

## **Kurzfassung**

Am 21.12.2006 wurde das Projekt einer inhaltlichen Bewertung von Leitlinien begonnen. Da zu dieser Zeit das Thema der präoperativen Diagnostik im Gemeinsamen Bundesausschuss von Interesse war, zog das IQWiG die präoperative Diagnostik als Thema für das Arbeitspapier zur inhaltlichen Bewertung von Leitlinien heran. Im Laufe der Bearbeitung stellte sich heraus, dass eine inhaltliche Bewertung vollständiger Leitlinien nicht durchführbar war. Die Bearbeitung dieser Fragestellung wurde daraufhin eingestellt. Das IQWiG entwickelte im Folgenden einen Kriterienkatalog zur Einschätzung der internen Validität von Leitlinienempfehlungen. Zur Erprobung dieses Kataloges wurde beispielhaft das Thema präoperative Diagnostik herangezogen.

Der Kriterienkatalog für die Einschätzung der internen Validität von Leitlinienempfehlungen und das geplante Vorgehen für die Erprobung wurden intern in einer nicht veröffentlichten Projektskizze beschrieben. Schließlich wurde unter Beteiligung externer Sachverständiger ein Bericht in Form eines Arbeitspapiers erstellt. Dieses Arbeitspapier wurde einer internen Qualitätssicherung unterzogen.

Das Arbeitspapier wird 4 Wochen nach Übermittlung der endgültigen Fassung an den G-BA auf der Website des IQWiG veröffentlicht.

## **Fragestellung**

Ziel der Untersuchung war die Erprobung der in einem Manual zusammengefassten Kriterien zur Einschätzung der internen Validität von Leitlinienempfehlungen an Empfehlungen aus evidenzbasierten Leitlinien zur präoperativen Diagnostik.

## **Methoden**

Die Einschätzung der internen Validität erfolgte unter der Maßgabe, dass die interne Validität einer Leitlinienempfehlung angenommen werden kann, solange keine die Validität infrage stellenden Aspekte gefunden werden oder die interne Validität aufgrund fehlender Angaben in der betreffenden Leitlinie als unklar einzuschätzen ist. Als intern valide gilt in diesem Zusammenhang eine Leitlinienempfehlung dann, wenn für diese angenommen werden kann, dass zum Zeitpunkt der Leitlinienerstellung die Empfehlung

- inhaltlich eindeutig formuliert ist,
- auf umfassend und systematisch recherchierten Informationen beruht,
- keine wesentlichen patientenrelevanten Zielgrößen ungenannt bleiben,
- keine wesentlichen Studien (insbesondere von höherer Evidenz) unberücksichtigt bleiben,
- die Bewertung der zugrunde gelegten Evidenz in der Leitlinie dargelegt wird und
- sowohl die der Empfehlung zugeordnete als auch die außerhalb der Leitlinie recherchierte Evidenz weder dem der Empfehlung zugeordneten LoE noch
- dem Inhalt der Empfehlung selbst widerspricht.

### *Zu erprobende Methode*

Die Überprüfung der als Ausgangspunkt dienenden Validitätsannahme erfolgt anhand eines Manuals, welches 7 unterschiedliche Kriterien zur Einschätzung der internen Validität von Leitlinienempfehlungen enthält. Die Kriterien bilden teilweise sich gegenseitig bedingende Aspekte der internen Validität ab, was dazu führen kann, dass die Verletzung eines Kriteriums die Verletzung weiterer Kriterien nach sich ziehen kann. Die Anwendung der Kriterien erfolgte sukzessive entsprechend der Nummerierung.

Das Kriterium 1 „Ist die Informationsbeschaffung vollständig und ohne offensichtliche Mängel?“ gibt Hinweise darauf, ob die verwendeten Suchquellen und die verwendeten Recherchestrategien dazu geeignet waren, die für die Leitlinienempfehlung relevanten Studien zu identifizieren.

Das Kriterium 2 „Ist die Empfehlung inhaltlich eindeutig?“ ermöglicht eine Einschätzung, ob die untersuchte Leitlinienempfehlung grundsätzlich dafür geeignet ist, klinisches Handeln anzuleiten.

Das Kriterium 3 „Werden in den Vergleichspublikationen im Vergleich zu den bei der Empfehlung genannten Zielgrößen weitere patientenrelevante Zielgrößen betrachtet?“ ermöglicht eine Einschätzung, ob von den Leitlinienautoren alle wichtigen patientenrelevanten Zielgrößen berücksichtigt wurden.

Mithilfe des Kriteriums 4 „Welche Studientypen wurden eingeschlossen und wurde die relevante Literatur für die Empfehlung identifiziert?“ wird ermittelt, ob die vorhandene Literatur von den Leitlinienautoren identifiziert wurde.

Kriterium 5 „Sind Angaben über die Bewertung der der Empfehlung zugrunde liegenden Literatur vorhanden?“ fragt, ob die Leitlinienautoren ihr Vorgehen zur Evaluation der eingeschlossenen Literatur sowie die Ergebnisse der Evaluation darlegen.

Mithilfe des Kriteriums 6 „Gibt es Hinweise auf einen Anpassungsbedarf für den LoE einzelner Empfehlungen?“ wird ermittelt, ob es Hinweise auf einen möglichen Anpassungsbedarf des LoE zur untersuchten Empfehlung gibt.

Kriterium 7 „Ergibt sich aus der Interpretation der Ergebnisse aus Sicht der Bewerter ein Anpassungsbedarf inhaltlicher Art für die Empfehlung?“ fragt schließlich, ob sich aufgrund der Ergebnisse der bei Kriterium 6 bewerteten systematischen Übersichten und Studien Hinweise darauf ergeben, dass der Inhalt der Leitlinienempfehlung angepasst werden muss.

Als Basis der Einschätzung der internen Validität von Leitlinienempfehlungen dienen zum einen Informationen der betreffenden Leitlinie beziehungsweise der betreffenden Leitlinienempfehlung. Außerdem wird zum anderen außerhalb der Leitlinie recherchierte Evidenz hinzugezogen. Die Empfehlung wird abgeglichen mit dieser sogenannten Vergleichsliteratur, der Evidenz aus hochwertigen systematischen Übersichten

beziehungsweise evidenzbasierten Leitlinien. Als Ergebnis kann bezüglich der internen Validität einer Leitlinienempfehlung festgestellt werden, dass es (1.) keine Aspekte gibt, die die interne Validität infrage stellen, oder (2.) es Aspekte gibt, die die interne Validität infrage stellen, oder (3.) wegen fehlender Angaben in der Leitlinie nicht ausgeschlossen werden kann, dass es Aspekte gibt, die die interne Validität infrage stellen. Die Aussagen zur internen Validität von Leitlinienempfehlungen beschränken sich auf den Zeitpunkt der Leitlinienerstellung, Nach dem Zeitpunkt der Leitlinienerstellung generierte Evidenz kann nur die Aktualität der Empfehlung als einen Aspekt der externen Validität infrage stellen.

### *Methode der Erprobung*

Die Kriterien des Manuals wurden am Beispiel von Leitlinienempfehlungen zur präoperativen Diagnostik erprobt. Untersucht wurden Empfehlungen zum Ruhe-Elektrokardiogramm (Ruhe-EKG), zum Röntgenthorax und zu Blutgerinnungstests für Patienten der Risikogruppe ASA-I und ASA-II der American Society of Anesthesiologists mit einer elektiven oder dringlichen Operationsindikation.

Dazu erfolgte eine systematische Recherche nach Leitlinien zur präoperativen Diagnostik in Leitliniendatenbanken sowie bei fachübergreifenden und fachspezifischen Leitlinienanbietern. Einschlusskriterien waren neben der Evidenzbasierung unter anderem der Publikationszeitraum von Januar 2003 bis Februar 2011, die Sprachen Deutsch und Englisch und die Übertragbarkeit der Leitlinie auf das deutsche Gesundheitssystem.

Um die methodische Qualität der eingeschlossenen Leitlinien abzuschätzen, wurden sie mit dem Appraisal of Guidelines for Research & Evaluation (AGREE)-Instrument von 2 Reviewern bewertet.

Die für ASA-I- und ASA-II-Patienten relevanten Empfehlungen zum präoperativen Ruhe-EKG, zum präoperativen Röntgenthorax und zu präoperativen Blutgerinnungstests wurden extrahiert. Anschließend erfolgte für ausgewählte Empfehlungen eine Einschätzung der internen Validität mithilfe der Kriterien des Manuals.

Für die Einschätzung der internen Validität der Empfehlungen wurden neben evidenzbasierten Leitlinien systematische Übersichten als Vergleichspublikationen herangezogen. Diese wurden in den Datenbanken MEDLINE, Embase, Cochrane Database of Systematic Reviews (Cochrane Reviews), Database of Abstracts of Reviews of Effects (Other Reviews) und Health Technology Assessment Database (Technology Assessments) recherchiert. Die Literaturrecherche umfasste den Zeitraum bis zum 28.07.2011. Die methodische Qualität wurde von 2 Gutachtern mithilfe des AMSTAR (Assessment of Multiple Systematic Reviews)-Instruments bewertet. Es wurden nur die systematischen Übersichten als Vergleichspublikation herangezogen, in denen die Ergebnisse der jeweils eingeschlossenen Studien auf systematische, reproduzierbare und nachvollziehbare Weise dargestellt sind.

### ***Ergebnisse der Erprobung***

Insgesamt wurden 6 Leitlinien eingeschlossen (ESC 2009, NICE 2003, ICSI 2010, AHA 2009, ASGE 2008, ACP 2006), die von Institutionen aus Europa (n = 2) und aus den USA (n = 4) herausgegeben wurden. 3 der eingeschlossenen Leitlinien geben Empfehlungen zu allen relevanten präoperativen diagnostischen Verfahren (NICE 2003, ICSI 2010, ASGE 2008), 2 Leitlinien geben nur Empfehlungen zum präoperativen Ruhe-EKG (ESC 2009, AHA 2009) und eine nur zum präoperativen Röntgenthorax (ACP 2006).

Bei der methodischen Bewertung der Leitlinien mit dem AGREE-Instrument erhielten die Leitlinien die höchsten standardisierten Domänenwerte in den Domänen 1 (Geltungsbereich und Zweck), 4 (Klarheit und Gestaltung) und 5 (Anwendbarkeit). Die Leitlinie NICE 2003 fiel bei dieser Bewertung positiv auf. Sie erzielte in den Domänen 1 (Geltungsbereich und Zweck), 2 (Interessengruppen) und 5 (Anwendbarkeit) die höchsten standardisierten Domänenwerte und in der Domäne 3 (Methodologische Exaktheit) den zweithöchsten Domänenwert. Die Leitlinie ESC 2009 erreichte in den Domänen 4 (Klarheit und Gestaltung) und 6 (Redaktionelle Unabhängigkeit) den höchsten standardisierten Domänenwert.

Über die Recherche nach systematischen Übersichten wurden insgesamt 14 Dokumente identifiziert. 12 der eingeschlossenen systematischen Übersichten zeigten bei der AMSTAR-Bewertung Mängel. Es erfolgte unter anderem keine systematische, reproduzierbare und nachvollziehbare Darstellung der Ergebnisse der Studienbewertung. Eine qualitativ gute systematische Übersicht konnte nicht als Vergleichspublikation herangezogen werden, da sie keine einzelnen diagnostischen Verfahren, sondern nur die Kombination dieser untersucht. Daher konnte nur eine systematische Übersicht zur Einschätzung der Empfehlungen zum Ruhe-EKG eingesetzt werden. Entsprechend der im Manual beschriebenen Vorgehensweise wurden als weitere Vergleichspublikationen evidenzbasierte Leitlinien herangezogen.

Beispielhaft wurde die interne Validität der Empfehlungen für das präoperative Ruhe-EKG der Leitlinie ESC 2009 mit allen Kriterien des Manuals eingeschätzt. Aufgrund des hohen Arbeitsaufwands wurden die Kriterien 6 und 7 nur für diese Empfehlungen erprobt. Für den Röntgenthorax wurde die interne Validität der Empfehlungen der Leitlinie ASGE 2008 und für die Blutgerinnungstests diejenige der Leitlinie ICSI 2010 mit den Kriterien 1 bis 5 eingeschätzt. Als Vergleichspublikationen wurden neben der identifizierten systematischen Übersicht die jeweilig anderen Leitlinien, die Empfehlungen zu den entsprechenden präoperativen Verfahren geben, genutzt. Die interne Validität aller Empfehlungen zu den verschiedenen relevanten präoperativen Verfahren wurde mit den Kriterien 1, 2 und 5 eingeschätzt.

#### *ESC 2009: Empfehlungen zum präoperativen Ruhe-EKG*

Insgesamt ergaben sich Hinweise auf eine fehlende interne Validität der 4 Empfehlungen der Leitlinie ESC 2009 zum präoperativen Ruhe-EKG.

Kriterium 1: Aufgrund fehlender Informationen zur Informationsbeschaffung der Leitlinienautoren konnte das Kriterium 1 nicht vollständig angewendet werden. Die fehlenden Informationen schwächen das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 2: Es zeigte sich, dass die Population, für die die Empfehlungen gelten, nicht eindeutig genug definiert ist. Diese fehlende inhaltliche Eindeutigkeit der Empfehlungen schwächt das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 3: In der Leitlinie wird eine patientenrelevante Zielgröße genannt. Im Vergleich zu ihr werden in den Vergleichspublikationen zusätzliche patientenrelevante Zielgrößen genannt. Anhand der Beschreibungen der Zielgrößen in der Leitlinie lässt sich jedoch nicht sicher bewerten, ob die den Empfehlungen hinterlegten Studien zeigen konnten, dass die Durchführung eines EKG einen positiven Effekt auf die in der Leitlinie genannte patientenrelevante Zielgröße hat. Das Ergebnis schwächt das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 4: Über die Vergleichspublikationen konnte höherwertige Evidenz identifiziert werden, deren Relevanz für die zu evaluierenden Empfehlungen jedoch erst mit Kriterium 6 geprüft wurde. Das Ergebnis schwächt zunächst das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 5: Das methodische Vorgehen der Leitlinienautoren wird insgesamt nicht transparent dargestellt. So werden die verwendeten Ein- und Ausschlusskriterien und die Evaluation der Studienergebnisse nicht detailliert berichtet. Dies schwächt das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 6: Da die Autoren der Leitlinie ihre Schlüsselfragen und dazu passende Ein- und Ausschlusskriterien für die Empfehlungen zum präoperativen Ruhe-EKG nicht angeben, wurden für die Bearbeitung des Kriteriums 6 eigene Schlüsselfragen formuliert. Diese Schlüsselfragen wurden zum Ein- und Ausschluss der Studien verwendet, welche zur Prüfung der LoE-Vergabe herangezogen wurden.

In diesem Pilotprojekt wurden zur Einschätzung der LoE-Vergabe durch die Leitlinienautoren insgesamt 5 Publikationen bewertet. 2 Übersichten konnten aufgrund ihrer geringen methodischen Qualität nicht weiter berücksichtigt werden. Die verbliebenen 3 Studien erfüllten die auf der Grundlage der Empfehlungen formulierten Einschlusskriterien nicht vollständig. Um im Pilotprojekt trotzdem eine vollständige Einschätzung mit dem Manual durchführen zu können, wurden diese 3 Studien zur Evaluation herangezogen, da sie den Einschlusskriterien am ehesten genügten. Anhand der zitierten Evidenz ist der von den Leitlinienautoren vergebene LoE B für alle 4 eingeschätzten Empfehlungen zum präoperativen Ruhe-EKG nicht nachvollziehbar und müsste jeweils herabgestuft werden.

Über die Vergleichspublikationen konnte ein RCT identifiziert werden, der für die Einschätzung von 2 der 4 Empfehlungen herangezogen werden konnte. Diese Studie würde

für diese beiden Empfehlungen den von den Leitlinienautoren vergebenen LoE B rechtfertigen. Diese Studie wird jedoch von den Leitlinienautoren nicht zitiert. Außerdem untersuchte sie die Kombination verschiedener präoperativer Tests, sodass sich keine separaten Aussagen zum präoperativen EKG alleine treffen lassen.

Diese Ergebnisse der Einschätzung zu Kriterium 6 schwächen das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 7: Nach Prüfung der Studien mit dem Kriterium 7 konnte keine Aussage über den Anpassungsbedarf inhaltlicher Art der Empfehlungen gemacht werden. Es konnten keine weiteren möglichen Handlungsoptionen identifiziert werden. Als Ergebnis der Einschätzung mit dem Kriterium 7 lässt sich feststellen, dass nicht genügend Informationen vorlagen, auf deren Grundlage über einen Anpassungsbedarf inhaltlicher Art entschieden werden konnte.

#### *ASGE 2008: Empfehlungen zum präoperativen Röntgenthorax*

Insgesamt ergaben sich Hinweise auf eine fehlende interne Validität der 2 Empfehlungen der Leitlinie ASGE 2008 zum präoperativen Röntgenthorax. Die Einschätzung erfolgte anhand der Kriterien 1 bis 5.

Kriterium 1: Aufgrund einer eingeschränkten Recherche der Leitlinienautoren wurden möglicherweise nicht alle für die Empfehlungen relevanten Studien identifiziert. Dieses Ergebnis schwächt das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 2: Eine Empfehlung der Leitlinie ASGE 2008 zum Röntgenthorax ist aufgrund ihrer Formulierung ohne den Hintergrundtext zweideutig. Bei der Prüfung der inhaltlichen Eindeutigkeit der zweiten Empfehlung fiel auf, dass die Population und ein Risikofaktor nicht eindeutig definiert sind. Als Ergebnis der Einschätzung mit dem Kriterium 2 ergibt sich, dass die Empfehlungen nicht eindeutig sind. Dieses Ergebnis schwächt das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 3: Sowohl in der Leitlinie ASGE 2008 als auch in den Vergleichspublikationen werden keine eindeutig patientenrelevanten Zielgrößen genannt. Dieses Ergebnis schwächt das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 4: Über die Vergleichspublikationen wurde keine höherwertige Evidenz identifiziert. Dieses Ergebnis stärkt das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 5: Das methodische Vorgehen der Leitlinienautoren ist nicht transparent dargestellt. Dieses Ergebnis schwächt das Vertrauen in die interne Validität der Empfehlungen.

#### *ICSI 2010: Empfehlungen zu präoperativen Blutgerinnungstests*

Insgesamt ergaben sich Hinweise auf eine fehlende interne Validität der 4 Empfehlungen der Leitlinie ICSI 2010 zu präoperativen Blutgerinnungstests. Die Einschätzung erfolgte anhand der Kriterien 1 bis 5.

Kriterium 1: Das Kriterium 1 konnte aufgrund fehlender Informationen nicht als vollständig erfüllt bewertet werden. Die fehlenden Informationen schwächen das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 2: Für eine Empfehlung fehlt die Definition, welche abnormen Befunde oder spezifischen Indikationen zur Durchführung von Blutgerinnungstests führen sollten. Als Ergebnis der Einschätzung mit dem Kriterium 2 ergibt sich, dass 3 der 4 Empfehlungen eindeutig formuliert sind. Dies stärkt das Vertrauen in die interne Validität der inhaltlich eindeutigen 3 Empfehlungen.

Kriterium 3: Im Gegensatz zur Leitlinie ICSI 2010 nennt eine Vergleichspublikation eine patientenrelevante Zielgröße. Dieses Ergebnis schwächt das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 4: Über die Vergleichspublikationen konnte höherwertige Evidenz identifiziert werden, deren Relevanz jedoch erst mit Kriterium 6 geprüft wurde. Das Ergebnis schwächt zunächst das Vertrauen in die interne Validität der Empfehlungen.

Kriterium 5: Das methodische Vorgehen der Leitlinienautoren ist nicht transparent dargestellt. Dieses Ergebnis schwächt das Vertrauen in die interne Validität der Empfehlungen.

### ***Möglichkeiten der Weiterentwicklung des Manuals***

Die Erprobung des Manuals hat gezeigt, dass die Kriterien des Manuals grundsätzlich dazu geeignet sind, Aspekte der inhaltlichen Validität von Leitlinienempfehlungen zu erfassen.

Erschwert wird die Anwendung der Kriterien jedoch dadurch, dass eine Vielzahl der Informationen, die gemäß den Kriterien zu extrahieren und für eine Einschätzung der Belastbarkeit einer Empfehlung unerlässlich sind, in den untersuchten Leitlinien nicht abgebildet war. Dies kann durch das Thema der Erprobung begründet sein.

Bei einer Weiterentwicklung der Kriterien des Manuals zu einem Algorithmus sollte geprüft werden, an welchen Stellen eine Einschätzung der internen Validität von Leitlinienempfehlungen trotz fehlender Informationen vorgenommen werden kann oder wann aufgrund fehlender Informationen die Einschätzung abgebrochen werden sollte. So könnte die Einschätzung der internen Validität zum Beispiel frühzeitig abgebrochen werden, wenn nicht deutlich wird, für wen die Empfehlung gilt oder wie bei der Umsetzung der Empfehlung vorgegangen werden soll, oder wenn zu der Empfehlung keine Schlüsselfrage vorliegt beziehungsweise keine Schlüsselfrage entwickelt werden kann. Bei einem Abbruch der Bewertung der internen Validität könnte diese dann frühzeitig als unklar eingestuft werden.

Deshalb sollte bei der Weiterentwicklung des Manuals die Reihenfolge der Bearbeitung der verschiedenen Kriterien entsprechend den Erfahrungen aus diesem Pilotprojekt angepasst werden.

Im Manual sind unterschiedliche Schritte vorgesehen, die den Aufwand einer Einschätzung der Validität reduzieren sollen. Zu diesen zählten beispielsweise die Verwendung von Vergleichsliteratur zur Erfassung der relevanten Evidenz oder die Begrenzung der zu evaluierenden Studien und systematischen Übersichten durch die Festlegung der Reihenfolge ihrer Bewertung. Dennoch ist die Bearbeitung der Kriterien mit einem erheblichen Aufwand verbunden. Daher sollte die Bearbeitung auf ausgewählte Fragestellungen begrenzt werden, beispielsweise wenn gegensätzliche Leitlinienempfehlungen zu einer identischen Fragestellung vorliegen. Die Anwendung der Kriterien auf eine gesamte Leitlinie scheint zum jetzigen Zeitpunkt nicht sinnvoll zu sein.

### ***Fazit***

Bei der Erprobung der 7 Kriterien des Manuals am Beispiel von Empfehlungen zur präoperativen Diagnostik konnten jeweils für die Kriterien 1 bis 6 Einschätzungen zur internen Validität der Empfehlungen abgegeben werden:

- Mit Kriterium 1 wurde eingeschätzt, ob die Suchquellen vollständig waren und die verwendete Suchstrategie ohne offensichtliche Mängel war. Unterschiede in der Qualität der Dokumentation der Informationsbeschaffung konnten abgebildet werden.
- Mit Kriterium 2 wurde eingeschätzt, ob die Empfehlungen inhaltlich eindeutig sind. Fehlende Definitionen zu den verschiedenen Aspekten der Leitlinienempfehlungen konnten identifiziert werden.
- Mit Kriterium 3 sollten fehlende, von den Leitlinienautoren nicht berücksichtigte Zielgrößen identifiziert werden. Die Anwendung dieses Kriteriums war mit Schwierigkeiten verbunden, da die berücksichtigten Zielgrößen von den Leitlinienautoren häufig nicht explizit benannt wurden.
- Mit Kriterium 4 wurde dargestellt, welche Studientypen von den Leitlinienautoren eingeschlossen wurden und wie die Auswahl der Studientypen begründet wird. Außerdem wurde die Vollständigkeit der bei der Empfehlung zitierten Literatur eingeschätzt. Die Identifizierung der Studien, die in der Leitlinie einer Empfehlung zugeordnet sind, und die Bestimmung des Studientyps auf Basis der Angaben in den Leitlinien waren mit Schwierigkeiten verbunden.
- Mit Kriterium 5 wurde eingeschätzt, ob die Leitlinienautoren ihr Vorgehen zur Evaluation der eingeschlossenen Literatur sowie die Ergebnisse ihrer Evaluation transparent darstellen. Die Einschätzung war möglich. Die im Manual vorgesehene Vorgehensweise war jedoch aufwendig, da die Leitlinienautoren teilweise ihre Methoden sehr detailliert darstellen.
- Die beim Kriterium 6 vorgesehene Einschätzung, ob die LoE-Vergabe der Leitlinienautoren nachvollzogen werden kann, war grundsätzlich möglich. Probleme entstanden jedoch, da in den Leitlinien oft nicht im Detail dargestellt wurde, wie auf Basis der



einzelnen in der Leitlinie zitierten Studien der LoE für eine konkrete Empfehlung bestimmt wurde.

- Mit Kriterium 7 sollte eingeschätzt werden, ob sich auf Basis der im Projekt evaluierten Studien Hinweise auf einen inhaltlichen Anpassungsbedarf der Empfehlung ergeben. Im Pilotprojekt konnte mit Kriterium 7 kein Anpassungsbedarf inhaltlicher Art identifiziert werden.

Erschwert wurde die Einschätzung der internen Validität insbesondere durch fehlende Informationen in den Leitlinien und eine geringe Anzahl von hochwertigen Studien zum Thema präoperative Diagnostik. Zudem war die Bearbeitung der Kriterien mit einem hohen Arbeitsaufwand verbunden.

Die Erprobung des Manuals hat gezeigt, dass einzelne Kriterien des Manuals grundsätzlich dazu geeignet sind, verschiedene Aspekte der internen Validität von Leitlinienempfehlungen zu erfassen. Aufgrund des hohen Arbeitsaufwands eignet sich das Manual aktuell jedoch nicht für die Einschätzung aller Empfehlungen einer Leitlinie. Es sollte nur zur Bearbeitung einzelner, besonders begründeter Fragestellungen eingesetzt werden.

Bei einer Weiterentwicklung des Kriterienkatalogs zu einem Algorithmus sollten Stoppregeln für einen frühzeitigen Abbruch der Einschätzung eingebaut werden, wenn zum Beispiel bestimmte Informationen fehlen oder bestimmte Kriterien nicht erfüllt sind. Die interne Validität der Empfehlungen würde in diesen Fällen dann als unklar eingestuft werden. Dafür sollte die Hierarchie der unterschiedlichen Kriterien geprüft werden.

Ebenso sollte die Reihenfolge der unterschiedlichen Kriterien geprüft werden.

**Schlagwörter:** Chirurgische Verfahren – Operative, Präoperative Behandlung, Diagnostische Routineprüfungen, Elektrokardiographie, Röntgenuntersuchung – Thorakale, Hämatologische Untersuchungen, Methodische Leitlinienbewertung

**Keywords:** Surgical Procedures – Operative, Preoperative Care, Diagnostic Tests – Routine, Electrocardiography, Radiography – Thoracic, Hematologic Tests, Methodological Guideline Appraisal