

## **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 Validität 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.

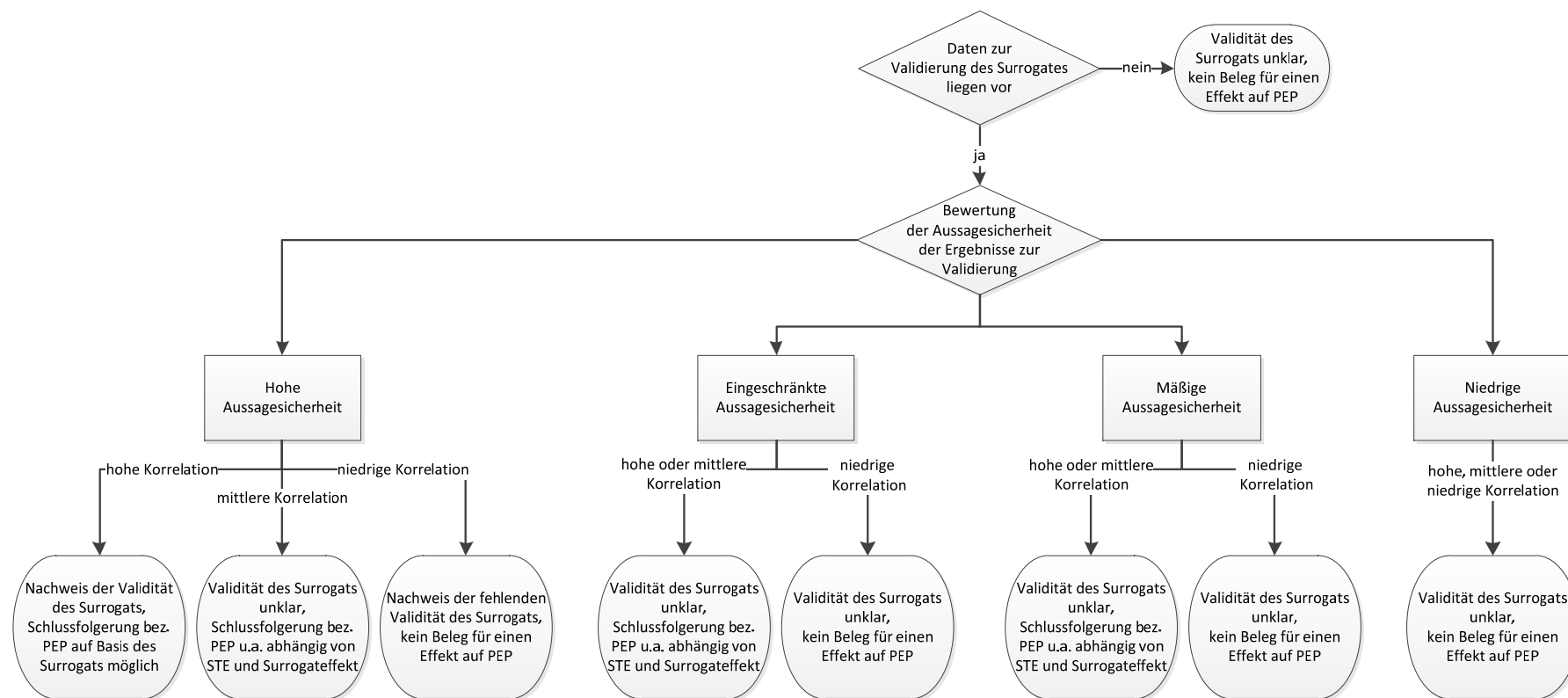


Abbildung A: Einfluss der Aussagesicherheit der Validierungsstudien und der Korrelation der Effekte von Surrogat und patientenrelevantem Endpunkt auf die Validität des Surrogats

PEP: patientenrelevanter Endpunkt; STE: Surrogate Threshold Effect

Bei hoher Aussagesicherheit der Validierungsstudie kann aus einer hohen Korrelation ( $R \geq 0,85$ ) 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. 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 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 (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.

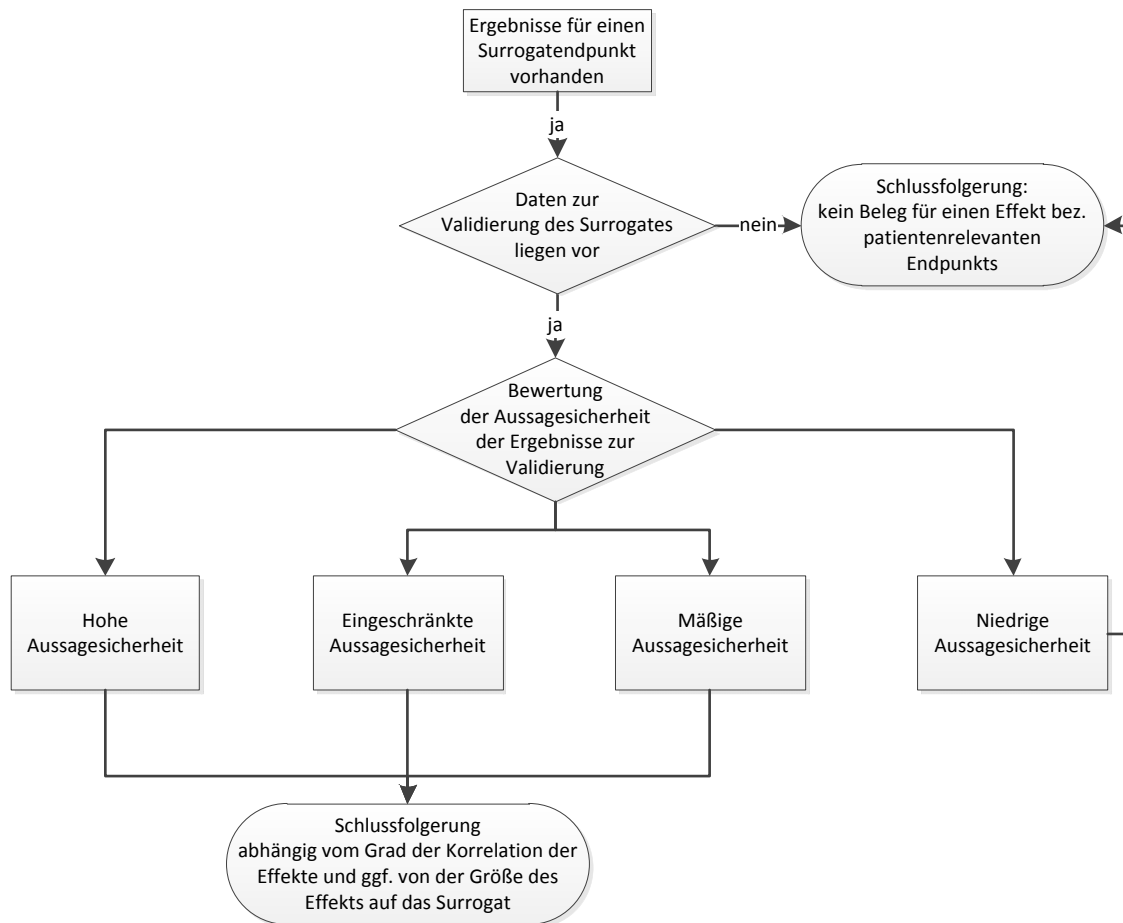


Abbildung B: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: Stellenwert von Validierungsstudien und deren Aussagesicherheit

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.

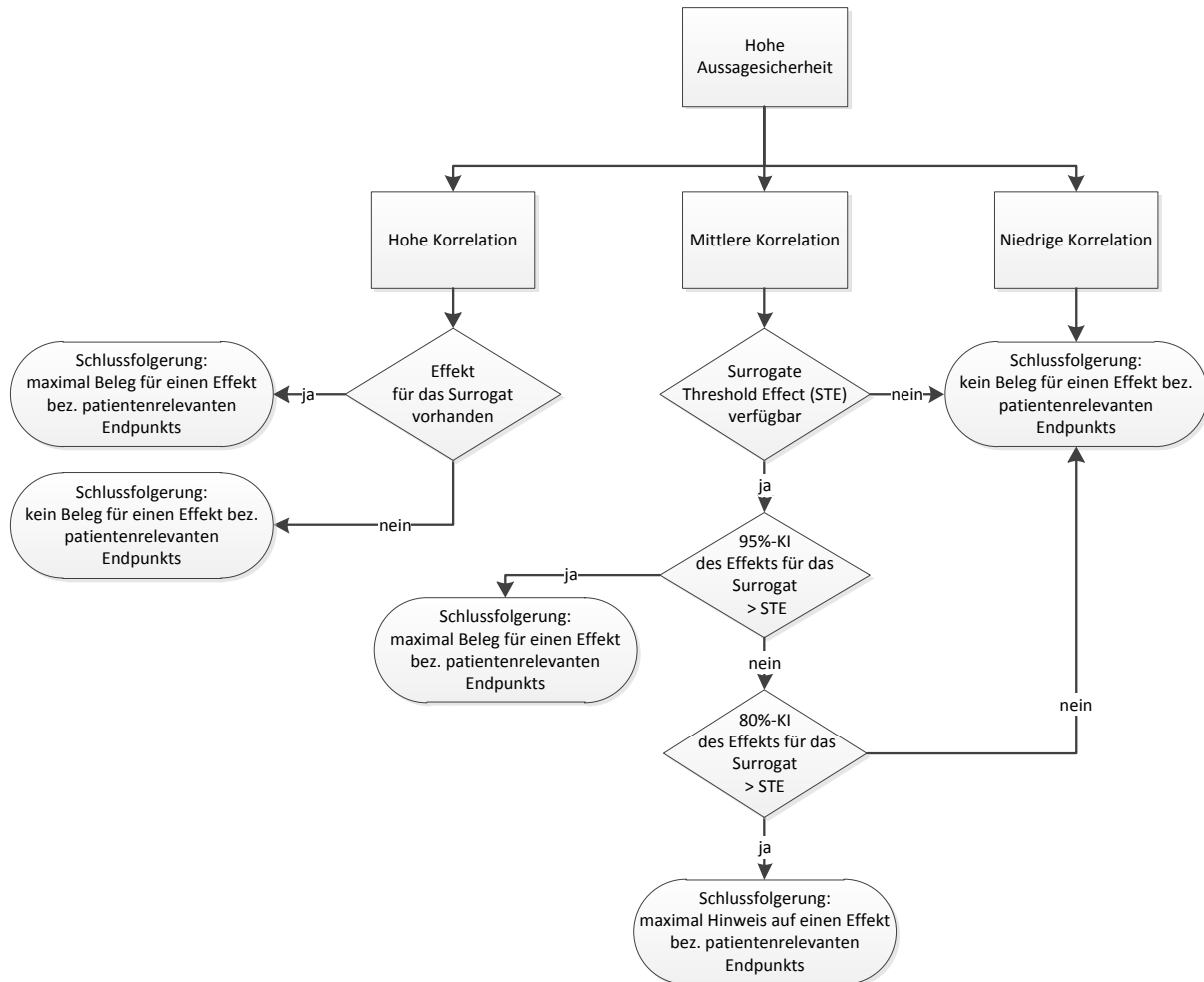


Abbildung C: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: mögliche Schlussfolgerungen bei hoher Aussagesicherheit der Validierungsstudien

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.

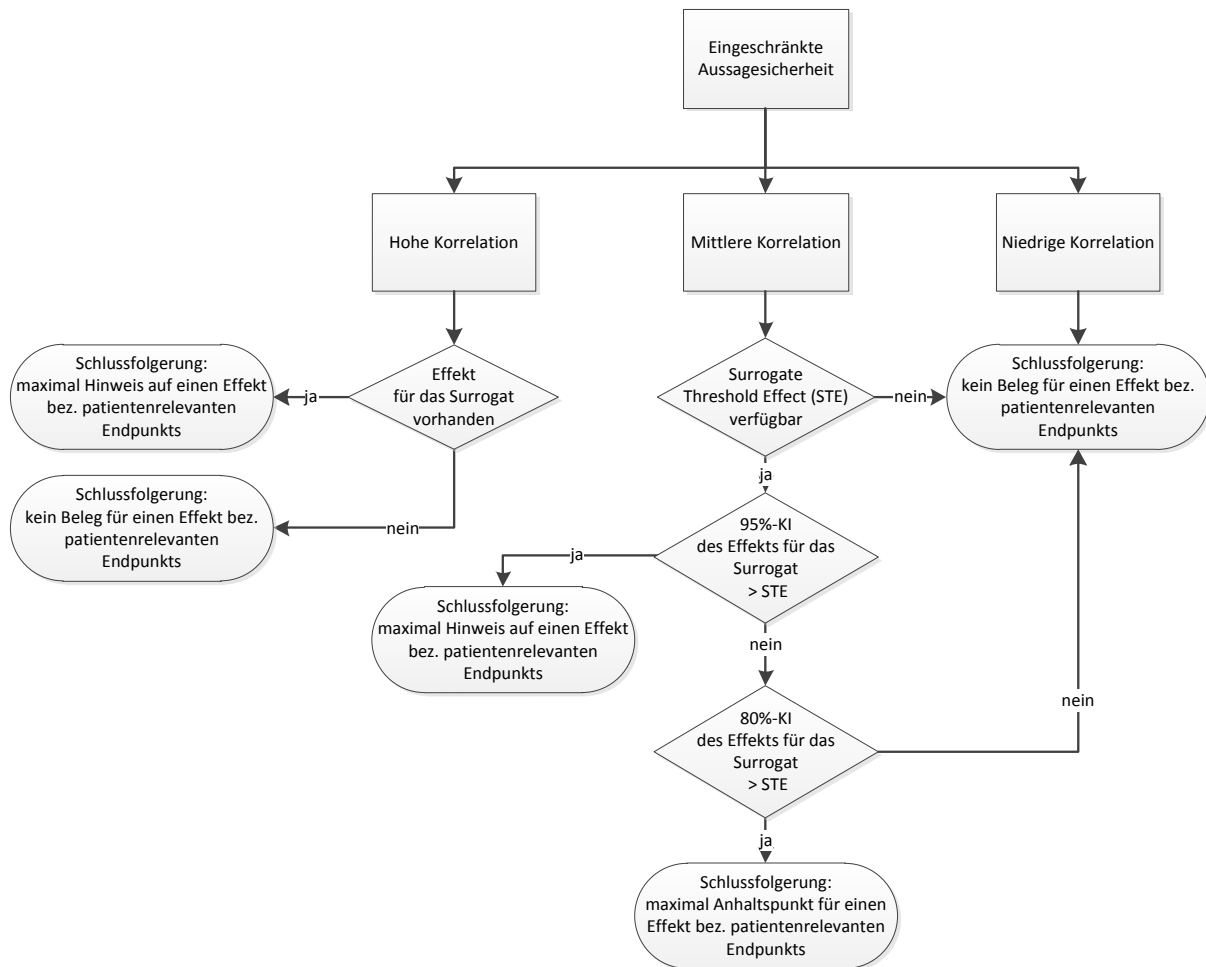


Abbildung D: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: mögliche Schlussfolgerungen bei eingeschränkter Aussagesicherheit der Validierungsstudien

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.

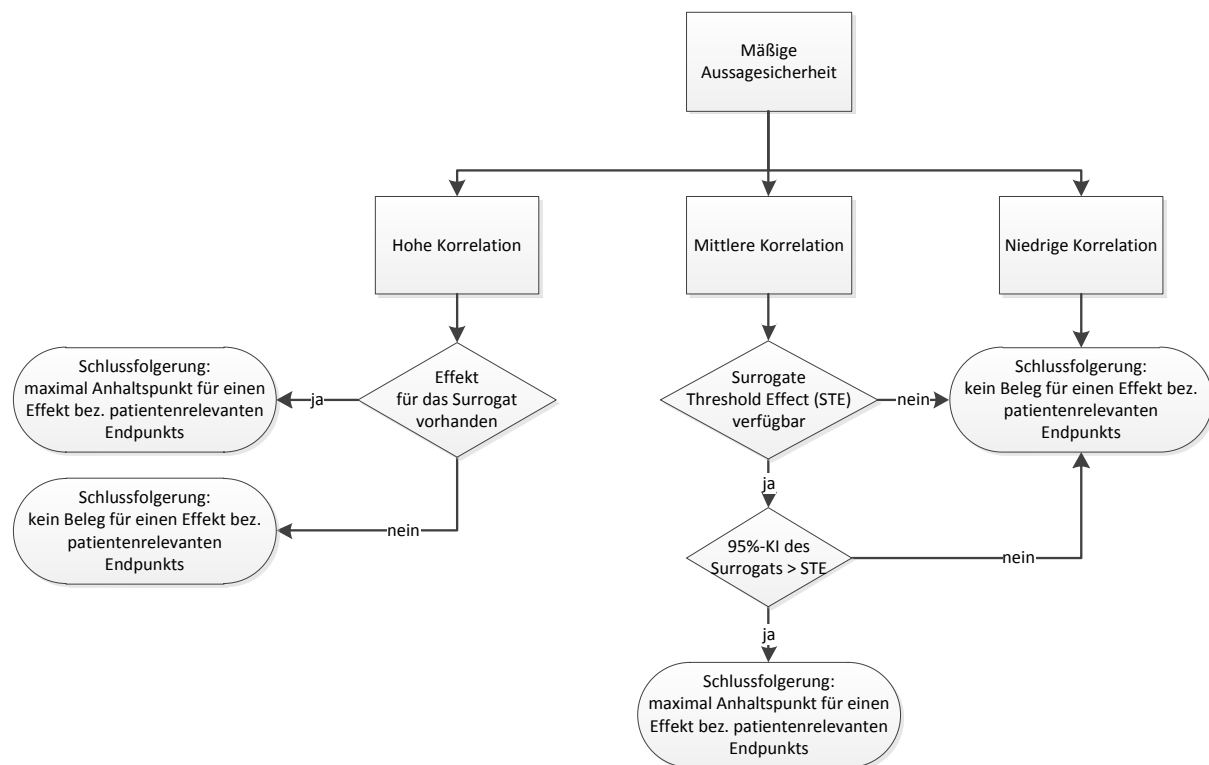


Abbildung E: Berücksichtigung von Surrogatendpunkten in der Nutzenbewertung: mögliche Schlussfolgerungen bei mäßiger Aussagesicherheit der Validierungsstudien

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