meta-analytischer Ansatz auf Studienebene für aggregierte Daten: ▪ Lineare Regression von Effektschätzer des OS auf Effektschätzer des SEP (jeweils logHR). ▪ Berechnung des Bestimmtheitsmaßes R^2 ▪ Modellvalidierung durch Kreuzvalidierung ▪ Sensitivitätsanalysen a) unter Verwendung berichteter statt selbst berechneter Effektschätzer, b) nur unter Einbeziehung von Studien, die SEP gemäß FDA-Definition verwenden.	2

(Fortsetzung)

Tabelle 11: Methodik zur Validierung der Surrogatendpunkte – Brustkrebs (Fortsetzung)

Studie	Datengewinnung	Datengrundlage	Methodik zur Gegenüberstellung von Surrogat und patientenrelevantem Endpunkt	Methoden-kategorie gemäß Abschnitt 5.1.2.2
Metastasierender Brustkrebs				
Hackshaw 2005	Systematische Suche in MEDLINE	42 RCTs Patientenzahl: 9163 Beobachtungsdauer: k. A.	Meta-analytischer Ansatz für aggregierte Daten (auf Studienebene) nach A'Hern (1988) [95]: ▪ Lineare Regression von Effektschätzer des OS (HR) auf Effektschätzer des SEP (OR). Berechnung des Bestimmtheitsmaßes R^2.	2
Sherrill 2008	Systematische Suche in Pubmed, Handsuche in Referenzlisten eingeschlossener Studien und wichtiger Übersichtsarbeiten	67 RCTs Patientenzahl: 17 081 Beobachtungsdauer: k. A.	Meta-analytischer Ansatz für aggregierte Daten (auf Studienebene): ▪ Lineare Regression von Effektschätzer des OS auf Effektschätzer des SEP (jeweils HR). Berechnung des Bestimmtheitsmaßes R^2.	2

Abkürzungen: ASCO: American Society of Clinical Oncology, DC: Disease control, FDA: Food and Drug Administration, HR: Hazard Ratio, k. A.: keine Angabe, OR: Odds ratio, OS: Overall survival, PFS: Progression-free survival, RCT: randomisierte kontrollierte Studie, SEP: Surrogatendpunkt, TR: Tumor response, TTP: Time to tumor progression, SABCS: San Antonio Breast Cancer Symposium.

[1] Bei mehrarmigen Studien ging der Kontrollarm mehrfach in die Auswertung ein.
[2] Kann als Formulierung der Prentice-Kriterien für die gegebene Situation angesehen werden, auch wenn Prentice (1989) [47] an dieser Stelle nicht genannt oder zitiert wird.
[3] Nicht weiter spezifiziert.

5.2.2.3 Einschätzung der Aussagesicherheit der eingeschlossenen Studien

Anhand der in Abschnitt 4.2.3 genannten Kriterien wurde die Aussagesicherheit der Ergebnisse in den eingeschlossenen Validierungsstudien bewertet.

Darmkrebs

Tabelle 12 gibt einen Überblick über die Bewertung der Aussagesicherheit für die einzelnen Validierungsstudien zu Darmkrebs.

Insgesamt wurde die Aussagesicherheit für die eingeschlossenen Validierungsstudien ausnahmslos als niedrig bewertet, da in allen Fällen mehrere Kriterien in großem Ausmaß von der jeweils höchstmöglichen Aussagesicherheit abwichen.

Zwar wurden in den meisten Studien (12 von 15 Studien) anerkannte und in der Literatur beschriebene Verfahren zur Validierung verwendet, jedoch wurden nur für einen kleinen Teil der Studien Sensitivitätsanalysen zur Prüfung von Robustheit und / oder Generalisierbarkeit durchgeführt. In 5 Studien wurden Sensitivitätsanalysen präsentiert: Burzykowski 2008b untersuchte die Validität bei veränderter zeitlicher Operationalisierung des Surrogatendpunktes DFS. Sargent 2007 analysierte die Validität sowohl für eine gemeinsame Population von Stadium-II- und -III-Patienten wie auch für getrennte Populationen. Burzykowski 2004 rechnete Analysen unter Verwendung verschiedener Landmark-Zeiten und für verschiedene Messniveaus des Surrogatendpunktes. Buyse 2007 bezog Daten aus verschiedenen Zeiträumen ein. Cortiñas Abrahantes 2008 untersuchte verschiedene Zusammenhangsstrukturen zwischen den Effekten von Surrogat und patientenrelevantem Endpunkt. Von diesen Studien untersuchten außerdem 3 (Burzykowski 2008, Buyse 2007, Cortiñas Abrahantes 2008) die Generalisierbarkeit, indem sie entweder überprüften, ob die Ergebnisse der Validierung auf weitere Datensätze übertragbar waren oder Kreuzvalidierung anwendeten.

Lediglich in 3 Studien beruhte die Datengewinnung auf einer systematischen Suche. Für einen Großteil der eingeschlossenen Studien (12 der 15 Studien) war jedoch die Art der Datengewinnung nicht systematisch oder zumindest unklar. Zwar beruhten die Validierungen bei diesen Studien zum Teil auf einer sehr hohen Patientenzahl (z. B. über 18 000 Patienten aus dem ACCENT-Datensatz bei Sargent 2005), die Repräsentativität dieses Studienpools ließ sich anhand der vorliegenden Informationen aber nicht beurteilen.

Weiterhin lagen für 11 der 15 Validierungsstudien keine ausreichenden Informationen über die eingeschlossenen Patientenpopulationen vor. Zwar wurde die Erkrankung der Patienten grob beschrieben als zum Beispiel „fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)“, aus den Publikationen ging aber in keinem Fall hervor, wie hoch der Anteil der Patienten mit Rektumkarzinom oder der Anteil der Patienten mit Metastasen war. Für 3 Studien (Burzykowski 2008b; Sargent 2005; Sertdemir 2009) wurde die untersuchte Patientenpopulation nicht hinreichend eingegrenzt, sodass Patienten mit unterschiedlichen Prognosen zusammen ausgewertet wurden. Für keine dieser 3 Studien standen geeignete Analysen für Teilpopulationen zur Verfügung. Für die Studie Sargent 2007

wurde die Aussagesicherheit bezogen auf dieses Kriterium zweigeteilt: Zum einen wurde die Validität von 3-Jahres-DFS als Surrogat für 5-Jahres-OS getrennt für Patienten mit Stadium II und III durchgeführt; diese wurden als hinreichend eingegrenzt angesehen. Zum anderen wurde die Validität für unterschiedliche Operationalisierungen des Endpunkts DFS als Surrogat für OS untersucht, jedoch an einer Gesamtpopulation von Patienten mit Stadium II und III. Für diese Analysen wurde die Indikation als nicht hinreichend eingegrenzt angesehen, auch deshalb, weil sich in den getrennten Auswertungen deutlich unterschiedliche Korrelationen für Patienten mit Stadium II und Stadium III ergaben.

In lediglich 2 Studien bei fortgeschrittenem Kolorektalkarzinom wurden die untersuchte Intervention und die jeweilige Kontrollintervention auf einen einzelnen Wirkstoff eingegrenzt (Sertdemir 2009 und Alonso 2008a). Bei Sertdemir wurden allerdings nur 4 einzelne Studien ausgewertet und nicht zusammengefasst. Bei 11 der 15 Validierungsstudien wurden für die jeweiligen Interventionsgruppen unterschiedliche Wirkstoffklassen zusammengefasst. Für keine dieser Studien lagen getrennte Auswertungen für einzelne Wirkstoffe oder zumindest Wirkstoffklassen vor. Somit kann keine Aussage darüber getroffen werden, auf welche Situationen die Validierung anwendbar ist. Bei allen übrigen Studien waren die Informationen in den Studien für eine Bewertung der Eingrenzung der Interventionen bzw. Kontrollinterventionen unzureichend. Für keine der 15 Validierungsstudien standen geeignete Sensitivitätsanalysen zur Verfügung.

Ebenso waren nur für 6 der 15 Validierungsstudien zum Krankheitsbild einheitliche Definitionen für die untersuchten Surrogatendpunkte oder die jeweiligen patientenrelevanten Endpunkte angewendet.

Brustkrebs

Tabelle 13 gibt einen Überblick über die Bewertung der einzelnen zur Bewertung der Aussagesicherheit berücksichtigten Kriterien für die einzelnen Validierungsstudien zu Brustkrebs.

Auch in den Validierungsstudien zu Brustkrebs wurde die Aussagesicherheit ausnahmslos als niedrig bewertet, da in allen Fällen mehrere Kriterien in großem Ausmaß von der jeweils höchstmöglichen Aussagesicherheit abwichen.

So wurden nur in 2 der 6 Studien anerkannte und in der Literatur beschriebene Verfahren zur Validierung verwendet. Darüber hinaus wurden auch nur in einem geringen Anteil der Studien Analysen zur Prüfung der Robustheit des angewendeten Verfahrens und / oder die Generalisierbarkeit der Ergebnisse zur Validierung überprüft.

In allen 6 Studien erfolgte die Datengewinnung mittels systematischer Suche in bibliografischen Datenbanken.

Während die in die Validierungsstudien eingeschlossenen Patientenpopulation in 2 Studien hinreichend eingegrenzt war, waren die Informationen zur Patientenpopulation in 4 weiteren Studien für eine solche Einschätzung unzureichend.

In 5 der 6 Studien wurden sowohl für die Intervention als auch für die Kontrollintervention unterschiedliche Wirkstoffklassen zusammengefasst. Für keine dieser Studien lagen getrennte Auswertungen für einzelne Wirkstoffe oder zumindest Wirkstoffklassen vor. Somit kann keine Aussage darüber getroffen werden, auf welche Situationen die Validierung anwendbar ist. In der Studie Miksad 2008 wurden hingegen Anthrazyklin- und Taxan-basierte Therapien getrennt ausgewertet. Dennoch waren in diesem Fall keine Informationen zur jeweiligen Kontrollintervention angegeben, sodass auch hier von einer niedrigen Aussagesicherheit ausgegangen werden musste. Für keine der 6 Validierungsstudien standen geeignete Sensitivitätsanalysen zur Verfügung.

In der Hälfte der Studien zum Krankheitsbild Brustkrebs waren die Informationen zu den Definitionen für die untersuchten Endpunkte unzureichend. In 2 Studien wurden durchgehend einheitliche Definitionen angegeben, in einer Studie waren diese uneinheitlich.

Tabelle 12: Bewertung der Aussagesicherheit der eingeschlossenen Validierungsstudien – Darmkrebs

Studie	Anerkanntes, beschriebenes Verfahren verwendet	Analysen zur Robustheit / Generalisier-barkeit	Datengrundlage	Eingrenzung der Indikationen und / oder Schweregrade	Eingrenzung der Interventionen / Kontroll-interventionen	Endpunkt-Definitionen	Aussage-sicherheit
Kolonkarzinom im frühen Krankheitsstadium							
Alonso 2008a	ja	nein / nein	unklar	unklar	unklar / unklar	unklar	niedrig
Buyse 2008	ja	nein / nein	nicht systematisch	unklar	unklar / unklar	einheitlich	niedrig
Green 2008	ja	nein / nein	unklar	unklar	unklar / unklar	einheitlich	niedrig
Nicht metastasierendes Kolonkarzinom							
Burzykowski 2008b	ja	ja / ja	nicht systematisch	nein	unterschiedliche Wirkstoffklassen / unterschiedliche Wirkstoffklassen	unklar	niedrig
Sargent 2005	ja	nein / nein	nicht systematisch	nein	unterschiedliche Wirkstoffklassen / unterschiedliche Wirkstoffklassen	unklar	niedrig
Sargent 2007	ja	ja / nein	nicht systematisch	ja bzw. nein[a]	unterschiedliche Wirkstoffklassen / unterschiedliche Wirkstoffklassen	unklar	niedrig
Sertdemir 2009	ja	nein /nein	nicht systematisch	nein	Wirkstoff / Wirkstoff	einheitlich	niedrig
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)							
Alonso 2008a	ja	nein / nein	unklar	unklar	Wirkstoff / Wirkstoff	unklar	niedrig
Baker 2006	ja	nein / nein	unklar	unklar	unterschiedliche Wirkstoffklassen / Wirkstoffklasse	unklar	niedrig

(Fortsetzung)

Tabelle 12: Bewertung der Aussagesicherheit der eingeschlossenen Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Anerkanntes, beschriebenes Verfahren verwendet	Analysen zur Robustheit / Generalisier-barkeit	Datengrundlage	Eingrenzung der Indikationen und / oder Schweregrade	Eingrenzung der Interventionen / Kontroll-interventionen	Endpunkt-Definitionen	Aussage-sicherheit
Burzykowski 2004	ja	ja / nein	unklar	unklar	unterschiedliche Wirkstoffklassen / Wirkstoffklasse	einheitlich	niedrig
Buyse 2000b	ja	nein / nein	unklar	unklar	unterschiedliche Wirkstoffklassen / Wirkstoffklasse	einheitlich	niedrig
Buyse 2007	ja	ja / ja	unklar	unklar	unterschiedliche Wirkstoffklassen / Wirkstoff	unklar	niedrig
Buyse 2008	ja	nein / nein	nicht systematisch	unklar	unterschiedliche Wirkstoffklassen / Wirkstoff	einheitlich	niedrig
Cortiñas Abrahantes 2008	ja	ja / ja	unklar	unklar	unklar / unklar	unklar	niedrig
Green 2008	ja	nein / nein	unklar	unklar	unklar / unklar	einheitlich	niedrig
Metastasierendes Kolorektalkarzinom							
Johnson 2006	nein	nein / nein	systematisch	unklar	unterschiedliche Wirkstoffklassen / unterschiedliche Wirkstoffklassen	unklar	niedrig
Montagnani 2009	nein	nein / nein	systematisch	unklar	unterschiedliche Wirkstoffklassen / unterschiedliche Wirkstoffklassen	unklar	niedrig

(Fortsetzung)

Tabelle 12: Bewertung der Aussagesicherheit der eingeschlossenen Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Anerkanntes, beschriebenes Verfahren verwendet	Analysen zur Robustheit / Generalisier-barkeit	Datengrundlage	Eingrenzung der Indikationen und / oder Schweregrade	Eingrenzung der Interventionen / Kontroll-interventionen	Endpunkt-Definitionen	Aussage-sicherheit
Tang 2007	nein	nein / nein	systematisch	unklar	unterschiedliche Wirkstoffklassen / unterschiedliche Wirkstoffklassen	einheitlich	niedrig
a: In der Studie Sargent 2007 fanden sich zum einen Analysen getrennt für Patienten mit Stadium II und III, diese wurden als hinreichend eingegrenzt angesehen. Zum anderen wurde die Validität für unterschiedliche Operationalisierungen des Endpunkts DFS als Surrogat für OS untersucht, jedoch an einer Gesamtpopulation von Patienten mit Stadium II und III. Für diese Analysen wurde die Indikation als nicht hinreichend eingegrenzt angesehen.							

Tabelle 13: Bewertung der Aussagesicherheit der eingeschlossenen Validierungsstudien – Brustkrebs

Studie	Anerkanntes, beschriebenes Verfahren verwendet	Analysen zur Robustheit / Generalisier-barkeit	Datengrundlage	Eingrenzung der Indikationen und / oder Schweregrade	Eingrenzung der Interventionen / Kontroll-interventionen	Endpunkt-Definitionen	Aussage-sicherheit
Nicht metastasierender oder lokal fortgeschrittener Brustkrebs							
Ng 2007	nein	nein / nein	systematisch	ja	unterschiedliche Wirkstoffklassen / unterschiedliche Wirkstoffklassen	uneinheitlich	niedrig
Fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses)							
Bruzzi 2005	ja	nein / nein	systematisch	unklar	unterschiedliche Wirkstoffklassen / unterschiedliche Wirkstoffklassen	einheitlich	niedrig
Burzykowski 2008c	ja	nein / nein	systematisch	unklar	unterschiedliche Wirkstoffklassen / unterschiedliche Wirkstoffklassen	unklar	niedrig
Miksad 2008	nein	ja / ja	systematisch	unklar	Anthrazyklin-basierte Therapie: Wirkstoffklasse / unklar. Taxan-basierte Therapie: unterschiedliche Wirkstoffklassen / unklar	einheitlich	niedrig

(Fortsetzung)

Tabelle 13: Bewertung der Aussagesicherheit der eingeschlossenen Validierungsstudien – Brustkrebs (Fortsetzung)

Studie	Anerkanntes, beschriebenes Verfahren verwendet	Analysen zur Robustheit / Generalisier-barkeit	Datengrundlage	Eingrenzung der Indikationen und / oder Schweregrade	Eingrenzung der Interventionen / Kontroll-interventionen	Endpunkt-Definitionen	Aussage-sicherheit
Metastasierender Brustkrebs							
Hackshaw 2005	nein	nein / nein	systematisch	unklar	unterschiedliche Wirkstoffklassen / unterschiedliche Wirkstoffklassen	unklar	niedrig
Sherrill 2008	nein	ja / nein	systematisch	ja	unterschiedliche Wirkstoffklassen / unterschiedliche Wirkstoffklassen	unklar	niedrig

5.2.3 Ergebnisse zur Validierung von Surrogatendpunkten für patientenrelevante Endpunkte

In den folgenden Abschnitten werden die Ergebnisse der Studien zur Validierung von Surrogatendpunkten für die Krankheitsbilder Darmkrebs und Brustkrebs präsentiert. Dabei werden zunächst die numerischen Ergebnisse aus den Publikationen sowie die Schlussfolgerung der Autoren der Validierungsstudien zur Validität der Surrogatparameter dargestellt.

Zusätzlich wurden die Ergebnisse der Studien aus Sicht des IQWiG zusammengefasst und interpretiert. Die Interpretation der Ergebnisse erfolgte dabei zunächst getrennt bezüglich der statistischen Resultate und bezüglich der Aussagesicherheit. Auf Basis dieser Bewertung der statistischen Ergebnisse wurde unter Berücksichtigung der Aussagesicherheit der jeweiligen Studie zusätzlich eine Gesamtinterpretation der Daten zur Validität der untersuchten Surrogatendpunkte vorgenommen.

Die Bewertung der statistischen Ergebnisse erfolgte zunächst in Abhängigkeit der Kategorie der verwendeten Methodik (vgl. 5.1): Bei Verfahren aus der Kategorie 1 (testbasierte Verfahren), die die Prentice-Kriterien untersuchen, wurde die Validität nur beim Nachweis aller 4 Kriterien als gegeben betrachtet [47]. In allen anderen Fällen blieb die Validität unklar.

Bei Verfahren aus der Kategorie 2 (korrelationsbasierte Verfahren), die das Ausmaß des Zusammenhangs von Surrogat bzw. Surrogateffekt mit dem des Endpunktes bzw. Effekts auf den Endpunkt untersuchen, wurde zunächst geprüft, ob Angaben über die Assoziation sowohl auf individueller wie auch auf Studienebene (Effektebene) vorlagen (siehe Abschnitt 5.1.2.2). Im Hinblick auf eine Nutzenbewertung des IQWiG wird die Korrelation auf individueller Ebene nicht berücksichtigt (siehe Abschnitt 6.1 der Diskussion). Lagen diese Angaben nicht in vollständiger Form von Punktschätzern mit zugehörigen Konfidenzintervallen vor, gab es keine statistisch beurteilbaren Resultate, sodass auch die Validität des Surrogats unklar blieb. Gab es jedoch Angaben, z. B. in Form von Bestimmtheitsmaß und / oder Korrelationskoeffizienten jeweils mit Konfidenzintervall, konnte die Validität ggf. genauer bewertet werden: Dazu wurde das Ausmaß des Zusammenhangs durch Einstufung der Korrelation in die drei Kategorien hoch, mittel und niedrig bewertet: Bei Korrelationskoeffizienten (R, r oder ρ) $\geq 0{,}85$ bzw. Bestimmtheitsmaßen (R^2) $\geq 0{,}72$, gemessen durch die jeweils untere 95%-Konfidenzintervallgrenze, wurde die Korrelation als hoch bezeichnet. In der biometrischen Literatur gibt es keine einheitliche klare Regelung zur Beurteilung der Korrelationen und damit der Validität. Es wird in den meisten Fällen nur unspezifisch eine „hohe“ Korrelation verlangt. In der Publikation Burzykowski (2005) werden Bestimmtheitsmaße von „around 0,9“ in diesem Sinne als ausreichend bezeichnet; Lassere (2008) leitet aus Bestimmtheitsmaßen > 0,9 „excellent statistical validity“ und aus solchen > 0,75 „very good statistical validity“ ab. (Die Wahl der Grenze von 0,85 für den Korrelationskoeffizienten erfolgte unter der Überlegung, dass bei „wahrer“ Korrelation von 0,9 auch genaue Schätzungen zwangsläufig eine untere Konfidenzintervallgrenze kleiner als 0,9 aufweisen.) Ebenso besteht Einigkeit darüber, dass niedrige Korrelationen unvereinbar mit der Validität eines Surrogatendpunktes sind. So kann die Aussage von Piedbois (2008),

dass die Akzeptanz eines Surrogats bei einem Korrelationskoeffizienten von mindestens 0,7 möglich ist, als Vorschlag einer Mindestanforderung angesehen werden. Buyse (2009) bezeichnet bei einem Beispiel ein Bestimmtheitsmaß von 0,47 als „unacceptably low“ [17]. Daher wurde die Korrelation als niedrig bezeichnet, wenn der Korrelationskoeffizient (R, r oder ρ) $\leq 0,7$ bzw. das Bestimmtheitsmaß (R^2) $\leq 0,49$ betrug, jeweils unter Berücksichtigung der oberen Konfidenzintervallgrenze. Die übrigen Fälle, für die auch in der Literatur keine allgemeinen Bewertungsrichtlinien existieren, ließen zunächst keine eindeutigen Schlussfolgerungen zu. Da eine Korrelation zwischen 0,85 und 0,7 in diesen Fällen möglich war, wurden entsprechende Ergebnisse als mittlere Korrelation bezeichnet.

Die Bewertung der Aussagesicherheit erfolgte gemäß Tabelle 12 und Tabelle 13 in Abschnitt 5.2.2.3.

Für die zusammenfassende Interpretation der Ergebnisse wurden die Bewertungen der statistischen Ergebnisse sowie der Aussagesicherheit gemeinsam betrachtet: Bei hoher Aussagesicherheit wäre das abschließende Urteil der Bewertung allein unter Berücksichtigung der statistischen Ergebnisse erfolgt. In diesem Fall hätte bei korrelationsbasierten Verfahren eine hohe Korrelation den Nachweis der Validität des Surrogatendpunktes bedeutet. Eine niedrige Korrelation weist bei hoher Aussagesicherheit dagegen das Fehlen einer Validität nach. Bei niedriger Aussagesicherheit blieb die Validität des Surrogatendpunktes insgesamt unklar. Bei eingeschränkter und mäßiger Aussagesicherheit ist die Validität des Surrogatendpunktes ebenfalls zunächst unklar. In diesen Fällen kann ein Surrogat bei ausreichend hohem Behandlungseffekt aber dennoch sinnvoll eingesetzt werden, zum Beispiel durch Anwendung des Surrogate Threshold Effects (STE). Bei der Bewertung der Validität wurden Angaben zum Surrogat Threshold Effect (STE) jedoch nicht berücksichtigt, da dies in Abhängigkeit vom beobachteten Behandlungseffekt des Surrogats bezüglich der betrachteten Intervention zu erfolgen hat.

5.2.3.1 Darmkrebs

Tabelle 14 zeigt die Ergebnisse und Schlussfolgerungen aus den Studien zur Validierung von Surrogatparametern des Krankheitsbildes Darmkrebs.

Von den 3 Studien zur Indikation Kolonkarzinom in frühem Krankheitsstadium untersuchte 1 Studie TTR und 2 Studien untersuchten DFS als Surrogat für Gesamtüberleben. Bei TTR wurden zwar hohe Punktschätzer des Bestimmtheitsmaßes erreicht, aber die statistische Validität konnte aufgrund breiter Konfidenzintervalle nicht beurteilt werden. Verbunden mit der niedrigen Aussagesicherheit der zu Grunde liegenden Validierungsstudie ist insgesamt die Validität des Surrogats nicht nachgewiesen. Bezüglich der Ergebnisse zum DFS waren in beiden Studien die Korrelationskoeffizienten bzw. Bestimmtheitsmaße hoch (jeweils untere Grenze des 95 %-Konfidenzintervalls über 0,85 bzw. 0,72). Wegen der ebenfalls niedrigen Aussagesicherheit bedingt durch durchweg unzureichende Informationen bleibt die Validität insgesamt jedoch unklar.

Alle 4 Studien mit Ergebnissen zum nicht metastasierten Kolonkarzinom untersuchten den Surrogatendpunkt DFS, teilweise in verschiedenen Operationalisierungen, für denselben Datensatz (ACCENT-Datensatz). In den meisten Fällen wurde der Zusammenhang auf individueller Ebene nicht berichtet. Zur Bedeutung der Korrelation auf individueller Ebene für eine Nutzenbewertung des IQWiG siehe Abschnitt 6.1 im Kapitel „Diskussion“. Auf Studienebene waren die unteren Grenzen der Konfidenzintervalle für die Korrelationskoeffizienten bzw. Bestimmtheitsmaße zumeist kleiner als 0,85 bzw. 0,72, sodass in diesen Fällen mittlere Korrelationen möglich waren. Vereinzelt zeigten sich auf Studienebene auch hohe Korrelationen. Die Aussagesicherheit war in allen Fällen niedrig, insbesondere deswegen, weil jeweils unterschiedliche Wirkstoffklassen und Patientenpopulationen mit unterschiedlichen Prognosen zusammengefasst ausgewertet wurden, ohne dass entsprechende Sensitivitätsanalysen vorlagen. Lediglich die Studie Sargent 2007 berichtete eine getrennte Auswertung für Patienten mit jeweils Stadium II und III. Hier zeigten sich für die verschiedenen Stadien deutlich unterschiedliche Ergebnisse, was die Validität einer zusammengefassten Auswertung infrage stellt. Damit ist die Validität auch insgesamt unklar. Ergänzend sei bemerkt, dass in der Studie Sargent 2007 verschiedene Operationalisierungen für den Endpunkt DFS untersucht wurden, indem die Daten zu verschiedenen Zeitpunkten (nach 1, 2 bzw. 3 Jahren Beobachtungsdauer) zensiert wurden. Hier wurden bei später erfolgenden Zensierungen höhere Zusammenhänge beobachtet.

Von den 8 Studien mit Ergebnissen zum Indikationsbereich fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms) untersuchte 1 den Surrogatendpunkt TTR, 4 den Surrogatendpunkt TR und 3 den Surrogatendpunkt PFS. Für TTR kann eine hohe Korrelation auf Studienebene ausgeschlossen werden. Auch unter Berücksichtigung der niedrigen Aussagesicherheit bleibt die Validität des Surrogats unklar. Für TR, die jeweils am selben Datensatz untersucht wurde, konnte Baker 2006 unter Nutzung des selbst entwickelten Ansatzes keine Validität zeigen. Die niedrigen Punktschätzer von Burzykowski 2004 und Buyse 2000 sind ebenfalls Anzeichen gegen die Validität, allerdings kann wegen der breiten Konfidenzintervalle die Situation nicht eindeutig bewertet werden. (Allein für den Fall der ordinal skalierten TR gibt es wegen des niedrigen Bestimmtheitsmaßes, auch unter Berücksichtigung der Konfidenzintervalle, Anzeichen dafür, dass der Surrogatendpunkt nicht valide ist.) Ähnlich sind die Ergebnisse von Cortiñas Abrahantes 2008 zu sehen, die zusätzlich noch durch die Analysen zur Generalisierbarkeit weitere Argumente gegen die Validität liefern. Erschwerend für die Gesamtbeurteilung kommt bei allen genannten Studien hinzu, dass möglicherweise Daten von Patienten mit unterschiedlichen Prognosen zusammengefasst ausgewertet wurden und dabei oft auch Interventionen aus unterschiedlichen Wirkstoffklassen zusammengefasst wurden. Damit bleibt die Valididät von TR letztlich unklar. Die Analysen zum Surrogatendpunkt PFS ergaben mittlere (Green 2008) bis hohe Zusammenhänge (Buyse 2007; Buyse 2008) auf Studienebene. Allerdings ergab sich in allen 3 Studien eine niedrige Aussagesicherheit, insbesondere durch unzureichende Angaben zu den in die Auswertung einbezogenen Patientenpopulationen und zum Teil die Zusammenfassung unterschiedlicher Wirkstoffklassen ohne entsprechende Sensitivitätsanalyen. Dadurch ist insgesamt die

Validität von PFS als Surrogat für OS unklar. Aus diesem Grund sollten auch die Ergebnisse zum STE, die Buyse 2007 und Buyse 2008 präsentieren, mit Vorsicht interpretiert werden.

Für metastasierendes Kolorektalkarzinom wurden die Surrogatendpunkte TR, TTP und PFS in jeweils 2 Studien untersucht. Für TR berichten Johnson 2006 und Tang 2007 niedrige bis mittlere Korrelationen bei unbekannter bzw. hoher Variabilität. Ähnliche Beobachtungen können an denselben Studien und mit denselben Schlussfolgerungen auch für TTP gemacht werden. In Johnson 2006 werden auch STEs angegeben, jedoch sind diese wegen fehlender Konfidenzintervalle nicht zu bewerten. Für den Surrogatendpunkt PFS zeigen Montagnani 2009 und Tang 2007 mittlere Korrelationen. Da die Aussagesicherheit in allen Fällen v. a. wegen interventions- und indikationsübergreifender Datensätze wiederum niedrig ist, ist die Validität dieser Surrogatendpunkte insgesamt als unklar anzusehen.

Übergreifend über die Indikationen und Interventionen kann festgestellt werden, dass die Ergebnisse in den Studien aus statistischer Sicht in den meisten Fällen keine eindeutige Schlussfolgerung über die Validität der Surrogate erlauben bzw. der Surrogatendpunkt wegen geringen Zusammenhangs mit dem klinischen Endpunkt als nicht valide anzusehen ist. Die Gründe hierfür liegen einerseits in oftmals breiten 95 %-Konfidenzintervallen für die zugehörigen Punktschätzer, bei denen die obere Intervallgrenze um mehr als 0,3 größer ist als die untere. Andererseits gibt es häufig keine Angaben zur Variabilität, sodass die Genauigkeit der Schätzer von Korrelation oder Bestimmtheitsmaß unklar bleibt. Problematisch ist weiterhin, dass in Studien, die mehrere Verfahren bzw. Operationalisierungen eines Endpunktes verwenden, teilweise deutlich verschiedene Punktschätzer resultieren (diese diskrepanten Ergebnisse erschweren die Beurteilung der Validität).

Erschwert wird die Beurteilung in allen Fällen auch durch eine niedrige Aussagesicherheit. Für keine Studie kann sichergestellt werden, dass nicht Daten von Patienten mit deutlich unterschiedlichen Prognosen zusammengefasst ausgewertet wurden. In den meisten Fällen wurde darüber hinaus auch nicht zwischen Interventionen unterschiedlicher Wirkstoffklassen und damit auch völlig unterschiedlichen Wirkprinzipien unterschieden. Geeignete Sensitivitätsanalysen, die die Auswirkungen der verschiedenen zusammengefassten Indikationen oder Interventionen untersuchten, lagen für nahezu keine Studie vor. Weitere nachteilige Bewertungsaspekte kamen in allen Fällen noch dazu. Vor diesem Hintergrund ist die Validität für alle untersuchten Surrogatendpunkte für alle Indikationsbereiche letztlich nicht nachgewiesen.

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Kolonkarzinom in frühem Krankheitsstadium			
Alonso 2008a	*TTR als Surrogat für OS[1]:* ▪ Individuelle Ebene: R^2_h = 0,84 (95%-KI [0,83; 0,85]) ▪ Studienebene: 1. R^2_h = 0,85 (95%-KI [0,53; 0,96]) 2. R^2_{trial} = 0,82 (95%-KI [0,44; 0,95])	„Der große Wert [..] für R^2 zeigt, dass die Zeit bis zum Wiederauftreten des Krebses über 84% der Unsicherheit bezüglich des Überlebens des Patienten erklärt. […] breite Konfidenzintervalle behindern unsere Interpretation dieser Punktschätzer." ("The large value […] for the R^2 indicates that time-to-cancer recurrence can explain more than 84% of the uncertainty about the survival of the patient. […] wide confidence intervals hamper our interpretation of these point estimates")	▪ Interpretation der statistischen Ergebnisse Hohe Korrelation auf individueller Ebene[2]; auf Studienebene ist mittlere Korrelation möglich. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.
Buyse 2008	*DFS als Surrogat für OS:* ▪ Rangkorrelationskoeffizient von DFS und OS = 0,96 (95%-KI [0,95; 0,97]) ▪ Korrelationskoeffizient R der Behandlungseffekte für DFS und OS (jeweils als logHR) = 0,94 (95%-KI [0,87;1,01]) ▪ Surrogate threshold effect (HR_{DFS}) = 0,93 (Nutzen) bzw. 1,07 (geringerer Nutzen)	„DFS ist ein hervorragendes Surrogat für OS […] beim frühen kolorektalen Karzinom" „DFS (und PFS bei fortgeschrittener Erkrankung) sind akzeptable Surrogate für OS bei kolorektalem Karzinom" ("In early colorectal cancer […] DFS is an excellent surrogate for OS" "DFS (and PFS in advanced disease) are acceptable surrogates for OS in colorectal cancer")	▪ Interpretation der statistischen Ergebnisse Es liegen auf individueller[2] und Studienebene hohe Korrelationen vor. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Kolonkarzinom in frühem Krankheitsstadium			
Green 2008	*DFS als Surrogat für OS:* 1. Studienarmebene: R^2 = 0,88; Spearman-Rangkorrelationskoeffizient = 0,93 Studienebene: R^2 = 0,80; Spearman-Rangkorrelationskoeffizient = 0,85 2. PTE: > 100 % 3. Studienebene: τ = 0,90 (95%-KI [0,89; 0,90]) indiv. Ebene: R^2_{indiv} = 0,85 (95%-KI [0,72; 0,99]) 4. Übereinstimmung der Studienergebnisse in 12 von 13 Studien (in 10 dieser Studien jeweils kein signifikanter Unterschied)	„die formalen Maße von Freedman und Burzykowski et al. deuten eine Surrogatfunktion an“ „deuten konsistent eine Surrogatfunktion zwischen beiden Endpunkten an“ (“the formal measures of Freedman and Burzykowski et al. suggest surrogacy” “consistently suggest surrogacy between the two endpoints“)	▪ Interpretation der statistischen Ergebnisse Es liegen auf individueller[2] und Studienebene hohe Korrelationen vor. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Nicht metastasierendes Kolonkarzinom			
Burzykowski 2008b	*Analyse auf Grundlage aller Informationen (vollständige Nachbeobachtungszeit)* ▪ Korrelationskoeffizient zwischen Behandlungseffekten für DFS und OS (jeweils als logHR[3]): 0,95 (95%-KI [0,91; 0,99]) ▪ Surrogate threshold effect $logHR_{DFS}$ = -0,09 (damit HR_{DFS}= 0,91 bzw. 1,09) ▪ Sensitivitätsanalyse: Vorhersage des Behandlungseffekts durch Surrogat innerhalb des Vorhersageintervalls bei 4 von 8 Teilpopulationen[4]. *Analyse auf Grundlage der zensierten Informationen (Zensierung OS nach 5 Jahren und DFS wie angegeben nach 1, 2 oder 3 Jahren):* ▪ Korrelationskoeffizienten zwischen den Behandlungseffekten von DFS und OS (jeweils als logHR): 1-Jahres-DFS: 0,86 (95%-KI [0,76; 0,96]) 2-Jahres-DFS: 0,92 (95%-KI [0,90; 0,94]) 3-Jahres-DFS: 0,92 (95%-KI [0,90; 0,95]) ▪ *Surrogate threshold effect:* 1-Jahres-DFS: $logHR_{DFS}$ = -0,49 (damit HR_{DFS} = 0,61 bzw. 1,63) 2-Jahres-DFS: $logHR_{DFS}$ = -0,26 (damit HR_{DFS} = 0,77 bzw. 1,30) 3-Jahres-DFS: $logHR_{DFS}$ = -0,21 (damit HR_{DFS} = 0,81 bzw. HR_{DFS} = 1,23) ▪ Sensitivitätsanalyse: Vorhersage des Behandlungseffekts durch Surrogat innerhalb des Vorhersageintervalls bei 8 von 8 Studien – für alle der untersuchten Zensierungsarten[5].	„Die Ergebnisse suggerieren, dass DFS, nach 3 Jahren zensiert, als ein Surrogat für OS, nach 5 Jahren zensiert, für kolorektales Karzinom angesehen werden kann.“ „wenn eine Studie einer neuen experimentellen Behandlung DFS als ein Surrogat für OS verwendet, wird eine erhebliche Unsicherheit bleiben bezüglich des echten Überlebensvorteils, den diese Behandlung bietet“ „die Verwendung von DFS nach 2 oder 3 Jahren wäre die beste Option für die Vorhersage des OS nach 5 Jahren“ (“Results suggest that DFS censored at 3 years can be considered a surrogate for OS censored at 5 years for colorectal cancer” “if a trial of a new experimental treatment uses DFS as a surrogate for OS, there will remain substantial uncertainty regarding the true survival benefit provided by that treatment” “using DFS at 2 or 3 years would be the best option for the prediction of OS at 5 years“)	▪ Interpretation der statistischen Ergebnisse Korrelationen auf individueller Ebene sind nicht berichtet.[2] Auf Studienebene liegt hohe Korrelation (DFS, 2-Jahres-DFS, 3-Jahres-DSF) bzw. mindestens mittlere Korrelation (1-Jahres-DFS) vor. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität der Surrogatendpunkte unklar.

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Nicht metastasierendes Kolonkarzinom			
Sargent 2005	*Median 3-Jahres-DFS (alle Patienten) als Surrogat für OS:* 1. Studienarmebene: R^2 = 0,85;r = 0,88 Studienebene: R^2 = 0,90; r = 0,94 Kreuzvalidierung: 24 von 25 beobachteten HRs für OS innerhalb der 95%-Vorhersageintervalle 2. Studienebene: Korrelation = 0,78 (95%-KI [0,60; 0,96]); individuelle Ebene: Korrelation = 0,873 (95%-KI [0,869; 0,877]) 3. PTE > 100 % 4. Konkordanz[10] in 23 von 25 Studien (in 5 Studien beide Endpunkte signifikant)	„Die Ergebnisse suggerieren, dass nach 3 Jahren medianer Beobachtungszeit DFS in klinischen Studien mit adjuvanter Therapie des Kolonkarzinoms mit einem Fluorouracil-basierten Regime ein angemessener Endpunkt ist, obwohl marginal signifikante Verbesserungen im DFS sich möglicherweise nicht in einen signifikanten OS-Nutzen übersetzen“ (”These results suggest that DFS after 3 years of median follow-up is an appropriate end point for adjuvant colon cancer clinical trials of fluorouracilbased regimens, although marginally significant DFS improvements may not translate into significant OS benefits.”)	▪ Interpretation der statistischen Ergebnisse Hohe Korrelation auf individueller Ebene.[2] Auf Studienebene ist mittlere Korrelation möglich. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Nicht metastasierendes Kolonkarzinom			
Sargent 2007	*DFS[6,7] als Surrogat für 5-Jahres-OS[8]:* *3-Jahres(Median)-DFS (Stadium II):* 1. Studienebene: R^2 = 0,70 (95%-KI [0,44; 0,80]); ρ = 0,70[9] 2. trial-level association = 0,70 (95%-KI [0,47; 0,93]) 3. Konkordanz[10] in 22 von 25 Studien (in 1 Studie beide Endpunkte signifikant) *3-Jahres(Median)-DFS (Stadium III):* 1. Studienebene: R^2 = 0,92 (95%-KI [0,83; 0,95]); ρ = 0,98 2. trial-level association = 0,88 (95%-KI [0,78; 0,98]) 3. Konkordanz[10] in 20 von 25 Studien *3-Jahres(Median)-DFS (Stadien II/III)*: 1. Studienebene: R^2 = 0,90 (95%-KI [0,80; 0,94]) 2. trial-level association = 0,78 (95%-KI [0,60; 0,96]) 3. Konkordanz[10] in 23 von 25 Studien *3-Jahres(Minimum)-DFS (Stadien II/III):* 1. Studienebene: R^2 = 0,92 (95%-KI [0,82; 0,94]) 2. trial-level association = 0,88 (95%-KI [0,66; 0,97]) 3. Konkordanz[10] in 23 von 25 Studien *2-Jahres(Median)-DFS (Stadien II/III):* 1. Studienebene: R^2 = 0,81 (95%-KI [0,63; 0,88]) 2. trial-level association = 0,72 (95%-KI [0,51; 0,94]) 3. Konkordanz[10] in 23 von 25 Studien	„DFS-Endpunkte nach 2 oder 3 Jahren medianem Follow-up sind ausgezeichnete Prädiktoren für 5-Jahres-Gesamtüberleben.“ „…in dieser erweiterten Analyse der ACCENT-Datenbank haben wir validiert, dass für Studien in Patienten mit Stadium-III-Erkrankung oder einer Mischung aus Stadium-II- und -III-Erkrankung DFS nach einer dreijährigen medianen Beobachtungszeit ein valider und angemessener primärer Endpunkt ist.“ „Die früheren Endpunkte von 1-Jahres (Minimum)-DFS als ein früher Indikator eines negativen Ergebnisses und 2-Jahres (Median)-DFS als primärer Endpunkt sind vielversprechend und sollten in zukünftigen Studien validiert werden.“ (“DFS outcomes after 2- or 3-year median follow-up are excellent predictors of 5-year OS.” ”...in this expanded analysis of the ACCENT database, we have validated that for trials in patients with stage III disease, or a mix of stage II and III disease, DFS after 3-year median follow-up is a valid and appropriate primary end point.” “The earlier end points of 1-year minimum DFS, as an early indicator of a negative outcome, and 2-year median DFS, as a primary end point, are promising and should be validated in future trials.”)	▪ Interpretation der statistischen Ergebnisse Korrelationen auf individueller Ebene sind nicht berichtet.[2] Unter der Annahme, dass „trial-level association“ ein quadriertes Korrelationsmaß ist[11]: mittlere Korrelation möglich (Stadium II), hohe Korrelation (Stadium III). Mindestens mittlere Korrelation (für alle speziellen Operationalisierungen bei Stadien II/III). ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität aller Surrogatendpunkte unklar.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Nicht metastasierendes Kolonkarzinom			
Sargent 2007 (Fortsetzung)	*1-Jahres(Minimum)-DFS (Stadien II/III):* 1. Studienebene: R^2 = 0,86 (95%-KI [0,72; 0,91]) 2. trial-level association = 0,48 (95%-KI [0,43; 0,92]) *3.* Konkordanz[10] in 22 von 25 Studien		
Sertdemir 2009	*DFS als Surrogat für OS:* Studie C-01: 1. R^2_{trial} = 0,641 (95%-KI [0,432; 0,782]) 2. PTE = 0,361 (95%-KI [-4,65; 5,31]) Studie C-02: 1. R^2_{trial} = 0,223 (95%-KI [0,008; 0,503]) 2. PTE = 1,271 (95%-KI [-1,28; 4,95]) Studie C-03: 1. R^2_{trial} = 0,761 (95%-KI [0,550; 0,872]) 2. PTE = 0,864 (95%-KI [0,18; 1,64]) Studie C-04: 1. R^2_{trial} = 0,560 (95%-KI [0,404; 0,687]) 2. PTE = 1,67 (95%-KI [-1,11; 6,15])	„RFS ist kein gutes Surrogat für das Gesamtüberleben in klinischen Studien mit nicht-signifikanten Behandlungseffekten und moderat für signifikante Behandlungseffekte" (Ziel: nicht die Validität von RFS zu beweisen…aufzeigen, wie diese Kriterien durch den Behandlungseffekt beeinflusst werden)" ("recurrence free survival [RFS] is not a good surrogate for overall survival in clinical trials with non significant treatment effects and moderate for significant treatment effects" (aim: not to prove validity of RFS … show how these criteria are affected by treatment effect))	▪ Interpretation der statistischen Ergebnisse Korrelationen auf individueller Ebene sind nicht berichtet.[2] Auf Studienebene ist mittlere Korrelation möglich. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)			
Alonso 2008a	*TTR als Surrogat für OS:* ▪ Individuelle Ebene: R^2_h = 0,76 (95%-KI [0,74; 0,78]) ▪ Studienebene: 1. R^2_h = 0,82 (95%-KI [0,41; 0,96]) 2. R^2_{trial} = 0,88 (95%-KI [0,52; 0,97])	„Wir […] beobachteten einen großen Wert für R^2_h, der zeigte, dass das Surrogat einen großen Anteil (über 76%) unserer Unsicherheit über den wahren Endpunkt erklären kann." "we […] observed a large value of R^2_h indicating that the surrogate can explain a large proportion (more than 76%) of our uncertainty about the true endpoint."	▪ Interpretation der statistischen Ergebnisse Auf individueller Ebene mittlere Korrelation.[2] Auf Studienebene mittlere Korrelation möglich. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.
Baker 2006	*TR als Surrogat für Überleben[12]:* APEP = 9 (95%-KI [6; 12]) ACMD = 21 APEO = 8 (95%-KI [5; 11]) damit APEP < ACMD, APEP ~ APEO	„Tumoransprechen ist kein guter Surrogat-Endpunkt, um den Effekt der Behandlung auf den Unterschied in den medianen Überlebenszeiten vorherzusagen." ("tumor response is not a good surrogate endpoint for predicting the effect of treatment on the difference in median survival times.")	▪ Interpretation der statistischen Ergebnisse Die Schätzung des APEP liegt nicht unterhalb der des APEO (Validitätskriterium nicht erfüllt). ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)			
Burzykowski 2004	*TR als Surrogat für OS[13]:* *SEP binär (Landmark-Zeit 0 Monate[14]):* ▪ *Studienebene* 1. $R^2_{trial}(\alpha) = 0,44$ (95%-KI [0,13; 0,69]) 2. $R^2_{trial}(\alpha,\eta) = 0,46$ (95%-KI [0,12; 0,69]) ▪ *Individuelle Ebene:* $\theta = 4,91$ (95%-KI [4,16; 5,67]) *SEP binär (Landmark-Zeit 6 Monate):* ▪ *Studienebene* 1. $R^2_{trial}(\alpha) = 0,29$ (95%-KI [0,03; 0,57]) 2. $R^2_{trial}(\alpha,\eta) = 0,31$ (95%-KI [0,02; 0,58]) ▪ *Individuelle Ebene:* $\theta = 2,71$ (95%-KI [2,28; 3,14]) *SEP ordinal (4 Kategorien)* ▪ *Studienebene:* $R^2_{trial}(\alpha) = 0,16$ (95%-KI [0,00; 0,42]) ▪ *Individuelle Ebene:* $\theta = 6,78$ (95%-KI [6,01; 7,55])	„Tumoransprechen ist ein schwaches Surrogat für eine Vorhersage des Behandlungseffektes auf Studienebene, auch wenn Response hochgradig prädiktiv für das Überleben […] auf individueller Ebene ist“ „Diese Ergebnisse suggerieren, dass die Verwendung von Tumoransprechen als ein Surrogat für das Überleben in Studien, die den Effekt einer 5FU oder FUDR-einschließenden Behandlung bei fortgeschrittenem kolorektalen Karzinom untersuchen, zu nicht validen Ergebnissen führen kann.“ „Weitere Forschung ist notwendig“ (“tumour response is a poor surrogate for a prediction of the treatment effect at the trial level, even though the response is highly prognostic of survival […] at the individual level” „These results suggest that using tumour response as a surrogate for survival in trials investigating the effect of treatment involving 5FU or FUDR in advanced colorectal cancer may lead to invalid results.” “Further research is required”)	▪ Interpretation der statistischen Ergebnisse Auf Studienebene mittlere Korrelation möglich (binärer SEP) bzw. niedrige Korrelation (ordinaler SEP). Für die Bewertung auf individueller Ebene fehlt ein Bewertungsmaßstab.[2] ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)			
Buyse 2000b	*TR als Surrogat für OS:* ▪ p < 0,0001 („tumor response was a highly significant predictor of survival, irrespective of the landmark time chosen or adjustment for performance status“) ▪ RE = 0,12 [0,03; 0,20][15] ▪ R^2 = 0,38 [0,09; 0,68] ▪ PTE = 0,95 [0,52; 4,00]	„In der Praxis impliziert solch ein niedriges Bestimmtheitsmaß, dass Behandlungseffekte auf das Überleben nicht zuverlässig vorhergesagt werden können aus Behandlungseffekten auf das Ansprechen in individuellen klinischen Studien.“ „Die Zunahme der Tumoransprechrate lässt sich übersetzen in eine Zunahme des Gesamtüberlebens […]. Das Wissen, dass eine Behandlung einen Nutzen bezogen auf das Tumoransprechen hat, erlaubt keine genaue Vorhersage des letztendlichen Nutzens bezüglich des Überlebens.“ (“In practice, such a low coefficient of determination[16] implies that treatment effects on survival cannot be predicted reliably from treatment effects on response in individual clinical trials” “Increase in tumor response rate translates into an increase in overall survival […] Knowledge that a treatment has benefits on tumor response does not allow accurate prediction of the ultimate benefit on survival.”)	▪ Interpretation der statistischen Ergebnisse Korrelation auf individueller Ebene ist nicht berichtet.[2] Mittlere Korrelation auf Studienebene möglich.. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)			
Buyse 2007	*PFS als Surrogat für OS:* ▪ *Individuelle Ebene:* 1. verkürzter Zeitraum: ρ = 0,32 (95%-KI [-0,14; 0,67]) 2. gesamter Zeitraum: ρ = 0,82 (95%-KI [0,82; 0,83]) ▪ *Korrelation zwischen den Behandlungseffekten für PFS und OS:* 1. verkürzter Zeitraum: R = 0,94 (95%-KI [0,87; 1,01]) 2. gesamter Zeitraum: R = 0,99 (95%-KI [0,94; 1,04][17]) ▪ STE = 0,86 bzw. 1,16[18] *Vergleich der beobachteten Effekte in OS mit den 95%-Vorhersageintervallen für die vorhergesagten Effekte in OS auf Basis der beobachteten Effekte in PFS in 3 Validierungsstudien[19]:* Studie 1[20] HR_{OS} beobachtet: 1,31 HR_{OS} vorhergesagt: 1,25 (95%-VI [1,00; 1,55] Studie 2[21]: HR_{OS} beobachtet: 1,24 HR_{OS} vorhergesagt: 1,17 (95%-VI [0,96; 1,43] Studie 3[22]: HR_{OS} beobachtet: 1,21 HROS vorhergesagt: 1,40 (95%-VI [1,12; 1,75])	„PFS ist ein akzeptables Surrogat für OS beim fortgeschrittenen kolorektalen Karzinom.“ (“PFS is an acceptable surrogate for OS in advanced colorectal cancer.”)	▪ Interpretation der statistischen Ergebnisse Mittlere Korrelation auf individueller Ebene[2], hohe Korrelation auf Studienebene. (verkürzter Zeitraum: Mittlere Korrelation möglich auf individueller Ebene, hohe Korrelation auf Studienebene) ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)			
Buyse 2008	*PFS als Surrogat für OS:* ▪ Rangkorrelationskoeffizient von PFS und OS = 0,82 (95%-KI [0,82;0,83]) ▪ Korrelationskoeffizient R der Behandlungseffekte für PFS und OS (jeweils als logHR) = 0,99 [0,94;1,04][23] ▪ Surrogate threshold effect (HRPFS) = 0,86 (Nutzen) bzw. 1,16 (geringerer Nutzen)[24]	„PFS ist ein gutes Surrogat für OS […] bei fortgeschrittenem kolorektalem Karzinom“ „DFS (und PFS bei fortgeschrittener Erkrankung) sind akzeptable Surrogate für OS bei kolorektalem Karzinom“ (“In advanced colorectal cancer […] PFS is a good surrogate for OS” “DFS (and PFS in advanced disease) are acceptable surrogates for OS in colorectal cancer“)	▪ Interpretation der statistischen Ergebnisse Mittlere Korrelation auf individueller Ebene[2], hohe Korrelation auf Studienebene. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)			
Cortiñas Abrahantes 2008	*TR als Surrogat für OS:* *Studienebene:* 1. R^2_h = 0,298 [0,109; 0.673] ▪ ohne Kreuzvalidierung: 2. R^2_h = 0,151 [0,014; 0,461] 3. R^2_h = 0,151 [-0,113; 0,415] 4. R^2_h = 0,151 [0,000; 0,438] 5. RD_{tree} = 0,472 [0,305; 0,851] 6. RD_{tree} = 0,567 [0,441; 0,734] 7. R^2_h = 0,623 [0,454; 0,833] 8. RD_{SVMR} = 0,450 [0,157; 0,738] ▪ mit Kreuzvalidierung 2. R^2_h = 0,003 [0,000; 0,297] 3. R^2_h = 0,003 [-0,041; 0,047] 4. R^2_h = 0,003 [0,000; 0,156] 5. RD_{tree} = 0,293 [0,021; 0,654] 6. RD_{tree} = 0,279 [0,099; 0,654] 7. R^2_h = 0,344 [0,046; 0,696] 8. RD_{SVMR} = 0,294 [0,022; 0,630]	"die Surrogatfunktion auf der Studienebene unterscheidet sich erheblich von dem, was in der Literatur berichtet wurde … größere Werte und engere Konfidenzintervalle" ("the trial level surrogacy is considerably different form what has been reported in the literature … larger values and narrower confidence intervals")	▪ Interpretation der statistischen Ergebnisse Korrelationen auf individueller Ebene sind nicht berichtet.[2] Auf Studienebene sind mittlere Korrelationen nur bei manchen Verfahren möglich, ansonsten liegen niedrige Korrelationen vor. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)			
Green 2008	*PFS als Surrogat für OS:* 1. Studienarmebene: R^2 = 0,002; Spearman-Rangkorrelationskoeffizient = 0,13 Studienebene: R^2 = 0,84; Spearman-Rangkorrelationskoeffizient = 0,92 2. PTE: > 100 % 3. Studienebene: τ = 0,58 (95%-KI [0,18; 0,98]) indiv. Ebene: R^2_{indiv} = 0,61 (95%-KI [0,59; 0,64]) 4. Übereinstimmung der Studienergebnisse in 8 von 10 Studien (in 5 dieser Studien jeweils kein signifikanter Unterschied)	„Wir schlussfolgern auf der Basis dieser Assoziationen, dass es mäßige Evidenz für eine Surrogatfunktion zwischen Einjahres-PFS und Zweijahres-OS gibt." "Based on these associations, we conclude that there is modest evidence for surrogacy between one-year PFS and two-year OS."	▪ Interpretation der statistischen Ergebnisse Mittlere Korrelation auf individueller Ebene[2], auf Studienebene mittlere Korrelation möglich. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.
Metastasierendes Kolorektalkarzinom			
Johnson 2006	*TR und TTP als Surrogate für OS:* ▪ *Lineare Regression:* TR: R^2 = 0,10 TTP: R^2 = 0,33 ▪ *Surrogate Threshold (für verschiedene Fallzahlen benötigter Behandlungseffekt bez. SEP, um einen Effekt bez. OS zu zeigen):* TR (Responderanteil in %): 38 (250 Patienten), 26 (500 Patienten), 20 (750 Pat.); TTP (in Monaten): 3,2 (250 Patienten), 2,3 (500 Patienten), 1,9 (750 Pat.)	„Studien zum […] kolorektalen Karzinom sollten Überleben als ihren primären Endpunkt messen, es sei denn, es wird erwartet, dass die Differenz der Ergebnisse des Surrogats den STE überschreitet" "trials in […] colorectal cancer should measure survival as their primary outcome unless the surrogate outcome difference is anticipated to exceed the threshold effect size"	▪ Interpretation der statistischen Ergebnisse Wegen fehlender Variabilitätsangaben können die Korrelationen nicht verlässlich beurteilt werden. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes nicht nachgewiesen.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Metastasierendes Kolorektalkarzinom			
Montagnani 2009	*PFS als Surrogat für OS:* ▪ *Studienarmebene:* Kontroll-Arme: $\rho = 0,87$ ($p = 0,0128$) BV-Arme: $\rho = 0,63$ ($p = 0,0709$) ▪ *Studienebene:* $\rho = 0,52$ ($p = 0,1362$)	„die Korrelation zwischen PFS und OS und zwischen ΔPFS und ΔOS für Patienten mit metastasierendem kolorektalem Karzinom, die mit einer BV-basierten Bio-Chemotherapie behandelt wurden, wurde nicht nachgewiesen.“ (“the correlation between PFS and OS and between *ΔPFS and ΔOS* for patients with metastatic colorectal cancer treated with BV-based bio-chemotherapy was not demonstrated.”)	▪ Interpretation der statistischen Ergebnisse Wegen fehlender Variabilitätsangaben können die Korrelationen nicht verlässlich beurteilt werden. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.
Tang 2007	*SEP als Surrogate für OS:* 1. *Studienarmebene:* PFS: $r_s = 0,79$ (95%-KI [0,65; 0,87]) TTP: $r_s = 0,24$ (95%-KI [-0,31;0,55]) TR: $r_s = 0,59$ (95%-KI [0,42; 0,72]) *Studienebene:* PFS: $r_s = 0,74$ (95%-KI [0,47; 0,88]) TTP: $r_s = 0,52$ (95%-KI [0,004; 0,81]) TR: $r_s = 0,39$ (95%-KI [0,08; 0,63]) 2. *Lineare Regression:* $R^2 = 0,65$	„Die Verwendung von PFS als ein Surrogatendpunkt in randomisierten kontrollierten Studien bei First-line-Chemotherapie für MCRC könnte angemessen sein“ „PFS könnte ein angemessenes Surrogat für OS sein.“ (“the usage of PFS as a surrogate endpoint in randomized controlled trials in first-line chemotherapy for MCRC may be appropriate” “PFS may be an appropriate surrogate for OS.”)	▪ Interpretation der statistischen Ergebnisse Korrelationen auf individueller Ebene sind nicht berichtet.[2] Auf Studienebene sind mittlere Korrelationen möglich (PFS, TTP) bzw. liegt niedrige Korrelation vor (TR). ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität aller Surrogatendpunkte unklar.

(Fortsetzung)

Tabelle 14: Ergebnisse der Validierungsstudien – Darmkrebs (Fortsetzung)

Abkürzungen: ACCENT: Adjuvant Colon Cancer End Points Group, ACMD: average clinically meaningful difference, APEP: average prediction error of the predicted effect, APEO: average prediction error of other true endpoints, BV: Bevacizumab, DFS: Disease-free survival, 5-FU: Fluorouracil, FUDR: Floxuridin, HR: Hazard ratio, KI: Konfidenzintervall, MCRC: metastatic colorectal cancer, OS: Overall survival, PFS: Progression-free survival, PTE: proportion of treatment effect, RE: relativer Effekt, RFS: Recurrence-free survival, RR: Response rate, SEP: Surrogatendpunkt, STE: Surrogat Threshold Effect, TTP: Time to progression, TTR: Time to tumor recurrence, TR: Tumor response.

[1] Ergebnisse unter Ausschluss einer Studie, die zu Berechnungsproblemen bei R^2_{trial} führte. Unter Einschluss dieser Studie resultierten auf individueller Ebene R^2_h = 0,84 [0,82; 0,85], auf Studienebene R^2_h = 0,82 [0,40; 0,95].
[2] Für eine Nutzenbewertung des IQWiG wäre der Zusammenhang auf individueller Ebene nicht von ausschlaggebender Bedeutung (siehe auch Abschnitt 6.1 im Kapitel „Diskussion“). Für eine uneingeschränkte Surrogatvaliedierung ist die Korrelation auf individueller Ebene Voraussetzung (siehe Abschnitt 5.1.2.1)
[3] Keine Angabe des Rangkorrelationskoeffizienten auf individueller Ebene.
[4] Diese Teilpopulationen stammen aus 3 Studien mit den Interventionen orale Fluoropyrimidine vs. unbehandelte Kontrollgruppe. Somit sind sie vergleichbar mit den zur Validierung herangezogenen Studien. Neben Patienten mit Kolonkarzinom wurden auch Patienten mit Rektumkarzinom eingeschlossen.
[5] In der Publikation sind keine zusammenfassenden Informationen durch Maßzahlen o. Ä. gegeben, es werden nur Einzelergebnisse bzw. Abbildungen dargestellt.
[6] Beobachtungen werden so zensiert, dass die mediane (bzw. minimale) Nachbeobachtungszeit 1, 2 bzw. 3 Jahre beträgt.
[7] Publikation liefert außerdem weitere Maße zur Evaluierung.
[8] Beobachtungen werden so zensiert, dass die mediane (bzw. minimale) Nachbeobachtungszeit 5 Jahre beträgt.
[9] Gleiche Werte von R^2 und ρ sind widersprüchlich. Möglicherweise wird in der Publikation ein falscher Wert für ρ angegeben.
[10] Gleiche Testentscheidungen des logRank-Tests für SE und OS führen zu Konkordanz.
[11] Aus der Publikation geht nicht hervor, ob ein quadriertes oder nicht quadriertes Maß gemeint ist. Die in der Tabelle vorgenomme Interpretation stellt damit den liberalen Ansatz dar.
[12] Die Angaben sind Abbildungen entnommen und damit nur ungefähr.
[13] Dargestellt sind die Ergebnisse ohne Adjustierung nach Performance-Status. Die Ergebnisse mit Adjustierung sind sehr ähnlich
[14] Weitere Ergebnisse zu Landmark-Zeiten 3, 4 und 5 Monate sind nicht dargestellt. Diese sind jedoch ähnlich zu den Ergebnissen mit 0 und 6 Monaten.
[15] Dargestellt ist vermutlich das Konfidenzintervall, was in der Publikation aber nicht spezifiziert wird.
[16] Gemeint ist R^2.
[17] Ausschluss einer Studie führte zu verändertem Wert R = 0,74 [0,44; 1,04].
[18] Ausschluss einer Studie führte zu verändertem STE = 0,77 bzw. 1,3.
[19] Interventionen in den Validierungsstudien 5-FU/LV ± Irinotecan oder Oxaliplatin. Abweichend von den Interventionen in den zur Validierung herangezogenen Studien.
[20] Irinotecan-EU
[21] Irinotecan-US
[22] Oxaliplatin-EU
[23] Ausschluss einer einflussreichen Studie führte zu R = 0,74 (95%-KI [0,44-1,04]).
[24] Ausschluss einer einflussreichen Studie führte zu STE = 0,77 bzw. 1,30.

5.2.3.2 Brustkrebs

Tabelle 15 zeigt die Ergebnisse und Schlussfolgerungen aus den Studien zur Validierung von Surrogatparametern des Krankheitsbildes Brustkrebs.

Die einzige Studie (Ng 2007) für nicht metastasierenden oder lokal fortgeschrittenen Brustkrebs untersuchte den Surrogatendpunkt DFS. Unabhängig vom gewählten Modell lagen nur Ergebnisse ohne Konfidenzintervall vor, sodass die Validität des Surrogats nicht zu beurteilen war.

In den 3 Studien für den Bereich fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses) wurden die Surrogatendpunkte TR, DC, PFS und TTP betrachtet. TR wurde in 2 Studien untersucht, wovon 1 Studie (Bruzzi 2005) ein Validierungsverfahren nach Prentice (1989) [47] anwandte, bei dem jedoch bereits der Nachweis des zweiten Kriteriums misslang. In der anderen Studie (Burzykowski 2008c) wurde durch korrelationsbasierte Verfahren ein nur mittlerer Zusammenhang mit hoher Variabilität von Surrogat und klinischem Endpunkt berichtet. In beiden Fällen bleibt die Validität von TR als Surrogat für OS damit unklar. Burzykowski 2008c präsentierte auch Ergebnisse zu den Surrogatendpunkten DC und TTP. Hier zeigten sich auf Studienebene ebenfalls nur mittlere Korrelationen bezogen auf die Punktschätzer bei geringer Präzision, sodass auch hier die Validität unklar bleibt. Die gleichen Schlussfolgerungen ergeben sich für PFS aus den Analysen von Burzykowski 2008c und Miksad 2008, über die ebenfalls mittlere Korrelationen und, sofern überhaupt vorhanden, breite Konfidenzintervalle berichtet werden. Die von Miksad durchgeführten Sensitivitätsanalysen zu möglichen Unterschieden bei Anthrazyklinen und Taxanen ließen wegen fehlender Variabilitätsangaben keine eindeutigen Interpretationen zu. Die stets niedrige Aussagesicherheit führte schließlich bei allen Surrogatendpunkten dazu, dass die Validität insgesamt nicht nachgewiesen ist.

Für metastasierenden Brustkrebs untersuchte die Studie Hackshaw 2005 die Surrogatendpunkte TR, PD und TTP. Sherrill 2008 fasste aufgrund uneinheitlicher Definitionen der Endpunkte TTP und PFS in den eingeschlossenen Studien PFS und TTP als einheitliches Surrogat zusammen. Für TR und PD gibt Hackshaw 2005 niedrige Bestimmtheitsmaße an. Wegen fehlender Variabilitätsangaben kann die Validität jedoch nicht genauer beurteilt werden. Für TTP bzw. PFS/TTP werden mittlere Bestimmtheitsmaße berichtet. Weil auch hier Angaben zur Präzision der Schätzer fehlen, bleibt die Validität unklar. Die Studie von Sherrill 2008 zeigte weiterhin im Rahmen einer Subgruppenanalyse bei HER2-positiven Patienten ein hohes Bestimmtheitsmaß von 0,93. Die Unsicherheit dieser Schätzung konnte jedoch wegen fehlender Angaben nicht eingeschätzt werden, sodass auch hier die Validität unklar ist.

Ähnlich wie bei den anderen Studien führte eine niedrige Aussagesicherheit dazu, dass die Validität nicht umfassend bewertet werden kann und somit insgesamt unklar ist. Die niedrige Aussagesicherheit war vor allem dadurch bedingt, dass unterschiedliche Wirkstoffklassen zusammengefasst ausgewertet wurden, ohne dass der Einfluss der unterschiedlichen Interventionen durch entsprechende geeignete Sensitivitätsanalysen untersucht worden wäre.

Übergreifend über die Indikationen, Interventionen und betrachteten Surrogatendpunkte wurden – ohne Berücksichtigung von Subgruppen- bzw. Sensitivitätsanalysen – Korrelationen von höchstens 0,69 (Burzykowski 2008) und Bestimmtheitsmaße von höchstens 0,56 (Miksad 2008) berichtet. Angaben zu zugehörigen Konfidenzintervalle fanden sich nur teilweise. Damit wurde für keinen der untersuchten Surrogatendpunkte ein hohes Maß des Zusammenhangs zwischen Surrogat und patientenrelevantem Endpunkt berichtet. Die Validität der Surrogate blieb unklar.

Tabelle 15: Ergebnisse der Validierungsstudien – Brustkrebs

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Nicht metastasierender oder lokal fortgeschrittener Brustkrebs			
Ng 2007	*DFS als Surrogat für OS:* 1. Basis-Modell: a) $r = 0,62$; $R^2 = 0,38$ b) $r = 0,61$; $R^2 = 0,37$ 2. Vollständiges Modell: a) $r = 0,65$; $R^2 = 0,42$; b) $r = 0,63$; $R^2 = 0,39$	„es gibt eine statistisch signifikante Korrelation von moderater Stärke zwischen Unterschieden im 2-Jahres-DFS zwischen Behandlungsarmen sowie Unterschieden im 5-Jahres-OS, aufgrund der breiten Vorhersageintervalle ist aber die Korrelation für die Verwendung als Prädiktor nicht hinreichend stark.“ (“there is a statistically significant correlation, of moderate strength, between differences in 2-year DFS between treatment arms and differences in 5-year OS, but due to the wide prediction intervals, the correlation is not sufficiently strong enough to be used as a predictor.”)	▪ Interpretation der statistischen Ergebnisse Wegen fehlender Variabilitätsangaben können die Korrelationen nicht verlässlich beurteilt werden. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 15: Ergebnisse der Validierungsstudien – Brustkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses)			
Bruzzi 2005	*TR als Surrogat für OS:* ▪ *Behandlungseffekt auf SEP[1]:* OR 0,60 (95%-KI [0,51; 0,72]); p < 0,0001 ▪ *Behandlungseffekt auf OS:* OR 0,94 (95%-KI [0,86; 1,04]); p = 0,22 ▪ *SEP-Effekt auf OS:* SEP partial response: OR 0,69 (95%-KI [0,62; 0,77]) SEP complete response: OR 0,48 (95%-KI [0,40; 0,57]) in beiden Fällen p < 0.0001 (Cox-Modell mit Tumoransprechen als zeitabhängiger Kovariable) ▪ *Behandlungseffekt auf OS adjustiert für SEP* HR 1,005 (95%-KI [0,91; 1,11]); p = 0,92 (Cox-Modell)	„Diese Ergebnisse liefern keinen direkten und/oder beweiskräftigen Beleg bezüglich der allgemeinen Validität des objektiven Ansprechens auf eine Chemotherapie als ein Surrogat für das Überleben bei fortgeschrittenem Brustkrebs.“ „es ist nicht möglich zu schlussfolgern, dass das objektive Ansprechen auf eine der First-line-Chemotherapien […] mit einem Überlebensvorteil assoziiert ist“ „Die potenzielle Rolle des objektiven Ansprechens als ein Surrogatendpunkt für das Überleben in Chemotherapie-Studien zu metastatischem Brustkrebs rechtfertigt weitere Untersuchungen“ (“These results do not provide direct and/or conclusive evidence on the general validity of objective response to chemotherapy as a surrogate of survival in advanced breast cancer.” “it is not possible to conclude that objective response to any first-line chemotherapy […] is associated with a survival benefit” “The potential role of objective response as a surrogate end point for survival in chemotherapy trials of metastatic breast cancer warrants further investigation”)	▪ Interpretation der statistischen Ergebnisse Die Prentice-Kriterien wurden nicht erfüllt (Validitätskriterium nicht erfüllt) ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 15: Ergebnisse der Validierungsstudien – Brustkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses)			
Burzykowski 2008c	*SEP als Surrogate für OS:* ▪ *Individuelle Ebene:* TR: OR = 2,9 (95%-KI [2,6; 3,3])[2] DC: OR = 5,5 (95%-KI [4,8; 6,3])[3] PFS: Rangkorrelationskoeffizient = 0,688 (95%-KI [0,686; 0,690]) TTP: Rangkorrelationskoeffizient = 0,682 (95%-KI [0,680; 0,684]) ▪ *Studienebene:* TR: $\rho = 0{,}57$[4] (95%-KI [-0,31; 1,44]) DC: $\rho = 0{,}47$[5] (95%-KI [-0,46; 1,40]) PFS: Rangkorrelationskoeffizient = 0,48 (95%-KI [-0,34; 1,30]) TTP: Rangkorrelationskoeffizient = 0,49 (95%-KI [-0,32; 1,30])	„[PFS und Überleben:] eine Null-Korrelation konnte nicht ausgeschlossen werden“ „die Daten bieten keinen Beleg für die Behauptung […], dass Tumoransprechen ein Surrogat für Überleben sein könnte“ „die Ergebnisse und Schlussfolgerungen für die Krankheitskontrolle und für TTP sind jenen ähnlich, die jeweils für Tumoransprechen und PFS erzielt wurden“ „in diesen Studien konnte kein Endpunkt als ein gutes Surrogat für OS aufgezeigt werden“ (“[PFS and survival:] a zero correlation could not be excluded” “data offer no evidence for [the] claim […] that tumour response could be a surrogate for survival” “the results and conclusions for disease control and for TTP are similar to those obtained for tumour response and PFS, respectively” “No end point could be demonstrated as a good surrogate for OS in these trials.”)	▪ Interpretation der statistischen Ergebnisse Auf individueller Ebene liegen niedrige Korrelationen (PFS, TTP) vor.[6] Für die Beurteilung von TR und DC fehlten Vergleichsmaßstäbe. Auf Studienebene sind mittlere Korrelationen möglich. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität aller Surrogatendpunkte unklar.

(Fortsetzung)

Tabelle 15: Ergebnisse der Validierungsstudien – Brustkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses)			
Miksad 2008	*PFS als Surrogat für OS:* 1.*Übereinstimmung (Kappa) der Richtung der Behandlungseffekte[7]:* Anthrazykline: 0,71 (95%-KI [0,36; 1,00]); p = 0,0029 Taxane: 0,75 (95%-KI [0,42; 1,00]); p = 0,0029 2.*Meta-analytischer Ansatz auf Studienebene:* ▪ *Bestimmtheitsmaß:* Anthrazykline: R^2 = 0,49 Taxane: R^2 = 0,35 ▪ *Kreuzvalidierung (Anteil der Studien, für die unter Ausschluss aus der Modellierung das beobachtete OS im Vorhersageintervall lag):* Anthrazykline: 16/16 Taxane: 15/17 ▪ *Sensitivitätsanalysen:* a) Anthrazykline: R^2 = 0,56 Taxane: R^2 = 0,25 b) Anthrazykline: R^2 = 0,46 Taxane: R^2 = 0,59 ▪ *Subgruppenanalysen[8]::* Anthracycline: Jahr der Rekrutierung des letzten Patienten[9]: p = 0,44 Taxane: First-line-Therapie vs. Second-line-Therapie: p = 0,60	„Diese meta-analytische, studienbasierte Analyse von Anthrazyklin- und Taxan-basierter Chemotherapie für fortgeschrittenen Brustkrebs suggeriert, dass, während der Behandlungseffekt auf Studienebene bez. PFS mit dem Behandlungseffekt auf Studienebene bez. OS signifikant assoziiert ist, Vorhersagen, die auf PFS basieren, für diese Behandlungen bei fortgeschrittenem Brustkrebs unsicher sind. Die Verwendung von standardisierten Endpunkt-Definitionen könnte jedoch die Reliabilität und Validität von Daten zu Surrogatendpunkten bei fortgeschrittenem Brustkrebs steigern.“ „PFS könnte ein angemessener Endpunkt sein, auch wenn es kein formaler Surrogatendpunkt ist mit der Fähigkeit, OS vorherzusagen“ (“This meta-analytic, trial-based analysis of anthracycline- and taxane-based chemotherapy for advanced breast cancer suggests that while the trial-level treatment effect on PFS is significantly associated with the trial-level treatment effect on OS, predictions based on trial-level PFS are surrounded with uncertainty for these treatments in advanced breast cancer. However, use of standardized endpoint definitions may increase the reliability and validity of surrogate endpoint data in advanced breast cancer.” “PFS may be an appropriate endpoint even if is not a formal surrogate endpoint with an ability to predict OS”)	▪ Interpretation der statistischen Ergebnisse Wegen fehlender Variabilitätsangaben können die Korrelationen nicht verlässlich beurteilt werden. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

Tabelle 15: Ergebnisse der Validierungsstudien – Brustkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Metastasierender Brustkrebs			
Hackshaw 2005	*TR (vollständig oder teilweise) als Surrogat für OS*[10]*:* ▪ $p < 0.0001$, $R^2 = 34\%$ *TR (vollständig) als Surrogat für OS:* ▪ $p = 0.02$, $R^2 = 12\%$, *PD als Surrogat für OS:* ▪ $p < 0{,}0001$, $R^2 = 38\%$ *TTP als Surrogat für OS:* ▪ $p < 0{,}001$, $R^2 = 56\%$ ▪ ($p = 0{,}017$, $R^2 = 58\%$)[11]	„Die Zeit bis zur Progression könnte ein nützlicher Surrogat-Marker für die Vorhersage des Überlebens bei Frauen sein, die eine First-line-Chemotherapie mit Anthrazyklinen erhalten, und könnte in zukünftigen Studien verwendet werden, um den Überlebensvorteil der First-line-Chemotherapie, verglichen mit FAC oder FEC, abzuschätzen. Die anderen Marker, Tumor-anprechen und progressive Erkrankung, waren weniger gut.“ (“Time to progression may be a useful surrogate marker for predicting survival in women receiving first-line anthracycline chemotherapy and could be used to estimate the survival benefit in future trials of first-line chemotherapy compared to FAC or FEC. The other markers, tumour response and progressive disease, were less good.”)	▪ Interpretation der statistischen Ergebnisse Wegen fehlender Variabilitätsangaben können die Korrelationen nicht verlässlich beurteilt werden. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität des Surrogatendpunktes unklar.

(Fortsetzung)

Tabelle 15: Ergebnisse der Validierungsstudien – Brustkrebs (Fortsetzung)

Studie	Ergebnisse	Schlussfolgerung der Autoren hinsichtlich der inhaltlichen Aussagen	Zusammenfassung und Interpretation der Ergebnisse
Metastasierender Brustkrebs			
Sherrill 2008	*TTP/PFS als Surrogat für OS:* ▪ $R^2 = 0,30$ ▪ *Subgruppenanalysen:* Studien mit hormoneller Therapie: $R^2 = 0,24$ Anthracyclinen: $R^2 = 0,43$ First-line-Behandlungen: $R^2 = 0,28$ Non-First-line-Behandlungen $R^2 = 0,32$ nur HER2+-Patienten: $R^2 = 0,93$ ▪ *Sensitivitätsanalysen* Studien, in denen HRs berichtet wurden: $R^2 = 0,52$ TTP $\geq$ 6 Monate in der Kontrollgruppe: $R^2 = 0,35$ Studien mit mehr als 100 Patienten pro Gruppe: $R^2 = 0,31$	„Wir glauben nicht, dass diese Ergebnisse ausreichend sind, um TTP als vollwertigen Surrogatendpunkt für das Gesamtüberleben bei metastasierendem Brustkrebs zu akzeptieren.“ (“We do not believe that the current findings are sufficient for accepting time-to-progression as a fully validated surrogate marker for overall survival in MBC.”)	▪ Interpretation der statistischen Ergebnisse Wegen fehlender Variabilitätsangaben können die Korrelationen nicht verlässlich beurteilt werden. ▪ Aussagesicherheit Die Aussagesicherheit wird als niedrig bewertet. ▪ Gesamtinterpretation Insgesamt ist die Validität der Surrogatendpunkte unklar.
Abkürzungen: DC: Disease control, DFS: Disease-free survival, FAC: Fluorouracil, Adriamycin, Cyclophosphamid, FEC: Fluorouracil, Epirubicin, Cyclophosphamid, MBC: metastatic breast cancer, HER 2+: human epidermal growth factor receptor-2, HR: Hazard ratio, KI: Konfidenzintervall, OR: Odds ratio, OS: Overall survival, PD: Progressive disease, PFS: Progression-free survival, SEP: Surrogatendpunkt, TR: Tumor response, TTP: Time to tumor progression.			

[1] Unklar, ob SE partial response oder SE complete response gemeint ist.
[2] Untere Grenze des KI für alle Landmark-Zeiten > 1,6.
[3] Untere Grenze des KI für alle Landmark-Zeiten > 2,3.
[4] Bei Verwendung verschiedener Landmark-Zeiten variieren die Werte des Punktschätzers zwischen 0,54 und 0,63.
[5] Bei Verwendung verschiedener Landmark-Zeiten variieren die Werte des Punktschätzers zwischen 0,39 und 0,55.
[6] Für eine Nutzenbewertung des IQWiG wäre der Zusammenhang auf individueller Ebene nicht von ausschlaggebender Bedeutung (siehe auch Abschnitt 6.1 im Kapitel „Diskussion“). Für eine uneingeschränkte Surrogatvaliedierung ist die Korrelation auf individueller Ebene Voraussetzung (siehe Abschnitt 5.1.2.1)
[7] p-Werte für den Test der Nullhypothese, dass keine Übereinstimmung existiert.
[8] p-Wert für die jeweilige Kovariable im Meta-Regressionsmodell.

(Fortsetzung)

Tabelle 15: Ergebnisse der Validierungsstudien – Brustkrebs (Fortsetzung)

[9] Als Proxy für den Einfluss der nach dem Studienregime gegebenen Behandlungen.
[10] p-Werte beziehen sich im Folgenden auf die Hypothese, dass im Regressionsmodell kein Behandlungseffekt durch SE existiert.
[11] Ergebnis für die 9 Studien, in denen Tod als Ereignis bzgl. TTP angesehen wurde.

5.2.4 Sensitivitätsanalyse

Sensitivitätsanalysen, die innerhalb einzelner Validierungsstudien vorgenommen wurden, wurden bei der Interpretation der Ergebnisse berücksichtigt (siehe Abschnitt 5.2.3). Darüber hinaus waren auf Basis der verfügbaren Informationen keine Sensitivitätsanalysen hinsichtlich methodischer Faktoren möglich.

5.2.5 Faktoren, die einen Einfluss auf die Validierung haben können

Eine Untersuchung von Faktoren, die einen Einfluss auf die Validierung eines Surrogats haben könnten, war in fast allen Fällen nicht möglich, da entsprechende Informationen nicht zur Verfügung standen. Insbesondere waren die Daten für eine sinnvolle Untersuchung des Einflusses von Charakteristika der Tumorerkrankung bzw. der therapeutischen Intervention nicht ausreichend. Die einzigen Ausnahmen sind die Studien Miksad 2008 und Sherrill 2008 für das Krankheitsbild Brustkrebs: In Miksad 2008 ist der Einfluss der Kovariablen Rekrutierungsjahr und Therapieart (First-line vs. Second-line) in adjustierten Analysen jeweils nicht signifikant. Sherrill führte getrennte Analysen durch für Patienten, die mit First-line- bzw. Second-line-Therapien behandelt wurden. Die Ergebnisse erlauben keine eindeutigen Schlüsse (die Korrelationen betragen etwa 0,30, es fehlen aber Konfidenzintervalle).

5.2.6 Zusammenfassung der Beleglage

Eine zusammenfassende Übersicht über die Beleglage zur Validität der verschiedenen Surrogatendpunkte in Studien zu Darm- und Brustkrebs ist in Tabelle 16 und Tabelle 17 dargestellt.

Auf Grundlage der eingeschlossenen Validierungsstudien wurde die Validität der Surrogatendpunkte sowohl für Brustkrebs als auch für Darmkrebs in jedem Fall als insgesamt unklar bewertet. In den meisten Fällen waren die Korrelationen der jeweiligen Effekte nicht ausreichend hoch, sodass damit die notwendige Bedingung für den Nachweis der Validität nicht erfüllt war. In allen Studien wurde außerdem die Aussagesicherheit als niedrig bewertet. Daher blieb die Validität bei allen untersuchten Surrogatendpunkten unabhängig von den statistischen Ergebnissen letztlich unklar.

Tabelle 16: Landkarte der Beleglage – Darmkrebs

Publikation	Intervention	Surrogat	Patientenrelevanter Endpunkt	Ergebnis der Autoren	Interpretation der statistischen Ergebnisse[1]	Aussage-sicherheit	Gesamt-interpretation
Kolonkarzinom im frühen Krankheitsstadium							
Alonso 2008a	unklar	TTR	OS	↗	Korrelation: hoch / mittel	niedrig	Validität unklar
Buyse 2008	unklar	DFS	OS	↑	Korrelation: hoch / hoch	niedrig	Validität unklar
Green 2008	unklar	DFS	OS	↑	Korrelation: hoch / hoch	niedrig	Validität unklar
Nicht metastasierendes Kolonkarzinom							
Burzykowski 2008b	Unterschiedliche Chemotherapien; 5-FU in mindestens einem Studienarm	DFS 1-Jahres-DFS[2] 2-Jahres-DFS[2] 3-Jahres-DFS[2]	5-Jahres-OS[2] 5-Jahres-OS[2] 5-Jahres-OS[2] 5-Jahres-OS[2]	? ? ? ↗	Korrelationen: n. b. / hoch n. b. / mittel n. b. / hoch n. b. / hoch	niedrig niedrig niedrig niedrig	Validität in allen Fällen unklar
Sargent 2005	Unterschiedliche Chemotherapien; 5-FU in mindestens einem Studienarm	Median 3-Jahres-DFS[3]	Median 5-Jahres-OS[3]	↗	Korrelation: hoch / mittel	niedrig	Validität unklar

(Fortsetzung)

Tabelle 16: Landkarte der Beleglage – Darmkrebs (Fortsetzung)

Publikation	Intervention	Surrogat	Patientenrelevanter Endpunkt	Ergebnis der Autoren	Interpretation der statistischen Ergebnisse[1]	Aussage-sicherheit	Gesamt-interpretation
Nicht metastasierendes Kolonkarzinom							
Sargent 2007	Unterschiedliche Chemotherapien; 5-FU in mindestens einem Studienarm				Korrelationen:		
		Median-3-Jahres-DFS (Stadium II)	Median 5-Jahres-OS[3]	nicht verfügbar	n. b. / mittel	niedrig	Validität in allen Fällen unklar
		Median-3-Jahres-DFS (Stadium III)	Median 5-Jahres-OS[3]	↑	n. b. / hoch	niedrig	
		Minimum 1-Jahres-DFS[3]	Median 5-Jahres-OS[3]	↗[4]	n. b. / mittel	niedrig	
		Median 2-Jahres-DFS[3]	Median 5-Jahres-OS[3]	↗[4]	n. b. / mittel	niedrig	
		Median 3-Jahres-DFS[3]	Median 5-Jahres-OS[3]	↑[5]	n. b. / mittel	niedrig	
		Minimum 3-Jahres-DFS[3]	Median 5-Jahres-OS[3]	nicht verfügbar	n. b. / mittel	niedrig	
Sertdemir 2009	Unterschiedliche Therapien (OP, 5-FU-enthaltende Therapien, Immuntherapien)	DFS	OS	?	Korrelation: n. b. / mittel	niedrig	Validität unklar

(Fortsetzung)

Tabelle 16: Landkarte der Beleglage – Darmkrebs (Fortsetzung)

Publikation	Intervention	Surrogat	Patientenrelvanter Endpunkt	Ergebnis der Autoren	Interpretation der statistischen Ergebnisse[1]	Aussage-sicherheit	Gesamt-interpretation
Fortgeschrittenes Kolorektalkarzinom (inkl. metastasierenden Kolorektalkarzinoms)							
Alonso 2008a	5-FU (± INF-α) vs. 5-FU/LV (± INF-α)	TTR	OS	↗	Korrelation: mittel / mittel	niedrig	Validität unklar
Baker 2006	Fluoropyrimidin Monotherapie vs. Fluoropyrimidin Mono- oder Kombinationstherapie	TR	Überleben	↓	APEP < APEO: nicht erfüllt	niedrig	Validität unklar
Burzykowski 2004	Fluoropyrimidin Monotherapie vs. Fluoropyrimidin Mono- oder Kombinationstherapie	TR binär: TR ordinal	OS OS	↘[6] ↘[6]	Korrelationen: n. b. / mittel n. b. / niedrig	niedrig niedrig	Validität in allen Fällen unklar
Buyse 2000b	Fluoropyrimidin Monotherapie vs. Fluoropyrimidin Mono- oder Kombinationstherapie	TR	OS	↘	Korrelation: n. b. / mittel	niedrig	Validität unklar
Buyse 2007	5-FU/LV vs. 5-FU oder Raltitrexed	PFS	OS	↗	Korrelation: mittel / hoch	niedrig	Validität unklar
Buyse 2008	5-FU/LV vs. 5-FU oder Raltitrexed	PFS	OS	↑	Korrelation: mittel / hoch	niedrig	Validität unklar
Cortiñas Abrahantes 2008	unklar	TR	OS	nicht verfügbar	Korrelation: n. b. / mittel	niedrig	Validität unklar
Green 2008	unklar	PFS	OS	?	Korrelation: mittel / mittel	niedrig	Validität unklar

(Fortsetzung)

Tabelle 16: Landkarte der Beleglage – Darmkrebs (Fortsetzung)

Publikation	Intervention	Surrogat	Patientenrelvanter Endpunkt	Ergebnis der Autoren	Interpretation der statistischen Ergebnisse[1]	Aussage-sicherheit	Gesamt-interpretation
Metastasierendes Kolorektalkarzinom							
Johnson 2006	Unterschiedliche Chemotherapien	TR TTP	OS OS	nicht verfügbar nicht verfügbar	Korrelationen: n. b. / n. b. n. b. / n. b.	niedrig niedrig	Validität in allen Fällen unklar
Montagnani 2009	Fluoropyrimidinbasierte Mono- oder Kombinationstherapien mit Bevacizumab vs. Kontrolltherapien ohne Bevacizumab	PFS	OS	↓	Korrelation: n. b.	niedrig	Validität unklar
Tang 2007	Unterschiedliche Chemotherapien	TR TTP PFS	OS OS OS	nicht verfügbar nicht verfügbar ↗	Korrelationen: n. b. / niedrig n. b. / mittel n. b. / mittel	niedrig niedrig niedrig	Validität in allen Fällen unklar

Symboldefinitionen: ↑: Nachweis der Validität des Surrogatendpunktes, ↗: Anzeichen für die Validität des Surrogatendpunktes, ?: Validität unklar, ↘: Anzeichen für die Nicht-Validität des Surrogatendpunktes, ↓: Nachweis der Nicht-Validität des Surrogatendpunktes

DFS: Disease-free survival, 5-FU: Fluorouracil, INF-α: Interferon-alpha, LV: Leucovorin, PFS: Progression-free survival, TR: Tumor response, TTP: Time to tumor progression, TTR: Time to tumor recurrence, OP: Operation, OS: Overall survival, vs.: versus. n. b.: nicht beurteilbar.

[1] Bei Korrelationen bezieht sich die erste Angabe auf die individuelle Korrelation, die zweite auf die Korrelation auf Studienebene. Für eine Nutzenbewertung des IQWiG wäre der Zusammenhang auf individueller Ebene nicht von ausschlaggebender Bedeutung (siehe auch Abschnitt 6.1 im Kapitel „Diskussion“). Für eine uneingeschränkte Surrogatvaliedierung ist die Korrelation auf individueller Ebene Voraussetzung (siehe Abschnitt 5.1.2.1).
[2] Beobachtungen werden nach 1, 2, 3 bzw. 5 Jahren zensiert
[3] Beobachtungen werden so zensiert, dass die mediane (bzw. minimale) Nachbeobachtungszeit 1, 2, 3 bzw. 5 Jahre beträgt.
[4] Bewertung bezieht sich auf Ergebnisse einer vereinigten Population (Stadien II/III)

(Fortsetzung)

Tabelle 16: Landkarte der Beleglage – Darmkrebs (Fortsetzung)

[5] Bewertung bezieht sich auf Ergebnisse einer vereinigten Population (Stadien II/III) sowie Ergebnisse für die Population in Stadium III. Für die Ergebnisse der Population in Stadium II liefern die Autoren keine Interpretation
[6] Autoren bewerten die Ergebnisse zusammenfassend ohne nach dem Messniveau zu unterscheiden

Tabelle 17: Landkarte der Beleglage – Brustkrebs

Publikation	Intervention	Surrogat	Patientenrelevanter Endpunkt	Ergebnis der Autoren	Interpretation der statistischen Ergebnisse[a]	Aussage-sicherheit	Gesamt-interpretation
Nicht metastasierender oder lokal fortgeschrittener Brustkrebs							
Ng 2007	Unterschiedliche Chemo-, Hormon- und/oder Immuntherapien in beiden Studienarmen	DFS	OS	?	Korrelation: n. b. / n. b.	niedrig	Validität unklar
Fortgeschrittener Brustkrebs (inkl. metastasierenden Brustkrebses)							
Bruzzi 2005	Intensivierte Epirubicin-enthaltende Chemotherapien vs. Epirubicin-enthaltende Standardchemotherapien	TR	OS	?	Prentice-Kriterien: nicht erfüllt	niedrig	Validität unklar
Burzykowski 2008c	Anthrazyklin-Mono- oder -Kombinationschemotherapien vs. Taxan-Mono- oder Anthrazyklin + Taxan-Kombinationschemotherapien	TR DC PFS TTP	OS OS OS OS	? ? ? ?	Korrelationen: n. b. / mittel n. b. / mittel niedrig /mittel niedrig / mittel	niedrig niedrig niedrig niedrig	Validität in allen Fällen unklar
Miksad 2008	Anthrazyklin- oder Taxan-Kombinationschemotherapien vs. unklare Vergleichstherapien	PFS	OS	?	Korrelation: n. b. / n. b.	niedrig	Validität unklar

(Fortsetzung)

Tabelle 17: Landkarte der Beleglage – Brustkrebs (Fortsetzung)

Publikation	Intervention	Surrogat	Patientenrelevanter Endpunkt	Ergebnis der Autoren	Interpretation der statistischen Ergebnisse[a]	Aussage-sicherheit	Gesamt-interpretation
Metastasierender Brustkrebs							
Hackshaw 2005	Anthrazyklin-Kombinationschemotherapien (mit und ohne Hormontherapie) vs. Kombinationstherapien (mit und ohne Anthrazyklin)	TTP TR PD	OS OS OS	↗ ? ?	Korrelationen: n. b. / n. b. n. b. / n. b. n. b. / n. b.	niedrig niedrig niedrig	Validität in allen Fällen unklar
Sherrill 2008	Unterschiedliche Chemo-, Hormon-, Biologika- oder sonstige Therapien, als Mono- oder Kombinationstherapie in beiden Studienarmen	TTP/PFS	OS	?	Korrelation: n. b. / n. b.	niedrig	Validität unklar

a: Bei Korrelationen bezieht sich die erste Angabe auf die individuelle Korrelation, die zweite auf die Korrelation auf Studienebene. Für eine Nutzenbewertung des IQWiG wäre der Zusammenhang auf individueller Ebene nicht von ausschlaggebender Bedeutung (siehe auch Abschnitt 6.1 im Kapitel „Diskussion“). Für eine uneingeschränkte Surrogatvaliedierung ist die Korrelation auf individueller Ebene Voraussetzung (siehe Abschnitt 5.1.2.1)

Symboldefinitionen: ↑: Nachweis der Validität des Surrogatendpunktes, ↗: Anzeichen für die Validität des Surrogatendpunktes, ?: Validität unklar, ↘: Anzeichen für die Nichtvalidität des Surrogatendpunktes, ↓: Nachweis der Nichtvalidität des Surrogatendpunktes

DC: Disease control, DFS: Disease-free survival, OS: Overall survival, PD: Progressive disease, PFS: Progression-free survival, TR: Tumor response, TTP: Time to tumor progression, vs.: versus. n. b.: nicht beurteilbar.

6 Diskussion

Ziel des vorliegenden Projekts war es, Empfehlungen für die Validierung von Surrogatendpunkten in der Onkologie darzustellen. Hierzu wurden zwei Teilziele bearbeitet, 1) die Darstellung und Bewertung methodischer Verfahren, die zur Validierung von Surrogatendpunkten eingesetzt werden (zur Diskussion der Ergebnisse siehe Abschnitt 6.1), sowie 2) die Darstellung und Bewertung von Studien zur Validierung von Parametern des Tumoransprechens als Surrogate für patientenrelevante Endpunkte, insbesondere für das Überleben von Patienten. Die Bearbeitung des zweiten Teilziels wurde beispielhaft für die Krankheitsbilder Darmkrebs und Brustkrebs vorgenommen (zur Diskussion der Ergebnisse siehe Abschnitt 6.2).

Die Ergebnisse des Projekts sollen dem Gemeinsamen Bundesausschuss als Diskussionsgrundlage für die Verwendung von Surrogatendpunkten bei der Nutzenbewertung von onkologischen Behandlungen dienen.

Bei der Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung sind 2 Ebenen zu trennen. Auf der ersten Ebene muss zunächst die Validität des Surrogats beurteilt werden. Auf der zweiten Ebene stellt sich die Frage, welche Schlussfolgerungen bezüglich patientenrelevanter Endpunkte aus den Surrogateffekten abhängig von der Ausprägung der Validität des Surrogats gezogen werden können.

Ausgehend von den Ergebnissen zu den zwei Teilzielen des Projekts werden im Folgenden das Vorgehen zur Validierung von Surrogatendpunkten sowie mögliche Schlussfolgerungen zur Validität diskutiert (Ebene 1, Abschnitt 6.1). Darüber hinaus wird ein Algorithmus für den Umgang mit Surrogatendpunkten in der Nutzenbewertung entwickelt, der die möglicherweise unterschiedlich ausgeprägte Validität von Surrogatendpunkten berücksichtigt (Ebene 2, Abschnitt 6.3).

6.1 Ergebnis des Methodenreviews: mögliches Vorgehen zur Validierung von Surrogatendpunkten

Aus den vorhandenen Übersichtsarbeiten sowie den in der Literaturrecherche gefundenen methodischen Arbeiten zur Validierung von Surrogaten lassen sich folgende grundsätzliche Aussagen bzw. folgendes Vorgehen zur Validierung von Surrogatendpunkten ableiten:

1. Der Nachweis einer Korrelation zwischen Surrogat und klinischem Endpunkt reicht alleine zur Surrogatvalidierung nicht aus. Dies wurde bereits von Fleming 1996 erläutert [96].

2. Die Übertragbarkeit von Aussagen zur Validität von Surrogaten zwischen verschiedenen Erkrankungen oder Ausprägungen einer Erkrankung bzw. zwischen verschiedenen Interventionen ist nicht ohne Weiteres möglich. So können z. B. unterschiedliche Prognosen der Erkrankung oder unterschiedliche Wirkmechanismen von Interventionen den Zusammenhang zwischen Surrogat und patientenrelevantem Endpunkt beeinflussen.

Die Validität eines Surrogats kann deshalb sowohl krankheits- als auch interventionsspezifisch sein, d. h., eine abschließende Surrogatvalidierung mithilfe statistischer Methoden kann ggf. nur innerhalb einer Indikation und einer Intervention erfolgen. In welchem Ausmaß Aussagen zur Validität von Surrogaten zwischen verschiedenen Erkrankungen bzw. Interventionen übertragen werden können, muss geprüft und begründet werden. Der Einschluss von Studien mit unterschiedlichen Erkrankungen oder Interventionen in eine Validierungsstudie hat ggf. einen Einfluss auf die Sicherheit der Aussage zur Validität des Surrogats. Zur Beurteilung der Übertragbarkeit sollten bei Validierungsstudien, die verschiedene Krankheitsentitäten oder Interventionen einschließen, mindestens Untersuchungen zur Heterogenität vorliegen.

In den eingeschlossenen methodischen Arbeiten wird die Indikations- und Interventionsspezifität der Surrogatvalidierung unterschiedlich stark betont. Einige Autoren vertreten die Meinung, dass eine Extrapolation von Validierungsergebnissen auf andere Patientenpopulationen bzw. Interventionen eher nicht angemessen ist bzw. die Validität von Surrogaten für jeden Wirkstoff separat geprüft werden sollte [6,41]. Andere Autoren schlagen vor, ggf. Validierungen innerhalb einer Wirkstoffklasse zu akzeptieren, wenn diese Wirkstoffklasse begründet abgegrenzt ist, also das gleiche biologische Wirkprinzip besitzt [4,7,8].

3. Zur Validierung eines Surrogats werden umfangreiche Daten benötigt, vorzugsweise eine Meta-Analyse mit individuellen Patientendaten von mehreren randomisierten Studien mit ausreichender Ergebnissicherheit. In diesen Studien müssen sowohl das Surrogat als auch der klinische Endpunkt erhoben worden sein.

4. Zur Surrogatvalidierung gibt es weder ein universell anzuwendendes Maß noch eine allgemein beste Schätzmethode noch eine allgemein akzeptierte Grenze, deren Überschreitung den Nachweis der Validität bedeuten würde.

5. Zurzeit werden in der Praxis primär korrelationsbasierte Verfahren zur Surrogatvalidierung mit Schätzung von Korrelationsmaßen auf Studien- und individueller Ebene angewendet. In der Literatur wird auf Studienebene für den Nachweis der Validität in der Regel eine „hohe“ Korrelation zwischen den Effekten für das Surrogat und den patientenrelevanten Endpunkt verlangt, verschiedene Autoren legen eine Korrelation von 0,9 als möglichen Grenzwert nahe (Burzykowski 2005, Lassere 2008). Ebenso besteht Einigkeit, dass niedrige Korrelationen unvereinbar mit der Validität eines Surrogatendpunkts sind, hier werden Werte von unter 0,7 vorgeschlagen (Piedbois 2008, Buyse 2009). Bei Korrelationen zwischen 0,9 und 0,7 kann die Validität des Surrogats damit als unklar eingestuft werden. Bei der Beurteilung der Korrelation kommen jeweils die Konfidenzintervalle der geschätzten Korrelationskoeffizienten zum Einsatz.

   Insbesondere wenn unter Berücksichtigung der zugehörigen Konfidenzintervalle keine hohe Korrelation vorliegt (untere Konfidenzgrenze der geschätzten Korrelationskoeffizienten < 0,85), die Validität des Surrogats also unklar bleibt, können

durch Anwendung des STE-Konzepts nach Burzykowski & Buyse (2005, 2006) bei hinreichend großen Effekten für das Surrogat noch Aussagen bezüglich patientenrelevanter Endpunkte getroffen werden. Das STE-Konzept dient dabei zur Festlegung von Schwellenwerten für den Entscheidungsprozess, ob ein beobachteter Effekt auf das Surrogat mit ausreichender Sicherheit auch mit einem Effekt auf den interessierenden Endpunkt einhergeht [49,50]. Hierzu muss die untere Konfidenzgrenze des Behandlungseffekts bezüglich des Surrogats größer sein als der STE.

Die Anwendung von Methoden, die nicht zum Bereich der korrelationsbasierten Verfahren gehören, bedarf einer ausreichenden Begründung, insbesondere dann, wenn als Datengrundlage nur eine Studie verwendet werden soll.

Besondere Validierungsanforderungen im Rahmen von Nutzenbewertungen

In der Nutzenbewertung gemäß den Anforderungen der evidenzbasierten Medizin und des IQWiG [Methodenpapier S. 8] werden Aussagen zum Nutzen bezogen auf eine Gruppe von Patienten formuliert. Aussagen für individuelle Patienten werden in der Regel nicht getroffen. Verfahrenselemente, mit denen die Surrogatvalidität für einen individuellen Patienten bestimmt werden kann, die aber keine Rückschlüsse auf Populationsebene erlauben, sind daher für die Nutzenbewertung nicht von Bedeutung. Aus diesem Grund kann die Komponente der individuellen Korrelation von Surrogat und patientenrelevantem Endpunkt, die Teil korrelationsbasierter Validierungsverfahren ist, hinsichtlich der Frage der Übertragbarkeit des Surrogateffekts auf den Effekt für den patientenrelevanten Endpunkt unberücksichtigt bleiben. Um Aussagen über den Nutzen hinsichtlich eines patientenrelevanten Endpunkts auf Populationsebene zu ermöglichen, genügt es, wenn die Validität des diesen Endpunkt ersetzenden Surrogats durch ausreichend hohe Korrelation der Behandlungseffekte für die Endpunkte nachgewiesen wurde.

Differenzierte abschließende Beurteilung der Validität eines Surrogats

Die Ergebnisse des Methodenreviews zeigen, dass Aussagen zur Validität eines Surrogats durch eine Reihe von Faktoren beeinflusst werden. Abhängig von ihren Charakteristika hat eine Validierungsstudie eine unterschiedlich ausgeprägte Aussagesicherheit, die bei der abschließenden Aussage zur Validität des Surrogats berücksichtigt werden muss (Ebene 1 der Verwendung von Surrogaten in der Nutzenbewertung, s. u.). Aus unterschiedlichen Konstellationen kann sich ergeben, dass die Validität des Surrogats unklar bleibt. Auch in dieser Situation sind differenzierte Schlussfolgerungen bezüglich patientenrelevanter Endpunkte aus Surrogateffekten möglich (Ebene 2 der Verwendung von Surrogaten in der Nutzenbewertung, siehe Abschnitt 6.3).

Die weitere Diskussion beschränkt sich auf korrelationsbasierte Validierungsverfahren, da diese in den identifizierten Validierungspublikationen fast ausnahmslos verwendet wurden und daher als derzeitiger Standard angesehen werden können.

Abbildung 4 fasst die Überlegungen zum Vorgehen bei der Surrogatvalidierung und zur Verbindung der Validierungsergebnisse zur Nutzenbewertung zusammen.

Faktoren, die in die Bewertung der Aussagesicherheit einer Validierungsstudie eingehen sollten, sind das für die Validierung verwendete statistische Verfahren, die Robustheit und Generalisierbarkeit der Ergebnisse, die Datengrundlage, die Eingrenzung von Indikationen und Interventionen sowie Endpunktdefinitionen. Bei hoher Aussagesicherheit der Validierungsstudie kann aus einer hohen Korrelation ($R \geq 0{,}85$ gemessen an der unteren Grenze des 95 %-Konfidenzintervalls) der Effekte von Surrogat und patientenrelevantem Endpunkt ein Nachweis für die Validität des Surrogats abgeleitet werden. Aus einer niedrigen Korrelation ($R \leq 0{,}7$ gemessen an der oberen Grenze des 95 %-Konfidenzintervalls) ergibt sich ein Nachweis für das Fehlen der Validität des Surrogats. Bei hoher Aussagesicherheit der Validierungsstudie und mittlerer Korrelation ($R < 0{,}85$ bis $> 0{,}7$) bleibt die Validität des Surrogats unklar. Ebenfalls unklar, d. h. nicht abschließend zu beurteilen, ist die Validität des Surrogats bei eingeschränkter oder mäßiger Aussagesicherheit der Validierungsstudie. Liegen keine Daten zur Validierung des Surrogats vor oder sind die vorliegenden Daten nur von geringer Aussagesicherheit, kann auf Basis der Ergebnisse zum Surrogat keine Schlussfolgerung bezüglich des patientenrelevanten Endpunkts getroffen und damit auch kein Beleg für einen Effekt auf den patientenrelevanten Endpunkt abgeleitet werden.

Das eigentliche Ziel der Surrogatvalidierung ist die Klärung der Frage, welche Schlussfolgerungen bezüglich patientenrelevanter Endpunkte aus den Surrogatendpunkten gezogen werden können. Aus Effekten, die für ein valides Surrogat gemessen werden, kann direkt auf die Effekte bezüglich des patientenrelevanten Endpunkts geschlossen werden. Effekte von Surrogaten, die nachgewiesen nicht valide sind, erlauben keine Aussage bezüglich patientenrelevanter Endpunkte. Ist die Validität des Surrogats unklar, können unter Anwendung des STE-Konzepts Aussagen zu patientenrelevanten Endpunkten gemacht werden, die in ihrer Ergebnissicherheit abgestuft sind und so der Datenlage differenziert Rechnung tragen. Ein Vorschlag zur Interpretation von Surrogateffekten bei unklarer Validität wird in Abschnitt 6.3 vorgestellt.

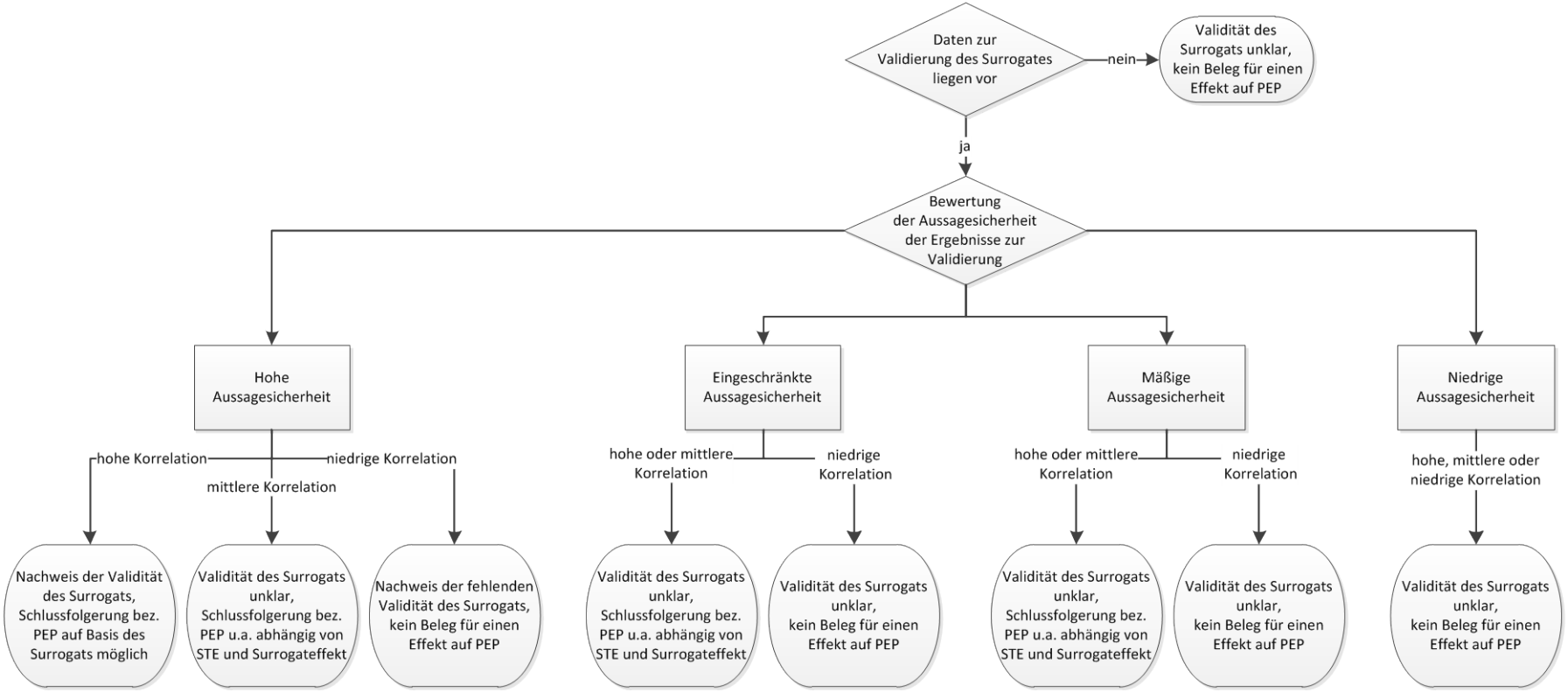

Abbildung 4: Einfluss der Aussagesicherheit der Validierungsstudien und der Korrelation der Effekte von Surrogat und patientenrelevantem Endpunkt auf die Validität des Surrogats

PEP: patientenrelevanter Endpunkt; STE: Surrogate Threshold Effect

6.2 Methodik und Ergebnisse der Untersuchungen zur Validierung von Parametern des Tumoransprechens bei den Krankheitsbildern Darmkrebs und Brustkrebs

Das Methodenreview ergab, dass methodische Fragen der Surrogatvalidierung intensiv bearbeitet werden und die Methodik laufend erweitert wird. International besteht in der Literatur jedoch ein allgemeiner Konsens bezüglich der in Abschnitt 6.1 beschriebenen methodischen Eckpunkte. In den im vorliegenden Bericht identifizierten Studien zur Validierung von Surrogaten bei Darm- und Brustkrebs bilden sich diese Eckpunkte wie folgt ab:

Der Nachweis einer Korrelation zwischen Surrogat und patientenrelevantem Endpunkt reicht allein für eine Validierung nicht aus

Beim Einschluss der Studien zur Validierung von Parametern des Tumoransprechens bei Darm- und Brustkrebs wurde dieses Ergebnis des Methodenreviews dahin gehend berücksichtigt, dass Studien, die ausschließlich eine Korrelation zwischen Surrogat und patientenrelevantem Endpunkt untersuchten, nicht weiter betrachtet wurden (siehe auch Abschnitt 5.2.1.2). Die Autoren dieser Studien beschreiben ihre Untersuchungen in der Regel auch nicht als Analysen zur Validierung von Surrogaten. In zukünftigen Recherchen nach Valdierungsstudien sollten Studien, die ausschließlich die oben beschriebenen Korrelationen betrachten, bereits bei der Selektion ausgeschlossen werden, d.h. dieser methodische Ansatz sollte ein Ausschlusskriterium darstellen.

Die Validierung eines Surrogats ist ggf. sowohl krankheits- als auch interventionsspezifisch – die Übertragbarkeit muss zumindest geprüft und begründet werden

Die Frage, ob eine Validierung eines Surrogats zwischen verschiedenen Indikationen und Interventionen übertragen werden kann, ergibt sich aus den generellen Überlegungen zum Zusammenhang von Surrogat und patientenrelevantem Endpunkt [97]. So kann der kausale Zusammenhang zwischen Surrogat und patientenrelevantem Endpunkt abhängig von der Erkrankung variieren. Möglicherweise kann dieser Zusammenhang auch vom Schweregrade bzw. Stadium einer Erkrankung abhängig sein, z. B. wenn diese zu einer unterschiedlichen Prognose führen. Verschiedene Interventionen können u. a. in unterschiedlichem Maße auf einen patientenrelevanten Endpunkt, aber nicht unterschiedlich auf das Surrogat wirken.

Werden in einem Studienpool zur Validierung eines Surrogats Studien mit verschiedenen Erkrankungen oder verschiedenen Interventionen eingeschlossen, kann aus dem Gesamtergebnis ohne weitere Untersuchungen keine Schlussfolgerung zum Zusammenhang zwischen Surrogat und patientenrelevantem Endpunkt für alle eingeschlossenen Interventionen und Indikationen gezogen werden. Hier wäre in jedem Fall der Einfluss der verschiedenen Interventionen bzw. Indikationen auf das Gesamtergebnis zu prüfen, z. B. durch eine Untersuchung der Heterogenität. Ohne eine Prüfung der Übertragbarkeit können Untersuchungen auf Basis von Studien mit unterschiedlichen Erkrankungen oder Interventionen die Validität eines Surrogats nicht klären.

In der Mehrzahl der eingeschlossenen Validierungsstudien war unklar, welche Patientengruppen in den Studien, die zur Validierung herangezogen wurden, untersucht wurden. In weiteren Validierungsstudien wurden verschiedene Erkrankungen bzw. Schweregrade von Erkrankungen zusammengefasst. Auch bezüglich der in den eingeschlossenen Studien verabreichten Interventionen wurde die Interpretation der Ergebnisse der Validierung durch die Zusammenfassung mehrerer Interventionen erschwert. Eine mögliche Heterogenität wurde in den Validierungsstudien nicht untersucht. Die Zusammenfassung von Daten aus Studien mit verschiedenen Indikationen und Interventionen ohne Prüfung der Übertragbarkeit muss als Mangel der eingeschlossenen Arbeiten zur Surrogatvalidierung bezeichnet werden. Dieser Mangel führt zu einer geringen Sicherheit der Aussagen zur Validität der untersuchten Surrogate.

Zur Validierung eines Surrogats werden vorzugsweise Meta-Analysen mehrerer randomisierter Studien eingesetzt

Das Methodenreview ergab, dass die Validierung von Surrogaten vorzugsweise mithilfe von Meta-Analysen mehrerer randomisierter Studien vorgenommen wird. Dieser Befund wird durch die Recherchen nach Validierungsstudien für die Krankheitsbilder Darm- und Brustkrebs bestätigt. In der überwiegenden Mehrzahl der Validierungsstudien wurden korrelationsbasierte Verfahren verwendet, in denen die Daten mehrerer randomisierter Studien eingesetzt wurden. Einzelne Validierungsstudien setzten testbasierte Verfahren ein, die auch auf mehrere randomisierte Studien zurückgriffen.

Für die meisten Validierungsstudien in der Indikation „Darmkrebs“ blieb allerdings unklar, wie die zugrundeliegenden randomisierten Studien zusammengestellt wurden, und insbesondere, ob diese Zusammenstellung systematisch erfolgte. Dies führte zu einer Verringerung der Aussagesicherheit dieser Validierungsstudien.

Zur Validierung eines Surrogats werden primär korrelationsbasierte Verfahren eingesetzt, eindeutige Grenzen für die Validität von Surrogaten sind nicht definiert

In allen 15 Validierungsstudien von Surrogaten für das Krankheitsbild Darmkrebs wurden korrelationsbasierte Verfahren verwendet, in einer der 15 Validierungsstudien wurde zusätzlich auch ein testbasiertes Verfahren verwendet. Zur Validierung von Surrogaten für das Krankheitsbild Brustkrebs kamen in 5 der 6 Studien korrelationsbasierte Verfahren und in einer der 6 Studien testbasierte Verfahren zur Anwendung.

Die abschließenden Aussagen der Studienautoren zur Validität der untersuchten Surrogate blieben qualitativ. In keiner der Validierungsstudien wurde eine Interpretation der Ergebnisse anhand begründeter Grenzwerte vorgenommen.

Interpretation der statistischen Ergebnisse

Aus den statistischen Ergebnissen der Validierungsstudien blieb bei Anwendung der in Abschnitt 6.1 vorgeschlagenen Grenzwerte für Korrelationen die Validität der Parameter des

Tumoransprechens als Surrogat für das Gesamtüberleben zumeist unklar. Eine Ausnahme stellen folgende Situationen dar:

In beiden Studien zum Kolonkarzinom im frühen Krankheitsstadium, die DFS als Surrogat für OS untersuchten (Buyse 2008 und Green 2008), zeigten sich hohe Korrelationen für den Zusammenhang zwischen den Behandlungseffekten bezüglich der beiden Endpunkte. Für beide Studien blieb aber unklar, ob eine hinreichende Eingrenzung sowohl der untersuchten Indikationen als auch der untersuchten Interventionen vorgenommen wurde. Daher ist auch unklar, für welche Konstellationen dieses Ergebnis gültig ist. Darüber hinaus war es für beide Studien zumindest unklar, ob die für die Validierung verwendeten Studien auf Grundlage einer systematischen Suche eingeschlossen wurden. Dadurch lässt sich nicht beurteilen, ob für die Validierung ein repräsentativer Datensatz zugrunde liegt. Wegen einer niedrigen Aussagesicherheit ist die Datenlage für die Validierung des Surrogats und damit für Aussagen über die Validität des Surrogats in der Gesamtinterpretation unzureichend.

In den Studien Buyse 2007 und Buyse 2008 ergaben sich für das Kolonkarzinom im fortgeschrittenen Krankheitsstadium hohe Korrelationen von PFS und Gesamtüberleben. Da jeweils die Interventionen in den eingeschlossenen Studien unklar blieben, kann nicht beschrieben werden, für welche Konstellation dieses Ergebnis gültig ist. Wegen einer niedrigen Aussagesicherheit ist die Datenlage für die Validierung des Surrogats und damit für Aussagen über die Validität des Surrogats in der Gesamtinterpretation unzureichend. Ergänzend ist anzumerken, dass eine weitere Studie zum Surrogat PFS in dieser Indikation vorliegt, aus der sich keine hohe Korrelation ergab, die Ergebnisse zwischen den Validierungsstudien also nicht konsistent sind.

In der Studie Sargent 2007 (nicht metastasierendes Kolonkarzinom) zeigte sich bei der Subgruppenanalyse der Patienten mit Stadium III eine hohe Korrelation von Median-3-Jahres-DFS und Median-5-Jahres-OS. Da für diese Untersuchung unterschiedliche Interventionen zusammengefasst ausgewertet wurden, ohne dass geeignete Untersuchungen zur Prüfung von Heterogenität zwischen verschiedenen Wirkstoffen bzw. Wirkstoffklassen vorlagen, lässt sich nicht beurteilen, für welche Wirkstoffklassen das Ergebnis gültig ist. Wegen einer niedrigen Aussagesicherheit ist die Datenlage für die Validierung des Surrogats und damit für Aussagen über die Validität des Surrogats in der Gesamtinterpretation unzureichend. Ergänzend ist anzumerken, dass weitere Studien zum Surrogat DFS in dieser Indikation vorliegen, aus denen sich keine hohe Korrelation ergab, die Ergebnisse zwischen den Validierungsstudien also nicht konsistent sind.

In der Studie von Burzykowski 2004 zeigte sich bei der Validierung von ordinal gemessener Tumor Response als Surrogat für das Gesamtüberleben bei fortgeschrittenem Kolorektalkarzinom unter Behandlung mit einer Monotherapie mit Fluoropyrimidinen als intravenöse Bolusgabe versus eine Behandlung mit Fluoropyrimidinen als Monotherapie (kontinuierliche Infusion oder hepatoarterielle Infusion) oder als Kombinationstherapie eine geringe Korrelation. Auch diese Studie hatte eine niedrige Aussagesicherheit, sodass die Datenlage für die Validierung des Surrogats Tumor Response in der Gesamtinterpretation ebenfalls unzureichend war.

Abschließende Einschätzung

Zusammenfassend kann festgestellt werden, dass die eingeschlossenen Validierungsstudien die Ergebnisse der Methodenreviews in großen Teilen bestätigen, jedoch keine der Studien die anerkannten Gütekriterien konsequent erfüllte. Dies mag auch dadurch bedingt sein, dass einige der Validierungsstudien nicht primär auf die Validierung eines Surrogats ausgerichtet waren, sondern methodische Arbeiten zur Anwendung der Validierungsmethodik darstellten. Entsprechend finden sich diverse der Arbeiten aus Teilziel 2 auch im Methodenreview wieder (Teilziel 1). In den Validierungsstudien wurden häufig verschiedene Interventionen oder Indikationen gemeinsam ausgewertet, oder es blieb unklar, ob sich die Datenbasis auf eine Intervention und Indikation beschränkte. Die Übertragbarkeit der Ergebnisse der Surrogatvalidierung zwischen den eingeschlossenen Interventionen oder Indikationen wurde nicht adressiert. Auch blieb zumeist unklar, wie die Datenbasis zusammengestellt war, oder die Zusammenstellung war nachweislich nicht systematisch. Die Aussagesicherheit der Validierungsstudien war insgesamt niedrig.

Auch aufgrund der statistischen Ergebnisse der Validierungsstudien blieb die Validität der untersuchten Surrogate meist unklar. Eine Ausnahme bildet das krankheitsfreie Überleben (DFS) bei Kolonkarzinom im frühen Krankheitsstadium, wo sich in beiden Validierungsstudien eine hohe Korrelation zeigte. Die Aussagesicherheit dieser beiden Studien war jedoch ebenfalls niedrig, sodass sich aus ihnen kein Beleg für die Validiät des DFS für das Gesamtüberleben aus diesen Studien ableiten lässt.

Zusammenfassend bleibt die Validität der Parameter des Tumoransprechens als Surrogate für patientenrelevante Endpunkte (hier Gesamtüberleben) in den Krankheitsbildern Darm- und Brustkrebs aus der vorliegenden Bewertung deshalb unklar.

6.3 Vorschlag für einen Algorithmus zum Umgang mit Surrogatendpunkten in onkologischen Studien in der Nutzenbewertung

Ausgehend von den Ergebnissen der Recherche und Bewertung methodischer Verfahren, deren Ziel die Validierung von Surrogatendpunkten ist (Teilziel 1 des Berichts), sowie von den Ergebnissen von Studien zur Validierung von Surrogatendpunkten bei Darm- und Brustkrebs (Teilziel 2 des Berichts) wird im Folgenden ein Vorschlag zum Umgang mit Surrogatendpunkten in onkologischen Studien für die Nutzenbewertung entwickelt (Ebene 2 der Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung, siehe Abschnitt 6).

Dieser Vorschlag berücksichtigt die Aussagesicherheit der Ergebnisse zur Validierung der Surrogatendpunkte, die Korrelation zwischen Surrogateffekten und Effekten bezüglich patientenrelevanter Endpunkte aus Validierungsstudien sowie die Größe der Effekte, die für das Surrogat in den Studien zur Nutzenbewertung gemessen wurden. Aus der Kombination dieser Faktoren wird im Folgenden abgeleitet, welche Schlussfolgerungen in der Nutzenbewertung aus Studienergebnissen zu Surrogatendpunkten gezogen werden können. Es wird davon ausgegangen, dass Validierungsverfahren eingesetzt werden, mit denen die

Korrelation der Behandlungseffekte von Surrogat und patientenrelevantem Endpunkt geschätzt werden kann.

Abbildung 5 zeigt die ersten Schritte des Algorithmus zur Berücksichtigung von Ergebnissen zu Surrogatendpunkten in der Nutzenbewertung.

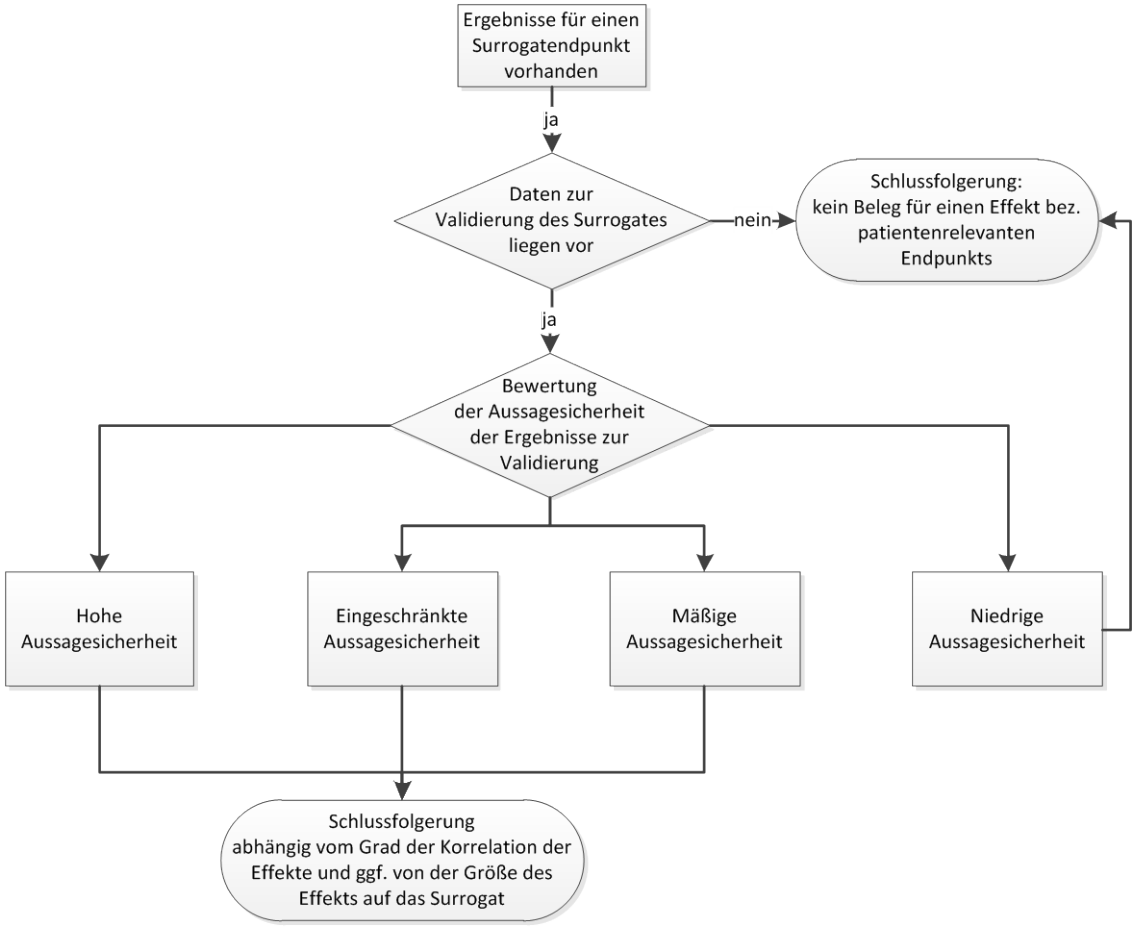

Abbildung 5: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: Stellenwert von Validierungsstudien und deren Aussagesicherheit

PEP: patientenrelevanter Endpunkt

Werden für eine Nutzenbewertung Ergebnisse für einen Surrogatendpunkt vorgelegt, stellt sich die Frage, ob Daten zur Validierung des Surrogats zur Verfügung stehen. Ist das nicht der Fall, liegt aus den Ergebnissen zu dem Surrogatendpunkt kein Beleg (oder Hinweis oder Anhaltspunkt) für einen Effekt bezüglich eines patientenrelevanten Endpunkts vor. Wenn Daten zur Validierung des Surrogats verfügbar sind, werden diese in einem ersten Schritt bezüglich ihrer Aussagesicherheit bewertet. Dabei werden die Kriterien der Aussagesicherheit wie in Abschnitt 4.2.3 beschrieben eingesetzt. Bei niedriger Aussagesicherheit der Validierungsergebnisse gibt es aus den Ergebnissen zu dem Surrogatendpunkt grundsätzlich keinen Beleg für einen Effekt bezüglich eines patientenrelevanten Endpunkts. Bei hoher, eingeschränkter oder mäßiger Aussagesicherheit der Validierungsergebnisse ist die Schlussfolgerung zu patientenrelevanten Endpunkten abhängig vom Grad der Korrelation der

Effekte von Surrogat und patientenrelevantem Endpunkt in den Validierungsstudien sowie ggf. von der Größe des Effekts für das Surrogat in den Studien zur Nutzenbewertung. Das Grundprinzip ist dabei, dass die Sicherheit der Aussage in der Nutzenbewertung (z. B. Beleg für oder Hinweis auf einen Effekt auf einen patientenrelevanten Endpunkt) umso höher ist, je sicherer ein Effekt auf das Surrogat auch einen Effekt auf den patientenrelevanten Endpunkt abbildet. Für diese Abstufung der Sicherheit der Aussage werden auch unterschiedliche Konfidenzintervalle (95 %-Konfidenzintervall und 80 %-Konfidenzintervall) zur Beurteilung der Größe des Effekts auf das Surrogat verwendet (siehe Abbildung 6 bis Abbildung 8).

Abbildung 6 zeigt die möglichen Schlussfolgerungen für die Nutzenbewertung bei sehr hoher Aussagesicherheit der Validierungsstudien.

Abbildung 6: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: mögliche Schlussfolgerungen bei hoher Aussagesicherheit der Validierungsstudien

KI: Konfidenzintervall; STE: Surrogate Threshold Effect

Abhängig von der Korrelation zwischen Surrogat und patientenrelevantem Endpunkt aus Validierungsstudien sind unterschiedliche Schlussfolgerungen aus den Ergebnissen des Surrogatendpunkts in den Studien zur Nutzenbewertung ableitbar.

- Aus Validierungsstudien mit hoher Aussagesicherheit kann bei hoher Korrelation ($R \geq 0{,}85$; $R^2 \geq 0{,}72$, gemessen jeweils durch die untere Grenze des 95%-Konfidenzintervalls) und einem Effekt für das Surrogat maximal auf einen Beleg für einen Effekt bezüglich des patientenrelevanten Endpunkts geschlossen werden. Unter einem Effekt wird ein statistisch signifikanter Gruppenunterschied in einer vergleichenden Interventionsstudie bzw. einer Meta-Analyse solcher Studien verstanden. Die Aussage, dass maximal auf einen Beleg geschlossen werden kann, bedeutet, dass zusätzlich zur Prüfung der Validität des Surrogats in Validierungsstudien die Ergebnissicherheit der Daten zum Surrogat in den Studien zur Nutzenbewertung berücksichtigt wird, z. B. das Verzerrungspotenzial der Studien oder die Bestätigung der Ergebnisse in einer zweiten Studie. Bei hohem Verzerrungspotenzial oder Fehlen einer bestätigenden Studie würde die Aussage z. B. von einem Beleg auf einen Hinweis für einen Effekt bezüglich des patientenrelevanten Endpunkts herabgestuft.

  Zeigen die Studien zur Nutzenbewertung keinen Effekt für das Surrogat, d. h. liegt in diesen Studien bzw. in einer Meta-Analyse der Studien kein statistisch signifikanter Gruppenunterschied vor, ist generell ein Effekt bezüglich des patientenrelevanten Endpunkts nicht belegt (gilt auch für den Pfad mit mittlerer Korrelation).

- Bei hoher Aussagesicherheit der Validierungsstudien und mittlerer Korrelation ($R < 0{,}85$; $R^2 < 0{,}72$ bis $R > 0{,}7$; $R^2 > 0{,}49$) zwischen Surrogat und patientenrelevantem Endpunkt soll der Effekt für das Surrogat aus den Studien zur Nutzenbewertung einem Surrogate Threshold Effect (STE) gegenübergestellt werden. Ist ein solcher STE nicht verfügbar, gilt ein Effekt bezüglich des patientenrelevanten Endpunkts als nicht belegt. Ist ein STE verfügbar, wird zunächst das 95 %-KI des Effekts für das Surrogat aus den Studien zur Nutzenbewertung zu diesem STE in Beziehung gesetzt. Liegt das 95 %-KI vollständig oberhalb des STE, liegt maximal ein Beleg für einen Effekt bezüglich des patientenrelevanten Endpunkts vor (bei erhöhter Ergebnisunsicherheit in den Studien zur Nutzenbewertung, z. B. durch ein hohes Verzerrungspotenzial, kann dieses Ergebnis wiederum zu einem Hinweis herabgestuft werden). Liegt das 95 %-KI nicht vollständig oberhalb des STE, wird das 80 %-KI des Effekts für das Surrogat zu dem STE in Beziehung gesetzt. Liegt das 80 %-KI vollständig oberhalb des STE, kann maximal auf einen Hinweis für einen Effekt bezüglich des patientenrelevanten Endpunkts geschlossen werden. Bei erhöhter Ergebnisunsicherheit, z. B. weil das Ergebnis nicht in einer zweiten Studie bestätigt wurde, wird der Hinweis auf einen Effekt herabgestuft. Überschreitet das 80 %-KI des Surrogateffekts den STE nicht vollständig, ist bei hoher Aussagesicherheit der Validierungsstudien und mittlerer Korrelation ein Effekt bezüglich des patientenrelevanten Endpunkts nicht belegt, und es liegt auch kein Hinweis und kein Anhaltspunkt für dafür vor.

- Bei hoher Aussagesicherheit der Validierungsstudien und niedriger Korrelation ($R \leq 0{,}7$; $R^2 \leq 0{,}49$, gemessen durch die obere Grenze des 95%-Konfidenzintervalls) zwischen Surrogat und patientenrelevantem Endpunkt gilt das Surrogat als nicht valide. Aus einem

Effekt auf das Surrogat ergibt sich damit grundsätzlich kein Beleg (sowie kein Hinweis und kein Anhaltspunkt) für einen Effekt bezüglich des patientenrelevanten Endpunkts.

Abbildung 7 zeigt die möglichen Schlussfolgerungen für die Nutzenbewertung bei eingeschränkter Aussagesicherheit der Validierungsstudien.

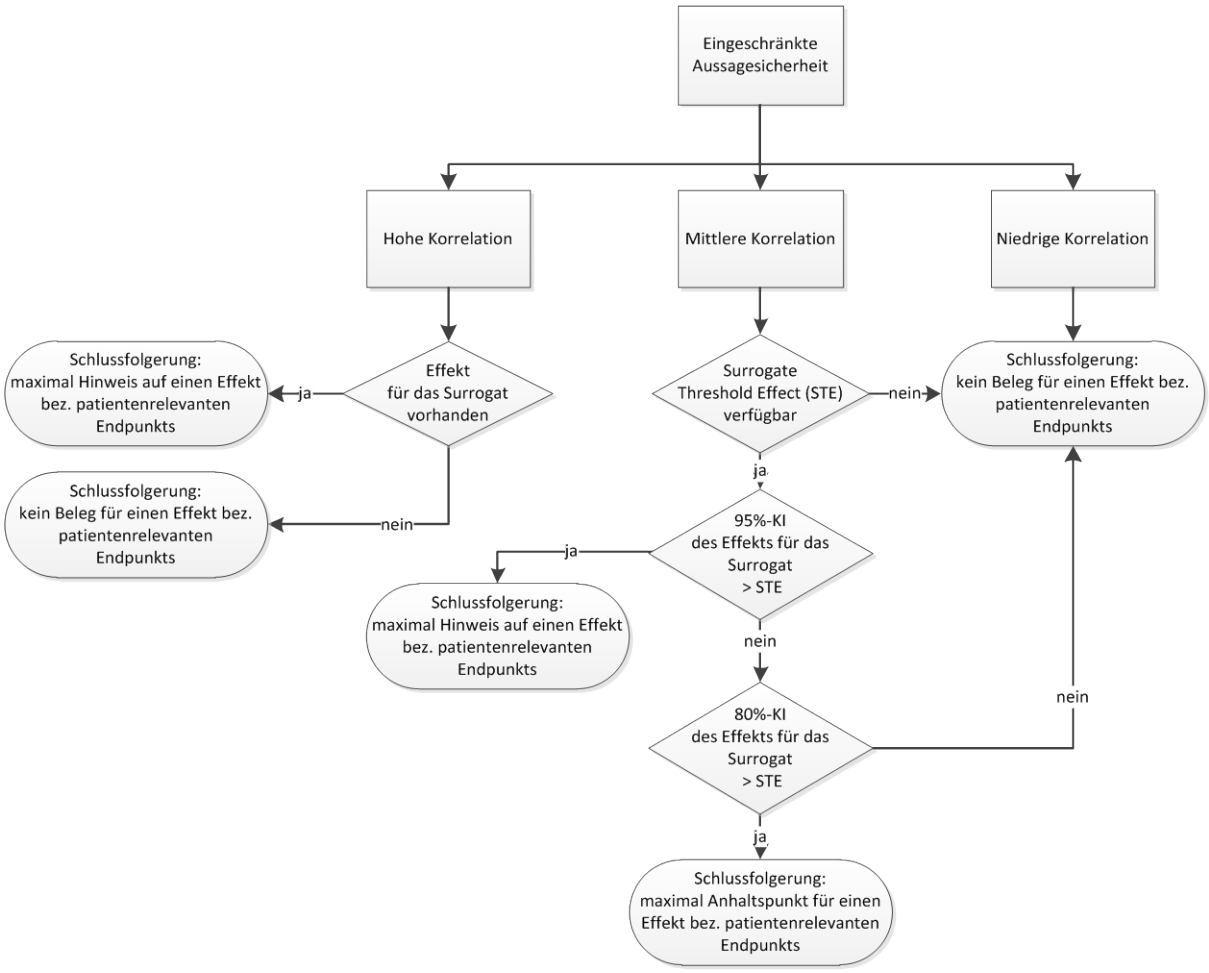

Abbildung 7: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: mögliche Schlussfolgerungen bei eingeschränkter Aussagesicherheit der Validierungsstudien

KI: Konfidenzintervall; STE: Surrogate Threshold Effect

Im Vergleich zu den möglichen Schlussfolgerungen bei hoher Aussagesicherheit der Validierungsstudien haben in der Situation, in der die Aussagesicherheit dieser Studien eingeschränkt ist, alle Schlussfolgerungen zu Effekten bezüglich patientenrelevanter Endpunkte eine geringere Ergebnissicherheit.

- Bei eingeschränkter Aussagesicherheit und hoher Korrelation von Surrogat und patientenrelevantem Endpunkt kann aus Surrogateffekten in den Studien zur

Nutzenbewertung maximal auf einen Hinweis auf Effekte bezüglich patientenrelevanter Endpunkte geschlossen werden.

- Bei eingeschränkter Aussagesicherheit und mittlerer Korrelation kann bei Überschreiten des STE durch das 95 %-KI des Surrogateffekts maximal ein Hinweis, bei Überschreiten des STE durch das 80 %-KI maximal ein Anhaltspunkt für einen Effekt bezüglich des patientenrelevanten Endpunkts abgeleitet werden.

- Bei eingeschränkter Aussagesicherheit und niedriger Korrelation liegt grundsätzlich kein Beleg für einen Effekt bezüglich des patientenrelevanten Endpunkts vor.

Für alle abschließenden Schlussfolgerungen wird wie oben beschrieben zusätzlich die Ergebnissicherheit der Studien zur Nutzenbewertung berücksichtigt, d. h., eine Einschränkung der Ergebnissicherheit führt zu einer Herabstufung der Aussage.

Abbildung 8 zeigt die möglichen Schlussfolgerungen für die Nutzenbewertung bei mäßiger Aussagesicherheit der Validierungsstudien.

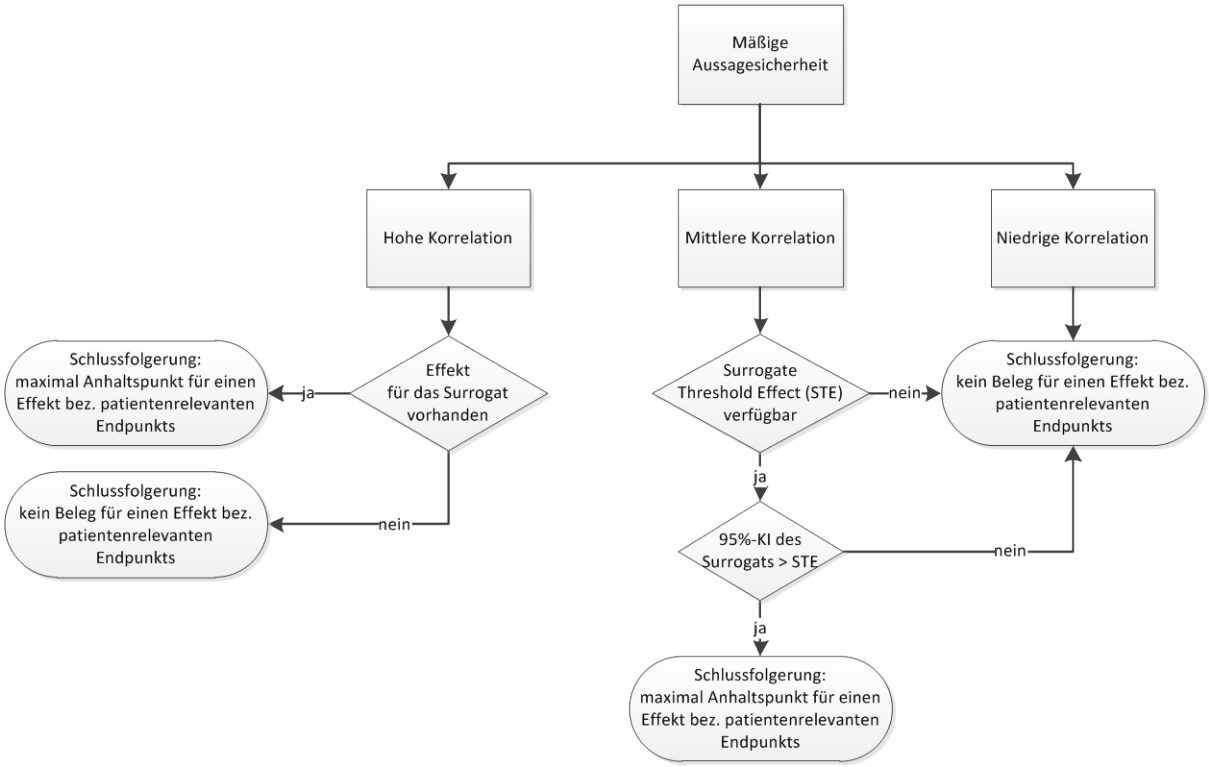

Abbildung 8: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: mögliche Schlussfolgerungen bei mäßiger Aussagesicherheit der Validierungsstudien

KI: Konfidenzintervall; STE: Surrogate Threshold Effect

Bei mäßiger Aussagesicherheit der Validierungsstudien können aus den Ergebnissen zu Surrogaten in den Studien zur Nutzenbewertung maximal Anhaltspunkte für Effekte bezüglich patientenrelevanter Endpunkte abgeleitet werden. Bei mittlerer Korrelation wird hierfür ausschließlich das 95 %-KI zu dem STE in Beziehung gesetzt. Wie zuvor wird auch bei mäßiger Aussagesicherheit nach der Beurteilung der maximal abzuleitenden Aussage zum Effekt bezüglich eines patientenrelevanten Endpunkts die Ergebnissicherheit der Ergebnisse zu dem Surrogat aus den Studien zur Nutzenbewertung berücksichtigt. Bei einer Einschränkung der Ergebnissicherheit wird die Bewertung „Anhaltspunkt“ zu „kein Beleg (sowie kein Hinweis und kein Anhaltspunkt)“ herabgestuft.

6.4 Abschließende Anmerkungen

Die Frage der Validität von Surrogatendpunkten in onkologischen Studien hat für die Bewertung von Interventionen zur Behandlung onkologischer Erkrankungen eine große Bedeutung. Das gilt insbesondere, weil Entscheidungen für oder gegen eine Therapie von Patienten mit onkologischen Erkrankungen wegen des lebensbedrohlichen Charakters der Erkrankung und der häufig schweren Nebenwirkungen der Therapie weitreichende Folgen haben.

Das systematische Review der Methodik zur Surrogatvalidierung hat gezeigt, dass adäquate Vorschläge für Validierungsmethoden vorliegen und in der Praxis erprobt sind. Eine Validierung von Surrogaten erscheint also aus methodischer Sicht grundsätzlich möglich.

Die systematische Recherche nach Validierungsstudien in den Krankheitsbildern Darmkrebs und Brustkrebs hat eine Reihe von Arbeiten zur Validierung von Parametern des Tumoransprechens als Surrogate für das Gesamtüberleben identifiziert. Auch in der Onkologie ist die Anwendung der Methodik zur Surrogatvalidierung also in der Praxis erprobt. Nach den Kriterien des vorliegenden Berichts war die Aussagesicherheit jedoch bei allen Publikationen beeinträchtigt. Dazu trug wesentlich bei, dass in den Untersuchungen Studien mit verschiedenen Indikationen gepoolt wurden oder dass verschiedene Interventionen zusammengefasst wurden, ohne dass untersucht oder begründet wurde, ob diese Zusammenfassung berechtigt war und ob die Ergebnisse hierdurch wesentlich beeinflusst wurden. Häufig blieb auch unklar, welche Interventionen überhaupt zusammengefasst wurden. Darüber hinaus wurde beim Krankheitsbild Darmkrebs die Datengrundlage häufig unsystematisch zusammengestellt oder das diesbezügliche Vorgehen blieb unklar. Es wäre deshalb dringend erforderlich, innerhalb der Validierungsstudien getrennte Auswertungen für einzelne Indikationen und Interventionen durchzuführen und die Ergebnisse transparent zu berichten, sodass die Interpretation der Ergebnisse erleichtert wird, zumal sich die statistischen Ergebnisse (Angaben zur Korrelation) zwischen den Validierungsstudien zum Teil deutlich unterschieden.

Als ein Ergebnis des Methodenreviews kann festgehalten werden, dass die Validierung eines Surrogats ggf. sowohl krankheits- als auch interventionsspezifisch (zumindest wirkstoffklassenspezifisch) ist und deshalb die Validierung mithilfe statistischer Methoden nur unter bestimmten Voraussetzungen abschließend indikations- und interventionsübergreifend

erfolgen kann. Das erschwert insbesondere die Validierung von Surrogaten für die Bewertung neuer Wirkstoffe. Trotz dieser methodischen Anforderung, die auf Basis grundsätzlicher Überlegungen zum Zusammenhang von Surrogat und patientenrelevantem Endpunkt nachvollzogen werden kann [97], wären Daten zur Validität von Surrogaten für Interventionen, deren Stellenwert bereits anhand patientenrelevanter Endpunkte geklärt ist, für ein allgemeines Verständnis des Zusammenhangs zwischen Parametern des Tumoransprechens und Gesamtüberleben innerhalb einer Indikation von großer Bedeutung. Ggf. könnten umfangreiche Validierungsdaten für etablierte Therapien eine Hilfestellung bei der Interpretation von Ergebnissen neuer Wirkstoffe geben und damit die Unsicherheit bezüglich des Einsatzes neuer Wirkstoffe bei onkologischen Patienten verringern. Solche Validierungsdaten könnten von den Sponsoren klinischer Studien mit onkologischen Arzneimitteln adäquat analysiert (d. h. u. a. stratifiziert nach Indikation und Intervention) in einem öffentlich zugänglichen Register zur Verfügung gestellt werden. Dabei sollten die in diesem Bericht beschriebenen Kriterien der Aussagesicherheit von Validierungsstudien beachtet werden, um eine Interpretation der Ergebnisse zu ermöglichen. Auch wenn die bislang vorliegenden Validierungsstudien in den Bereichen Darmkrebs und Brustkrebs in diesem Bericht mit einer niedrigen Aussagesicherheit bewertet wurden, zeigt eine Gesamtschau dieser Studien, dass diese anerkannten Gütekriterien grundsätzlich erfüllbar sind.

- Die meisten Validierungsstudien verwendeten ein anerkanntes, beschriebenes Verfahren zur Validierung.
- Es liegen Validierungsstudien vor, in denen eine Eingrenzung auf Wirkstoffe oder Wirkstoffklassen begründet vorgenommen wurde oder der Einfluss unterschiedlicher Wirkstoffe / Wirkstoffklassen untersucht wurde.
- Ebenfalls liegen Validierungsstudien vor, in denen eine Eingrenzung auf eine Indikation oder einen Schweregrad begründet vorgenommen wurde oder der Einfluss unterschiedlicher Indikationen bzw. Schweregrade untersucht wurde.
- In mehreren Validierungsstudien wurde die Operationalisierung der Endpunkte thematisiert, und es wurden einheitliche Endpunktdefinitionen verwendet.
- Bei allen Validierungsstudien im Bereich Brustkrebs und einigen im Bereich Darmkrebs erfolgte die Zusammenstellung der Datengrundlage systematisch.
- Analysen zur Robustheit und Generalisierbarkeit wurden ebenfalls in einzelnen Validierungsstudien durchgeführt.

Im Interesse einer möglichst informierten Therapie von Patienten mit onkologischen Erkrankungen erscheint eine gemeinsame Anstrengung aller Beteiligten zur Verbesserung der Datenlage zum Stellenwert von Surrogatendpunkten sehr wünschenswert.

Mit dem vorliegenden Bericht stellt das IQWiG auf Basis einer systematischen Recherche zur Methodik zur Surrogatvalidierung sowie der beispielhaften systematischen Aufarbeitung von

Validierungsstudien in den Krankheitsbildern Darm- und Brustkrebs einen Algorithmus zum Umgang mit Surrogatendpunkten in der Nutzenbewertung von onkologischen Behandlungen zur Diskussion. Dieser Vorschlag hat das Ziel, ausgehend von den verfügbaren Daten und Informationen zu einer differenzierten Betrachtung der Ergebnisse von Surrogatendpunkten zu kommen, die die Ergebnissicherheit der verschiedenen Konstellationen berücksichtigt.

7 Fazit

Methoden zur Surrogatvalidierung

Es liegen adäquate Vorschläge für das methodische Vorgehen bei der Validierung von Surrogatendpunkten vor, die auch in der Praxis erprobt sind. Eine Validierung von Surrogaten erscheint daher grundsätzlich möglich. Dabei besteht zu folgenden grundsätzlichen Aussagen international ein allgemeiner Konsens:

- Der Nachweis einer Korrelation zwischen Surrogat und klinischem Endpunkt reicht alleine zur Surrogatvalidierung nicht aus.
- Die Übertragbarkeit von Aussagen zur Validität von Surrogaten zwischen verschiedenen Erkrankungen oder Ausprägungen einer Erkrankung bzw. zwischen verschiedenen Interventionen ist nicht ohne Weiteres möglich. Zur Beurteilung der Übertragbarkeit sollten bei Validierungsstudien, die verschiedene Krankheitsentitäten oder Interventionen einschließen, mindestens Untersuchungen zur Heterogenität vorliegen.
- Zur Validierung eines Surrogats werden umfangreiche Daten benötigt, vorzugsweise eine Meta-Analyse von mehreren randomisierten Studien mit ausreichender Ergebnissicherheit. In diesen Studien müssen sowohl das Surrogat als auch der klinische Endpunkt erhoben worden sein.
- Zur Surrogatvalidierung gibt es weder ein universell anzuwendendes Maß noch eine allgemein beste Schätzmethode noch eine allgemein akzeptierte Grenze, deren Überschreitung den Nachweis der Validität bedeuten würde.
- Zurzeit werden in der Praxis primär korrelationsbasierte Verfahren zur Surrogatvalidierung mit Schätzung von Korrelationsmaßen auf Studien- und individueller Ebene angewendet. Das Konzept eines Surrogate Threshold Effects (STE) ist dabei hilfreich zur Festlegung von Schwellenwerten für den Entscheidungsprozess.

Validierungsstudien für die Krankheitsbilder Darmkrebs und Brustkrebs

Die zu den Krankheitsbildern Darmkrebs und Brustkrebs vorliegenden Validierungsstudien erlauben keine abschließende Aussage zur Validität von Parametern des Tumoransprechens für das Gesamtüberleben bei diesen Erkrankungen. In den meisten Fällen waren die Korrelationen der jeweiligen Effekte nicht ausreichend hoch, sodass damit die notwendige Bedingung für den Nachweis der Validität nicht erfüllt war. In allen Studien wurde außerdem die Aussagesicherheit als niedrig bewertet. Daher blieb die Validität bei allen untersuchten Surrogatendpunkten unabhängig von den statistischen Ergebnissen letztlich unklar. Für die niedrige Aussagesicherheit der Validierungsstudien waren insbesondere folgende Faktoren maßgeblich:

- Die Surrogatvalidierung wurde zumeist nicht interventionsspezifisch durchgeführt, und der Einfluss der verschiedenen Interventionen auf das Gesamtergebnis wurde nicht untersucht.
- Häufig wurden verschiedene Indikationen oder verschiedene Schweregrade der Erkrankung zusammengefasst, und der Einfluss dieser verschiedenen Behandlungssituationen auf das Gesamtergebnis wurde nicht untersucht.
- Im Krankheitsbild Darmkrebs erfolgte die Zusammenstellung der Datengrundlage häufig unsystematisch oder das Vorgehen blieb unklar.

Die Gesamtschau der Validierungsstudien zeigt, dass diese Mängel grundsätzlich behebbar sind. Es wäre im Interesse einer möglichst informierten Therapie von Patienten mit onkologischen Erkrankungen wünschenswert, wenn systematisch Validierungsstudien unter Beachtung der genannten methodischen Standards durchgeführt und transparent berichtet würden. Dies erfordert eine gemeinsame Anstrengung aller Beteiligten. Die damit verbundene Verbesserung der Erkenntnisse zum Stellenwert von Surrogatendpunkten wäre auch geeignet, Entscheidungswege für die Bewertung von neuen Arzneimitteln für onkologische Erkrankungen trotz begrenzter Datenlage zu diesen Arzneimitteln aufzuzeigen.

Vorschlag des IQWiG für den Umgang mit Surrogatendpunkten in der Nutzenbewertung onkologischer Behandlungen

Bei der Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung sind 2 Ebenen zu trennen. Auf der ersten Ebene muss zunächst die Validität des Surrogats beurteilt werden. Auf der zweiten Ebene stellt sich die Frage, welche Schlussfolgerungen bezüglich patientenrelevanter Endpunkte aus den Surrogateffekten abhängig von der Ausprägung der Validität des Surrogats gezogen werden können.

Werden für eine Nutzenbewertung Ergebnisse für einen Surrogatendpunkt vorgelegt, ist zunächst zu prüfen, ob Daten zur Validierung des Surrogats zur Verfügung stehen. Ist das nicht der Fall, liegt aus den Ergebnissen zu dem Surrogatendpunkt kein Beleg (oder Hinweis oder Anhaltspunkt) für einen Effekt bezüglich eines patientenrelevanten Endpunkts vor. Wenn Daten zur Validierung des Surrogats verfügbar sind, werden diese in einem ersten Schritt bezüglich ihrer Aussagesicherheit bewertet. Bei niedriger Aussagesicherheit der Validierungsergebnisse gibt es aus den Ergebnissen zu dem Surrogatendpunkt grundsätzlich keinen Beleg für einen Effekt bezüglich eines patientenrelevanten Endpunkts. Bei hoher, eingeschränkter oder mäßiger Aussagesicherheit der Validierungsergebnisse ist die Schlussfolgerung zu patientenrelevanten Endpunkten abhängig vom Grad der Korrelation der Effekte von Surrogat und patientenrelevantem Endpunkt in den Validierungsstudien sowie ggf. von der Größe des Effekts für das Surrogat in den Studien zur Nutzenbewertung. Hierzu wird der Effekt auf das Surrogat, der sich aus den Studien zur Nutzenbewertung ergibt, einem Surrogate Threshold Effect gegenübergestellt. Das Grundprinzip ist dabei, dass die Sicherheit der Aussage in der Nutzenbewertung (z. B. Beleg für oder Hinweis auf einen Effekt auf einen patientenrelevanten Endpunkt) umso höher ist, je sicherer ein Effekt auf das Surrogat auch einen Effekt auf den patientenrelevanten Endpunkt abbildet.

8 Liste der eingeschlossenen Studien

8.1 Teilziel 1

Übersichtsarbeiten

Burzykowski T, Molenberghs G, Buyse M (Ed). The evaluation of surrogate endpoints. New York: Springer; 2005.

Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. Stat Methods Med Res 2008; 17(3): 303-340.

Mangiapane S, Velasco Garrido M. Surrogatendpunkte als Parameter der Nutzenbewertung [online]. (DIMDI Schriftenreihe Health Technology Assessment; Band 91). URL: http://portal.dimdi.de/de/hta/hta_berichte/hta250_bericht_de.pdf.

Molenberghs G, Burzykowski T, Alonso A, Assam P, Tilahun A, Buyse M. A unified framework for the evaluation of surrogate endpoints in mental-health clinical trials. Stat Methods Med Res 2010; 19(3): 205-236.

Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. Stat Med 2006; 25(2): 183-203.

Originalarbeiten

Abrahantes JC, Shkedy Z, Molenberghs G. Alternative methods to evaluate trial level surrogacy. Clin Trials 2008; 5(3): 194-208.

Alonso A, Molenberghs G. Surrogate marker evaluation from an information theory perspective. Biometrics 2007; 63(1): 180-186.

Alonso A, Molenberghs G. Evaluating time to cancer recurrence as a surrogate marker for survival from an information theory perspective. Stat Methods Med Res 2008; 17(5): 497-504.

Alonso A, Molenberghs G. Surrogate end points: hopes and perils. Expert Rev Pharmacoecon Outcomes Res 2008; 8(3): 255-259.

Baker SG. A simple meta-analytic approach for using a binary surrogate endpoint to predict the effect of intervention on true endpoint. Biostatistics 2006; 7(1): 58-70.

Baker SG. Two simple approaches for validating a binary surrogate endpoint using data from multiple trials. Stat Methods Med Res 2008; 17(5): 505-514.

Burzykowski T. Surrogate endpoints: wishful thinking or reality? Stat Methods Med Res 2008; 17(5): 463-466.

Buyse M. Use of meta-analysis for the validation of surrogate endpoints and biomarkers in cancer trials. Cancer J 2009; 15(5): 421-425.

Buyse M, Sargent DJ, Grothey A, Matheson A, De Gramont A. Biomarkers and surrogate end points: the challenge of statistical validation. Nat Rev Clin Oncol 2010; 7(6): 309-317.

Deslandes E, Chevret S. Assessing surrogacy from the joint modelling of multivariate longitudinal data and survival: application to clinical trial data on chronic lymphocytic leukaemia. Stat Med 2007; 26(30): 5411-5421.

Ditlevsen S, Keiding N. A comment on: statistical evaluation of biomarkers as surrogate endpoints: a literature review by C. J. Weir and R. J. Walley, Statistics in Medicine 2006; 25:183-203. Stat Med 2007; 26(6): 1415-1416.

Duffy SW, Treasure FP. Potential surrogate endpoints in cancer research: some considerations and examples. Pharm Stat 2009.

Dunning AJ. Comment on 'Evaluating a surrogate endpoint at three levels, with application to vaccine development' by Peter B. Gilbert, Li Qin and Stefen G. Self, Statistics in Medicine, DOI: 10.1002/sim.3122. Stat Med 2008; 27(29): 6268-6270.

Fleischer F, Gaschler-Markefski B, Bluhmki E. A statistical model for the dependence between progression-free survival and overall survival. Stat Med 2009; 28(21): 2669-2686.

Freedman L. Quantitative science methods for biomarker validation in chemoprevention trials. Cancer Biomark 2007; 3(3): 135-140.

Ghosh D. On assessing surrogacy in a single trial setting using a semicompeting risks paradigm. Biometrics 2009; 65(2): 521-529.

Ghosh D, Elliott MR, Taylor JM. Links between analysis of surrogate endpoints and endogeneity. Stat Med 2010; 29(28): 2869-2879.

Gilbert PB, Qin L, Self SG. Evaluating a surrogate endpoint at three levels, with application to vaccine development. Stat Med 2008; 27(23): 4758-4778.

Gilbert PB, Qin L, Self SG. Response to Andrew Dunning's comment on 'Evaluating a surrogate endpoint at three levels, with application to vaccine development'. Stat Med 2009; 28(4): 716-719.

Green E, Yothers G, Sargent DJ. Surrogate endpoint validation: statistical elegance versus clinical relevance. Stat Methods Med Res 2008; 17(5): 477-486.

Huang J, Huang B. Evaluating the proportion of treatment effect explained by a continuous surrogate marker in logistic or probit regression models. Stat Biopharm Res 2010; 2(2): 229-238.

Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. Biometrics 2009; 65(2): 530-538.

Lassere M, Johnson K, Hughes M, Altman D, Buyse M, Galbraith S et al. Simulation studies of surrogate endpoint validation using single trial and multitrial statistical approaches. J Rheumatol 2007; 34(3): 616-619.

Lassere MN, Johnson KR, Boers M, Tugwell P, Brooks P, Simon L et al. Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. J Rheumatol 2007; 34(3): 607-615.

Li Y, Taylor JM, Elliott MR. A Bayesian approach to surrogacy assessment using principal stratification in clinical trials. Biometrics 2010; 66(2): 523-531.

MacKinnon DP, Lockwood CM, Brown CH, Wang W, Hoffman JM. The intermediate endpoint effect in logistic and probit regression. Clin Trials 2007; 4(5): 499-513.

Piedbois P, Buyse M. Endpoints and surrogate endpoints in colorectal cancer: a review of recent developments. Curr Opin Oncol 2008; 20(4): 466-471.

Pryseley A, Tilahun A, Alonso A, Molenberghs G. Information-theory based surrogate marker evaluation from several randomized clinical trials with continuous true and binary surrogate endpoints. Clin Trials 2007; 4(6): 587-597.

Qu Y, Case M. Quantifying the effect of the surrogate marker by information gain. Biometrics 2007; 63(3): 958-962.

Sarkar S, Qu Y. Quantifying the treatment effect explained by markers in the presence of measurement error. Stat Med 2007; 26(9): 1955-1963.

Sertdemir Y, Burgut R. Does the decision in a validation process of a surrogate endpoint change with level of significance of treatment effect? A proposal on validation of surrogate endpoints. Contemp Clin Trials 2009; 30(1): 8-12.

Shi Q, Sargent DJ. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. Int J Clin Oncol 2009; 14(2): 102-111.

Tilahun A, Maringwa JT, Geys H, Alonso A, Raeymaekers L, Molenberghs G et al. Investigating association between behavior, corticosterone, heart rate, and blood pressure in rats using surrogate marker evaluation methodology. J Biopharm Stat 2009; 19(1): 133-149.

Tilahun A, Pryseley A, Alonso A, Molenberghs G. Information theory-based surrogate marker evaluation from several randomized clinical trials with binary endpoints, using SAS. J Biopharm Stat 2008; 18(2): 326-341.

Van Walraven C, Oake N, Coyle D, Taljaard M, Forster AJ. Changes in surrogate outcomes can be translated into clinical outcomes using a Monte Carlo model. J Clin Epidemiol 2009; 62(12): 1306-1315.

Wolfson J, Gilbert P. Statistical identifiability and the surrogate endpoint problem, with application to vaccine trials. Biometrics 22.01.2010 [Epub ahead of print].

8.2 Teilziel 2

8.2.1 Krankheitsbild Darmkrebs

Eingeschlossene Studien

Abrahantes JC, Shkedy Z, Molenberghs G. Alternative methods to evaluate trial level surrogacy. Clin Trials 2008; 5(3): 194-208.

Alonso A, Molenberghs G. Evaluating time to cancer recurrence as a surrogate marker for survival from an information theory perspective. Stat Methods Med Res 2008; 17(5): 497-504.

Baker SG. A simple meta-analytic approach for using a binary surrogate endpoint to predict the effect of intervention on true endpoint. Biostatistics 2006; 7(1): 58-70.

Burzykowski T, Buyse M, Yothers G, Sakamoto J, Sargent D. Exploring and validating surrogate endpoints in colorectal cancer. Lifetime Data Anal 2008; 14(1): 54-64.

Burzykowski T, Molenberghs G. The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer. J R Statist Soc A 2004; 167(1): 103-124.

Buyse M, Burzykowski T, Carroll K, Michiels S, Sargent DJ, Miller LL et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. J Clin Oncol 2007; 25(33): 5218-5224.

Buyse M, Burzykowski T, Michiels S, Carroll K. Individual- and trial-level surrogacy in colorectal cancer. Stat Methods Med Res 2008; 17(5): 467-475.

Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Lancet 2000; 356(9227): 373-378.

Green E, Yothers G, Sargent DJ. Surrogate endpoint validation: statistical elegance versus clinical relevance. Stat Methods Med Res 2008; 17(5): 477-486.

Johnson KR, Ringland C, Stokes BJ, Anthony DM, Freemantle N, Irs A et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. Lancet Oncol 2006; 7(9): 741-746.

Montagnani F, Migali C, Fiorentini G. Progression-free survival in bevacizumab-based first-line treatment for patients with metastatic colorectal cancer: is it a really good end point? J Clin Oncol 2009; 27(28): e132-e133.

Sargent DJ, Patiyil S, Yothers G, Haller DG, Gray R, Benedetti J et al. End points for colon cancer adjuvant trials: observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT Group. J Clin Oncol 2007; 25(29): 4569-4574.

Sargent DJ, Wieand HS, Haller DG, Gray R, Benedetti JK, Buyse M et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. J Clin Oncol 2005; 23(34): 8664-8670.

Sertdemir Y, Burgut R. Does the decision in a validation process of a surrogate endpoint change with level of significance of treatment effect? A proposal on validation of surrogate endpoints. Contemp Clin Trials 2009; 30(1): 8-12.

Tang PA, Bentzen SM, Chen EX, Siu LL. Surrogate end points for median overall survival in metastatic colorectal cancer: literature-based analysis from 39 randomized controlled trials of first-line chemotherapy. J Clin Oncol 2007; 25(29): 4562-4568.

Korrelationsstudien

Blazer DG 3rd, Kishi Y, Maru DM, Kopetz S, Chun YS, Overman MJ et al. Pathologic response to preoperative chemotherapy: a new outcome end point after resection of hepatic colorectal metastases. J Clin Oncol 2008; 26(33): 5344-5351.

Buyse M, Piedbois P. On the relationship between response to treatment and survival time. Stat Med 1996; 15(24): 2797-2812.

Dhani N, Tu D, Sargent DJ, Seymour L, Moore MJ. Alternate endpoints for screening phase II studies. Clin Cancer Res 2009; 15(6): 1873-1882.

Louvet C, De Gramont A, Tournigand C, Artru P, Maindrault-Goebel F, Krulik M. Correlation between progression free survival and response rate in patients with metastatic colorectal carcinoma. Cancer 2001; 91(11): 2033-2038.

8.2.2 Krankheitsbild Brustkrebs

Eingeschlossene Studien

Bruzzi P, Del Mastro L, Sormani MP, Bastholt L, Danova M, Focan C et al. Objective response to chemotherapy as a potential surrogate end point of survival in metastatic breast cancer patients. J Clin Oncol 2005; 23(22): 5117-5125.

Burzykowski T, Buyse M, Piccart-Gebhart MJ, Sledge G, Carmichael J, Luck HJ et al. Evaluation of tumor response, disease control, progression-free survival, and time to

progression as potential surrogate end points in metastatic breast cancer. J Clin Oncol 2008; 26(12): 1987-1992.

Hackshaw A, Knight A, Barrett-Lee P, Leonard R. Surrogate markers and survival in women receiving first-line combination anthracycline chemotherapy for advanced breast cancer. Br J Cancer 2005; 93(11): 1215-1221.

Miksad RA, Zietemann V, Gothe R, Schwarzer R, Conrads-Frank A, Schnell-Inderst P et al. Progression-free survival as a surrogate endpoint in advanced breast cancer. Int J Technol Assess Health Care 2008; 24(4): 371-383.

Ng R, Pond GR, Tang PA, MacIntosh PW, Siu LL, Chen EX. Correlation of changes between 2-year disease-free survival and 5-year overall survival in adjuvant breast cancer trials from 1966 to 2006. Ann Oncol 2008; 19(3): 481-486.

Sherrill B, Amonkar M, Wu Y, Hirst C, Stein S, Walker M et al. Relationship between effects on time-to-disease progression and overall survival in studies of metastatic breast cancer. Br J Cancer 2008; 99(10): 1572-1578.

Korrelationsstudien

Brain E, Garrino C, Misset JL, Carbonero IG, Itzhaki M, Cvitkovic E et al. Long-term prognostic and predictive factors in 107 stage II/III breast cancer patients treated with anthracycline-based neoadjuvant chemotherapy. Br J Cancer 1997; 75(9): 1360-1367.

Budd GT, Cristofanilli M, Ellis MJ, Stopeck A, Borden E, Miller MC et al. Circulating tumor cells versus imaging: predicting overall survival in metastatic breast cancer. Clin Cancer Res 2006; 12(21): 6403-6409.

Carey LA, Metzger R, Dees EC, Collichio F, Sartor CI, Ollila DW et al. American Joint Committee on Cancer tumor-node-metastasis stage after neoadjuvant chemotherapy and breast cancer outcome. J Natl Cancer Inst 2005; 97(15): 1137-1142.

Chang J, Powles TJ, Allred DC, Ashley SE, Clark GM, Makris A et al. Biologic markers as predictors of clinical outcome from systemic therapy for primary operable breast cancer. J Clin Oncol 1999; 17(10): 3058-3063.

Decensi A, Guerrieri-Gonzaga A, Gandini S, Serrano D, Cazzaniga M, Mora S et al. Prognostic significance of Ki-67 labeling index after short-term presurgical tamoxifen in women with ER-positive breast cancer. Ann Oncol 17.08.2010 [Epub ahead of print].

Dunnwald LK, Gralow JR, Ellis GK, Livingston RB, Linden HM, Lawton TJ et al. Residual tumor uptake of [^{99m}Tc]-sestamibi after neoadjuvant chemotherapy for locally advanced breast carcinoma predicts survival. Cancer 2005; 103(4): 680-688.

Edwards MJ, Bonadonna G, Valagussa P, Gamel JW. End points in the analysis of breast cancer survival: relapse versus death from tumor. Surgery 1998; 124(2): 197-202.

Eralp Y, Smith TL, Altundag K, Kau SW, Litton J, Valero V et al. Clinical features associated with a favorable outcome following neoadjuvant chemotherapy in women with localized breast cancer aged 35 years or younger. J Cancer Res Clin Oncol 2009; 135(1): 141-148.

Hennessy BT, Hortobagyi GN, Rouzier R, Kuerer H, Sneige N, Buzdar AU et al. Outcome after pathologic complete eradication of cytologically proven breast cancer axillary node metastases following primary chemotherapy. J Clin Oncol 2005; 23(36): 9304-9311.

Kimmick GG, Cirrincione C, Duggan DB, Bhalla K, Robert N, Berry D et al. Fifteen-year median follow-up results after neoadjuvant doxorubicin, followed by mastectomy, followed by adjuvant cyclophosphamide, methotrexate, and fluorouracil (CMF) followed by radiation for stage III breast cancer: a phase II trial (CALGB 8944). Breast Cancer Res Treat 2009; 113(3): 479-490.

Kountourakis P, Missitzis I, Doufexis D, Zobolas V, Pissakas G, Arnogiannaki N et al. Neoadjuvant sequential epirubicin and docetaxel followed by surgery-radiotherapy and post-operative docetaxel or gemcitabine/vinorelbine combination based on primary response: a multimodality approach for locally advanced breast cancer. J Cancer Res Clin Oncol 13.04.2010 [Epub ahead of print].

Lee SH, Lee J, Park J, Park SH, Lee KE, Lee SI et al. Capecitabine monotherapy in patients with anthracycline- and taxane-pretreated metastatic breast cancer. Med Oncol 2004; 21(3): 223-231.

Mazouni C, Peintinger F, Wan-Kau S, Andre F, Gonzalez-Angulo AM, Symmans WF et al. Residual ductal carcinoma in situ in patients with complete eradication of invasive breast cancer after neoadjuvant chemotherapy does not adversely affect patient outcome. J Clin Oncol 2007; 25(19): 2650-2655.

Nahleh Z, Sivasubramaniam D, Dhaliwal S, Sundarajan V, Komrokji R. Residual cancer burden in locally advanced breast cancer: a superior tool. Curr Oncol 2008; 15(6): 17-24.

Nakamura S, Yagata H, Ohno S, Yamaguchi H, Iwata H, Tsunoda N et al. Multi-center study evaluating circulating tumor cells as a surrogate for response to treatment and overall survival in metastatic breast cancer. Breast Cancer 2010; 17(3): 199-204.

Pierga JY, Mouret E, Laurence V, Dieras V, Savigioni A, Beuzeboc P et al. Prognostic factors for survival after neoadjuvant chemotherapy in operable breast cancer: the role of clinical response. Eur J Cancer 2003; 39(8): 1089-1096.

Tampellini M, Berruti A, Bitossi R, Gorzegno G, Alabiso I, Bottini A et al. Prognostic significance of changes in CA 15-3 serum levels during chemotherapy in metastatic breast cancer patients. Breast Cancer Res Treat 2006; 98(3): 241-248.

Vincent-Salomon A, Rousseau A, Jouve M, Beuzeboc P, Sigal-Zafrani B, Freneaux P et al. Proliferation markers predictive of the pathological response and disease outcome of patients

with breast carcinomas treated by anthracycline-based preoperative chemotherapy. Eur J Cancer 2004; 40(10): 1502-1508.

9 Literatur

1. Wittes J, Lakatos E, Probstfield J. Surrogate endpoints in clinical trials: cardiovascular diseases. Stat Med 1989; 8(4): 415-425.

2. Food and Drug Administration. Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics [online]. 05.2007 [Zugriff: 05.10.2010]. URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071590.pdf

3. European Medicines Agency. Guideline on the evaluation of anticancer medicinal products in man [online]. 14.12.2005 [Zugriff: 05.10.2010]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/12/WC500017748.pdf.

4. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. Stat Med 2006; 25(2): 183-203.

5. Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. Stat Methods Med Res 2008; 17(3): 303-340.

6. Mangiapane S, Velasco Garrido M. Surrogatendpunkte als Parameter der Nutzenbewertung [online]. (DIMDI Schriftenreihe Health Technology Assessment; Band 91). URL: http://portal.dimdi.de/de/hta/hta_berichte/hta250_bericht_de.pdf.

7. Burzykowski T, Molenberghs G, Buyse M (Ed). The evaluation of surrogate endpoints. New York: Springer; 2005.

8. Molenberghs G, Burzykowski T, Alonso A, Assam P, Tilahun A, Buyse M. A unified framework for the evaluation of surrogate endpoints in mental-health clinical trials. Stat Methods Med Res 2010; 19(3): 205-236.

9. Saad ED, Katz A, Buyse M. Overall survival and post-progression survival in advanced breast cancer: a review of recent randomized clinical trials. J Clin Oncol 2010; 28(11): 1958-1962.

10. Alonso A, Molenberghs G, Burzykowski T, Renard D, Geys H, Shkedy Z et al. Quantifiying the effect of the surrogate marker by informatioon gain: reply. Biometrics 2007; 63(3): 962-963.

11. Qu Y, Case M. Quantifying the effect of the surrogate marker by information gain. Biometrics 2007; 63(3): 958-962.

12. Alonso A, Molenberghs G. Surrogate marker evaluation from an information theory perspective. Biometrics 2007; 63(1): 180-186.

13. Alonso A, Molenberghs G. Evaluating time to cancer recurrence as a surrogate marker for survival from an information theory perspective. Stat Methods Med Res 2008; 17(5): 497-504.

14. Alonso A, Molenberghs G. Surrogate end points: hopes and perils. Expert Rev Pharmacoecon Outcomes Res 2008; 8(3): 255-259.

15. Baker SG. Two simple approaches for validating a binary surrogate endpoint using data from multiple trials. Stat Methods Med Res 2008; 17(5): 505-514.

16. Burzykowski T. Surrogate endpoints: wishful thinking or reality? Stat Methods Med Res 2008; 17(5): 463-466.

17. Buyse M. Use of meta-analysis for the validation of surrogate endpoints and biomarkers in cancer trials. Cancer J 2009; 15(5): 421-425.

18. Buyse M, Sargent DJ, Grothey A, Matheson A, De Gramont A. Biomarkers and surrogate end points: the challenge of statistical validation. Nat Rev Clin Oncol 2010; 7(6): 309-317.

19. Abrahantes JC, Shkedy Z, Molenberghs G. Alternative methods to evaluate trial level surrogacy. Clin Trials 2008; 5(3): 194-208.

20. Deslandes E, Chevret S. Assessing surrogacy from the joint modelling of multivariate longitudinal data and survival: application to clinical trial data on chronic lymphocytic leukaemia. Stat Med 2007; 26(30): 5411-5421.

21. Ditlevsen S, Keiding N. A comment on: statistical evaluation of biomarkers as surrogate endpoints: a literature review by C. J. Weir and R. J. Walley, Statistics in Medicine 2006; 25:183-203. Stat Med 2007; 26(6): 1415-1416.

22. Duffy SW, Treasure FP. Potential surrogate endpoints in cancer research: some considerations and examples. Pharm Stat 22.12.2009 [Epub ahead of print].

23. Dunning AJ. Comment on 'Evaluating a surrogate endpoint at three levels, with application to vaccine development' by Peter B. Gilbert, Li Qin and Stefen G. Self, Statistics in Medicine, DOI: 10.1002/sim.3122. Stat Med 2008; 27(29): 6268-6270.

24. Fleischer F, Gaschler-Markefski B, Bluhmki E. A statistical model for the dependence between progression-free survival and overall survival. Stat Med 2009; 28(21): 2669-2686.

25. Freedman L. Quantitative science methods for biomarker validation in chemoprevention trials. Cancer Biomark 2007; 3(3): 135-140.

26. Ghosh D. On assessing surrogacy in a single trial setting using a semicompeting risks paradigm. Biometrics 2009; 65(2): 521-529.

27. Ghosh D, Elliott MR, Taylor JM. Links between analysis of surrogate endpoints and endogeneity. Stat Med 2010; 29(28): 2869-2879.

28. Gilbert PB, Qin L, Self SG. Evaluating a surrogate endpoint at three levels, with application to vaccine development. Stat Med 2008; 27(23): 4758-4778.

29. Gilbert PB, Qin L, Self SG. Response to Andrew Dunning's comment on 'Evaluating a surrogate endpoint at three levels, with application to vaccine development'. Stat Med 2009; 28(4): 716-719.

30. Green E, Yothers G, Sargent DJ. Surrogate endpoint validation: statistical elegance versus clinical relevance. Stat Methods Med Res 2008; 17(5): 477-486.

31. Huang J, Huang B. Evaluating the proportion of treatment effect explained by a continuous surrogate marker in logistic or probit regression models. Stat Biopharm Res 2010; 2(2): 229-238.

32. Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. Biometrics 2009; 65(2): 530-538.

33. Lassere M, Johnson K, Hughes M, Altman D, Buyse M, Galbraith S et al. Simulation studies of surrogate endpoint validation using single trial and multitrial statistical approaches. J Rheumatol 2007; 34(3): 616-619.

34. Lassere MN, Johnson KR, Boers M, Tugwell P, Brooks P, Simon L et al. Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. J Rheumatol 2007; 34(3): 607-615.

35. Li Y, Taylor JM, Elliott MR. A Bayesian approach to surrogacy assessment using principal stratification in clinical trials. Biometrics 2010; 66(2): 523-531.

36. MacKinnon DP, Lockwood CM, Brown CH, Wang W, Hoffman JM. The intermediate endpoint effect in logistic and probit regression. Clin Trials 2007; 4(5): 499-513.

37. Piedbois P, Buyse M. Endpoints and surrogate endpoints in colorectal cancer: a review of recent developments. Curr Opin Oncol 2008; 20(4): 466-471.

38. Pryseley A, Tilahun A, Alonso A, Molenberghs G. Information-theory based surrogate marker evaluation from several randomized clinical trials with continuous true and binary surrogate endpoints. Clin Trials 2007; 4(6): 587-597.

39. Sarkar S, Qu Y. Quantifying the treatment effect explained by markers in the presence of measurement error. Stat Med 2007; 26(9): 1955-1963.

40. Sertdemir Y, Burgut R. Does the decision in a validation process of a surrogate endpoint change with level of significance of treatment effect? A proposal on validation of surrogate endpoints. Contemp Clin Trials 2009; 30(1): 8-12.

41. Shi Q, Sargent DJ. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. Int J Clin Oncol 2009; 14(2): 102-111.

42. Tilahun A, Maringwa JT, Geys H, Alonso A, Raeymaekers L, Molenberghs G et al. Investigating association between behavior, corticosterone, heart rate, and blood pressure in rats using surrogate marker evaluation methodology. J Biopharm Stat 2009; 19(1): 133-149.

43. Tilahun A, Pryseley A, Alonso A, Molenberghs G. Information theory-based surrogate marker evaluation from several randomized clinical trials with binary endpoints, using SAS. J Biopharm Stat 2008; 18(2): 326-341.

44. Van Walraven C, Oake N, Coyle D, Taljaard M, Forster AJ. Changes in surrogate outcomes can be translated into clinical outcomes using a Monte Carlo model. J Clin Epidemiol 2009; 62(12): 1306-1315.

45. Wolfson J, Gilbert P. Statistical identifiability and the surrogate endpoint problem, with application to vaccine trials. Biometrics 22.01.2010 [Epub ahead of print].

46. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001; 69(3): 89-95.

47. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989; 8(4): 431-440.

48. Frangakis CE, Rubin DB. Principal stratification in cause inference. Biometrics 2002; 58(1): 21-29.

49. Burzykowski T, Buyse M. An alternative measure for mety-analytic surrogate endpoint validation. In: Burzykowski T, Molenberghs G, Buyse M (Ed). The evaluation of surrogate endpoints. New York: Springer; 2005. S. 323-340.

50. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006; 5(3): 173-186.

51. Alonso A, Geys H, Molenberghs G, Kenward MG, Vangeneugden T. Validation of surrogate markers in multiple randomized clinical trials with repeated measurements. Biom J 2003; 45(8): 931-945.

52. Alonso A, Molenberghs G, Geys H, Buyse M, Vangeneugden T. A unifying approach for surrogate marker validation based on Prentice's criteria. Stat Med 2006; 25(2): 205-221.

53. Ellenberg S, Hamilton JM. Surrogate endpoints in clinical trials: cancer. Stat Med 1989; 8(4): 405-413.

54. Alonso A, Geys H, Vangeneugden T. Repeated measures and surrogate endpoint validation. In: Burzykowski T, Molenberghs G, Buyse M (Ed). The evaluation of surrogate endpoints. New York: Springer; 2005. S. 231-251.

55. Alonso A, Molenberghs G, Burzykowski T, Renard D, Geys H, Shkedy Z et al. Prentice's approach and the meta-analytic paradigm: a reflection on the role of statistics in the evaluation of surrogate endpoints. Biometrics 2004; 60(3): 724-728.

56. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology 1992; 3(2): 143-155.

57. Xu J, Zeger SL. The evaluation of multiple surrogate endpoints. Biometrics 2001; 57(1): 81-87.

58. Ditlevsen S, Christensen U, Lynch J, Damsgaard MT, Keiding N. The mediation proportion: a structural equation approach for estimating the proportion of exposure effect on outcome explained by an intermediate variable. Epidemiology 2005; 16(1): 114-120.

59. Burzykowski T, Molenberghs G. The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer. J R Statist Soc A 2004; 167(1): 103-124.

60. Baker SG. A simple meta-analytic approach for using a binary surrogate endpoint to predict the effect of intervention on true endpoint. Biostatistics 2006; 7(1): 58-70.

61. Burzykowski T, Buyse M, Yothers G, Sakamoto J, Sargent D. Exploring and validating surrogate endpoints in colorectal cancer. Lifetime Data Anal 2008; 14(1): 54-64.

62. Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Lancet 2000; 356(9227): 373-378.

63. Buyse M, Burzykowski T, Carroll K, Michiels S, Sargent DJ, Miller LL et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. J Clin Oncol 2007; 25(33): 5218-5224.

64. Buyse M, Burzykowski T, Michiels S, Carroll K. Individual- and trial-level surrogacy in colorectal cancer. Stat Methods Med Res 2008; 17(5): 467-475.

65. Johnson KR, Ringland C, Stokes BJ, Anthony DM, Freemantle N, Irs A et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. Lancet Oncol 2006; 7(9): 741-746.

66. Montagnani F, Migali C, Fiorentini G. Progression-free survival in bevacizumab-based first-line treatment for patients with metastatic colorectal cancer: is it a really good end point? J Clin Oncol 2009; 27(28): e132-e133.

67. Sargent DJ, Wieand HS, Haller DG, Gray R, Benedetti JK, Buyse M et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. J Clin Oncol 2005; 23(34): 8664-8670.

68. Sargent DJ, Patiyil S, Yothers G, Haller DG, Gray R, Benedetti J et al. End points for colon cancer adjuvant trials: observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT Group. J Clin Oncol 2007; 25(29): 4569-4574.

69. Tang PA, Bentzen SM, Chen EX, Siu LL. Surrogate end points for median overall survival in metastatic colorectal cancer: literature-based analysis from 39 randomized controlled trials of first-line chemotherapy. J Clin Oncol 2007; 25(29): 4562-4568.

70. Bruzzi P, Del Mastro L, Sormani MP, Bastholt L, Danova M, Focan C et al. Objective response to chemotherapy as a potential surrogate end point of survival in metastatic breast cancer patients. J Clin Oncol 2005; 23(22): 5117-5125.

71. Burzykowski T, Buyse M, Piccart-Gebhart MJ, Sledge G, Carmichael J, Luck HJ et al. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. J Clin Oncol 2008; 26(12): 1987-1992.

72. Hackshaw A, Knight A, Barrett-Lee P, Leonard R. Surrogate markers and survival in women receiving first-line combination anthracycline chemotherapy for advanced breast cancer. Br J Cancer 2005; 93(11): 1215-1221.

73. Miksad RA, Zietemann V, Gothe R, Schwarzer R, Conrads-Frank A, Schnell-Inderst P et al. Progression-free survival as a surrogate endpoint in advanced breast cancer. Int J Technol Assess Health Care 2008; 24(4): 371-383.

74. Ng R, Pond GR, Tang PA, MacIntosh PW, Siu LL, Chen EX. Correlation of changes between 2-year disease-free survival and 5-year overall survival in adjuvant breast cancer trials from 1966 to 2006. Ann Oncol 2008; 19(3): 481-486.

75. Sherrill B, Amonkar M, Wu Y, Hirst C, Stein S, Walker M et al. Relationship between effects on time-to-disease progression and overall survival in studies of metastatic breast cancer. Br J Cancer 2008; 99(10): 1572-1578.

76. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics 2000; 1(1): 49-67.

77. Burzykowski T, Buyse M, Geys H, Renard D. Validation of surrogate endpoints in randomised clinical trials with failure time endpoints. Appl Stat 2001; 50(4): 405-422.

78. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Stat Med 1992; 11(2): 167-178.

79. Begg C, Leung D. On the use of surrogate end points in randomized trials. J R Stat Soc Ser A Stat Soc 2000; 163(1): 15-28.

80. Wolmark N, Fisher B, Rockette H, Redmond C, Wickerham DL, Fisher ER et al. Postoperative adjuvant chemotherapy or BCG for colon cancer: results from NSABP protocol C-01. J Natl Cancer Inst 1988; 80(1): 30-36.

81. Wolmark N, Rockette H, Wickerham DL, Fisher B, Redmond C, Fisher ER et al. Adjuvant therapy of Dukes' A, B, and C adenocarcinoma of the colon with portal-vein fluorouracil hepatic infusion: preliminary results of National Surgical Adjuvant Breast and Bowel Project protocol C-02. J Clin Oncol 1990; 8(9): 1466-1475.

82. Wolmark N, Rockette H, Fisher B, Wickerham DL, Redmond C, Fisher ER et al. The benefit of leucovorin-modulated fluorouracil as postoperative adjuvant therapy for primary colon cancer: results from National Surgical Adjuvant Breast and Bowel Project protocol C-03. J Clin Oncol 1993; 11(10): 1879-1887.

83. Wolmark N, Rockette H, Mamounas E, Jones J, Wieand S, Wickerham DL et al. Clinical trial to assess the relative efficacy of fluorouracil and leucovorin, fluorouracil and levamisole, and fluorouracil, leucovorin, and levamisole in patients with Dukes' B and C carcinoma of the colon: results from National Surgical Adjuvant Breast and Bowel Project C-04. J Clin Oncol 1999; 17(11): 3553-3559.

84. Advanced Colorectal Cancer Meta-Analysis Project. Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: evidence in terms of response rate. J Clin Oncol 1992; 10(6): 896-903.

85. Advanced Colorectal Cancer Meta-Analysis Project. Meta-analysis of randomized trials testing the biochemical modulation of fluorouracil by methotrexate in metastatic colorectal cancer. J Clin Oncol 1994; 12(5): 960-969.

86. Meta-Analysis Group in Cancer. Reappraisal of hepatic arterial infusion in the treatment of nonresectable liver metastases from colorectal cancer. J Natl Cancer Inst 1996; 88(5): 252-258.

87. Meta-analysis Group in Cancer. Efficacy of intravenous continuous infusion of fluorouracil compared with bolus administration in advanced colorectal cancer. J Clin Oncol 1998; 16(1): 301-308.

88. Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. Biometrics 1998; 54(3): 1014-1029.

89. Lin DY, Fleming TR, De Gruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. Stat Med 1997; 16(13): 1515-1527.

90. Corfu-A Study Group. Phase III randomized study of two fluorouracil combinations with either interferon alfa-2a or leucovorin for advanced colorectal cancer. J Clin Oncol 1995; 13(4): 921-928.

91. Greco FA, Figlin R, York M, Einhorn L, Schilsky R, Marshall EM et al. Phase III randomized study to compare interferon alfa-2a in combination with fluorouracil versus fluorouracil alone in patients with advanced colorectal cancer. J Clin Oncol 1996; 14(10): 2674-2681.

92. Ding CG. On the computation of the distribution of the square of the sample multiple correlation coefficient. Comput Stat Data Anal 1996; 22(4): 345-350.

93. Chu TM, Weir B, Wolfinger R. A systematic statistical linear modeling approach to oligonucleotide array experiments. Math Biosci 2002; 176(1): 35-51.

94. Piccart-Gebhart MJ, Burzykowski T, Buyse M, Sledge G, Carmichael J, Luck HJ et al. Taxanes alone or in combination with anthracyclines as first-line therapy of patients with metastatic breast cancer. J Clin Oncol 2008; 26(12): 1980-1986.

95. A'Hern RP, Ebbs SR, Baum MB. Does chemotherapy improve survival in advanced breast cancer? A statistical overview. Br J Cancer 1988; 57(6): 615-618.

96. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med 1996; 125(7): 605-613.

97. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. Health Aff (Millwood) 2005; 24(1): 67-78.

Anhang A – Suchstrategien

Teilziel 1

1. MEDLINE

Suchoberfläche: Ovid

- Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations
- Ovid MEDLINE(R) Daily
- Ovid MEDLINE(R) 1950 to Present

Die Strategie beruht teilweise auf Weir 2006 [4]

#	**Searches**
1	(surrogate or endpoint or end point).ab,ti.
2	Endpoint Determination/
3	or/1-2
4	Clinical Trials as Topic/mt, sn
5	Randomized Controlled Trials as Topic/mt, sn
6	Endpoint Determination/mt, sn
7	(valid or validity or validat*).ab,ti.
8	Biological Markers/
9	Tumor Markers, Biological/
10	biomarker*.ab,ti.
11	7 and (8 or 9 or 10)
12	or/4-6
13	11 or 12
14	3 and 13
15	(statist* and (validat* or evaluat*) and (surrogate or biomarker)).mp.
16	or/14-15
17	limit 16 to (english or german)
18	limit 17 to yr="2007 -Current"
19	*[Abgleich mit Pubmed]*
20	18 NOT 19

2. Pubmed

Suchoberfläche: NLM

- PubMed - as supplied by publisher
- PubMed - in process
- PubMed - OLDMEDLINE

Search	Most Recent Queries
#1	Search surrogate[TIAB] OR endpoint[TIAB] OR end point[TIAB]
#2	Search valid[TIAB] OR validity[TIAB] OR validat*[TIAB]
#3	Search biomarker*[TIAB]
#4	Search #1 AND #2 AND #3
#5	Search statist* AND (validat* OR evaluat*) AND (surrogate OR biomarker)
#6	Search #4 OR #5
#7	Search #6 NOT medline[sb]
#8	Search #7 Limits: English, German, Publication Date from 2007 to 2011

3. The Cochrane Library

Suchoberfläche: Wiley

- Cochrane Methodology Register (Methods Studies)
- Health Technology Assessment Databse (Technology Assessments)

ID	Search
#1	(surrogate or endpoint or end point):ab,ti
#2	MeSH descriptor Endpoint Determination explode all trees
#3	(#1 OR #2)
#4	MeSH descriptor Clinical Trials as Topic explode all trees with qualifiers: MT,SN
#5	MeSH descriptor Randomized Controlled Trials as Topic explode all trees with qualifiers: MT,SN
#6	(valid or validity or validat*):ab,ti
#7	MeSH descriptor Biological Markers, this term only
#8	MeSH descriptor Tumor Markers, Biological, this term only
#9	biomarker*:ab,ti
#10	(#6 AND (#7 OR #8 OR #9))
#11	(#4 OR #5)
#12	(#10 OR #11)
#13	(#3 AND #12)

ID	Search
#14	(statist* and (validat* or evaluat*) and (surrogate or biomarker))
#15	(#13 OR #14)
#16	(#15), from 2007 to 2010
#17	(#16)

Teilziel 2 – Krankheitsbild Darmkrebs

1. EMBASE

Suchoberfläche: Ovid

EMBASE (1980 to 2010 August 23)

#	**Searches**
1	exp Colon Cancer/
2	exp Rectum Cancer/
3	((colorectal* or rectal* or colon* or sigma* or sigmo* or rectum*) and (cancer? or metastas* or tumor? or tumour? or carcinom*)).ab,ti.
4	or/1-3
5	(surrogate or endpoint* or end point*).ab,ti.
6	overall survival/
7	disease free survival/
8	progression free survival.ab,ti.
9	time to progression.ti,ab.
10	disease free survival.ab,ti.
11	objective response rate.ab,ti.
12	time to treatment failure.ab,ti.
13	overall survival.ab,ti.
14	or/6-13
15	validation process/
16	validation study/
17	(valid or validity or validat*).ab,ti.
18	or/15-17
19	biological marker/
20	disease marker/
21	(biomarker* or marker*).ab,ti.
22	or/19-21
23	and/18,22
24	(statist* and (validat* or evaluat*) and (surrogate or biomarker)).mp.
25	5 and (14 or 23)
26	24 or 25
27	4 and 26

2. MEDLINE

Suchoberfläche: Ovid

- Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations
- Ovid MEDLINE(R) Daily
- Ovid MEDLINE(R) 1950 to Present

Die Strategie beruht teilweise auf Weir 2006 [4]

#	**Searches**
1	exp Colorectal Neoplasms/
2	((colorectal* or rectal* or colon* or sigma* or sigmo* or rectum*) and (cancer? or metastas* or tumor? or tumour? or carcinom*)).ab,ti.
3	or/1-2
4	(surrogate or endpoint* or end point*).ab,ti.
5	Endpoint Determination/
6	or/4-5
7	progression free survival.ab,ti.
8	time to progression.ti,ab.
9	disease free survival.ab,ti.
10	Disease-Free Survival/
11	objective response rate.ab,ti.
12	time to treatment failure.ab,ti.
13	overall survival.ab,ti.
14	or/7-13
15	Clinical Trials as Topic/mt [Methods]
16	Clinical Trials as Topic/sn [Statistics & Numerical Data]
17	Randomized Controlled Trials as Topic/mt
18	Randomized Controlled Trials as Topic/sn
19	Endpoint Determination/mt
20	Endpoint Determination/sn
21	(valid or validity or validat*).ab,ti.
22	Biological Markers/
23	Tumor Markers, Biological/
24	(biomarker* or marker*).ab,ti.
25	21 and (22 or 23 or 24)
26	or/15-20
27	25 or 26
28	(statist* and (validat* or evaluat*) and (surrogate or biomarker)).mp.

#	**Searches**
29	6 and (14 or 27)
30	28 or 29
31	3 and 30

3. Pubmed

Suchoberfläche: NLM

- PubMed - as supplied by publisher
- PubMed - in process
- PubMed - OLDMEDLINE

Search	**Most Recent Queries**
#5	Search ((colorectal* OR rectal* OR colon* OR sigma* OR sigmo* OR rectum*) and (cancer* OR metastas* OR tumor OR tumour* OR carcinom*))
#6	Search (surrogate or endpoint* or end point*)
#7	Search progression free survival
#8	Search time to progression
#9	Search disease free survival
#10	Search objective response rate
#11	Search time to treatment failure
#12	Search overall survival
#13	Search #7 OR #8 OR #9 OR #10 OR #11 OR #12
#14	Search valid or validity or validat*
#15	Search biomarker* or marker*
#16	Search #14 AND #15
#17	Search (statist* AND (validat* OR evaluat*) AND (surrogate OR biomarker))
#18	Search #6 AND (#13 OR #16)
#19	Search #17 OR #18
#20	Search #5 AND #19
#21	Search #20 NOT medline[sb]

4. The Cochrane Library

Suchoberfläche: Wiley

- Cochrane Methodology Register (Methods Studies)

ID	Search
#1	MeSH descriptor Colorectal Neoplasms explode all trees
#2	((colorectal* or rectal* or colon* or sigma* or sigmo* or rectum*) and (cancer* or metastas* or tumor* or tumour* or carcinom*)):ti,ab
#3	(#1 OR #2)
#4	surrogate or endpoint* or end point* or biomarker* or marker*
#5	Survival
#6	(#4 OR #5)
#7	(#3 AND #6)

Teilziel 2 – Krankheitsbild Brustkrebs

1. EMBASE

Suchoberfläche: Ovid

EMBASE (1980 to 2010 October 06)

#	**Searches**
1	((breast or mamma*) adj6 (cancer* or carcinom* or tumor* or tumour* or neoplasm*)).ab,ti.
2	exp breast cancer/
3	1 or 2
4	(surrogate or endpoint* or end point*).ab,ti.
5	overall survival/
6	disease free survival/
7	overall survival.ab,ti.
8	(progression free survival or disease free survival or event free survival).ab,ti.
9	(complete response or partial response or objective resonse).ab,ti.
10	time to treatment failure.ab,ti.
11	(time to progression or time to tumo?r progression).ab,ti.
12	or/5-11
13	validation process/
14	validation study/
15	(valid or validity or validat* or evaluating).ab,ti.
16	or/13-15
17	biological marker/
18	disease marker/
19	biomarker*.ab,ti.
20	or/17-19
21	and/16,20
22	3 and 4 and (12 or 21)

2. MEDLINE

Suchoberfläche: Ovid

- Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations
- Ovid MEDLINE(R) Daily
- Ovid MEDLINE(R) 1950 to Present

Die Strategie beruht teilweise auf Weir 2006 [4]

#	**Searches**
1	((breast or mamma*) adj6 (cancer* or carcinom* or tumor* or tumour* or neoplasm*)).ab,ti.
2	exp Breast Neoplasms/
3	1 or 2
4	(surrogate or endpoint* or end point*).ab,ti.
5	Endpoint Determination/
6	or/4-5
7	Disease-Free Survival/
8	overall survival.ab,ti.
9	(progression free survival or disease free survival or event free survival).ab,ti.
10	(complete response or partial response or objective resonse).ab,ti.
11	time to treatment failure.ab,ti.
12	(time to progression or time to tumo?r progression).ab,ti.
13	or/7-12
14	Clinical Trials as Topic/mt, sn
15	Randomized Controlled Trials as Topic/mt, sn
16	Endpoint Determination/mt, sn
17	(valid or validity or validat* or evaluating).ab,ti.
18	Biological Markers/
19	Tumor Markers, Biological/
20	biomarker*.ab,ti.
21	17 and (18 or 19 or 20)
22	or/14-16
23	13 or 21 or 22
24	and/3,6,23
25	*[Abgleich mit Pubmed]*
26	24 not 25

3. Pubmed

Suchoberfläche: NLM

- PubMed - as supplied by publisher
- PubMed - in process
- PubMed - OLDMEDLINE

Search	Most Recent Queries
#1	Search ((breast[TIAB] or mamma*[TIAB]) AND (cancer*[TIAB] or carcinom*[TIAB] or tumor*[TIAB] or tumour*[TIAB] or neoplasm*[TIAB]))
#2	Search (surrogate or endpoint* or end point*)
#3	Search overall survival[TIAB]
#4	Search "progression free survival"[TIAB] OR "disease free survival"[TIAB] OR "event free survival"[TIAB]
#5	Search "complete response"[TIAB] OR "partial response"[TIAB] OR "objective resonse"[TIAB]
#6	Search "time to treatment failure"[TIAB]
#7	Search "time to progression"[TIAB] OR "time to tumor progression"[TIAB] OR "time to tumour progression"[TIAB]
#8	Search valid[TIAB] OR validity[TIAB] OR validat*[TIAB] OR evaluating[TIAB]
#9	Search biomarker*[TIAB]
#10	Search #8 AND #9
#12	Search #1 AND #2 AND (#3 OR #4 OR #5 OR #6 OR #7 OR #10)
#13	Search #12 NOT medline[sb]
#14	Search #13 Limits: published in the last 2 years

4. The Cochrane Library

Suchoberfläche: Wiley

- Cochrane Methodology Register (Methods Studies)

ID	Search
#1	(surrogate or endpoint* or end point*):ab,ti
#2	MeSH descriptor Endpoint Determination explode all trees
#3	(#1 OR #2)
#4	MeSH descriptor Clinical Trials as Topic explode all trees with qualifiers: MT,SN
#5	MeSH descriptor Randomized Controlled Trials as Topic explode all trees with qualifiers: MT,SN
#6	(valid or validity or validat* or evaluating):ab,ti
#7	MeSH descriptor Biological Markers, this term only
#8	MeSH descriptor Tumor Markers, Biological, this term only
#9	biomarker*:ab,ti
#10	(#6 AND (#7 OR #8 OR #9))
#11	(#4 OR #5)
#12	MeSH descriptor Disease-Free Survival explode all trees
#13	(overall survival):ab,ti
#14	(progression free survival or disease free survival or event free survival):ab,ti
#15	(complete response or partial response or objective resonse):ab,ti
#16	(time to treatment failure):ab,ti
#17	(time to progression or time to tumor progression or time to tumour progression):ab,ti
#18	(#12 OR #13 OR #14 OR #15 OR #16 OR #17)
#19	MeSH descriptor Breast Neoplasms explode all trees
#20	((breast or mamma*) NEAR/6 (cancer* or carcinom* or tumor* or tumour* or neoplasm*)):ab,ti
#21	(#19 OR #20)
#22	(#3 AND #21 AND (#10 OR #11 OR #18))
#23	(#22)

Anhang B – Liste der ausgeschlossenen Dokumente zum Thema mit Ausschlussgründen

B.1 – Liste der ausgeschlossenen Dokumente zu Teilziel 1 (methodische Verfahren zur Validierung von Surrogatendpunkten)

Nicht ME1

1. Bucher HC. Studien mit Surrogatendpunkten: Nutzen und Grenzen in der klinischen Entscheidungsfindung. Internist 2008; 49(6): 681-687.

2. Bucher HC. Nutzen und Grenzen von Studien mit Surrogat-Endpunkten in der klinischen Entscheidungsfindung. Z Evid Fortbild Qual Gesundhwes 2010; 104(3): 230-237.

3. Chow SC, Lu Q, Tse SK, Chi E. Statistical methods for assessment of biosimilarity using biomarker data. J Biopharm Stat 2010; 20(1): 90-105.

4. Cleophas TJ, Zwinderman AH, Chaib AH. Novel procedures for validating surrogate endpoints in clinical trials. Curr Clin Pharmacol 2007; 2(2): 123-128.

5. Freemantle N, Calvert M. Composite and surrogate outcomes in randomised controlled trials. BMJ 2007; 334(7597): 756-757.

6. Johnson KR, Freemantle N, Anthony DM, Lassere MND. LDL-cholesterol differences predicted survival benefit in statin trials by the surrogate threshold effect (STE). J Clin Epidemiol 2009; 62(3): 328-336.

7. Piedbois P, Miller Croswell J. Surrogate endpoints for overall survival in advanced colorectal cancer: a clinician's perspective. Stat Methods Med Res 2008; 17(5): 519-527.

8. Prentice RL. Surrogate and mediating endpoints: current status and future directions. J Natl Cancer Inst 2009; 101(4): 216-217.

9. Rigatto C, Barrett BJ. Biomarkers and surrogates in clinical studies. In: Parfrey P, Barrett B (Ed). Clinical epidemiology: practice and methods. New York: Humana; 2009. S. 137-154. (Methods in molecular biology; Band 473).

10. Shariat SF, Lotan Y, Vickers A, Karakiewicz PI, Schmitz-Drager BJ, Goebell PJ et al. Statistical consideration for clinical biomarker research in bladder cancer. Urol Oncol 2010; 28(4): 389-400.

11. Velasco Garrido M, Mangiapane S. Surrogate outcomes in health technology assessment: an international comparison. Int J Technol Assess Health Care 2009; 25(3): 315-322.

12. Wang X, Wu Y, Zhou H. Outcome- and auxiliary-dependent subsampling and its statistical inference. J Biopharm Stat 2009; 19(6): 1132-1150.

B.2 – Liste der ausgeschlossenen Dokumente zu Teilziel 2 - Krankheitsbild Darmkrebs

Nicht VE1

1. Bowen PE, Mobarhan S. Evidence from cancer intervention and biomarker studies and the development of biochemical markers. Am J Clin Nutr 1995; 62(6 Suppl): 1403S-1409S.

2. Folprecht G, Grothey A, Alberts S, Raab HR, Kohne CH. Neoadjuvant treatment of unresectable colorectal liver metastases: correlation between tumour response and resection rates. Ann Oncol 2005; 16(8): 1311-1319.

3. Grambsch P, Louis TA, Bostick RM, Grandits GA, Fosdick L, Darif M et al. Statistical analysis of proliferative index data in clinical trials. Stat Med 1994; 13(16): 1619-1634.

4. Mir O, Ropert S, Alexandre J, Goldwasser F. Hypertension as a surrogate marker for the activity of anti-VEGF agents. Ann Oncol 2009; 20(5): 967-970.

5. Murukesh N, Dive C, Jayson GC. Biomarkers of angiogenesis and their role in the development of VEGF inhibitors. Br J Cancer 2010; 102(1): 8-18.

6. Sanderson P, Johnson IT, Mathers JC, Powers HJ, Downes CS, McGlynn AP et al. Emerging diet-related surrogate end points for colorectal cancer: UK Food Standards Agency diet and colonic health workshop report. Br J Nutr 2004; 91(2): 315-323.

7. Sobrero A, Aprile G. Resectability of initially unresectable liver metastases from colorectal cancer should not be the primary end point of clinical trials. Gastrointest Cancer Res 2009; 3(1): 41-42.

8. Tochetto SM, Rezai P, Rezvani M, Nikolaidis P, Berggruen S, Atassi B et al. Does multidetector CT attenuation change in colon cancer liver metastases treated with 90Y help predict metabolic activity at FDG PET? Radiology 2010; 255(1): 164-172.

Nicht VE2

1. Bruzzi P. Objective response to treatment as a potential surrogate marker of survival in breast cancer. Ann N Y Acad Sci 2002; 963: 144-147.

2. Burzykowski T, Bang Y. Disease-free survival as a surrogate endpoint for overall survival in an adjuvant trial of curatively resected stomach cancer using individual patient data meta-analysis. J Clin Oncol 2009; 27(15 Suppl 1): 4517.

3. Johnson K, Hill S, Stokes B, Berlin J, Freemantle N, Anthony D. Developing methodology for validation and comparison of surrogate markers in randomized clinical trials [P-066] [online]. In: Cochrane Colloquium Abstracts: 12th Cochrane Colloquium: bridging the gaps; 02.-04.10.2004; Ottawa, Kanada. 2004 [Zugriff: 07.10.2010]. URL: http://www.imbi.uni-freiburg.de/OJS/cca/index.php/cca/article/view/2678.

4. Methy N, Bedenne L, Conroy T, Bouche O, Chapet O, Ducreux M et al. Surrogate end points for overall survival and local control in neoadjuvant rectal cancer trials: statistical evaluation based on the FFCD 9203 trial. Ann Oncol 2010; 21(3): 518-524.

5. Sherrill B, Amonkar M, Wu Y, Hirst C, Stein S, Walker M et al. Relationship between effects on time-to-disease progression and overall survival in studies of metastatic breast cancer. Br J Cancer 2008; 99(10): 1572-1578.

6. Tanemura A, Van Hoesel AQ, Mori T, Yu T, Hoon DSB. The role of estrogen receptor in melanoma. Expert Opin Ther Targets 2007; 11(12): 1639-1648.

Nicht VE5

1. Sargent DJ, Yothers G, Van Cutsem E, Cassidy J, Saltz L, Wolmark N et al. Use of two-year disease-free survival (DFS) as a primary endpoint in stage III adjuvant colon cancer trials with fluoropyrimidines with or without oxaliplatin or irinotecan: new data from 12,676 patients from MOSAIC, X-ACT, PETACC-3, NSAPB C-06 and C-07, and C89803. J Clin Oncol 2009; 27(15 Suppl 1): 4011.

B.3 – Liste der ausgeschlossenen Dokumente zu Teilziel 2 - Krankheitsbild Brustkrebs

Nicht VE1

1. Akashi-Tanaka S, Omatsu M, Shimizu C, Ando M, Terada K, Shien T et al. Favorable outcome in patients with breast cancer in the presence of pathological response after neoadjuvant endocrine therapy. Breast 2007; 16(5): 482-488.

2. Basik M, Aguilar-Mahecha A, Chapman JW, Jahan K, Hassan S, Han L et al. Use of serum SDF-1 as a predictive biomarker for outcomes in the NCIC CTG MA.14 trial of octreotide, a somatostatin analogue in postmenopausal breast cancer. J Clin Oncol 2010; 28(15 Suppl 1): 551.

3. Bidard FC, Mathiot C, Degeorges A, Etienne-Grimaldi MC, Delva R, Pivot X et al. Clinical value of circulating endothelial cells and circulating tumor cells in metastatic breast cancer patients treated first line with bevacizumab and chemotherapy. Ann Oncol 2010; 21(9): 1765-1771.

4. Bidard FC, Mathiot C, Degeorges A, Formento P, Delva R, Pivot X et al. CEC and CTC in stage IV breast cancers treated with bevacizumab (Bv) and chemotherapy (CT). European Journal of Cancer Supplements 2010; 8(3): 90.

5. Byrne GJ, Ghellal A, Iddon J, Blann AD, Venizelos V, Kumar S et al. Serum soluble vascular cell adhesion molecule-1: role as a surrogate marker of angiogenesis. J Natl Cancer Inst 2000; 92(16): 1329-1336.

6. Cheng JP, Yan Y, Wang XY, Lu YL, Yuan YH, Wang XL et al. Circulating tumor cells in metastatic breast cancer: monitoring response to chemotherapy and predicting progression-free survival. Chin J Cancer Res 2010; 22(3): 201-210.

7. Dose Schwarz J, Bader M, Jenicke L, Hemminger G, Janicke F, Avril N. Early prediction of response to chemotherapy in metastatic breast cancer using sequential 18F-FDG PET. J Nucl Med 2005; 46(7): 1144-1150.

8. Fisher B, Bryant J, Wolmark N, Mamounas E, Brown A, Fisher ER et al. Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. J Clin Oncol 1998; 16(8): 2672-2685.

9. Harry VN, Semple SI, Parkin DE, Gilbert FJ. Use of new imaging techniques to predict tumour response to therapy. Lancet Oncol 2010; 11(1): 92-102.

10. Jacobs MA, Stearns V, Wolff AC, Macura K, Argani P, Khouri N et al. Multiparametric magnetic resonance imaging, spectroscopy and multinuclear (^{23}Na) imaging monitoring of preoperative chemotherapy for locally advanced breast cancer. Acad Radiol 2010; 17(12): 1477-1485.

11. Jang SH, Hwang YI, Lee JC, Yang SH, Park SH, Kim DG et al. Early-onset chemotherapy-induced neutropenia predicts better prognosis in patients with advanced non-small-cell lung cancer. Clin Lung Cancer 2009; 10(5): E7.

12. Lee KS, Ro J, Nam BH, Lee ES, Kwon Y, Kwon HS et al. A randomized phase-III trial of docetaxel/capecitabine versus doxorubicin/cyclophosphamide as primary chemotherapy for patients with stage II/III breast cancer. Breast Cancer Res Treat 2008; 109(3): 481-489.

13. Lucidarme O. New non-invasive imaging techniques in lung cancer (excluding PET). J Thorac Oncol 2010; 5(5 Suppl 1): S30.

14. Lumachi F, Basso SMM. Serum tumor markers in patients with breast cancer. Expert Rev Anticancer Ther 2004; 4(5): 921-931.

15. Luoma ML, Hakamies-Blomqvist L, Sjostrom J, Pluzanska A, Ottoson S, Mouridsen H et al. Prognostic value of quality of life scores for time to progression (TTP) and overall survival time (OS) in advanced breast cancer. Eur J Cancer 2003; 39(10): 1370-1376.

16. Miller KD, Soule SE, Calley C, Emerson RE, Hutchins GD, Kopecky K et al. Randomized phase II trial of the anti-angiogenic potential of doxorubicin and docetaxel: primary chemotherapy as biomarker discovery laboratory. Breast Cancer Res Treat 2005; 89(2): 187-197.

17. Nunes RA, Li X, Kang SP, Burstein H, Roberts L, Carney W et al. Circulating tumor cells in HER-2 positive metastatic breast cancer patients treated with trastuzumab and chemotherapy. Int J Biol Markers 2009; 24(1): 1-10.

18. Ramaswamy B, Elias AD, Kelbick NT, Dodley A, Morrow M, Hauger M et al. Phase II trial of bevacizumab in combination with weekly docetaxel in metastatic breast cancer patients. Clin Cancer Res 2006; 12(10): 3124-3129.

19. Rugo HS, Dickler M, Franco S, Stopeck A, Lyandres J, Melisko M et al. Circulating tumor cell and endothelial cell data from a phase II evaluation of lapatinib and bevacizumab in HER2-overexpressing metastatic breast cancer. Cancer Res 2009; 69(2 Suppl S): Abstract 3154.

20. Specht JM, Tam SL, Kurland BF, Gralow JR, Livingston RB, Linden HM et al. Serial 2-[^{18}F]fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) to monitor treatment of bone-dominant metastatic breast cancer predicts time to progression (TTP). Breast Cancer Res Treat 2007; 105(1): 87-94.

21. Urtishak S, Alpaugh RK, Weiner LM, Swaby RF. Clinical utility of circulating tumor cells: a role for monitoring response to therapy and drug development. Biomarkers in Medicine 2008; 2(2): 137-145.

22. Vincenzi B, Santini D, Dicuonzo G, Battistoni F, Gavasci M, La Cesa A et al. Zoledronic acid-related angiogenesis modifications and survival in advanced breast cancer patients. J Interferon Cytokine Res 2005; 25(3): 144-151.

23. Vyzula R, Dusek L, Zaloudik J, Demlova R, Klimes D, Selvekerova S. Breast cancer and neoadjuvant therapy: any predictive marker? Neoplasma 2004; 51(6): 471-480.

24. Zhao X, Xu X, Guo L, Ragaz J, Guo H, Wu J et al. Biomarker alterations with metronomic use of low-dose zoledronic acid for breast cancer patients with bone metastases and potential clinical significance. Breast Cancer Res Treat 2010; 124(3): 733-743.

Nicht VE2

1. Berry W, Eisenberger M. Achieving treatment goals for hormone-refractory prostate cancer with chemotherapy. Oncologist 2005; 10(Suppl 3): 30-39.

2. Kayser S, Doehner K, Krauter J, Horst HA, Von Lilienfeld-Toal M, Germing U et al. Treatment-related AML is an independent adverse prognostic factor for relapse-free and overall survival: an analysis of 2,868 adult patients with newly diagnosed AML enrolled on seven AMLSG treatment trials. Blood 2009; 114(22): Abstract 578.

3. Tanemura A, Van Hoesel AQ, Mori T, Yu T, Hoon DSB. The role of estrogen receptor in melanoma. Expert Opin Ther Targets 2007; 11(12): 1639-1648.

Nicht VE5

1. De Bono J. Circulating tumour cells as biomarkers in clinical trials. European Journal of Cancer Supplements 2009; 7(2-3): 42-43.

2. Dueck AC, Hillman DW, Liu H, Rowland KM, Palmieri FM, Suman VJ et al. Comparison of binary efficacy endpoints in 11 North Central Cancer Treatment Group phase II metastatic breast cancer clinical trials. Cancer Res 2009; 69(2 Suppl S): Abstract 6147.

3. Guerrero-Zotano A, Gavila J, Climent MA, Juan MJ, Guillem V, Ruiz A. Response to neoadjuvant chemotherapy and outcome based on breast cancer subtype. J Clin Oncol 2009; 27(15 Suppl 1): e11516.

4. Johnson K, Hill S, Stokes B, Berlin J, Freemantle N, Anthony D. Developing methodology for validation and comparison of surrogate markers in randomized clinical trials. In: 12th Cochrane Colloquium: bridging the gaps; 02.-06.10.2004; Ottawa, Canada. Cochrane Collaboration; 2004. S. 143.

5. Keyvanjah K, Carley W. Correlations between efficacy and early changes in biomarkers in a metastatic breast cancer (mBC) clinical trial with sunitinib malate. J Clin Oncol 2010; 28(15 Suppl 1): 1055.

6. Toh U, Otsuka H, Iwakuma N, Shirouzu K, Ogo E, Yamana H et al. A phase II study of personalized selection of peptide vaccines combined with conventional chemo- or endocrine therapy for refractory breast cancer. J Clin Oncol 2010; 28(15 Suppl 1): TPS173.

Anhang C – Offenlegung potenzieller Interessenkonflikte des externen Reviewers

Im Folgenden sind potenziellen Interessenkonflikte des externen Reviewers dargestellt. Alle Informationen beruhen auf Selbstangaben an Hand des „Formblatts zur Offenlegung potenzieller Interessenkonflikte“. Das Formblatt ist unter http://www.iqwig.de abrufbar. Die in diesem Formblatt aufgeführten Fragen finden sich im Anschluss an diese Zusammenfassung.

Externer Reviewer

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Velasco Garrido, Marcial	nein	nein	nein	ja	nein	nein

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ wurden folgende 6 Fragen gestellt:

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor bei einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter[1] abhängig (angestellt) beschäftigt? Falls ja, wo und in welcher Position?

Frage 2: Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor einen Interessenverband im Gesundheitswesen oder einen vergleichbaren Interessenvertreter direkt oder indirekt beraten? Falls ja, wen und wie hoch ist / war die Zuwendung / das Honorar?

Frage 3: Haben Sie abseits einer Anstellung oder Beratungstätigkeit innerhalb des laufenden Jahres oder der 3 Kalenderjahre davor im Auftrag eines Interessenverbands im Gesundheitswesen oder eines vergleichbaren Interessenvertreters Honorare für Vorträge, Stellungnahmen, Ausrichtung und / oder Teilnahme an Kongressen und Seminaren – auch im Rahmen von Fortbildungsveranstaltungen, oder für (populär-)wissenschaftliche oder sonstige Aussagen oder Artikel erhalten? Falls ja, von wem, für welche Tätigkeiten und wie hoch war die Zuwendung / das Honorar?

Frage 4: Haben Sie abseits einer Anstellung oder Beratungstätigkeit und / oder hat die Institution[2], bei der Sie angestellt sind bzw. die Sie vertreten, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten? Falls ja, von wem, für welche Tätigkeit und in welcher Höhe?

Frage 5: Haben Sie und / oder hat die Institution, bei der Sie angestellt sind bzw. die Sie vertreten, innerhalb des laufenden Jahres oder der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Reisekostenunterstützung ohne wissenschaftliche Gegenleistungen) von einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter erhalten? Falls ja, von wem, aus welchem Anlass und in welcher Höhe?

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile (auch in Fonds) von einer Firma oder Institution, die zu einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter gehört? Falls ja, von wem und welchen Wert haben diese aktuell?

---

[1] Dieses Formblatt erfasst finanzielle Beziehungen zu Interessenverbänden im Gesundheitswesen oder vergleichbaren Interessenvertretern, insbesondere der pharmazeutischen Industrie und der Medizinprodukteindustrie.

[2] Sofern Sie in einer ausgedehnten Institution tätig sind, ist es ausreichend, die geforderten Angaben auf Ihre Arbeitseinheit (z. B.: Klinikabteilung, Forschungsgruppe etc.) zu beziehen.