

# Methodische Grundlagen

## Würdigung der Stellungnahmen zum Entwurf für Version 2.1

# Inhaltsverzeichnis

Verzeichnis der eingegangenen Stellungnahmen .....	3
Stellungnahmeverfahren zu Version 2.1 .....	4
1 Kapitelübergreifende Themen .....	5
1 Grundlagen der Qualitätsmessung .....	6
2 Evaluation von Regelungen zur Qualitätssicherung .....	13
3 Literaturrecherchen und Evidenzsynthese .....	25
2 Auswahl von Inhalten bei der Entwicklung von Qualitätsmessungen .....	34
3 Eignungskriterien für Qualitätsmessungen .....	39
4 Bewertungskonzepte und Referenzbereiche .....	62
4 Methoden der Datenanalyse.....	69
Literatur .....	78
Impressum.....	83

## Verzeichnis der eingegangenen Stellungnahmen

- Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V. (AWMF)
- Bundesärztekammer (BÄK)
- Bundespsychotherapeutenkammer (BPtK)
- Deutsche Gesellschaft für Gesundheitsökonomie e. V. (dggö)
- Deutsche Gesellschaft für Hebammenwissenschaft e. V. (DGHWi)
- Deutsche Gesellschaft für Innere Medizin e. V. (DGIM)
- Deutsche Gesellschaft für Kardiologie – Herz- und Kreislaufforschung e. V. (DGK)
- Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS) und Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR)
- Deutsche Gesellschaft für Nephrologie e. V. (DGfN)
- Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde e. V. (DGPPN)
- Deutsche Krankenhausgesellschaft e. V. (DKG)
- Deutsches Netzwerk Versorgungsforschung e. V. (DNVF)
- GKV-Spitzenverband (GKV-SV)
- Kassenärztliche Bundesvereinigung (KBV)
- Kassenzahnärztliche Bundesvereinigung (KZBV)
- Netzwerk Evidenzbasierte Medizin e. V. (EbM-Netzwerk)
- Patientenvertretung: Maßgebliche Patientenorganisationen nach § 140f SGB V (PatV)

## Stellungnahmeverfahren zu Version 2.1

Am 8. April 2024 veröffentlichte das IQTIG den Entwurf der Version 2.1 seiner „Methodischen Grundlagen“ und leitete damit das öffentliche Stellungnahmeverfahren dazu ein. Die „Methodischen Grundlagen“ erläutern die Methoden und Kriterien, die das IQTIG bei der Erfüllung seines gesetzlichen Auftrags als fachlich unabhängiges, wissenschaftliches Institut nach § 137a SGB V zugrunde legt. Darin beschreibt das IQTIG seine allgemeinen Methoden für die Entwicklung, Weiterentwicklung und Anpassung von Maßnahmen der Qualitätssicherung durch das IQTIG. Damit tragen die „Methodischen Grundlagen“ zur Transparenz über die Arbeit des IQTIG sowie über das Zustandekommen seiner Arbeitsergebnisse bei.

Bis zum 31. Mai 2024 hatten alle interessierten Organisationen und Einzelpersonen die Möglichkeit, schriftliche Stellungnahmen zu den neuen oder geänderten Inhalten der Version 2.1 der „Methodischen Grundlagen“ einzureichen. Insgesamt gingen 17 Stellungnahmen fristgerecht ein. Die eingegangenen Stellungnahmen wurden gesichtet, fachlich geprüft und die „Methodischen Grundlagen“ wurden auf Basis der Stellungnahmen überarbeitet. Am 3. Dezember 2024 wurde die überarbeitete Version 2.1 der „Methodischen Grundlagen“ auf der Website des Instituts veröffentlicht.

Das IQTIG bedankt sich bei allen Stellungnehmenden für die fachlichen Hinweise zur Weiterentwicklung seiner „Methodischen Grundlagen“. Im vorliegenden Dokument werden zentrale Aspekte aus den Stellungnahmen zusammengefasst und es wird erläutert, wie das Institut mit den vorgebrachten Hinweisen umgegangen ist. Stellungnahmen, die sich nicht direkt auf die im Vergleich zur Version 2.0 geänderten Inhalte der „Methodischen Grundlagen, Version 2.1“ beziehen, wurden dabei nicht berücksichtigt. Die Würdigung der Hinweise aus den Stellungnahmen ist nach der Reihenfolge der Kapitel in den „Methodischen Grundlagen“ geordnet.

# 1 Kapitelübergreifende Themen

In verschiedenen Stellungnahmen wurden Vorschläge geäußert, weitere Themen in die „Methodischen Grundlagen“ aufzunehmen. Genannte Ergänzungsvorschläge waren z. B. die Beschreibung dynamischer Verfahren zur Qualitätsmessung, beispielsweise Methoden der statistischen Prozesskontrolle (GMDS/IBS-DR, S. 2), Methoden für die Umsetzung laienverständlicher Informationen (AWMF, S. 1; DNVF, S. 7; KZBV, S. 5) und die zielgruppengerechte Darstellung von statistischen Ergebnissen (GMDS/IBS-DR, S. 1), ein Vorgehen zur Identifikation von Themenbereichen, für die das IQTIG die Einführung einer Qualitätsmessung empfiehlt (DNVF, S. 5), sowie die Betrachtung von Versorgungsqualität unter der Perspektive der physiologisch orientierten Geburtshilfe (DGHWi, S. 2-4 und 6).

**IQTIG:** In den „Methodischen Grundlagen“ werden diejenigen Methoden des IQTIG geschildert, die das IQTIG bereits regelhaft anwendet. Methoden, die sich noch in Entwicklung befinden, sowie ggf. weitere der genannten wichtigen Themen wird das IQTIG in den „Methodischen Grundlagen“ adressieren, wenn zukünftig entsprechende methodische Weiterentwicklungen vorgenommen werden.

# 1 Grundlagen der Qualitätsmessung

Von mehreren stellungnehmenden Organisationen wurde die Darstellung von qualitätsindikatoren- und Peer-Review-basierten Methoden kritisiert. Es wurde zu Bedenken gegeben, dass Peer-Review-Methoden nicht nur zur Prüfung von Dokumenten eingesetzt würden, sondern auch dem kollegialen Austausch und der Qualitätsförderung dienen (KBV, S. 3). Der Begriff „Peer Review“ sei bereits „als Instrument der freiwilligen Qualitätssicherung etabliert“ und solle nicht im Kontext der externen Qualitätssicherung des Gemeinsamen Bundesausschusses (G-BA) verwendet werden (GKV-SV, S. 7; PatV, S. 5). Peer Reviews seien kein Instrument der Qualitätsmessung und keine Methode zur Darstellung von Unterschieden zwischen Leistungserbringern, sondern „ein Verfahren mit solitärer Methodik“, das der Qualitätsentwicklung und -förderung diene. Daher seien an Peer Reviews andere Gütekriterien anzulegen als an quantitative Erhebungsmethoden (BÄK, S. 2 f. und 5). Eine vergleichende Darstellung von qualitätsindikatoren- und Peer-Review-basierten Methoden und die Darstellung der datengestützten Qualitätssicherung als höherwertig sei daher unangemessen (BÄK, S. 2 f. und 5). Die Darstellung Peer-Review-basierter Methoden wurde als unausgewogen (GKV-SV, S. 7) und als voreingenommen und interessengeleitet (DKG, S. 4) wahrgenommen. Die Darstellung habe zum Ziel, einen vermehrten Einsatz qualitativer Ansätze der Qualitätsbeurteilung zu verhindern (DKG, S. 4).

**IQTIG:** Für den Begriff „Peer Review“ gibt es im Kontext der Gesundheitsversorgung keine einheitliche Definition. In der internationalen wissenschaftlichen Literatur werden insbesondere Begriffe wie *audit* oder *clinical audit* und *external peer review* weitgehend synonym verwendet (BÄK 2013: 15, Flodgren et al. 2016: 5, Sens et al. 2007: 38). Häufig wird der Begriff auf freiwillige Aktivitäten zur Qualitätsverbesserung beschränkt (Grol 1994, Sens et al. 2018: 58), wie beispielsweise im Curriculum Ärztliches Peer Review der Bundesärztekammer (BÄK 2013) propagiert. Der Begriff wird aber auch für Maßnahmen mit verpflichtendem Charakter verwendet (Bader et al. 2021, Bion und Alderman 2023, Hut-Mossel et al. 2021, Lilford et al. 2007). Teilweise wird der Begriff zusammenfassend für das gesamte Verfahren von Qualitätsmessung, gemeinsamer Diskussion und Ableitung von Maßnahmen gebraucht (vgl. BÄK, S. 3). Peer-Review-basierte Methoden werden folglich für verschiedene Ziele eingesetzt: Vereinfacht kann unterschieden werden zwischen dem Einsatz für die Beurteilung von Versorgungsqualität (Qualitätsmessung) und dem Einsatz für die Ableitung von Ideen zur Qualitätsverbesserung (Qualitätsförderung).

In Abschnitt 4.2 des Entwurfs zur Version 2.1 der „Methodischen Grundlagen“ ist unter „Peer-Review-basierten Methoden“ bisher ausschließlich der Einsatz zum Zweck der Qualitätsmessung angesprochen. Das IQTIG versteht unter Review-basierter Qualitätsmessung die Beurteilung von Versorgungsqualität durch externe Fachkolleginnen und Fachkollegen. Kennzeichnend ist die Beteiligung dieser Peers am Erhebungsprozess,

die damit einen Einfluss auf die erhobenen Informationen und deren Bewertung haben, im Gegensatz zu sog. nicht-reaktiven Verfahren, bei denen die Erhebenden nicht direkt an der Informationsbeschaffung beteiligt sind (Meyer 2007: 228 f.). Für die Qualitätsmessung durch Fachkolleginnen und Fachkollegen ist in der Literatur kein einheitlicher Begriff etabliert. Genannt werden in diesem Zusammenhang beispielsweise (*external inspection* (Flodgren et al. 2016), *assessment (made by inspectors)* (Boyd et al. 2017), *peer evaluation* (Thai et al. 2022), *medical record review* (Klein et al. 2018) und *audit* (Hut-Mossel et al. 2021).

Der in mehreren Stellungnahmen geschilderte Eindruck, das IQTIG vergleiche Peer Reviews mit der datengestützten Qualitätssicherung und komme dabei zu einem voreingenommenen Ergebnis, beruht somit auf einem Missverständnis: Eine Bewertung von Peer Reviews als Methode zur Qualitätsförderung ist nicht Gegenstand der Ausführungen in den „Methodischen Grundlagen“. Ziel des Abschnitts ist vielmehr eine Darstellung der Charakteristika von indikator- bzw. kennzahlgestützten Methoden im Vergleich zu Peer-Review-basierten Methoden, wenn sie als Grundlage für leistungserbringerbezogene Qualitätsmessungen eingesetzt werden. Der Abschnitt zu Peer-Review-basierten Methoden wurde daher überarbeitet und stellt den Unterschied zwischen dem Einsatz mit dem Ziel der Qualitätsmessung und dem Einsatz mit dem Ziel der Qualitätsförderung nun deutlicher dar. Werden Peer-Review-basierte Methoden für die Messung und Bewertung der Versorgungsqualität eingesetzt, geht dies mit den geschilderten Einschränkungen hinsichtlich Objektivität/Reliabilität einher, und diese können durch möglichst standardisierte Vorgehensweisen und Kriterien teilweise reduziert werden. Der Einsatz Peer-Review-basierter Methoden als qualitative Methode, beispielsweise für die Identifikation von Verbesserungsansätzen der Versorgung, ist davon unbenommen.

Von einer stellungnehmenden Organisation wurde der hypothesengenerierende Charakter qualitativer Methoden hinterfragt. Es wurde die Auffassung geäußert, qualitative Methoden würden bereits im Rahmen des Stellungnahmeverfahrens nach der Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung (DeQS-RL)<sup>1</sup> eingesetzt und hätten die Funktion, Hypothesen bezüglich der Nichterfüllung von Qualitätsanforderungen zu bestätigen oder zu widerlegen. Auch würden bei vielen Peer-Review-Konzepten nicht nur Rückschlüsse aus Einzelfällen gezogen, sondern Strukturen und Prozesse untersucht. Umgekehrt würden auch bei quantitativen Verfahren verallgemeinernde Schlüsse gezogen, z. B. wenn „aus einzelnen Komplikationen ohne weitere Analysen auf die Existenz eines Qualitätsdefizits geschlossen wird“ (GKV-SV, S. 6 f.).

---

<sup>1</sup> Richtlinie des Gemeinsamen Bundesausschusses zur datengestützten einrichtungsübergreifenden Qualitätssicherung. In der Fassung vom 19. Juli 2018, zuletzt geändert am 18. Januar 2024, in Kraft getreten am 1. September 2024. URL: <https://www.g-ba.de/informationen/richtlinien/105/> (abgerufen am: 13.09.2024).

**IQTIG:** Qualitative Methoden sind gekennzeichnet durch einen bewusst wenig strukturierten, induktiven Prozess, in dem mittels nicht standardisierter oder allenfalls teilstandardisierter Verfahren nicht numerische (meist verbale) Daten erhoben und interpretierend ausgewertet werden, um daraus neue Hypothesen und Theorien zu bilden (Döring und Bortz 2016: 32). Ziel dabei ist, Sachverhalte im Detail zu verstehen oder neue Sachverhalte aufzudecken (Maats 2007: 294). Im Kontext der Qualitätssicherung könnte ein qualitativer Ansatz beispielsweise darin bestehen, ausgehend von Interviews mit Versorgungenden Hypothesen über die Qualitätskultur der Einrichtung zu bilden und darüber, welche Ansätze sich daraus möglicherweise für Qualitätsverbesserungen ergeben. In diesem Fall wird von einzelnen Sachverhalten (Informationen in den Interviews z. B. zu persönlichen Erfahrungen bei der Arbeit) ein verallgemeinernder Schluss auf einen anderen Sachverhalt (Qualitätskultur, Verbesserungsansätze) gezogen.

Die Messung von Versorgungsqualität entspricht dagegen einer Quantifizierung von Merkmalen der Versorgung, beispielsweise der Quantifizierung, wie häufig ein Leistungserbringer Patientinnen und Patienten mit hüftgelenknaher Femurfraktur innerhalb von 24 Stunden nach Aufnahme ins Krankenhaus operiert. Einem „Objekt“ (z. B. der Erbringung einer oder eines Bündels an Leistungen durch einen Leistungserbringer) wird ein Wert zugeordnet. Die möglichen Ergebnisse der Messung sind – im Gegensatz zu qualitativen Methoden – a priori vorgegeben, und das Ziel solcher Verfahren besteht darin, alle Objekte (z. B. alle Leistungserbringer, die Femurfrakturen operativ behandeln) einer dieser Ergebniskategorien zuzuordnen. Dabei können die Ergebniskategorien auf verschiedenen Skalenniveaus vorliegen, etwa auf Nominalskalenniveau (z. B. Verbesserungsbedarf / kein Verbesserungsbedarf) oder auf Intervallskalenniveau (z. B. von 0 % bis 100 %). Regeln und Kriterien beschreiben dabei, wie Beobachtungen zu Ergebniskategorien zugeordnet werden. Solche Regeln und Bewertungskriterien bestehen bei indikatorgestützten Methoden in den Rechenregeln und ggf. Referenzbereichen und bei Peer-Review-basierten Methoden in den Kriterien für die Bewertung von Strukturen und Prozessen (z. B. in Form eines „Auditleitfadens“ oder „Kriterienkatalogs“).

Bei Qualitätsbeurteilungen (= Qualitätsmessungen) im bisherigen Stellungnahmeverfahren nach DeQS-RL wird auch auf Informationen zurückgegriffen, die durch weniger standardisierte Befragungen und Beobachtungen erhoben wurden und dabei bereits einer impliziten Urteilsbildung von Fachexpertinnen und Fachexperten unterlegen haben. Diese Erhebungen dienen bisher vorwiegend dem Ziel einer Qualitätsbeurteilung auf Nominalskalenniveau (z. B. „qualitativ auffällig“ / „qualitativ unauffällig“). Eine im originären Sinne qualitative Auswertung der Informationen (s. o.) würde dem Ziel einer standardisierten, fairen Qualitätsbeurteilung widersprechen.

Quantitative Erhebungs- und Auswertungsmethoden treffen eine Aussage nur über den untersuchten Sachverhalt. Die Aussage eines Qualitätsindikators oder einer Qualitätskennzahl bezieht sich ausschließlich auf das operationalisierte Qualitätsmerkmal (z. B. Vermeidung von Sterbefällen bei ambulant erworbener Pneumonie). Verallgemeinernde



Schlüsse auf andere Sachverhalte (z. B. Versorgungsqualität der pneumologischen Abteilung insgesamt) sollten daher nicht gezogen werden.

In mehreren Stellungnahmen wurde die mögliche Rolle Peer-Review-basierter Methoden im Sinne eines Messinstruments thematisiert. Von einer stellungnehmenden Organisation wurde kritisiert, dass eine solche Rolle aus Perspektive der DeQS-RL nicht nachvollziehbar sei (GKV-SV, S. 6). Es wurden weitere Ausführungen gewünscht, wie Qualitätsmerkmale mittels Peer-Review-basierter Verfahren operationalisiert werden (KBV, S. 3 und 5) und in welcher Form Peer-Review-basierte Methoden im Rahmen der Qualitätsmessung genutzt werden sollten (PatV, S. 5). In einer weiteren Stellungnahme wurden Erläuterungen gewünscht, in welchen Bereichen das IQTIG Peer Reviews empfehle (AWMF, S. 4). Eine stellungnehmende Organisation sprach sich grundsätzlich gegen den Einsatz von Peer Reviews im Rahmen der verpflichtenden Qualitätssicherung aus, da diese im vertrags(zahn)ärztlichen Bereich nicht umsetzbar seien und die intrinsische Motivation der Leistungserbringer verloren gehe (KZBV, S. 7).

**IQTIG:** Der Stellenwert von Peer-Review-basierten Methoden als „qualitative“ Methode im üblichen wissenschaftlichen Sinn (des qualitativen Paradigmas als verstehend-interpretative, hypothesengenerierende Methode, vgl. Döring und Bortz 2016: 63) ist für die verpflichtende externe Qualitätssicherung bisher nicht geklärt. Die in den „Methodischen Grundlagen“ genannten Vorteile legen einen Einsatz qualitativer Analysemethoden dort nahe, wo die Generierung von Annahmen und Ideen, etwa von Ansätzen zur Qualitätsverbesserung, im Vordergrund steht und nicht das Treffen eines Qualitätsurteils, also die Messung der Versorgungsqualität. Dieser mögliche Einsatzbereich entspricht weitgehend dem Verständnis von Peer Reviews als Instrument des kollegialen Austauschs und der kontinuierlichen Qualitätsförderung. An qualitative Methoden in diesem Sinne sind andere Gütekriterien anzulegen als an Qualitätsmessungen (Döring und Bortz 2016: 106-114). Das IQTIG prüft derzeit den Stellenwert von Peer-Review-basierten Methoden für die verschiedenen Einsatzzwecke - einerseits als Instrument zur Qualitätsbewertung und andererseits als qualitative Methode - im Rahmen eines G-BA-Auftrags (G-BA 2024b).

Von einer stellungnehmenden Organisation wurde vermutet, die „Methodischen Grundlagen“ böten „derzeit keine Möglichkeit, ein Nichtintervenieren als Qualitätsindikator abzubilden“, beispielsweise mit dem Ziel, unnötige Interventionen bei Geburten zu reduzieren (DGHWi, S. 2).

**IQTIG:** Qualitätsindikatoren können auch Aussagen dazu treffen, ob die Indikation zu einem medizinischen Eingriff (sog. „Indikationsindikator“) und ob eine gemeinsame Entscheidungsfindung angemessen vorgenommen wurden. Im Rahmenkonzept des IQTIG für Qualität sind solche Indikatoren der Qualitätsdimension „Angemessenheit“ zugeordnet.

Von zwei stellungnehmenden Organisationen wurde die Auffassung geäußert, Qualitätsindikatoren könnten nicht eigenständig den Grad der Erfüllung von Qualitätsanforderungen beschreiben, da nicht alle möglichen Einflussfaktoren durch einen Indikator berücksichtigt werden könnten (DKG, S. 4; GKV-SV, S. 6 und 8). Daher sei der Begriff „Qualitätsmessung“ falsch (DKG, S. 4). Das IQTIG setze auch Indikatoren ein, „die relevante Limitationen etwa bei der Operationalisierung“ aufwiesen, und die Qualitätsindikatoren könnten „bestenfalls als Annäherung an das Qualitätsmerkmal verstanden werden“ (GKV-SV, S. 6 und 8). Die tatsächlichen Sachverhalte müssten über qualitative Verfahren aufgeklärt (DKG, S. 4) und die rechnerischen Ergebnisse im Stellungnahmeverfahren nach DeQS-RL validiert (GKV-SV, S. 6) werden.

**IQTIG:** Nach wissenschaftlichem Verständnis handelt es sich bei Messungen im Allgemeinen um eine Zuordnung von Zahlen zu Objekten (Schnell et al. 2013: 128). Indikatoren machen einen theoretischen Sachverhalt durch beobachtbare Sachverhalte messbar (Meyer 2004: 28 und 51, Döring und Bortz 2016: 223 und 228, Schnell et al. 2013: 117 und 121). In der Literatur zur Qualitätssicherung in der Gesundheitsversorgung wird dementsprechend von „Maßen“ oder „measures“ gesprochen (Sens et al. 2018: 42 und 46, NQF 2015). Maßzahlen treffen grundsätzlich keine abschließende Aussage über einen Sachverhalt, sondern erlauben, einen Sachverhalt zu quantifizieren. Die Ergebnisse solcher Maßzahlen, Kennzahlen oder Indikatoren bedürfen immer einer Interpretation.

Dasselbe trifft für die Ergebnisse weniger standardisierter Messverfahren (hier irrtümlich als „qualitativ“ bezeichnet, siehe oben) auf Grundlage z. B. von Expertenurteilen zu: Auch solche Verfahren basieren stets auf einer begrenzten Informationsgrundlage und sind ebenfalls mit Interpretationen und Einschränkungen verbunden. Eine Berücksichtigung aller möglichen Einflüsse auf die Qualitätsmessung ist weder durch indikatorgestützte noch durch andere, nicht standardisierte Methoden möglich. Indikatoren mit „relevanten“ Einschränkungen der Operationalisierung, d. h. mit ungeeigneter Operationalisierung gemäß der Definition in Anhang A, empfiehlt das IQTIG nicht zum Einsatz in der externen Qualitätssicherung.

### **Kennzahlen**

Von einigen stellungnehmenden Organisationen wurde die Aufnahme des Begriffs und Konzepts der Qualitätskennzahlen in die „Methodischen Grundlagen“ kommentiert.

So solle das IQTIG vom Begriff „Qualitätskennzahlen“ Abstand nehmen (KBV, S. 3; DGfN, S. 4), da dieser weder im SGB V noch in den einschlägigen Richtlinien zur gesetzlichen Qualitätssicherung nach §§ 136ff SGB V normiert sei (KBV, S. 3). Es sei unklar, wie ohne Referenzwert und Auswertungskonzept überprüft werden könne, ob Qualitätsanforderungen erfüllt worden seien (DGfN, S. 3). Qualitätskennzahlergebnisse seien für eine vergleichende Veröffentlichung zur Patienteninformation ungeeignet, sondern könnten allenfalls zur Darstellung von Grundgesamtheiten oder von Zwischenergebnissen der

Berechnung komplexer Qualitätsindikatoren eingesetzt werden (DGfN, S. 4). Qualitätskennzahlen würden im Vergleich zu Qualitätsindikatoren eine geringere Reliabilität und Validität aufweisen. Sie könnten daher lediglich Hinweise für mögliches Potenzial zur Qualitätsverbesserung liefern, aber nicht Gegenstand eines Stellungnahmeverfahrens gemäß § 17 Teil 1 DeQS-RL sein (KZBV, S. 6). Es fehlten ein Konzept und eine eindeutige Definition von Kennzahlen (GKV-SV, S. 5; KBV, S. 3; KZBV, S. 6).

Andere stellungnehmende Organisationen begrüßten dagegen die Nutzung von Qualitätskennzahlen, da diese eine solide Grundlage dafür schaffen würden, die Versorgung in verschiedenen Einrichtungen vergleichend darzustellen (DGHWi, S. 2). Die Abgrenzung von Qualitätskennzahlen zu Qualitätsindikatoren sei nachvollziehbar (AWMF, S. 2).

**IQTIG:** Das IQTIG bezeichnet Kennzahlen, mit denen die Versorgungsqualität gemessen werden soll, als Qualitätskennzahlen. Die Ergebnisse von Qualitätskennzahlen können veröffentlicht werden, um dadurch Transparenz über Versorgungsqualität herzustellen. Damit die Kennzahlergebnisse für diesen Zweck geeignet sind, müssen Qualitätskennzahlen dieselben Kriterien für die Eignung zur öffentlichen Berichterstattung erfüllen wie Qualitätsindikatoren – das Qualitätsmerkmal muss also Bedeutung für die Patientinnen und Patienten haben, die Messung muss objektiv, reliabel und valide erfolgen, die Kennzahlergebnisse müssen angemessen risikoadjustiert sein usw. Ist dies der Fall, dann bilden Unterschiede in den Kennzahlergebnissen, beispielsweise zwischen zwei Leistungserbringern, Unterschiede in der Versorgungsqualität ab. Die Adressaten der jeweiligen Veröffentlichung – beispielsweise Patientinnen und Patienten – können so die Qualitätskennzahlergebnisse nutzen, um die Leistungserbringer hinsichtlich ihrer Versorgungsqualität zu vergleichen. Die Festlegung eines Referenzbereichs ist hierfür nicht notwendig. Dieses Konzept wurde in Kapitel 4 klarer herausgearbeitet sowie das Glossar um entsprechende Definitionen ergänzt.

Von einer stellungnehmenden Organisation wurde die Frage gestellt, ob alle Kennzahlen ohne Bewertungskonzept, d. h. ohne Referenzbereich, die Operationalisierungen von Qualitätsmerkmalen seien, als „Qualitätskennzahlen“ anzusehen seien. In derselben Stellungnahme wurden einige Beispiele für Kennzahlen genannt und es wurde gefragt, welche dieser Kennzahlen Gegenstand der Eignungsprüfung und Aufwand-Nutzen-Abwägung sein sollten (GKV-SV, S. 4-5):

- *Nicht-indikatorassoziierte Kennzahlen ohne Referenzbereich, beispielsweise 1-Jahres-Sterblichkeit nach PCI (...)*
- *Kennzahlen, die zur Berechnung von Index-Indikatoren verwendet werden, z.B. in der Neonatologie zum QI „Qualitätsindex der Frühgeborenenversorgung“ (ID 51901), wo die Anzahl von verschiedenen Komplikationen bzw. Schädigungen gezählt werden*
- *Nicht risikoadjustierte „rohe“ Ereignisraten zu Indikatoren, deren Ergebnis dann risikoadjustiert als O/E dargestellt wird, z.B. im Modul Koronarchirurgie zum Indikator „Sterblichkeit im Krankenhaus“ (ID 352007) (...)*

- Stratifizierungen zu Qualitätsindikatoren, z.B. im Modul Koronarchirurgie zum Indikator „Sterblichkeit im Krankenhaus“ (ID 352007) (...)
- Darüber hinaus sind auch deskriptive Angaben aus der Basisauswertung für die Berechnung von Indikatoren erforderlich und werden somit zur Operationalisierung von Qualitätsmerkmalen verwendet.

Von einer weiteren stellungnehmenden Organisation wurde angenommen, dass sich die Anwendung der Eignungskriterien auf Qualitätsindikatoren beschränken würde (KBV, S. 6).

**IQTIG:** Ob eine Kennzahl „als Qualitätskennzahl anzusehen“ ist, hängt davon ab, ob diese Kennzahl ein Qualitätsmerkmal abbilden soll. Für alle Kennzahlen (mit oder ohne Referenzbereich), die Operationalisierungen von Qualitätsmerkmalen darstellen, geht das IQTIG davon aus, dass diese die Funktion von Qualitätskennzahlen erfüllen sollen. Dies trifft auf die Kennzahlen in Beispiel 1 (nicht indikatorassoziierte Kennzahl ohne Referenzbereich), Beispiel 2 (Kennzahlen, die Bestandteile eines Indexindikators sind) sowie Beispiel 4 (Stratifizierungen zu Qualitätsindikatoren) zu. Um als Qualitätskennzahl eingesetzt zu werden, muss eine Kennzahl die Eignungskriterien für Qualitätsmessungen erfüllen und ein ausreichendes Aufwand-Nutzen-Verhältnis aufweisen, diese Kennzahlen müssen also Gegenstand einer Eignungsprüfung sein.

Die Kennzahl in Beispiel 3 (nicht risikoadjustierte „rohe“ Ereignisrate zu einem risikoadjustierten Qualitätsindikator) sollte nicht als Qualitätskennzahl eingesetzt werden, denn sie ist dazu aufgrund der fehlenden Risikoadjustierung weniger geeignet als der zugehörige risikoadjustierte Qualitätsindikator. Die Darstellung einer solchen Kennzahl könnte jedoch zum Zweck der besseren Nachvollziehbarkeit der Berechnung des Indikatorergebnisses sinnvoll sein, z. B., um den Einfluss der Risikoadjustierung auf das Indikatorergebnis zu veranschaulichen. Kennzahlen zur Nachvollziehbarkeit der Berechnung werden keiner separaten Eignungsprüfung unterzogen, da die Daten zu ihrer Berechnung bereits im Rahmen der Eignungsprüfung des entsprechenden Indikators Gegenstand der Prüfung sind. Das Gleiche gilt für die „deskriptiven Angaben aus der Basisauswertung“ (Beispiel 5): Diese Kennzahlen sollen ebenfalls kein Qualitätsmerkmal abbilden, sondern werden zur Beschreibung der Datengrundlage oder zur Nachvollziehbarkeit der Berechnung eingesetzt. Die ihnen zugrunde liegenden Daten werden im Rahmen der Eignungsprüfung der zugehörigen Indikatoren mitgeprüft, sodass keine zusätzliche Eignungsprüfung notwendig ist.

## 2 Evaluation von Regelungen zur Qualitätssicherung

### Allgemeine Rückmeldungen

Mehrere stellungnehmende Organisationen begrüßten die Aufnahme eines Kapitels zu den Methoden der Evaluation von Regelungen in der Qualitätssicherung in die „Methodischen Grundlagen“ (u. a. KBV, S. 4; PatV, S. 3; BPTK, S. 5; GKV-SV, S. 19) und bewerteten die Einordnung der Evaluationen von Regelungen zur Qualitätssicherung als komplexe Intervention als nachvollziehbar (DGfN, S. 5; KBV, S. 4). Eine stellungnehmende Organisation führte aus, dass das Kapitel „gut nachvollziehbar und seine Anwendung für Qualitätssicherungsrichtlinien des G-BA geeignet [sei].“ (BÄK, S. 3) Mehrere stellungnehmende Organisationen wünschten sich in dem Kapitel weitergehende Konkretisierungen, z. B. hinsichtlich der konkreten (methodischen) Vorgehensweisen des IQTIG (KBV, S. 4; KZBV, S. 8; GKV-SV, S. 23-25; DKG, S. 6; PatV, S. 6; DGfN, S. 5) sowie Abgrenzungen zu seinen anderen Aufgaben (KBV, S. 4).

**IQTIG:** Das vorliegende Kapitel legt erstmalig die Methoden des IQTIG in der Durchführung von Evaluationen von Regelungen der Qualitätssicherung mit einem Schwerpunkt auf Evaluationen der „Wirksamkeit“ entsprechend § 136d SGB V dar. Begleitevaluationen, Ergebnisse der multiperspektivischen Prozessbeurteilung in Verbindung mit § 20 Satz 3 Teil 1 DeQS-RL oder die methodische Überprüfung („Evaluation“) der bestehenden Qualitätsindikatoren im Rahmen des Beschlusses des G-BA über Eckpunkte zur Weiterentwicklung der datengestützten gesetzlichen Qualitätssicherung (G-BA 2022a) sind hiervon ausgenommen und nicht Gegenstand des Evaluationskapitels, das sich ausschließlich auf Wirkungsevaluationen bezieht.

Eine weitergehende Konkretisierung dieser Methoden ist in der nächsten Überarbeitung der „Methodischen Grundlagen“ auf Grundlage der dann bestehenden aktuellen wissenschaftlichen Literatur sowie der praktischen Erfahrungen des IQTIG vorgesehen.

### Studiendesigns und Limitationen

Eine stellungnehmende Organisation begrüßte ausdrücklich die Möglichkeit der Anwendung von Interrupted-Time-Series-Ansätzen und weist zusätzlich auf die Möglichkeit weiterer quasi-experimenteller Studiendesigns hin (dggö, S. 1 f.). Eine weitere stellungnehmende Organisation bewertete die Beschreibung der Studiendesigns als gut nachvollziehbar und über das bestehende Evaluations-Rahmenkonzept (E-RK) (Veit et al. 2013) hinausgehend, forderte aber auch, Studiendesigns näher zu erläutern sowie die konkret „notwendigen Voraussetzungen“ hierfür zu benennen (GKV-SV, S. 24). Zwei stellungnehmende Organisationen wünschten sich eine weitergehende Klärung, wie in der

Evaluation geprüft würde, ob eine Interrupted Time Series angewendet werden könne (BPtK, S. 6; GKV-SV, S. 24).

Eine stellungnehmende Organisation legte dar, dass Beobachtungsstudien bei der Evaluation von Regelungen der Qualitätssicherung die Regel sein werden. Die Organisation leitete ab, dass daher der „Fokus hinsichtlich der Studiendesigns von Evaluationen darin liegen [sollte], realistische Szenarien mit ausführlicher Analyse von Limitationen und Chancen zu beschreiben, darauf einzugehen wie Limitationen begegnet werden könnte, und die Evaluationsergebnisse transparent herzuleiten“ (GKV-SV, S. 24).

Eine stellungnehmende Organisation kritisierte, dass sich das IQTIG beim Evaluationsdesign auf Vorher-nachher-Vergleiche beschränke, „mit der Begründung, dass Kontrollstudien bei bundesweit einheitlichen „Interventionen“ nicht bzw. nur deutlich eingeschränkt möglich seien“ (KBV, S. 4). Die Organisation führte weitergehend aus, dass sie erwartet hätte, „dass das IQTIG auch Effekte von Maßnahmen zur Qualitätssicherung als Vergleichsgröße erwägt, die von der Ärzteschaft selbst initiiert worden sind“ (KBV, S. 4).

In einer Stellungnahme wurde darauf hingewiesen, dass die Verwendung der Begrifflichkeiten *Prozess- und Implementierungsevaluation* missverständlich sei und das IQTIG klarer ausführen sollte, wie die Begrifflichkeiten vom IQTIG definiert werden. Bei einer vollständigen Betrachtung der Implementierung aus implementierungswissenschaftlicher Sicht sei die gewählte Form des IQTIG nicht ausreichend (DNVF, S. 4).

**IQTIG:** Die in der Evaluation von Regelungen zur Qualitätssicherung bisher eingesetzten sowie potenziell möglichen Studiendesigns sind abhängig von den konkreten Fragestellungen der zu evaluierenden Regelungen und von ihren Gegebenheiten, wie beispielsweise der Verfügbarkeit von Sekundärdaten, der Einführung der Regelung bundesweit oder nur begrenzt (Modellregion), der Evaluation einer Neueinführung oder einer Anpassung der Regelung. Da in den „Methodischen Grundlagen“ des IQTIG nicht den Anspruch verfolgt wird, alle potenziell möglichen Studiendesigns und deren Anwendungskriterien darzulegen, fokussieren diese bisher auf die aus Sicht des IQTIG wahrscheinlichsten Designs, wozu insbesondere auch die Interrupted Time Series gehören. Wie im Evaluationskapitel aufgeführt, sind aber auch weitere quasi-experimentelle Studiendesigns möglich.

Annahmen und Bedingungen einer Interrupted Time Series werden in der entsprechenden im Kapitel zitierten Fachliteratur detailliert aufgeführt und entwickeln sich weiter. Daher hat das IQTIG keine eigenen Kriterien für die Entscheidung hinsichtlich des Studiendesigns festgelegt, sondern bezieht sich auf den jeweils aktuellen wissenschaftlichen Standard. Eine Erweiterung der Darstellung möglicher Studiendesigns kann abhängig von den bis dahin vorliegenden Erfahrungen in den Evaluationen in einer zukünftigen Überarbeitung der „Methodischen Grundlagen“ berücksichtigt werden.

An dieser Stelle sei auch darauf hingewiesen, dass Interrupted Time Series Anwendung finden, wenn die Implementierung sowie die daraus entstandenen Daten entsprechend einer Beobachtungsstudie mit Vorher-nachher-Design vorliegen (Bernal et al. 2017, Hategeka et al. 2020).

Der vorgeschlagene Vergleich mit „Maßnahmen zur Qualitätssicherung, die von der Ärzteschaft selbst initiiert worden sind“, stellte in den bisher durchgeführten Evaluationen aufgrund des Fehlens einer für den Vergleich geeigneten Maßnahme keine Option dar. Das IQTIG schließt eine solche Möglichkeit aber nicht grundsätzlich aus.

Die Begrifflichkeiten hinsichtlich der Betrachtung der Implementierung im Rahmen der Evaluation wurden in der Überarbeitung des Kapitels 9 präzisiert.

Eine stellungnehmende Organisation führte aus, dass die Herausforderungen und Grenzen der Evaluationen gut nachvollziehbar beschrieben wurden, jedoch eine weitere Darstellung der bestehenden Limitationen „wie Limitation der Ressourcen bei der Datenerfassung, Limitation bei den Kosten, durch den Datenschutz, zeitliche Limitationen und Limitationen der kleinen Fallzahl, ab der keine verlässlichen Aussagen mehr zu machen sind“ (GKV-SV, S. 25) ergänzt werden könnten. Eine weitere stellungnehmende Organisation führte aus, dass aufgrund der dargelegten Limitationen das IQTIG Möglichkeiten für die Einführung von Regelungen aufzeigen solle, die eine verbesserte Evaluation ermöglichen (DNVF, S. 4).

**IQTIG:** Die in den „Methodischen Grundlagen“ benannten Limitationen beziehen sich auf grundsätzlich bestehende Limitationen mit Blick auf Evaluationen von Regelungen des G-BA zur Qualitätssicherung. Projektspezifische Limitationen werden in den jeweiligen Evaluationsplänen dargelegt. Aussagen zu weiteren Limitationen beispielsweise mit Blick auf kleine Fallzahlen, Ressourcen oder datenschutzrechtlichen Grenzen sind aus Sicht des IQTIG eher projektspezifisch, da beispielsweise kleine Fallzahlen nicht generell, sondern nur in bestimmten Kontexten limitierend sind. Demnach ist es wenig zielführend, diese allgemein zu fassen.

Da in den „Methodischen Grundlagen“ das methodische Vorgehen des IQTIG darlegt wird, sind Hinweise für eine Verbesserung von Evaluationsmöglichkeiten an dieser Stelle nicht vorgesehen. Hinweise bezüglich der Gestaltung von Evaluationen sowie zum Abbau von Evaluationshindernissen werden aber in anderen Produkten des IQTIG aufgegriffen.

### Hinweise zu Konkretisierungen

Eine stellungnehmende Organisation wies darauf hin, dass in der Darstellung des Umfelds und der Zielgruppe der Regelung zur Qualitätssicherung neben den Patientinnen und Patienten sowie Leistungserbringern weitere Stakeholder in die Umsetzung der Regelungen einbezogen sind und diese somit ebenfalls in die Evaluation mit einbezogen werden sollten (DGfN, S. 5). Eine stellungnehmende Organisation führte aus, dass die ökonomische Betrachtung in der Evaluation im Rahmen einer Kosten-Nutzen-Analyse ein zentraler Bestandteil sei und dies entsprechend in den „Methodischen Grundlagen“ dargelegt werden müsse (KBV, S. 4).

Eine weitere stellungnehmende Organisation führte aus, dass der Prozess zur Operationalisierung der Evaluationsfragestellungen sowie die Grundlage zur Ableitung von Bewertungskriterien offenbleibe, und wünschte sich dahingehend eine stärkere Präzisierung (GKV-SV, S. 22).

In einer Stellungnahme wurde ausgeführt, dass eine reine Beschreibung des Umfelds nicht ausreichend sei, da viele Einflüsse auf die Zielgruppe bereits vor einer stationären Behandlung stattfänden und diese Einflussfaktoren in den Analysen und Ergebnissen berücksichtigt werden sollten (DKG, S. 6). Weitergehend führte die stellungnehmende Organisation aus, dass die Interaktion zwischen Umfeld und Zielgruppe näher erläutert werden müsse und dass am Ende dieses Kapitels weitere Ausführungen zum Vorher-nachher-Vergleich fehlten (DKG, S. 6).

**IQTIG:** Den Einbezug weiterer Akteure in der Evaluation sieht das IQTIG regelhaft in den Fragestellungen bezüglich der Umsetzung von Regelungen zur Qualitätssicherung vor. Diese Fragen beinhalten nicht ausschließlich die Umsetzung der definierten Anforderungen durch die Leistungserbringer, sondern ebenso die Umsetzung der Maßnahmen zur Sicherstellung (Messungen der Anforderungsumsetzung, Umsetzungen der Maßnahmen zur Durchsetzung). Die „Methodischen Grundlagen“ wurden dahingehend angepasst, um den Einbezug der weiteren Akteure noch deutlicher darzulegen.

Wie einleitend im Evaluationskapitel beschrieben, wird der konkrete Umfang einer Evaluation mit dem Auftraggeber abgestimmt und leitet sich für das IQTIG aus der Beauftragung ab. In diesem Zusammenhang sind auch Kosten-/Aufwand-Nutzen-Analysen denkbar, die bislang aber nicht Gegenstand der Beauftragungen waren.

Aus Sicht des IQTIG ist das Vorgehen zur Ableitung der Evaluationsfragestellungen sowie der Bewertungskriterien bereits detailliert ausgeführt. Es ist unklar, welche weiteren Informationen benötigt werden. Das projektspezifische Vorgehen, d. h. welche Expertinnen und Experten konkret konsultiert werden oder ob Bewertungskriterien direkt aus der zu evaluierenden Regelung entnommen werden können, wird zudem im Evaluationsplan dargelegt.

Die Beschreibung des Umfelds der Zielgruppe bildet eine der Grundlagen für die Erstellung des Wirkmodells sowie für die Identifikation von Faktoren, welche einen Einfluss



auf die Umsetzung sowie auf die Wirkung der Regelung zur Qualitätssicherung haben. Darauf basierend kann abgeleitet werden, welche Faktoren im Rahmen der Evaluation berücksichtigt werden müssen. Dies schließt Faktoren auf Ebene der Patientinnen und Patienten (z. B. vorherige Versorgung) ein. Ebenso werden in diesem Rahmen potenzielle Interaktionen zwischen Zielgruppe und Intervention ausgeführt und inwieweit dies einen Vorher-nachher-Vergleich beeinflusst. Die spezifische Ausführung erfolgt dann im Rahmen des jeweiligen Evaluationsplans, da dies abhängig vom konkreten Evaluationsgegenstand ist.

### **Primärdatenerhebungen**

In einer Stellungnahme wurde ausgeführt, dass das Vorgehen zur Entwicklung von Erhebungsinstrumenten konkret dargelegt werden sollte. Zudem wurde die Frage aufgeworfen, warum keine bestehenden Erhebungsinstrumente genutzt werden (GKV-SV, S. 21).

Eine weitere stellungnehmende Organisation merkte an, dass sich der Einsatz von Patientenbefragungen in Richtlinieevaluationen von deren Einsatz als Instrument in der datengestützten einrichtungsübergreifenden Qualitätssicherung unterscheidet. Dies müsse in den „Methodischen Grundlagen“ ausgeführt werden und die Anforderungen an Patientenbefragungen im Rahmen von Evaluationen sollten dargelegt werden. Darüber hinaus wurde ausgeführt, dass „die Datenquelle der Patientenbefragung [...] bereits in vom G-BA beauftragten Evaluationen eingesetzt [wird]“ (PatV, S. 6).

**IQTIG:** Ob eine Primärdatenerhebung notwendig ist oder nicht, ist abhängig von der jeweiligen Evaluation. Die Methodik zur Erstellung eines Erhebungsinstrumentes ist wiederum abhängig vom jeweiligen Ziel der Erhebung (z. B. Erhebung über den Stand der Strukturen und Prozesse entsprechend des Evaluationsgegenstands). Für die Erhebungen bei Leistungserbringern stehen nach den Erfahrungen des IQTIG in der Regel keine bereits entwickelten Erhebungsinstrumente zur Verfügung, da es sich um regelungsspezifische Inhalte handelt. Sofern diese existieren, würde das IQTIG vorzugsweise auf validierte Erhebungsinstrumente zurückgreifen. Dies wurde im Kapitel nun verdeutlicht.

Die Einbindung von Patientenbefragungen im Rahmen der Evaluation ist immer abhängig von der konkreten Fragestellung. Entsprechend wird die Notwendigkeit einer Patientenbefragung gleichermaßen geprüft wie die Einbindung anderer Datenquellen. In diesem Zusammenhang sollten vorzugsweise bereits entwickelte und validierte Erhebungsinstrumente zum Einsatz kommen.

### **Berücksichtigung nicht intendierter Effekte**

Zwei stellungnehmende Organisationen führten weitere Beispiele für negative nicht intendierte Effekte auf (DGfN, S. 6; DGIM, S. 3). In einer weiteren Stellungnahme wurde dargelegt, dass negative nicht intendierte Effekte bereits vor der Implementation einer Regelung zur Qualitätssicherung systematisch erfasst und berücksichtigt werden (DGPPN, S. 1):

*In den Ausführungen zu negativen nicht-intendierten Effekten von QS-Verfahren werden in den Methodischen Grundlagen 2.1 zwei bekannte Effekte genannt, nämlich die Einstellung der Leistungserbringung und die nachrangige Behandlung anderer Patientengruppen, die regelhaft geprüft werden sollen. Zu weiteren nicht intendierten Effekten wird lediglich gesagt, dass diese in explorativen Ansätzen bei der Evaluationsplanerstellung identifiziert und berücksichtigt werden sollen. Es scheint jedoch kein methodisches Vorgehen zu geben, (negative) nicht-intendierte Effekte im Vorhinein systematisch zu antizipieren, ihre Größenordnung abzuschätzen und Vorkehrungen dagegen zu treffen. Die DGPPN plädiert dafür, dass ein entsprechendes methodisches Vorgehen entwickelt und zur Anwendung gebracht wird, bevor ein QS-Verfahren ausgerollt wird. (DGPPN, S. 1)*

**IQTIG:** Das vom IQTIG dargelegte Vorgehen bezieht sich ausschließlich auf das Vorgehen im Rahmen der Evaluationsplanerstellung. Die benannten nicht intendierten Effekte stellen dabei die in den Evaluationen regelhaft geprüften Effekte dar. Diese Liste wird projektspezifisch in den Evaluationsplänen erweitert.

Der in der Stellungnahme geäußerte Wunsch, dass unerwünschte Effekte bereits vorab systematisch geprüft werden, bezieht sich aus Sicht des IQTIG bereits auf die Entwicklung oder ggf. die Ausgestaltung einer Regelung zur Qualitätssicherung.

### **Bezug zum Rahmenkonzept Evaluation der BQS Institut für Qualität und Patientensicherheit GmbH (E-RK)**

Mehrere stellungnehmende Organisationen äußerten sich zum Verhältnis zwischen E-RK und dem Evaluationskapitel in den „Methodischen Grundlagen“ und wiesen darauf hin, dass das Verhältnis zwischen E-RK und dem neuen Evaluationskapitel in den „Methodischen Grundlagen“ nicht ausreichend dargelegt sei (DKG, S. 5; GKV-SV, S. 19; KZBV, S. 4; PatV, S. 7). Außerdem wurde angemerkt, dass das E-RK dem G-BA bisher als Basis für Beauftragungen sowohl an das IQTIG als auch an weitere externe Auftragnehmer diene (DKG, S. 5; GKV-SV, S. 19; KZBV, S. 4; PatV, S. 7). Eine stellungnehmende Organisation führte weitergehend aus, dass „seitens des IQTIG stets bestätigt [wurde], sich bei der Durchführung der einzelnen Schritte im Rahmen der Evaluation auf die Inhalte dieses Rahmenkonzepts zu beziehen“ (DKG, S. 5). Einer anderen Organisation war unklar, „ob sich das IQTIG bei Evaluationsaufträgen an den bisherigen Standards zur Evaluation festhält, oder inwiefern diese modifiziert werden“ (GKV-SV, S. 20) und ob „die Mindestanforderungen des Rahmenkonzepts ebenfalls als „Prüfnorm“ verstanden werden, an

der die Produkte des IQTIG zu Evaluationsfragestellungen gemessen werden können“ (GKV-SV, S. 20).

Zwei stellungnehmende Organisationen verwiesen zudem auf Kernbegriffe des E-RK, welche weiterhin Verwendung finden sollten, u. a. um unterschiedliche Interpretationen und Anwendungen der Begrifflichkeiten zur Bestimmung der Umsetzung einer Qualitätssicherungsmaßnahme in Zukunft zu vermeiden und Ergebnisse von Evaluationen besser vergleichen zu können (GKV-SV, S. 20; PatV, S. 7). Hierzu führt eine der Organisationen weitergehend aus:

**Etablierte Kernbegriffe: Umsetzung, Implementierung und Durchdringung von QS-Maßnahmen**

*Zwischen dem Evaluationsrahmenkonzept des BQS-Instituts und den Methodischen Grundlagen des IQTIG finden sich relevante Unterschiede in den Kennzahlen zur Beurteilung der „Umsetzung“ einer QS-Maßnahme. Die Begriffe Umsetzungsgrad, Durchdringungsgrad und Implementierungsgrad werden nur im ERK verwendet und dort spezifisch beschrieben (ERK S. 6). Allerdings weisen diese Definitionen Unschärfen auf, wie anhand des Beispiels „Durchdringungsgrad“ klar wird:*

- „Anteil an Einrichtungen, die eine Maßnahme realisieren oder realisiert haben (in Abgrenzung zum Implementierungsgrad und zum Umsetzungsgrad). Institutionsintern bezieht sich der Durchdringungsgrad auf Abteilungen oder sonstige Untereinheiten“ (ERK S. 6).

*Diese Definition lässt im Wortlaut offen, ob bei einer QS-Maßnahme mit mehreren Anforderungen an die Einrichtung alle Anforderungen erfüllt sein müssen, oder ob Einrichtungen, die nur einen Teil der Anforderungen erfüllen, ebenfalls in den Zähler zur Berechnung eingehen.*

*Im Unterschied dazu misst das IQTIG die Umsetzung der Anforderungen durch die Leistungserbringer „in der Regel“ anhand folgender Parameter (S. 120):*

- Anteil der Leistungserbringer mit einer regelungskonformen Umsetzung der Anforderungen an allen Leistungserbringern welche unter die Regelung fallen.
- Anteil der Leistungserbringer mit einer Umsetzung der entsprechenden Einzelanforderungen (sofern die entsprechende Regelung eine teilweise Umsetzung vorsieht/erlaubt). Hierunter fällt auch die Betrachtung der Umsetzung nach ggf. vorgesehenen Ausnahme- und Übergangsregelungen.

*Primär erscheint die Beschreibung der Kennzahlen zur Bestimmung der „Umsetzung“ einer Maßnahme im Konzept des IQTIG damit in Teilen präziser als im Rahmenkonzept – allerdings werden die im Rahmenkonzept etablierten Begrifflichkeiten Implementierungsgrad und Durchdringungsgrad vom IQTIG gar nicht verwendet. Es bleibt auch unklar, unter welchen Bedingungen das IQTIG von den Kennzahlen zur Messung der Umsetzung einer QS-Maßnahme abweicht, da ihnen nur „in der Regel“ gefolgt werden soll. (GKV-SV, S. 20 f.)*

Weiter wurde von derselben stellungnehmenden Organisation festgestellt, dass Inhalte des E-RK mit den Inhalten des Evaluationskapitels vergleichbar und sachgerecht seien. Unterschiede gebe es bei den Ausführungen zur Formulierung von Fragestellungen. In diesem Zusammenhang sei die Formulierung von Leitfragen ein „gut anwendbarer Maßstab für die praktische Arbeit im G-BA zur Beurteilung der Evaluationsberichte und zur Nachvollziehbarkeit der Evaluation und ihrer Ergebnisse“ (GKV-SV, S. 23). Hierbei forderte die stellungnehmende Organisation eine Konkretisierung im Evaluationskapitel (GKV-SV, S. 23). Weitere Unterschiede seien auch hinsichtlich der Festlegung und Anwendung von Bewertungskriterien vorhanden (GKV-SV, S. 24), wobei in diesem Punkt unklar bleibt, welche Anforderungen an das Evaluationskapitel die stellungnehmende Organisation anstrebt.

Eine stellungnehmende Organisation wies darauf hin, dass das E-RK nicht öffentlich zugänglich ist und daher eine Veröffentlichung erforderlich sei (AWMF, S. 2).

**IQTIG:** Das E-RK stellt ein Rahmenkonzept dar, mit dem Ziel, allgemeine Standards und Mindestanforderungen für Evaluationen von Qualitätssicherungsmaßnahmen, die durch Regelungen und Richtlinien des G-BA vorgesehen sind, zu definieren. Darin wird aber auch klargestellt, dass die weitere inhaltliche und konzeptionelle Gestaltung grundsätzlich frei ist (Veit et al. 2013).

Das Kapitel „Evaluation von Regelungen zur Qualitätssicherung“ in den „Methodischen Grundlagen“ des IQTIG knüpft an diese Gestaltungsfreiheit an, bewegt sich hierbei aber innerhalb der Mindestanforderungen. Es konkretisiert das E-RK oder wählt, soweit vorgesehen, andere Ausgestaltungen als die Vorschläge des E-RK.

Mit Blick auf den aktuellen Stand der Evaluationsforschung nimmt das Evaluationskapitel Aktualisierungen wie beispielsweise die Standards für Evaluationen der Gesellschaft für Evaluation e. V. (DeGEval 2016) auf, die auch das E-RK einbezieht. Es entwickelt das E-RK dahingehend konzeptionell weiter, dass es den dort bereits benannten Bezug zu komplexen Interventionen aufgreift (Veit et al. 2013: 29) und mit Blick auf Studiendesigns und Analysemethoden ausbaut.

Die hierdurch vorgenommenen Weiterentwicklungen und Konkretisierungen stellen keinen Widerspruch zum E-RK dar, vielmehr ermöglichen sie dem IQTIG, das 2013 veröffentlichte E-RK mit aktuelleren Ansätzen in Verbindung zu bringen. Entsprechend stellt das Evaluationskapitel des IQTIG eine Aktualisierung und konzeptionelle Einordnung über das E-RK hinaus dar.

Das E-RK führt drei Kernbegriffe ein: den Umsetzungsgrad, den Implementierungsgrad und den Durchdringungsgrad. Diese Begrifflichkeiten sind dort jedoch nicht eindeutig definiert, sodass es zu unterschiedlichen Interpretationen und Operationalisierungen kommt (Stellungnahme GKV-SV S. 20 f., G-BA (2024a: 151)). Das IQTIG verwendet diese Begriffe daher bewusst nicht, sondern legt eine konkrete Operationalisierung für die Messung der Umsetzung einer Regelung vor. Darüber hinaus operationalisiert das IQTIG

die Gegenstände im Rahmen von Evaluationsplänen/Konzepten mit Bezug auf die jeweilige Fragestellung und den jeweiligen Evaluationsgegenstand, um so auch die Spezifika des jeweiligen Versorgungsbereichs berücksichtigen zu können.

Schließlich sieht auch das IQTIG die Notwendigkeit, dass das E-RK öffentlich verfügbar wird, um den wissenschaftlichen Diskurs hierüber zu ermöglichen.

### **Bezug auf konkrete Evaluationsaufträge**

Eine stellungnehmende Organisation äußerte die Erwartung, dass das IQTIG für konkrete Evaluationsaufträge Konzepte im Rahmen der „Methodischen Grundlagen“ erarbeite. Die Organisation führte aus:

*Das Vorgehen für bestimmte Evaluationsaufträge sollte beschrieben werden, z. B.:*

- *Evaluation einer (Rahmen-)Richtlinie (z. B. DeQS-Richtlinie, Mindestmengen-Regelungen)*
- *Evaluation eines Qualitätssicherungsverfahrens (z. B. QS-Verfahren PCI)*
- *Evaluation einzelner Qualitätsindikatoren*
- *Evaluation von Maßnahmen zur Qualitätsförderung*

*Derartige Evaluationskonzepte und -pläne fehlen in den „Methodischen Grundlagen“ und sollten nachgearbeitet werden. (KZBV, S. 8)*

Eine weitere stellungnehmende Organisation bezog sich auf das Thema Geburtshilfe und die dort zu berücksichtigenden Aspekte (DGHWi, S. 3).

In einer weiteren Stellungnahme wurde ein Bezug zur Begleitevaluation des Verfahrens *Ambulante Psychotherapie* hergestellt und es werden konkrete Anforderungen an diese Begleitevaluation formuliert (BPtK, S. 5 f.).

In einer anderen Stellungnahme wurde darum gebeten, näher auszuführen, in welchem Umfang die Erstellung eines Evaluationskonzepts zu geringeren Aufwänden führen könne sowie ob dies einen Effekt auf die Ressourcenplanung und -auslastung des IQTIG habe (DKG, S. 6). Eine andere stellungnehmende Organisation führte aus, dass anhand eines konkreten Beispiels beschrieben werden sollte, in welchen Fällen Evaluationskonzepte genutzt werden könnten (GKV-SV, S. 20). Eine weitere stellungnehmende Organisation sah einen Widerspruch zwischen der Aussage, dass die Forschungsperspektive und der Umfang der Evaluation abhängig von den konkreten Fragestellungen der Beauftragung sind, und der Empfehlung eines Evaluationskonzepts bei mehreren Evaluationen zu thematisch ähnlichen Regelungen (KZBV, S. 8).

**IQTIG:** In den „Methodischen Grundlagen“ des IQTIG wird der Anspruch verfolgt, die grundlegenden Methoden darzulegen, nach denen das IQTIG seine Empfehlungen entwickelt. Für die Evaluation von Regelungen der Qualitätssicherung entspricht dies der Darlegung, wie Evaluationspläne erstellt werden und grundsätzlich in der Durchführung

von Evaluationen vorgegangen wird. Diese Darlegungen sind generisch formuliert und gelten für unterschiedliche Arten von Regelungen zur Qualitätssicherung (beispielsweise Mindestmengen-Regelungen (Mm-R)) übergreifend. Themenspezifische Ausführungen für konkrete Evaluationsgegenstände sind nicht Teil der „Methodischen Grundlagen“, sondern von Evaluationskonzepten und -plänen. Die Entwicklung und Darlegung von Evaluationsplänen bzw. Evaluationskonzepten erfolgt nach den jeweiligen Beauftragungen und in gesonderten Berichten.

Evaluationskonzepte bieten dabei aus Sicht des IQTIG Synergieeffekte sowohl bei den Auftragnehmern als auch dem Auftraggeber. Dies wurde entsprechend im Kapitel ergänzt. Evaluationskonzepte sind sinnvoll, wenn es starke Überschneidungen gibt, z. B. dieselbe Regelung für verschiedene Versorgungsbereiche zugrunde liegt (Mm-R, Qualitätsverträge nach § 110a SGB V). Die grundsätzliche Aussage, dass die Forschungsperspektive und der Umfang der Evaluation abhängig von den konkreten Fragestellungen der Beauftragung sind, steht hierzu in keinem Widerspruch. In solchen Fällen überscheiden sich lediglich die interessierenden Fragestellungen und damit auch die Forschungsperspektive, sodass sich idealerweise Umfänge reduzieren lassen. Aussagen zu Kontingenten sind nicht Teil der „Methodischen Grundlagen“ und müssen an anderer Stelle berücksichtigt werden.

### **Unabhängigkeit der Evaluation**

Mehrere stellungnehmende Organisationen thematisierten die Frage, ob das IQTIG unabhängig in der Evaluation sei, sofern es selbst an der Entwicklung bzw. Umsetzung der entsprechenden Verfahren des G-BA beteiligt war, bzw. führten an, dass das IQTIG keine Evaluation einer Qualitätssicherungsmaßnahme und ihrer entsprechenden Regelungen vornehmen sollte, die es selbst entwickelt hat (AWMF, S. 3; KBV, S. 4; KZBV, S. 4; BPTK, S. 5; DGfN, S. 5). Das IQTIG versteht hier die Maßnahmen im Rahmen der DeQS-RL.

Eine stellungnehmende Organisation führte weitergehend aus, es bestehe die „Gefahr, dass die Validität entwickelter QS-Verfahren selbstreferenziell bestätigt und lediglich eine scheinbare Eignung von Qualitätsindikatoren und -indikatorensets suggeriert wird, bei der tatsächliche Defizite oder unerwünschte Nebenwirkungen durch die Einführung von QS-Verfahren bzw. einzelner Qualitätsindikatoren übersehen werden“ (BPTK, S. 5).

Eine stellungnehmende Organisation führte aus, dass im Kapitel die Hinweise fehlten, „dass das IQTIG eine Vielzahl der QS-Maßnahmen des G-BA selbst entwickelt hat und dass das IQTIG in unterschiedlichen Rollen selbst Verfahrensteilnehmer von QS-Maßnahmen des G-BA ist. Die Rollen des IQTIG sind beispielsweise „datenhaltende Stelle“, „umsetzende Stelle von QS-Maßnahmen mit dem Auftrag der Datenauswertung, Verfahrenspflege und Weiterentwicklung“, „datenbewertende und datenveröffentlichende Stelle“ oder „Geschäftsstelle der Bundesstelle für die bundesbezogene, externe QS“ (DGfN, S. 5).

Eine weitere stellungnehmende Organisation wünschte eine Aufzählung, „welche Bestandteile der Evaluation das Institut selbst durchführen kann und welche Bestandteile anderweitig vergeben werden müssten“ (KBV, S. 4), und darüber hinaus „eine kritische Auseinandersetzung mit den Limitationen einer Selbstevaluation“ (KBV, S. 4).

**IQTIG:** Die Unabhängigkeit von Evaluationen ist ein relevantes Thema und stellt u. a. eine der Mindestanforderungen der Gesellschaft für Evaluation e. V. (DeGEval 2016) dar, an denen sich auch das IQTIG orientiert. Auf dieser Grundlage stimmt das IQTIG zu, dass die Unabhängigkeit einer Evaluation insbesondere dann relevant ist, wenn das IQTIG an der Entwicklung bzw. Ausgestaltung der Regelung zur Qualitätssicherung beteiligt ist oder war. Die Einschätzung und Überprüfung der Güte („Eignung“) der vor allem im Rahmen der DeQS-RL eingesetzten Qualitätsindikatoren kann auch als „Evaluation“ der Qualitätsindikatoren verstanden werden. Dies ist jedoch ein zentraler Teil der Aufgaben des IQTIG im Bereich des Regelbetriebs.

Im Kapitel zu Evaluationen wird der Fokus auf Wirkungsevaluationen gesetzt. Bislang führt das IQTIG aber keine Wirkungsevaluation von Qualitätssicherungsmaßnahmen und Regelungen (im Sinne des vorliegenden Evaluationskapitels) durch, an deren Entwicklung oder Ausgestaltung es direkt beteiligt ist oder war.

Formal findet darüber hinaus die in den „Methodischen Grundlagen“ dargelegte Vorgehensweise erst nach einer möglichen Beauftragung Anwendung. Die Erfüllung der Anforderungen an die Unabhängigkeit des Evaluators (ebenso wie Anforderungen an Kompetenz, Integrität etc.) sind durch den Auftraggeber zu beurteilen und sollten vor einer Beauftragung definiert werden (DeGEval 2016, Widmer 2012). Der Evaluator selbst steht wiederum in der Verpflichtung, transparent über etwaige Interessenkonflikte zu informieren und ggf. Maßnahmen zur Sicherstellung der Unabhängigkeit mit dem Auftraggeber abzustimmen.

### **Veröffentlichung von Evaluationsergebnissen**

In einer Stellungnahme wurde die Frage der Veröffentlichung von Evaluationsergebnissen aufgeführt. Hierbei sah die stellungnehmende Organisation die Notwendigkeit, eine Veröffentlichung hinsichtlich erwünschter sowie unerwünschter Effekte zu prüfen:

*Vorbehaltlich einer Beschlussfassung durch den G-BA erfolgt die Veröffentlichung der Evaluationsergebnisse durch das IQTIG. Hierbei ist ein bewusster Umgang mit dem Aspekt der Veröffentlichung notwendig. Ähnlich wie die Qualitätssicherung im Methodenpapier als komplexe Intervention verstanden wird, kann auch die Veröffentlichung von Evaluationsergebnissen verschiedene Wirkungen haben. Diese können, bezogen auf die Richtlinienziele, sowohl erwünschte als auch unerwünschte Effekte beinhalten. (PatV, S. 6)*

**IQTIG:** Die Veröffentlichung von Ergebnissen ist nicht Teil der „Methodischen Grundlagen“ und wurde daher in dem Kapitel nicht aufgeführt. Grundsätzlich spricht sich das IQTIG für eine Veröffentlichung aller Evaluationsprodukte (Evaluationsplan, Evaluationskonzept, Zwischenbericht, Abschlussbericht) aus, um eine möglichst hohe Transparenz bezüglich der Evaluation zu schaffen und hierdurch Vertrauen in die Evaluation und ihre Ergebnisse zu fördern.



## 3 Literaturrecherchen und Evidenzsynthese

### Orientierende Recherche

In einer Stellungnahme wurde die Frage aufgeworfen, wie, wenn die orientierende Recherche der Einarbeitung bzw. der Themenerschließung dient, die Ergebnissicherung bezogen auf ein zu planendes Projekt stattfindet (PatV, S. 8 f.). Von einer stellungnehmenden Organisation wurde der Ausbau des Kapitels 10 grundsätzlich begrüßt. Es wurde jedoch angemerkt, dass die Ergebnisse von orientierenden Recherchen transparenter als bislang offenlegt werden sollten, da diese nicht nur eine zentrale Informationsgrundlage für die Entwicklung oder Weiterentwicklung von QS-Verfahren des IQTIG darstellen, sondern im Rahmen der Evaluation von QS-Regelungen auch als eine wesentliche Methode zur Aufdeckung von Einflussfaktoren fungieren, die die Wirkung einer Intervention auf Endpunkte überlagern können. Auch wenn sich eine Aufschlüsselung von Suchstrategien bei orientierenden Recherchen in der Regel nicht konsistent und damit sinnvoll darstellen lassen dürfte, sollte zumindest die dadurch ermittelte Literatur regelhaft im Anhang von Berichten aufgelistet werden (BPtK, S. 6 f.).

Weiterhin wurde in einer Stellungnahme angemerkt, dass in diesem Kapitel weitere ergänzende Quellen fehlten, wie etwa der regelmäßig veröffentlichte UPD-Monitor (PatV, S. 8 f.).

**IQTIG:** Im Rahmen von orientierenden Recherchen wird gezielt nach wissenschaftlichen Informationen gesucht, die zur Beantwortung der jeweiligen Fragestellung (z. B. Einarbeitung ins Thema) benötigt werden. Dabei werden die Quellen so ausgewählt, dass vorrangig vertrauenswürdige, sichere Informationen identifiziert werden. Da es sich in den „Methodischen Grundlagen“ nur um einige wenige Quellenbeispiele handelt, die je nach Fragestellung ausgewählt werden, wurde auf eine Erweiterung der genannten Quellen verzichtet.

Für orientierende Recherchen zur Themenerschließung erfolgt keine detaillierte Darstellung der Suche, die gefundene Literatur wird in den Berichten zitiert und im Literaturverzeichnis gelistet. Die Darstellung nicht verwendeter Literatur ohne die Nutzung von Ein- und Ausschlusskriterien ist unüblich und nicht sinnvoll.

### Systematische Recherchen

In einer Stellungnahme wurde angemerkt, dass an manchen Stellen leichte Unklarheiten den Lesefluss erschweren würden (PatV, S. 9). Von einer weiteren stellungnehmenden Organisation wurde vorgeschlagen, im Falle von systematischen Literaturrecherchen das Cochrane Handbuch als orientierenden, methodischen Rahmen zu referenzieren (EbM-Netzwerk, S. 1). Zudem wurde empfohlen, die aktuelle Publikation von PRISMA 2020 statt der Vorversion von Liberati et al. zu zitieren (EbM-Netzwerk, S. 1). Weiterhin wurde von einer stellungnehmenden Organisation gewünscht, dass die aktuelle Version

des AWMF-Regelwerks (Ständige Kommission Leitlinien der AWMF 2023) zitiert wird (AWMF, S. 1).

**IQTIG:** Aufgrund der Anmerkungen in den Stellungnahmen wurden im entsprechenden Abschnitt 10.2 Konkretisierungen in den methodischen Beschreibungen und Anpassungen einiger Referenzen vorgenommen. Eine systematische Literaturrecherche - im Sinne eines systematischen Reviews - ist eine Zusammenfassung der Literatur zu einer klar formulierten Frage nach einem a priori definierten Konzept, welches auf einem systematischen, reproduzierbaren Vorgehen basiert. Vorhandene Studien zu einer Frage werden recherchiert, auf Relevanz geprüft und kritisch bewertet. Aus den (aufgrund vorab definierter Kriterien) identifizierten Studien werden die Ergebnisse extrahiert und zusammengefasst.

### Scoping Review

In drei Stellungnahmen wurden ergänzende Ausführungen zum Scoping Review gewünscht (KBV, S. 4 f.; KZBV, S. 9 f.; PatV, S. 9). Eine stellungnehmende Organisation wies darauf hin, dass sich die Methodik nicht von der einer orientierenden Recherche unterscheiden ließe und es daher unklar bliebe, ob - wie bei der orientierenden Literaturrecherche - nur eine Person recherchieren würde (auch PatV, S. 9) und eine Dokumentation der Suchstrategie ebenfalls unterbliebe (KBV, S. 4 f.). Von der anderen stellungnehmenden Organisation wurde außerdem angemerkt, dass die skizzierten Herangehensweisen zwar verständlich und plausibel dargelegt würden, aber ausführlichere Darstellungen mit Beispielen empfohlen (KZBV, S. 9 f.).

**IQTIG:** Die Methodik beim Scoping Reviews ähnelt der einer systematischen Recherche (und nicht der einer orientierenden Recherche). Zu eher breiten Fragestellungen erfolgt eine umfangreiche Recherche, um einen systematischen Gesamtüberblick zum Thema zu erhalten. Bei einem zweifach unabhängigen Screeningprozess können die Ein- und Ausschlusskriterien ggf. im Prozess noch angepasst werden. Es erfolgen eine detaillierte Dokumentation der Recherche inkl. Flussdiagramm sowie eine deskriptive Beschreibung der gefundenen Literatur. Eine kritische Bewertung der Literatur erfolgt hier allerdings nicht regelhaft. Für eine bessere Unterscheidung zwischen systematischen Recherchen und Recherchen im Sinne von Scoping Reviews wurden einige Ergänzungen im Abschnitt 10.3 vorgenommen.

### Leitlinien

In einer Stellungnahme wurde erneut (wie bereits 2021) kritisiert, dass aktuelles Leitlinienwissen nicht ausreichend in den Indikatoren des IQTIG abgebildet wird (AWMF, S. 3). Die Gründe dafür sollten verfahrensbezogen in den Qualitätsreports dargelegt werden. Es wurde empfohlen, Leitlinienwissen für die angewandten Indikatoren jeweils aktuell zu halten (AWMF, S. 3).

**IQTIG:** Im Zuge der Eckpunktebeauftragung zur Weiterentwicklung der datengestützten gesetzlichen Qualitätssicherung (G-BA 2022a) wurden bzw. werden aktuell die Qualitätsindikatoren aller Leistungsbereiche vom IQTIG geprüft. Vor allem für das Eignungskriterium „Zusammenhang mit einem unmittelbar patientenrelevanten Merkmal“ wurden dabei nur aktuelle Leitlinien berücksichtigt. Die Qualitätsindikatoren, für die eine Weiterführung empfohlen wurde, basieren also i. d. R. auf aktuellem Leitlinienwissen. Auch nach Beendigung des Eckpunkteprojektes werden vom IQTIG alle Qualitätsindikatoren in regelmäßigen Abständen auf ihre Eignung im Regelbetrieb geprüft werden. Auch hier sieht die Methodik laut den „Methodischen Grundlagen“ (siehe z. B. Kapitel 8 bzw. Anhang A) aktuelles Leitlinienwissen als eine wichtige Säule bei der Entscheidung an, ob die Qualitätsindikatoren beibehalten oder zur Abschaffung empfohlen werden.

Von einer stellungnehmenden Organisation wurde angemerkt, dass es bei der Bewertung der Evidenzgrundlage von Qualitätsindikatoren auf Basis von Leitlinien durchaus zu Konflikten kommen könne, wenn statt einer in der Versorgung in Deutschland etablierten deutschen oder europäischen Leitlinie eine andere ausländische Leitlinie als gültiger Maßstab angelegt werde (BÄK, S. 3 f.). Es wurde vermutet, dass es für die Leistungserbringer schwer nachvollziehbar sei, wenn dieser externe Maßstab an die Versorgungsqualität angelegt werde (BÄK, S. 3 f.).

**IQTIG:** Im IQTIG wird projektspezifisch ausgehend von der Fragestellung entschieden, welche Evidenz mindestens vorliegen muss, um eine Recherchefrage mit möglichst hoher Aussagesicherheit und möglichst geringer Fehleranfälligkeit beantworten zu können (siehe Abschnitt 10.4). Um beispielsweise den Zusammenhang mit einem unmittelbar patientenrelevanten Merkmal zu belegen, werden evidenzbasierte Leitlinien berücksichtigt. Werden bei der Recherche dabei auch deutsche evidenzbasierte Leitlinien identifiziert, werden diese auch bei der Synthese der Ergebnisse bzw. zur Beantwortung der Recherchefrage herangezogen (wobei höherbewertete Leitlinien eine höhere Gewichtung bei der inhaltlichen Beurteilung erhalten). Widersprüche zu nicht evidenzbasierten Leitlinien, die im deutschen Versorgungskontext etabliert sind, werden diskutiert.

Von zwei stellungnehmenden Organisationen wurde kritisiert, dass die deutschen Expertenstandards der Pflege den evidenzbasierten Leitlinien gleichgesetzt werden. Dies sollte überdacht bzw. erläutert werden (AWMF, S. 3; DGfN, S. 6).

**IQTIG:** Unter evidenzbasierten Leitlinien versteht das IQTIG Leitlinien mit eindeutig identifizierbaren Empfehlungen, die auf einer systematischen Literaturrecherche aufbauen und mit den jeweiligen Evidenzleveln (LoE, Level of Evidence) bzw. Empfehlungsstärken (GoR, Grade of Recommendation) versehen sind. Da dies bei den deutschen Expertenstandards der Pflege nicht vorliegt, wurde der Satz in Abschnitt 10.2.1 „Deutsche

Expertenstandards der Pflege werden evidenzbasierten Leitlinien gleichgesetzt“ gestrichen. Da die Evidenzanforderungen, wie in Abschnitt 10.4.1 beschrieben projektspezifisch definiert werden können, kann für einzelne Projekte auch ein Expertenstandard der Pflege zur Beantwortung einer Recherchefrage herangezogen werden.

### Kritische Bewertung

Drei stellungnehmende Organisationen hatten Anmerkungen zur kritischen Bewertung von Leitlinien (BÄK, KBV, DGIM). Dabei wünschten sich zwei Organisationen eine tiefergehende Erklärung, warum nur noch die drei Domänen 2, 3 und 6 des AGREE-II-Instrumentes (Stakeholder Involvement, Rigour of Development, Editorial Independence) für die Bewertung herangezogen werden, obwohl die nicht beachteten Domänen 1, 4 und 5 (Scope and Purpose, Clarity of Presentation und Applicability) die Qualität der Leitlinien aus Anwenderperspektive beschreiben würden. Diese seien laut der Stellungnehmenden besonders wichtig, um für eine breite Akzeptanz in den Arztpraxen und Krankenhäusern zu sorgen (BÄK, S. 3). Zudem sollte klargestellt werden, ob sie gar nicht oder in anderer Form in die Punktwertung eingehen (KBV, S. 5). Weiterhin wurde die Frage aufgeworfen, ob sich Domäne 2 und Domäne 6 nicht überschneiden würden (DGIM, S. 3).

**IQTIG:** Die Beschränkung der kritischen Bewertung mit dem AGREE-II-Instrument auf die Domänen 2, 3 und 6 erfolgt mit Blick auf die Zielsetzung, eine evidenzbasierte Grundlage für die Qualitätsindikatoren zur Verfügung zu stellen. Vor allem bei mittelbar patientenrelevanten Qualitätsmerkmalen (also Ergebnissen, Prozessen oder Strukturen, denen zwar kein Wert an sich zugeschrieben wird, deren Vorliegen bzw. deren Durchführung aber mit einer höheren bzw. niedrigeren Wahrscheinlichkeit des Auftretens unmittelbar patientenrelevanter Ereignisse einhergeht) muss der Nachweis erbracht werden, dass der Zusammenhang mit einem unmittelbar patientenrelevanten Merkmal besteht. Dafür spielt die Evidenzbasierung (Domäne 3) bzw. die methodische Qualität der berücksichtigten Leitlinien eine entscheidende Rolle. Domäne 2 fokussiert dabei auf die multidisziplinäre Zusammensetzung der Leitliniengruppe mit Beteiligung von Patientinnen und Patienten, während Domäne 6 die finanzielle Unabhängigkeit der Leitlinienorganisation bzw. den Umgang mit Interessenskonflikten von Leitlinienautorinnen und -autoren adressiert. Eine solche Auswahl an berücksichtigten Domänen (2, 3, und 6) ist etabliert und wird beispielsweise auch vom Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) bei der Entwicklung von Disease-Management-Programmen vorgenommen (IQWiG 2023).

Die anderen Domänen (Scope and Purpose, Clarity of Presentation und Applicability) zielen eher auf die Implementierung und Anwendung der Leitlinien ab (und damit nicht auf die Evidenzbasierung bzw. die methodische Qualität) und werden daher zur Beantwortung der Recherchefragen im IQTIG nicht berücksichtigt.

Eine stellungnehmende Organisation stimmte den Ausführungen zur kritischen Bewertung der Literatur zu (AWMF, S. 3). Für eine andere stellungnehmende Organisation blieb allerdings unklar, wie viele Personen überhaupt Bewertungen vornehmen, und es wurde in Bezug darauf eine Präzisierung erbeten (DKG, S. 7). In einer weiteren Stellungnahme wurde eine Erklärung gewünscht, warum spezifische Instrumente für die Qualitätsbewertung von Originalarbeiten (systematische Übersichtsarbeiten, Primärliteratur) gewählt wurden (GMDS/IBS-DR, S. 2). Weiterhin wurde in einer Stellungnahme vorgeschlagen, als Orientierung für das Reporting PRISMA 2020 und für die Bewertung der Vertrauenswürdigkeit der Evidenz auf GRADE zu verweisen (EbM-Netzwerk, S. 1)<sup>2</sup>. Von einer weiteren stellungnehmenden Organisation wurde angemerkt, dass die Verwendung des Critical Appraisal Skills Programme (CASP 2013) geprüft und neu bewertet werden müsse (DNVF, S. 4 f.).

**IQTIG:** Wie der Screeningprozess wird auch die kritische Bewertung von Leitlinien, systematischen Reviews und Primärliteratur (z. B. randomisierte kontrollierte Studien, Kohortenstudien) von jeweils zwei Personen unabhängig voneinander vorgenommen. Diese Information findet sich im Abschnitt 10.4.2 der „Methodischen Grundlagen“.

Für die Bewertung der methodischen Qualität der Literatur verwendet das IQTIG validierte bzw. in der Wissenschaft etablierte Instrumente; für systematische Reviews beispielsweise das standardisierte Bewertungsinstrument AMSTAR-2. Für die Primärliteratur wurden ebenfalls mögliche Bewertungstools genannt, projektspezifisch können hier aber auch andere standardisierte Bewertungsinstrumente (Critical Appraisal Skills Programme) geprüft und verwendet werden.

Für die systematischen Recherchen und Recherchen nach dem Scoping-Ansatz wird der gesamte Recherche- und Auswahlprozess transparent nach PRISMA 2020 inklusive eines Flussdiagrammes dargestellt. Die aktuellere Referenz von Page et al. (2021) wurde in Abschnitt 10.2 ergänzt.

### **Beste verfügbare Evidenz**

In einer Stellungnahme wurde angeregt, den Prozess der Entscheidung genauer zu beschreiben, wann weniger hochwertige Leitlinien für eine Fragestellung berücksichtigt werden, um auf Recherchen nach systematischen Übersichtsarbeiten oder Primärliteratur zu verzichten (DKG, S. 7). Auch in einer zweiten Stellungnahme wurde gefordert, dass die Faktoren, die die Aufwandsangemessenheit einer bestimmten Recherche begründen, genauer beschrieben werden (PatV, S. 9 f.).

Von einer anderen stellungnehmenden Organisation wurde darauf hingewiesen, dass insbesondere für die Neu- und Weiterentwicklung von Qualitätsindikatoren und Qualitätssicherungsverfahren belastbare evidenzbasierte Qualitätsdefizite vorliegen müs-

---

<sup>2</sup> Die AWMF hat sich in diesem Punkt den Ausführungen des EbM-Netzwerks angeschlossen (AWMF, S. 3).

sen und daher das Nutzen von weniger hochwertigen Leitlinien und anderen Publikationen (im Sinne von bester verfügbaren Evidenz) sich nur bedingt erschließt (KBV, S. 5). In einer weiteren Stellungnahme wurde begrüßt, dass bezüglich des Evidenzniveaus ein projektspezifisches Herangehen vorgesehen sei und jeweils die beste verfügbare Evidenz herangezogen werde (PatV, S. 3).

In einer weiteren Stellungnahme wurde angemerkt, dass es keine nachvollziehbare Vorgehensweise für Sonderfälle gibt, wenn aus Leitlinien oder wissenschaftlichen Studien – d. h. in der Regel aus Quellen mit besonders hoher Aussagesicherheit – benötigte Informationen nicht in zufriedenstellendem Umfang gewonnen werden können (BPtK, S. 7 f.). Hier solle ein Entscheidungsalgorithmus im Falle fehlender Evidenz aus der systematischen Literaturrecherche bei der Auswahl und Gewichtung von Informationen zur Ableitung von Qualitätsmerkmalen formuliert werden (BPtK, S. 7 f.). Es wurde in dieser Stellungnahme weiter angeregt, jene Informationen, die als Grundlage zur Entwicklung von QS-Verfahren herangezogen werden, möglichst vollständig im Anhang offenzulegen (auch Ergebnisse aus Fokusgruppen oder Expertengremien).

**IQTIG:** Die Recherchefragen sollen im IQTIG mit möglichst hoher Aussagesicherheit und möglichst geringer Fehleranfälligkeit beantwortet werden, unter Berücksichtigung des Rechercheaufwands und -nutzens. Die Festlegung, welche Evidenz mindestens vorliegen muss, erfolgt dabei projektspezifisch und wird in den jeweiligen Berichten erläutert und begründet. Ein allgemeiner Entscheidungsalgorithmus existiert daher nicht.

Das IQTIG hat zur Ableitung von Qualitätsmerkmalen ein definiertes methodisches Vorgehen: Qualitätsmerkmale werden dabei nicht nur aus Leitlinien und Studien abgeleitet, sondern auch aus weiteren Wissensquellen (siehe Kapitel 12 der „Methodischen Grundlagen“). In den Entwicklungsberichten wird die Herleitung von Qualitätsmerkmalen bis zur Operationalisierung zu Qualitätsindikatoren – insbesondere auch die Ergebnisse und Limitationen der verschiedenen Wissensquellen – immer ausführlich dargestellt. Bei der Ableitung der Qualitätsmerkmale ist stets ein Expertengremium involviert, mit dem die Inhalte aus den Wissensquellen diskutiert werden. Entsprechend den „Methodischen Grundlagen“ hat das Expertengremium bei allen Entwicklungsschritten eine beratende Funktion für das IQTIG. Die Empfehlungen des Gremiums werden vom IQTIG nach jedem Treffen reflektiert und berücksichtigt. Das Expertengremium ist somit eine unter verschiedenen Wissensquellen, die das IQTIG bei der Entwicklung von Qualitätsindikatoren nutzt. Die Entscheidung, welche Qualitätsmerkmale in die weitere Indikatorenentwicklung eingehen, wird durch das IQTIG nach einer umfassenden Aufbereitung und Gegenüberstellung der Ergebnisse aus allen verfügbaren Wissensquellen getroffen. Das IQTIG hält eine tabellarische Auflistung aller Ratingergebnisse für nicht zielführend, da das IQTIG erst nach Aufbereitung der Einschätzungen des Expertengremiums und in der Zusammenschau mit den Ergebnissen aus den anderen Wissensquellen eine finale Einschätzung vornimmt.

## Evidenzsynthesen

Von einer stellungnehmenden Organisation wurde kritisiert, dass sich zu wenige Angaben zur Methodik der Evidenzsynthese, sprich Meta-Analysen im Kapitel 10 finden lassen (GMDS/IBS-DR, S. 2<sup>3</sup>). Diese Aspekte sollten weiter detailliert werden, insbesondere vor dem Hintergrund, dass die avisierten Anwendungen des IQTIG durch Heterogenität zwischen Datenquellen geprägt sein dürften, sowohl bezüglich der Populationen als auch der Studienmethodik, wie z. B. Studiendesigns und Modellbildung. Vor jeder formalen Zusammenfassung von Daten sollte immer erst geprüft werden, ob diese in der jeweiligen Situation angebracht ist. Hierzu sollten die Voraussetzungen im Methodenpapier geschildert werden (GMDS/IBS-DR, S. 2).

Auch in einer weiteren Stellungnahme wurde angemerkt, dass die Evidenzsynthese nicht ausreichend beschrieben sei (DNVF, S. 4 f.).

Eine andere stellungnehmende Organisation sah das Vorgehen bei der Extraktion am Ende des Prozesses, dass eingeschlossene Leitlinien und Literatur tabellarisch aufgelistet und stichprobenhaft durch eine weitere Person geprüft werden, als sinnvoll an (KZBV, S. 10). Von einer anderen stellungnehmenden Organisation wurde allerdings erfragt, warum die tabellarische Extraktion der eingeschlossenen Leitlinien und Literatur von einer zweiten Person nur stichprobenhaft und nicht in Gänze geprüft werde bzw. wie und durch wen der Umfang der stichprobenhaften Prüfung festgelegt werde (PatV, S. 10). Weiterhin sei am Ende des Unterkapitels 10.4.3 die Einfügung des Begriffs „Outcome“ neben „Ergebnissen“ nicht selbsterklärend (PatV, S. 10).

**IQTIG:** Die Evidenzsynthesen des IQTIG sind bisher ausschließlich deskriptiv erfolgt. Metaanalysen sind aufgrund der Anzahl (mindestens zwei Studien) und Heterogenität der eingeschlossenen Publikationen bisher nicht durchgeführt worden. Sollten Metaanalysen durchgeführt werden, wird das IQTIG sich an wissenschaftlichen Standards orientieren und seine Methode ausführlich darlegen.

Für die eingeschlossenen Publikationen erfolgt die Extraktion durch eine Person. Eine zweite Person überprüft stichprobenartig die extrahierten Angaben. Dabei wird vor allem geprüft, ob Empfehlungsstärken, Evidenzlevel und/oder Studiendesigns richtig übernommen bzw. angegeben wurden, da diese für die Aussagesicherheit zur Beantwortung der Recherchefragen eine wesentliche Rolle spielen. Der Umfang der Stichprobe ist abhängig von der Datenqualität der geprüften Extraktion.

Wie am Ende in Abschnitt 10.4.3 beschrieben, werden relevante Informationen, wie z. B. Publikationsjahr, Studiendesign, Studienpopulation, Intervention, Outcomes (also die untersuchten Endpunkte der Studie, z. B. Mortalität, Morbidität, Lebenszufriedenheit) und Ergebnisse (also die Endergebnisse der Studie, etwa Nutzen oder Schaden der Intervention auf die jeweiligen Outcomes/auf die jeweiligen Endpunkte) extrahiert.

---

<sup>3</sup> Die AWMF hat sich in diesem Punkt den Ausführungen der GMDS/IBS-DR angeschlossen (AWMF, S. 3).

## Ethikrichtlinien und Gesetze

In einer Stellungnahme wurde empfohlen, neben Leitlinien, systematischen Reviews und Primärstudien auch Ethikrichtlinien und Gesetze in die Recherche einzubeziehen (DGHWi, S. 3).

**IQTIG:** Für sämtliche Projekte bzw. zur Informationsgewinnung für die Beantwortung von Recherchefragen werden auch in Deutschland geltende Gesetze, untergesetzliche Normen und Richtlinien regelhaft berücksichtigt. Bei systematischen Recherchen im Sinne eines systematischen Reviews bzw. bei Recherchen im Sinne eines Scoping Reviews werden diese allerdings nicht systematisch recherchiert, sondern stehen aufgrund ihres bindenden Charakters über allen anderen systematischen Recherchen und Erkenntnissen. Unter Umständen kann bei Vorliegen eines Gesetzes auch auf eine eigentlich geplante systematische Recherche verzichtet werden, wie beispielsweise im Abschlussbericht der Eckpunktebeauftragung für die QS-Verfahren *Perkutane Koronarintervention und Koronarangiographie (QS PCI)* und *Versorgung mit Herzschrittmachern und implantierbaren Defibrillatoren (QS HSMDEF)* jeweils für die Indikatorgruppe „Dosis-Flächen-Produkt“ gezeigt (IQTIG 2023: 76 f. und 189 f.).

## Qualifikation

Von einer stellungnehmenden Organisation wurde angemerkt, dass es keine Spezifizierung darüber gäbe, welche Qualifikationen die Personen aufweisen, die die Literaturrecherche und wissenschaftliche Bewertung durchführen (DGHWi, S. 3). Es wurde empfohlen, dies in den „Methodischen Grundlagen“ zu vermerken (angelehnt beispielsweise an die DIN-Norm EN ISO 9001:2015, in welcher spezifiziert wird, dass die Qualifikationen von Beauftragten für Qualitätssicherung ausführlich dargestellt werden sollen). Im Falle einer geburtshilflichen Recherche sollten beispielsweise immer beide Berufsgruppen, also sowohl Hebammen als auch Ärztinnen und Ärzte hinzugezogen werden (DGHWi, S. 3).

**IQTIG:** Im IQTIG stellen alle Fachbereiche, die bei der Entwicklung und Pflege von Qualitätsindikatoren beteiligt sind, ihre spezifische Expertise zur Verfügung, sodass bei den Projekten immer ein multidisziplinäres Team zusammenarbeitet (aus den Bereichen der Informationswissenschaften, Gesundheitswissenschaften, Medizin, Pflege, Biometrie, Informatik usw.). Die systematischen Literaturrecherchen bzw. die Recherchen im Sinne eines Scoping Reviews sowie die Bewertungen der eingeschlossenen Literatur am IQTIG werden durch qualifizierte und geschulte Personen durchgeführt, die langjährige Erfahrungen bei der Durchführung von Recherchen bzw. in der Nutzung der Bewertungsinstrumente haben.



**Unterstützung durch Künstliche Intelligenz (KI)**

Von einer stellungnehmenden Organisation wurde angeregt, zu einzelnen Prozessschritten der Literaturrecherche und -bewertung ergänzend auch KI-gestützte Systeme zu prüfen (PatV, S. 10).

**IQTIG:** Die aktuellen und schnellen Entwicklungen bei KI-gestützten Systemen werden vom IQTIG beobachtet. Für das in der vorliegenden Version der „Methodischen Grundlagen“ dargelegte Vorgehen bei der Literaturrecherche und Evidenzsynthese ist eine verstärkte Nutzung von KI-gestützten Systemen über die in den genannten Tools bereits enthaltenen Elemente hinaus nicht vorgesehen. Sollten sich durch die internen Projekte veränderte Handlungsweisen ergeben oder sich Methoden dahingehend konkretisieren, werden diese in der nächsten Version der „Methodischen Grundlagen“ aufgezeigt.

## 2 Auswahl von Inhalten bei der Entwicklung von Qualitätsmessungen

Von einer stellungnehmenden Organisation wurden alternative Vorgehensweisen zur Auswahl von Inhalten für die Qualitätsmessung zur Diskussion gestellt. Beispielsweise könne überlegt werden, nicht von einem Qualitätsmodell auszugehen, sondern alternativ von der verfügbaren Datengrundlage. Ausgehend von Datenfeldern, „die typischerweise in einer elektronischen Patientenakte“ zur Prüfung von Leitlinienkonformität und Komplikationen verfügbar seien, könnten Qualitätsindikatoren formuliert werden (DNVF, S. 3). An anderer Stelle wurde die Anregung gegeben, zunächst aus allen Qualitätsdimensionen des Rahmenkonzepts Qualitätsaspekte abzuleiten und Analysen bezüglich Über-, Unter- und Fehlversorgung auf Ebene der Qualitätsaspekte und nicht erst auf Ebene der Qualitätsmerkmale durchzuführen (DNVF, S. 5). Weiterhin wurde die Einschätzung geäußert, dass bei der Indikatorenentwicklung auch iterativ vorgegangen werden könnte, wobei die Qualitätsaspekte erst nach der Benennung von Qualitätsmerkmalen definiert würden (DNVF, S. 5).

**IQTIG:** Eine datengetriebene Auswahl von Inhalten für die Qualitätsmessung, die vorwiegend auf automatisiert erhebbaren Daten aufbaut, könnte zwar mit einer hohen Praktikabilität der Messung einhergehen, führt aber voraussichtlich zu unausgewogenen Indikatorensets. Beispielsweise sind Daten, die für die Beurteilung der Qualitätsdimension „Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten“ benötigt werden, in der Regel nicht der Patientenakte zu entnehmen. Das IQTIG empfiehlt daher weiterhin, die Inhalte für Qualitätsmessungen ausgehend vom möglichen Patientennutzen abzuleiten und im zweiten Schritt zu prüfen, ob die als relevant erachteten Inhalte abbildbar sind. Nur so ist ein Abgleich möglich zwischen den Inhalten, die für die Messung der Versorgungsqualität in einem Themenbereich als relevant erachtet werden, und dem, was mit den bestehenden Datenquellen abgebildet werden kann. Das Ausmaß der Übereinstimmung wird als Inhaltsvalidität bezeichnet und stellt ein wichtiges Gütekriterium für Indikatorensets dar (siehe auch Kapitel 14 der „Methodischen Grundlagen“).

In Bezug auf die Ebene, auf der Inhalte für die Qualitätsmessung abgeleitet werden, zeigt die Erfahrung des IQTIG, dass sich konkrete Hinweise auf Über-, Unter- und Fehlversorgung vorwiegend auf dem Auflösungs niveau von Qualitätsmerkmalen ergeben und weniger auf Ebene von Qualitätsaspekten oder Qualitätsdimensionen. Aus diesem Grund wurde die Schilderung in Kapitel 12, wie Qualitätsmerkmale und Qualitätsaspekte entwickelt werden, gegenüber der Vorversion angepasst.

In einer Stellungnahme wurden mögliche Kriterien genannt, anhand derer Themenbereiche für die Einführung von Qualitätsmessungen ausgewählt werden könnten. Dies könnten z. B. „kürzlich empfohlene Veränderungen der Versorgung“ sein, die noch nicht umfassend umgesetzt würden, oder das Vorliegen „konsentierter Gesundheitsversorgungsziele“ (DNVF, S. 5).

**IQTIG:** Kapitel 12 der „Methodischen Grundlagen“ bezieht sich auf die Auswahl von Inhalten *innerhalb* eines Themenbereichs der externen Qualitätssicherung. Die Identifikation, für *welche* neuen Themenbereiche eine Qualitätsmessung empfehlenswert ist, ist ebenfalls eine wichtige Aufgabe für die externe Qualitätssicherung. Das IQTIG geht derzeit im Rahmen eines G-BA-Auftrags (G-BA 2023a) der Frage nach, mittels welcher Kriterien Themenbereiche für die Qualitätssicherung priorisiert werden könnten, und bezieht die Hinweise aus der Stellungnahme gerne in seine Überlegungen ein.

Von zwei stellungnehmenden Organisationen wurde um Erläuterung gebeten, warum bei der Entwicklung von Qualitätsmodellen nicht mehr wie in der Vorversion der „Methodischen Grundlagen“ im ersten Schritt Qualitätsaspekte identifiziert und im zweiten Schritt daraus Qualitätsmerkmale abgeleitet würden, sondern nun im ersten Schritt Qualitätsmerkmale entwickelt und Qualitätsaspekte als Zusammenfassung der Qualitätsmerkmale aufgefasst würden (GKV-SV, S. 5 f.; PatV, S. 11).

**IQTIG:** Qualitätsmerkmale und Qualitätsaspekte können als unterschiedliche Detailebenen von Qualitätsthemen verstanden werden. Ihre Entwicklung verläuft nicht in linearer zeitlicher Abfolge, weswegen die Darstellung des Entwicklungsprozesses immer mit einer gewissen Vereinfachung verbunden ist. Hinweise aus den verschiedenen Wissensbeständen, welche Themen für die Qualitätssicherung möglicherweise von Interesse sind, können einen unterschiedlichen Detailgrad aufweisen. Dies wird bei der Analyse der Hinweise berücksichtigt, und es wird untersucht, welche Themen ähnliche Inhalte adressieren, welche sinnvoll zusammengefasst werden sollten und welche getrennt voneinander betrachtet werden sollten. Die praktischen Erfahrungen aus der Entwicklung von Indikatorensets haben gezeigt, dass sich Hinweise auf relevante Qualitätsthemen innerhalb eines Themenbereichs vorwiegend zunächst auf niedrigerer Auflösungsebene ergeben, d. h. auf Ebene von Qualitätsmerkmalen. Die Benennung von Qualitätsaspekten hat dementsprechend auch eine induktive Komponente, d. h., ein Qualitätsaspekt repräsentiert den gemeinsamen Nenner solcher Hinweise. Die Darstellung wurde daher angepasst, um dies treffender als bisher abzubilden. Für den Inhalt des Qualitätsmodells für einen Themenbereich ergeben sich daraus keine Änderungen.

In einer Stellungnahme wurde die Zuordnung von Qualitätsaspekten, Qualitätsmerkmalen und Qualitätsindikatoren im Vergleich der Abbildung 9 (Kapitel zur Entwicklung von Patientenbefragungen) mit der Abbildung 17 (Kapitel zur Auswahl von Inhalten zur Qualitätsmessung) angesprochen. Es wurde gefragt, warum in der Abbildung bei der Entwicklung von Patientenbefragungen ein Qualitätsindikator mehreren Qualitätsmerkmalen (statt nur einem Qualitätsmerkmal) zugeordnet sei und einem Qualitätsaspekt nur ein einzelner Qualitätsindikator (statt mehrere Qualitätsindikatoren) gegenüberstehe. Da Inhaltsvalidität erfordere, dass ein Qualitätsaspekt über mehrere Qualitätsindikatoren abgebildet werde, sei unklar, für welche Form der Datenerhebung das Ziel der Inhaltsvalidität und der Ausgewogenheit der Qualitätsindikatoren gelte (KBV, S. 5).

**IQTIG:** Das Ziel der Inhaltsvalidität gilt für alle Indikatorensets und ist unabhängig von den Datengrundlagen der Indikatoren. Inhaltsvalidität erfordert kein bestimmtes Zahlenverhältnis zwischen Qualitätsaspekten, Qualitätsmerkmalen und Qualitätsindikatoren. Sie wird vielmehr danach beurteilt, ob im Indikatorenset alle interessierenden Inhalte und nur diese adressiert werden und ob kein Ungleichgewicht zwischen den Indikatoren besteht (z. B. ein einseitiger Fokus auf Indikatoren einer Qualitätsdimension).

Bei der Entwicklung von Indikatoren auf Basis von Patientenbefragungen ist es häufig sinnvoll, mehrere verwandte Merkmale zu einem Qualitätsindikator zusammenzufassen, um eine zu große Zahl von Indikatoren zu vermeiden und um die Übersicht zu erhöhen. Dieses Vorgehen wird auch bei Indikatoren auf Basis QS-Dokumentation angewandt, beispielsweise bei der Zusammenfassung mehrerer Komplikationen in einem Komplikationsindikator. Wird eine hinreichende Zahl von Qualitätsmerkmalen bei der Operationalisierung zusammengefasst, erreicht der betreffende Qualitätsindikator denselben inhaltlichen Umfang wie ein Qualitätsaspekt. Um zu verdeutlichen, dass dies bei Indikatoren auf Basis von Patientenbefragungen vergleichsweise häufig vorkommt, wurde die Abbildung entsprechend gestaltet.

In einer Stellungnahme wurde die Frage aufgeworfen, warum für Qualitätsmerkmale, die über Fokusgruppen identifiziert würden, eine zusätzliche Prüfung der Bedeutung für die Patientinnen und Patienten vorgenommen werde. Dies sei bei „legitimierte[n] und repräsentativ zusammengesetzte[n] Fokusgruppen mit Patientinnen und Patienten nicht erforderlich“ (PatV, S. 11). In diesem Zusammenhang wurde um genauere Erläuterung gebeten, wie die Fokusgruppen zusammengesetzt würden, wie die Relevanz und Legitimation der Gruppenmitglieder sichergestellt werde und welche Vorgaben für die Diskussion genutzt würden (PatV, S. 13).

**IQTIG:** Bei der Methode der Fokusgruppe steht vor allem die Rekonstruktion der Patientensicht im Mittelpunkt und nicht die quantitative Erfassung von Aussagen. Über die Fokusgruppen soll eine Breite an relevanten Themen gewonnen werden. Die Methode der Fokusgruppe zielt darauf ab, Erkenntnisse von einer gezielt zusammengestellten, der

Forschungsfrage folgenden Auswahl an bestimmten Personen zu gewinnen und nicht auf Basis einer repräsentativen Stichprobe (Nyumba et al. 2018, Scheuren 2004). Die Zusammensetzung der Fokusgruppe richtet sich demnach nach „inhaltlicher Repräsentativität“ der Zielgruppe, wodurch die gewünschte Themenbreite gewonnen werden kann. Die Komposition der Zielgruppe grenzt sich folglich von dem ab, was im Allgemeinen unter einer „statistischen Repräsentativität“ verstanden wird. Bei Fokusgruppen handelt es sich um eine wissenschaftliche Methode zum Erkenntnisgewinn, formale Abstimmungen oder Legitimationen sind hier nicht berührt.

Bei der Konzeption, Durchführung und Auswertung von Fokusgruppen folgt das IQTIG etablierten wissenschaftlichen Standards. Zum Vorgehen bei der Entwicklung von Diskussionsleitfäden wird auf einschlägige Quellen, in denen das Standardvorgehen ausführlich beschrieben ist, verwiesen (Literatur in Abschnitt 7.2.2 der „Methodischen Grundlagen“). Die genaue Konzeption der jeweiligen Fokusgruppen (Zusammensetzung, Moderation, Themen des Moderationsleitfadens, methodisches Vorgehen der Auswertung, Zusammenspiel mit Ergebnissen der anderen Wissensbestände) und damit auch die Charakteristika der Fokusgruppen (Alter, Geschlecht, Bildung, Erkrankung etc.) werden in den betreffenden Abschlussberichten der einzelnen Entwicklungen regelhaft beschrieben.

Die Beurteilung des Eignungskriteriums „Bedeutung für die Patientinnen und Patienten“ erfolgt insbesondere auf Grundlage der Fokusgruppen, bezieht aber auch weitere Informationsquellen mit ein, die auf die Bedeutung eines Qualitätsmerkmals hinweisen können. Dies sind beispielsweise Publikationen zu Patientenbedarf und -präferenzen sowie Erfahrung der Expertengruppen einschließlich der darin beteiligten Patientenvertreterinnen und Patientenvertreter.

Von einer stellungnehmenden Organisation wurde gefragt, welche Bedeutung die Zusammenfassung von Qualitätsmerkmalen zu Qualitätsaspekten sowie die Zuordnung zu Qualitätsdimensionen im Hinblick auf die Eignungsprüfung der Qualitätsmerkmale hätten. Es wurde die Sorge geäußert, Qualitätsmerkmale, die „einen klaren Patientennutzen aufweisen“, könnten durch die Nichtzuordnung zu einem Qualitätsaspekt keinen Eingang in das Qualitätsmodell finden (PatV, S. 12).

**IQTIG:** Bei der Zusammenfassung der Qualitätsmerkmale zu Qualitätsaspekten handelt es sich nicht um einen Prüfschritt in Bezug auf die Qualitätsmerkmale. Die Zusammenfassung dient vielmehr der besseren Übersichtlichkeit.

In einer Stellungnahme wurde um Klarstellung gebeten, dass die Beschränkung der Auswahl möglicher Qualitätsmerkmale auf den Regelungsbereich des SGB V nicht für Beauftragungen des IQTIG durch das Bundesministerium für Gesundheit (BMG) gelte (PatV, S. 12 f.). In einer anderen Stellungnahme wurde empfohlen, Themen aus dem Bereich von SGB IX und SGB XI nicht generell aus der Betrachtung auszuschließen (DNVF, S. 5).

**IQTIG:** Das IQTIG ist gemäß § 137a SGB V auf Themen im Regelungsbereich des G-BA beschränkt. Auch Aufträge, die unmittelbar vom BMG an das IQTIG gegeben werden, beziehen sich gemäß Absatz 4 Satz 2 auf Aufgaben nach Absatz 3, die sonst Aufgaben des G-BA sind.

Von einer stellungnehmenden Organisation wurde angeregt, „eine qualitative Bewertung der Abbildung von Qualitätsaspekten durch die entwickelten Qualitätsindikatoren“ in den „Methodischen Grundlagen“ festzuschreiben. Die bloße Abbildung jedes Qualitätsaspekts durch mindestens einen Qualitätsindikator könne die Vollständigkeit und Güte der Abbildung des Qualitätsmodells nicht sicherstellen (BPtK, S. 9).

**IQTIG:** Es geht aus der Stellungnahme leider nicht ganz klar hervor, was mit der „qualitativen Bewertung der Abbildung“ gemeint ist. Möglicherweise ist die Prüfung der Inhaltsvalidität des Indikatorensets angesprochen. Diese ist regelhaft Bestandteil der Entwicklung von Indikatorensets und ist in Abschnitt 6.3 und Kapitel 14 erläutert.

Von einer stellungnehmenden Organisation wurde um Erläuterung gebeten, ob Qualitätsmerkmale „entweder identifiziert oder doch entwickelt“ würden und welche Schritte der Entwicklungsprozess von Qualitätsmerkmalen enthalte (PatV, S. 11 f.).

**IQTIG:** Für das IQTIG ist leider nicht ganz klar, welcher Unterschied in der Stellungnahme zwischen den Begriffen gesehen wird; das IQTIG versteht „Identifizierung“ und „Entwicklung“ von Qualitätsmerkmalen hier synonym. Das Vorgehen ist das folgende: Es werden Hinweise und Informationen gesammelt, u. a. aus Fokusgruppen, und dabei wird Fragen nachgegangen wie: Welche Probleme in der Versorgung werden berichtet? Wie schwerwiegend sind sie für die Patientinnen und Patienten? Treten bestimmte Probleme gehäuft auf? Kann die Versorgung durch Anstrengungen der Leistungserbringer verbessert werden? Die Eigenschaften der Versorgung, die anhand der Hinweise und Fragen als für die Qualitätssicherung von Interesse identifiziert werden, werden dokumentiert und als Qualitätsmerkmale bezeichnet. Aus Sicht des IQTIG steht dabei - unabhängig davon, ob man diesen Ablauf als Identifikation oder als Entwicklung bezeichnet - im Vordergrund, dass die für Patientinnen und Patienten relevanten Themen gefunden und im Qualitätsmodell zusammengefasst werden.

### 3 Eignungskriterien für Qualitätsmessungen

#### Allgemeine Anmerkungen

Von mehreren stellungnehmenden Organisationen wurde die Ergänzung einer Operationalisierung der Eignungskriterien im Anhang positiv hervorgehoben (BPtK, S. 10; GKV-SV, S. 9; KBV, S. 9; KZBV, S. 15). Dennoch sei für einige Eignungskriterien (Validität der Messung, Potenzial zur Verbesserung, Angemessenheit der Risikoadjustierung) noch nicht hinreichend transparent, anhand welcher Informationen und Überlegungen die Eignung beurteilt werde (BPtK, S. 10 f.).

**IQTIG:** Für die Beurteilung einiger Eignungskriterien lassen sich nicht alle zu berücksichtigenden Informationen sowie ein Algorithmus sinnvoll vorab festlegen. Dies betrifft insbesondere die Frage, welche Sachverhalte in einer Rechenregel und ggf. Risikoadjustierung auf welche Weise zu berücksichtigen sind. Die vom IQTIG genutzten, in Anhang A dargestellten Leitfragen sind daher als eben solche zu verstehen: als Leitfragen, die den Informationssammlungs- und Beurteilungsprozess für das jeweilige Eignungskriterium strukturieren. Transparenz über die Beurteilung der Eignungskriterien ergibt sich insbesondere dadurch, dass das IQTIG die zugrunde gelegten Informationen und Überlegungen schildert, die zu den Einschätzungen geführt haben.

Von mehreren stellungnehmenden Organisationen wurde die Methodik des IQTIG zur Eignungsprüfung als „komplex“ wahrgenommen (KZBV, S. 15; PatV, S. 14). In einer Stellungnahme wurde die Methodik als „überkomplex“ und „nicht mehr nachvollziehbar“ bewertet. Es wurde kritisiert, es würden „neue Leitfragen und Kennzahlen beschrieben, die bisher noch nie angewendet wurden“. Diese führten zu „viel subjektivem Entscheidungsspielraum“ bei der abschließenden Empfehlung über den Umgang mit einem Indikator und machten die Empfehlung weniger transparent. Es sei eine Überarbeitung „mit dem Ziel der Vereinfachung“ (GKV-SV, S. 3, 9 f.) bzw. eine „Verschlankung“ zur Aufwandsreduktion (KZBV, S. 15) sinnvoll.

**IQTIG:** Die Beurteilung der Eignung eines Qualitätsindikators ist unabhängig von der dazu eingesetzten Methodik eine komplexe Aufgabe. Denn um aussagekräftige Qualitätsindikatoren zu erhalten, müssen viele verschiedene Aspekte beachtet werden. Eine Vereinfachung der Herangehensweise an diese Aufgabe, etwa durch Einsatz eines Abstimmungsverfahrens ohne hinreichend detaillierte Leitfragen, ist zwar mit einem vergleichsweise weniger aufwändigen und augenscheinlich einfachen Prozess verbunden, mindert aber die Qualität der resultierenden Beurteilungen und deren inhaltlich-fachliche Nachvollziehbarkeit. Ein wichtiger Grund dafür ist, dass ohne Leitfragen die Beurteilungen in besonderem Maße von den impliziten Annahmen und Urteilen der Beteiligten abhängen, also eine hohe Subjektivität aufweisen. In der Vergangenheit sind aus

solchen einfachen Herangehensweisen stammende Empfehlungen dementsprechend gehäuft als nicht nachvollziehbar kritisiert worden.

Die Strukturierung der Eignungsprüfung durch Eignungskriterien sowie Leitfragen zu deren Beurteilung führt dagegen – anders als in der Stellungnahme vermutet – gerade zu einer Verringerung des subjektiven Entscheidungsspielraums und erhöht die Hürden für und die Transparenz über subjektiv gefärbte Empfehlungen. Außerdem werden die Anwender der Methodik intensiv in deren Anwendung geschult. Dies ist – wie bei allen Methoden – wichtig, um die korrekte und einheitliche Anwendung sicherzustellen.

Gleichzeitig wird durch die Strukturierung die Komplexität reduziert, und es wird transparent gemacht, welche Aspekte des komplexen Konstrukts „Eignung“ für die Empfehlung bezüglich des Indikators maßgeblich waren.

In zwei Stellungnahmen wurde der Eindruck geschildert, die Einschätzung der Leitfragen zu den Eignungskriterien beruhe auf subjektiven Eindrücken und berge die Gefahr willkürlicher oder interessengeleiteter Beurteilungen. Daher sollte festgelegt werden, wer im IQTIG an den Entscheidungen beteiligt ist, und ein Vier-Augen-Prinzip festgeschrieben werden (DNVF, S. 8). Das IQTIG solle beschreiben, wie und durch welche Personen die Entscheidungen zustande kämen (GKV-SV, S. 11).

**IQTIG:** Das IQTIG strebt eine möglichst hohe Transparenz seiner Methoden und Prozesse an. Zu diesem Zweck erfolgt eine umfassende Schilderung der eingesetzten Methoden und Kriterien für die Eignungsprüfung, nicht nur in den „Methodischen Grundlagen“, sondern auch in den Berichten und in umfangreichen zusätzlichen Anhängen. An einigen Stellen einer Eignungsprüfung sind dabei Einschätzungen und Abwägungen notwendig, die schwer bis kaum standardisierbar sind. Bei den Empfehlungen des IQTIG handelt es sich grundsätzlich nicht um Empfehlungen einzelner Personen, sondern alle Einschätzungen werden durch ein multiprofessionelles Team und je nach fachlicher Anforderung organisationseinheitenübergreifend entwickelt und diskutiert. Selbstverständlich hat das IQTIG nicht nur entsprechende interne Prozesse zur Qualitätssicherung seiner Empfehlungen implementiert, sondern auch Schulungen der multiprofessionellen Teams, die die Methodik umsetzen. Insbesondere diese internen Schulungen haben eine hohe Bedeutung für die einheitliche Anwendung der Methodik.

Eine noch weitergehende Festlegung und Detaillierung von Prüfprozessen und -kriterien an diesen Stellen erhöht die Komplexität von Beurteilung und Darstellung, ist aber hinsichtlich der Transparenz nicht zielführend. Denn für die Nachvollziehbarkeit fachlicher Empfehlungen ist Transparenz über ihre inhaltlich-fachlichen Begründungen wichtiger noch als Transparenz über den Prüfprozess. Um seine Empfehlungen so transparent und nachvollziehbar wie möglich darzustellen, schlüsselt das IQTIG daher bei Eignungsprüfungen von Qualitätsindikatoren die Ergebnisse und Begründungen für seine Einschätzungen sehr detailliert auf, einschließlich der Ergebnisse aller Kennzahlen und Leitfragen.



Das Ergebnis einer Eignungsprüfung ist auch keine „Entscheidung“ seitens des IQTIG, sondern eine fachliche Empfehlung an den G-BA, welche Indikatoren unter Aufwand-Nutzen-Gesichtspunkten weitergeführt werden sollten und welche nicht. Der G-BA hat auf Basis der detaillierten Informationen zu den Aufwänden und Nutzen der einzelnen Indikatoren die Möglichkeit, vom IQTIG abweichende Einschätzungen und Entscheidungen zu treffen.

Von einer stellungnehmenden Organisation wurde gefragt, ob sich die Eignungskriterien des IQTIG für Qualitätsmessungen ausschließlich auf quantitative Methoden beziehen und was unter dem im Abschnitt 12.3. (Qualitätsmodell) der „Methodischen Grundlagen“ erwähnten Peer-Review-Katalog zu verstehen sei (KBV, S. 6).

**IQTIG:** Die Eignungskriterien des Qualitätsmerkmals und die Eignungskriterien der Operationalisierung beziehen sich auf alle Qualitätsmessungen, unabhängig davon, ob die Messung mit standardisierten Methoden wie z. B. Qualitätsindikatoren oder mit weniger standardisierten (z. B. Peer-Review-basierten) Methoden durchgeführt wird. Die Eignungskriterien des Bewertungskonzepts sind in den „Methodischen Grundlagen“ mit Blick auf Qualitätsindikatoren beschrieben und beziehen sich daher auf den Referenzbereich und die statistische Klassifikationsgüte. Sie dienen letztendlich der Auswertungs- und Interpretationsobjektivität und gelten damit sinngemäß auch für nicht indikatorbasierte Messungen. Aufgrund der in Kapitel 4 dargestellten Limitationen wenig standardisierter Messungen empfiehlt das IQTIG jedoch in erster Linie den Einsatz indikatorbasierter Methoden zur Qualitätsmessung. Bei nicht indikatorbasierten Messmethoden kann Auswertungs- und Interpretationsobjektivität ggf. durch die Festlegung von Bewertungskriterien, wie auf Grundlage der Informationen ein Qualitätsurteil getroffen wird, verbessert werden. Mit dem Peer-Review-Kriterienkatalog ist ein Katalog solcher Bewertungskriterien gemeint. Eignungskriterien für qualitative Methoden (im eigentlichen Sinn, s. o.) sind nicht Gegenstand der „Methodischen Grundlagen“.

In einer Stellungnahme wurde um Erläuterung gebeten, welche Unterschiede sich in der Anwendung der Eignungskriterien auf bestehende und auf neu zu entwickelnde Indikatoren ergeben (GKV-SV, S. 9).

**IQTIG:** Bei neu zu entwickelnden Indikatoren dienen die Eignungskriterien als Leitlinie für die Auswahl von Qualitätsmerkmalen und deren Operationalisierung über Qualitätsindikatoren. Das bedeutet, die Indikatoren werden von vorneherein so konstruiert, dass sie die Eignungskriterien möglichst gut erfüllen. Beispielsweise werden Datenfelder entwickelt, die möglichst eindeutige Informationen erfassen und die für das Qualitätsmerkmal passende Information enthalten (Objektivität und Validität der Messung).

Bei bestehenden Indikatoren wird die Eignung demgegenüber rückblickend geprüft, auf Basis der Entwicklungsdokumentation (z. B. dazu, welche Evidenz für ein mittelbar patientenrelevantes Merkmal vorliegt) und von Daten aus der Anwendung. Beispielsweise wird geprüft, ob sich aus der Anwendung Hinweise ergeben haben, dass die Validität der Messung wegen nicht berücksichtigter Einflussfaktoren geringer ist als ursprünglich gedacht.

### **Bedeutung für die Patientinnen und Patienten**

Von einer stellungnehmenden Organisation wurde auf das Wirtschaftlichkeitsgebot im SGB V hingewiesen, und es wurde vermutet, die „medizinische Relevanz“ der Qualitätsanforderungen werde der Patientenzentrierung untergeordnet und nicht angemessen berücksichtigt. Wirtschaftlichkeit und Notwendigkeit sollten in einem eigenen Eignungskriterium abgebildet werden (DGfN, S. 6).

**IQTIG:** Bei der Entwicklung von Qualitätsmessungen berücksichtigt das IQTIG, welche Qualitätsmerkmale der Versorgung durch die Leistungserbringer zu verantworten sind (Eignungskriterium „Beeinflussbarkeit durch den Leistungserbringer“) und welcher Erfüllungsgrad eines Qualitätsmerkmals durch die Leistungserbringer erreichbar ist (Eignungskriterien „Potenzial zur Verbesserung“ und „Angemessenheit des Referenzbereichs“). Dadurch trägt das IQTIG auch bisher schon den Handlungsmöglichkeiten der Leistungserbringer unter den Rahmenbedingungen der gesetzlichen Krankenversicherung Rechnung.

In einer Stellungnahme wurde darauf hingewiesen, dass im Themenbereich der Geburtshilfe Patientinnen ohne behandlungsbedürftige Erkrankung versorgt würden. Daher müsse bei der Auswahl geeigneter Qualitätsmerkmale darauf geachtet werden, dass nicht nur Verbesserungen patientenrelevanter Endpunkte, sondern auch Verschlechterungen patientenrelevanter Endpunkte bei zuvor Gesunden, z. B. aufgrund unnötiger Eingriffe, in die Qualitätssicherung einbezogen werden (DGHWi, S. 4).

**IQTIG:** Die Vermeidung unerwünschter Folgen einer Behandlung sowie die Vermeidung nicht notwendiger Behandlungen sind wichtige Aspekte der Versorgungsqualität und sind im Rahmenkonzept des IQTIG den Qualitätsdimensionen „Patientensicherheit“ und „Angemessenheit“ zugeordnet (siehe Kapitel 1). Bei der Entwicklung der Inhalte, die in einem Themenbereich durch die Qualitätssicherung adressiert werden sollen, strebt das IQTIG an, alle Qualitätsdimensionen angemessen zu berücksichtigen.

### **Potenzial zur Verbesserung**

Von zwei stellungnehmenden Organisationen wurde auf die Bedeutung empirischer Daten (BPtK, S. 9) sowie insbesondere internationaler Leitlinien (DGHWi, S. 4) hingewiesen, wenn bei der Auswahl neuer Qualitätsaspekte und -merkmale für die Qualitätssicherung das Potenzial zur Verbesserung beurteilt werden soll.

**IQTIG:** Das IQTIG teilt diese Auffassung und berücksichtigt daher insbesondere wissenschaftliche Literatur und Leitlinien, sofern diese für das entsprechende Thema verfügbar sind.

In einer Stellungnahme wurde um weitere Erläuterungen gebeten, welche Vorteile die statistische Methode des IQTIG zur Einschätzung der Anzahl verbesserbarer Ereignisse gegenüber der ABC-Methode (Weissman et al. 1999) hat, warum die Leistungserbringer in Gruppen mit höherer und geringerer Kompetenz aufgeteilt würden und inwieweit die Methode des IQTIG präziser sei (DKG, S. 10).

**IQTIG:** Eine Gruppenbildung basierend auf einer Rangreihenbildung anhand von geschätzter Kompetenz erfolgt bei beiden Methoden. Anders als die ABC-Methode berücksichtigt die Methode des IQTIG bei dieser Bildung einer Rangreihenfolge zusätzlich patientenbezogene Risikofaktoren. Bei der Methode des IQTIG wird statistische Unsicherheit nicht nur bei der Gruppenbildung berücksichtigt, sondern auch bei der darauf aufbauenden Schätzung der Anzahl an verbesserbaren Ereignissen. Beispielsweise wird durch dieses Vorgehen vermieden, dass aufgrund zufallsbedingt „guter“ Indikatorwerte bei Leistungserbringern mit kleiner Fallzahl die Anzahl verbesserbarer Ereignisse und der erreichbare Indikatorwert unangemessen hoch berechnet werden und in der Folge das Potenzial zur Verbesserung überschätzt wird.

Von einer stellungnehmenden Organisation wurde der Begriff „Kompetenzparameter“ in der Darstellung der statistischen Methode des IQTIG zur Einschätzung verbesserbarer Ereignisse kritisiert, da er suggeriere, es handle sich um die Fähigkeit bzw. den Sachverstand eines Leistungserbringers. Es wurde um Bezug zum statistischen Rahmenkonzept in Kapitel 20 der „Methodischen Grundlagen“ gebeten (KBV, S. 8).

**IQTIG:** Der Begriff „Kompetenzparameter“ wird konsistent zur Erläuterung in Abschnitt 20.1.1 der „Methodischen Grundlagen“ verwendet. Er bezieht sich - wie alle Qualitätsmessungen - nicht auf *sämtliche* Strukturen und Handlungen eines Leistungserbringers, sondern nur auf diejenigen, die für das durch einen Indikator abgebildete Qualitätsmerkmal relevant sind. In diesem Kontext bezeichnet das IQTIG damit den einem Indikator zugrunde liegenden (latenten) Parameter. Er hängt u. a. von der Operationalisierung des Indikators, z. B. über Rechenregeln, ab.

In einer Stellungnahme wurde als wichtig erachtet, bei der Beurteilung des möglichen Nutzens von Qualitätsindikatoren auch abzuschätzen, „welches Ausmaß an Qualitätsverbesserungen realistisch erwartet werden“ könne und wieviel davon „von den Leistungserbringern tatsächlich beeinflussbar“ sei (KZBV, S. 12).

**IQTIG:** Die Schätzung des Potenzials zur Verbesserung anhand des gleichnamigen Eignungskriteriums berücksichtigt das Ausmaß, in dem die Ergebnisse durch Anstrengun-

gen der Leistungserbringer realistisch beeinflusst werden können: Die Anzahl verbesserbarer Ereignisse und der erreichbare Indikatorwert werden vom IQTIG ausgehend von der Annahme geschätzt, dass alle Leistungserbringer das erreichen könnten, was die „obere“ Hälfte der Leistungserbringer (d. h., die Leistungserbringer mit höherer Qualität in Bezug auf das Qualitätsmerkmal) im Median bereits tatsächlich erreicht hat (siehe Anhang B der „Methodischen Grundlagen“).

Von einer stellungnehmenden Organisation wurden inkonsistente Beschreibungen der Einbindung von Expertinnen und Experten bei der Beurteilung zweier Eignungskriterien bemängelt. Für das „Potenzial zur Verbesserung“ werde einerseits eine Einschätzung nach RAND/UCLA-Methode in Kapitel 6 beschrieben und andererseits in Anhang A eine Einschätzung auf Basis von Informationen, die nur bei bereits eingesetzten Indikatoren zur Verfügung stünden. Für die „Beeinflussbarkeit durch die Leistungserbringer“ sei widersprüchlich, dass in Kapitel 6 eine Einschätzung nach RAND/UCLA-Methode und in Anhang A ein anderes Abstimmungsverfahren bezüglich der Zuschreibbarkeit geschildert sei (GKV-SV, S. 11 f.).

**IQTIG:** Für die Einschätzung des „Potenzials zur Verbesserung“ unterscheiden sich die Vorgehensweisen zwischen der Entwicklung neuer Qualitätsindikatoren (Kapitel 6) und der Beurteilung bereits eingesetzter Qualitätsindikatoren (Anhang B) aufgrund der unterschiedlichen Datengrundlagen, die zur Verfügung stehen. Dies wurde im Text nun deutlicher dargestellt. Betreffend die „Beeinflussbarkeit durch die Leistungserbringer“ wurde die widersprüchliche Passage in Anhang B gestrichen. Hier zeigten die praktischen Erfahrungen aus den Eignungsprüfungen, dass eine Einschätzung der maßgeblichen Zuschreibbarkeit eines Leistungserbringers gemeinsam mit den Expertinnen und Experten auch ohne Abstimmungsverfahren zuverlässig möglich ist und das Vorgehen zu einer besseren Fokussierung auf fachlich-inhaltliche Gründe führt. Das IQTIG wird bei künftigen Beauftragungen zur Neuentwicklung von QS-Verfahren prüfen, ob dieses Vorgehen auch bei der Entwicklung neuer Qualitätsindikatoren eingesetzt werden sollte.

In einer Stellungnahme wurde die Auffassung geäußert, dass die Methodik zur Bewertung des Potenzials zur Verbesserung nur im Falle „extremer Deckeneffekte“ zur Einschätzung eines geringen Potenzials zur Verbesserung führe. Demgegenüber würden „tatsächlich festgestellte“ Qualitätsdefizite nur unzureichend berücksichtigt. Für diese Auffassung werden mehrere Argumente angeführt. Erstens beachteten die Kennzahlen 1 und 2 zur Schätzung des Potenzials zur Verbesserung nicht die „Unschärfen der Modellierung“ des Qualitätsindikators, wie es das IQTIG sonst selbst bei der Festlegung von Referenzbereichen tue. Dies führe zu einer Überschätzung der Anzahl verbesserbarer Ereignisse. Zweitens stellten die Kennzahlen 1 und 2 nur „theoretische[s]“ Verbesserungspotential“ dar, während Kennzahl 3 „tatsächlich realisierte[s]“ Verbesserungspotential“ darstelle und regelmäßig geringer ausfalle als die Einschätzung über die anderen beiden Kennzahlen. Drittens sei die Betrachtung von Zeitverläufen der Indikator-

ergebnisse besser, um Deckeneffekte darzustellen. Viertens führe die Beurteilung von Kennzahl 1 anhand absoluter Zahlen zu einer systematischen Verzerrung und zur Unterschätzung des Verbesserungspotenzials von Indikatoren mit kleinen Fall- und Ereigniszahlen und bei Indikatoren mit großen Fallzahlen zu einer Überschätzung. Als Beispiel für diesen Punkt wurde der Indikator „Frühzeitige Blutgasanalyse oder Pulsoxymetrie“ im QS-Verfahren *Ambulant erworbene Pneumonie* angeführt (GKV-SV, S. 14-16).

**IQTIG:** Sowohl bei der Schätzung des Potenzials zur Verbesserung als auch bei seinen Empfehlungen zu Referenzbereichen berücksichtigt das IQTIG „Unschärfen der Modellierung“, d. h. Einflussfaktoren auf die Indikatorergebnisse, die nicht von den Leistungserbringern zu verantworten sind und nicht durch die Rechenregel aus den Indikatorergebnissen „herausgerechnet“ werden (können). Diese Einflussfaktoren sind in den empirischen Ergebnissen, die beim Einsatz eines Indikators im Regelbetrieb erhoben werden, enthalten. Sie sind gerade der Grund dafür, dass der gemäß Anhang B geschätzte erreichbare Indikatorwert nicht dem Ende der Indikatorkala entspricht, sondern davon abweicht. Da der erreichbare Indikatorwert auch den Ausgangspunkt für die Festlegung eines kriteriumsbezogenen Referenzbereichs bildet, berücksichtigt der Referenzbereich ebenfalls diese Einflussfaktoren. Anders als für die Einschätzung des Potenzials zur Verbesserung ist für die Festlegung eines kriteriumsbezogenen Referenzbereichs für die leistungserbringerbezogene Beurteilung eine *zusätzliche* tolerierte Abweichung sinnvoll. Diese bezieht sich *nicht* auf unberücksichtigte Einflussfaktoren/Unschärfen, sondern auf Qualitätsdefizite im Sinne der Abweichung vom Erreichbaren (siehe Abschnitt 15.3.2 der „Methodischen Grundlagen“). Verbesserungsmöglichkeiten finden sich somit auch bei Leistungserbringern innerhalb des Referenzbereichs, sodass die Orientierung an der Referenzbereichsgrenze das Potenzial zur Verbesserung unterschätzen würde.

Unklar bleibt, warum in der Stellungnahme die Kennzahlen 1 (Anzahl verbesserbarer Ereignisse) und 2 (Anteil der Leistungserbringer mit Abweichung vom erreichbaren Indikatorwert) als „theoretisches“ und Kennzahl 3 (Anteil der Leistungserbringer mit „qualitativ auffälliger“ Einstufung im Stellungnahmeverfahren nach § 17 Teil 1 DeQS-RL) als „tatsächliches“ Verbesserungspotenzial benannt werden. Gerade die Ableitung aus den *tatsächlich* erreichten Ergebnissen der Leistungserbringer spricht für die praktische Konsequenz der Kennzahlen 1 und 2. Im Vergleich zu „in der Literatur zu findenden Bezugsgrößen - d. h. erwartbaren Standards, Raten usw.“ (GKV-SV, S. 14) hat dieses Vorgehen darüber hinaus den Vorteil einer meist umfangreicheren empirischen Datengrundlage. Außerdem stellt sich die Frage nach der Übertragbarkeit von Studienergebnissen nicht. Demgegenüber ist Kennzahl 3 in hohem Maße vom Vorgehen im Stellungnahmeverfahren abhängig, wie auch in der Stellungnahme (GKV-SV, S. 15) angemerkt. Aufgrund der intransparenten Bewertungskriterien und der heterogenen Vorgehensweise können die Ergebnisse des Stellungnahmeverfahrens nach § 17 Teil 1 DeQS-RL nicht als Goldstandard dienen.

Ein weiterer Grund für die geringe Häufigkeit, mit der Verbesserungspotenziale im Stellungnahmeverfahren identifiziert werden, liegt darin, dass als zentrale Analysemethode häufig eine Ursachenanalyse einzelner Behandlungsfälle anhand von den Leistungserbringern selbst zusammengestellter Informationen vorgenommen wird. Wird dabei keine eindeutige Ursache für ein unerwünschtes Ereignis gefunden, wird dies zuweilen nicht als Qualitätsdefizit interpretiert. Dies stellt ein unangemessen konservatives Vorgehen dar, da ein Qualitätsdefizit auch vorliegen kann, wenn dessen Ursache nicht gefunden wurde – insbesondere auf Basis limitierter Informationen. Zudem wird im Stellungnahmeverfahren auch anderen Qualitätsmerkmalen als dem im Qualitätsindikator operationalisierten Qualitätsmerkmal nachgegangen, sodass aus Sicht des IQTIG eine komplementäre Berücksichtigung der jeweiligen Verbesserungspotenziale sachgerecht ist.

Unter „Deckeneffekt“ eines Messverfahrens wird in der Literatur die Beobachtung verstanden, dass Messergebnisse nahe am Ende der Skala liegen und daher nur eine geringe Varianz aufweisen können (Schermelleh-Engel und Werner 2012: 138, Döring und Bortz 2016: 738). Das in der Stellungnahme genannte Ziel, „Deckeneffekte darstellen zu können“, versteht das IQTIG so, dass identifiziert werden soll, wenn viele Leistungserbringer eine hohe Qualität erreichen und daher keine relevanten Qualitätssteigerungen für den betreffenden Indikator mehr zu erwarten sind. Dieses Ziel wird durch die drei Kennzahlen zum Potenzial zur Verbesserung bei der Beurteilung berücksichtigt und führt dort zu niedrigen Werten der jeweiligen Kennzahl. Dagegen trifft der Zeitverlauf von Indikatorergebnissen eine Aussage über den historischen Verlauf der Qualitätsentwicklung, d. h., wie viel Verbesserung bisher erreicht wurde. Er erlaubt keine Prognose, wie viel in Bezug auf das Qualitätsziel noch erreicht werden kann. Eine Extrapolation des historischen Verlaufs auf die zukünftige Entwicklung ist insbesondere deswegen nicht sinnvoll, weil jede Extrapolation von einer Konstanz der jeweiligen Rahmenbedingungen ausgeht, aber im Kontext der Qualitätssicherung eine wesentliche Rahmenbedingung, nämlich die Versorgungsqualität der Leistungserbringer, gerade nicht als konstant angenommen wird. So könnten die Indikatorergebnisse beispielsweise deswegen über mehrere Jahre stabil sein (Plateaubildung), weil eine eigentlich erreichbare Qualitätsverbesserung ausbleibt oder stagniert (persistierendes Qualitätsdefizit). Eine Orientierung am Zeitverlauf würde dann von der fälschlichen Annahme ausgehen, dass die Versorgungsqualität konstant bleibt bzw. bleiben muss. Das zum Prüfzeitpunkt gegebene Potenzial zur Verbesserung der Versorgung beurteilt das IQTIG daher anhand des aktuellen Stands der Qualitätsergebnisse.

Die Verwendung absoluter Zahlen ergibt sich aus der notwendigen Systemperspektive. Die vom G-BA angestrebte Effizienzsteigerung (G-BA 2022a, G-BA 2022b, G-BA 2023b) und die dazu erforderliche Aufwand-Nutzen-Betrachtung von Indikatoren und Indikatorensets entspricht einer Priorisierung der bestehenden Indikatoren. Sollen die für die externe Qualitätssicherung bundesweit eingesetzten Ressourcen einen möglichst hohen Patientennutzen erreichen, erfordert dies eine Betrachtung mit populationsbezogenen

Maßen (Heller und Dobson 2000, Gigerenzer und Edwards 2003). Populationsbezogene Maße ermöglichen eine Priorisierung anhand der Frage, mit welchen Indikatoren sich bei einem bestimmten Ressourceneinsatz möglichst viele Ereignisse für die Patientinnen und Patienten verbessern lassen. Bei der Aufwand-Nutzen-Betrachtung verwendet das IQTIG dementsprechend sowohl für die Beschreibung des Nutzens als auch des Aufwands absolute Zahlen. Indikatoren mit gleichem Aufwand-Nutzen-Verhältnis werden bei diesem Vorgehen verfahrensübergreifend gleich bewertet. Dies ermöglicht auch aus Aufwand-Nutzen-Perspektive begründete Aussetzungsempfehlungen für ganze Indikatorensets. Eine solche Aussetzungsempfehlung erfolgt dann, wenn alle Indikatoren des Sets ein ungünstiges Aufwand-Nutzen-Verhältnis aufweisen. Anderenfalls wäre nur eine Priorisierung von Indikatoren *innerhalb* eines Sets möglich. Aussetzungsempfehlungen für ganze Sets wären aus Aufwand-Nutzen-Überlegungen heraus praktisch nicht möglich.

Sofern die Bedeutung eines Ereignisses für die Patientinnen und Patienten bei verschiedenen Indikatoren als gleich eingeschätzt wird, ist eine Bevorzugung von Indikatoren mit höherer Zahl verbesserbarer Ereignisse daher sachgerecht und mit Blick auf die Aufwand-Nutzen-Optimierung explizit angestrebt. Denn auch bei in Relation zur betrachteten Population geringem Anteil verbesserbarer Ereignisse (im genannten Beispiel-Indikator z. B. 0,5 %) kann aufgrund der Größe der Population trotzdem eine große Zahl (im Beispiel: 1.480) an Patientinnen und Patienten von einer Verbesserung der Versorgungsqualität profitieren. Unklar bleibt, auf Grundlage welcher Kriterien die Quantifizierung der verbesserbaren Ereignisse in der Stellungnahme als „Überschätzung“ beurteilt wird.

Die Bedeutung eines Indikators wird vom IQTIG bei der Abwägung bereits berücksichtigt und über das Eignungskriterium „Bedeutung für die Patientinnen und Patienten“ adressiert. Bei Indikatoren mit hoher Bedeutung wird auch ein vergleichsweise schlechteres Aufwand-Nutzen-Verhältnis akzeptiert. Für Indikatoren in Leistungsbereichen mit geringen Fallzahlen und geringen Leistungserbringerzahlen kommt zudem zum Tragen, dass über die Kennzahlen 2 und 3 eine weitere Perspektive auf das Potenzial zur Verbesserung berücksichtigt wird, nämlich der Anteil der Leistungserbringer mit Verbesserungspotenzial bzw. Qualitätsdefizit. Für diese Indikatoren kann daher ebenfalls ein mehr als „geringes“ Potenzial zur Verbesserung festgestellt werden. Ein Beibehalten von Indikatoren, die weder eine ausreichende Zahl verbesserbarer Ereignisse noch ausreichend viele Leistungserbringer mit Verbesserungspotenzial und auch nicht ausreichend viele Leistungserbringer mit Qualitätsdefizit aufzeigen, wird vom IQTIG unter der Zielsetzung der Optimierung des Patientennutzens fachlich nicht empfohlen.

Eine systematische Verzerrung der Eignungsprüfung zugunsten von Prozessindikatoren entsteht durch die Verwendung absoluter Zahlen nicht. Eine ggf. zwischen zwei Indikatoren unterschiedliche Grundhäufigkeit der interessierenden Ereignisse ist gleichermaßen auch bei relativen Angaben, beispielsweise bei Bezug der verbesserbaren

Ereignisse auf die Zahl der Fälle, gegeben: Eine höhere Grundhäufigkeit der interessierenden Ereignisse, z. B. 100 statt 50 Ereignisse je 1.000 Indikatorfälle, schlägt sich (unter Annahme inhaltlich gleichen Verbesserungspotenzials, z. B. von 1/5 der Ereignisse) sowohl auf die absolute Häufigkeit verbesserbarer Ereignisse nieder, z. B. 20 statt 10 verbesserbare Ereignisse, als auch auf die relative Häufigkeit in Bezug auf die Indikatorfälle (2 % vs. 1 % verbesserbare Ereignisse). Ob eine Schwelle von 10 Ereignissen (absolut) oder 1 % der Indikatorfälle (relativ) verwendet wird, ändert die Bewertung nicht.

Das IQTIG versteht die Formulierung in der Stellungnahme so, dass Prozessindikatoren bei der Beurteilung des Nutzens ein geringeres Gewicht beigemessen werden sollte als Ergebnisindikatoren. Diese Auffassung teilt das IQTIG nicht. Das IQTIG sieht Prozess- und Ergebnisindikatoren als gleichermaßen wichtig an, sofern das Eignungskriterium „Bedeutung für die Patientinnen und Patienten“ gleich bewertet wurde: Bei Prozessindikatoren beziehen sich zwar die verbesserbaren Ereignisse (Kennzahl 1 des Potenzials zur Verbesserung) auf Prozessmerkmale und meist ist nur ein Anteil der Fälle mit verbessertem Prozessmerkmal auch mit einem verbesserten unmittelbar patientenrelevanten Ergebnis verbunden. Demgegenüber sind die verbesserbaren Ereignisse bei Ergebnisindikatoren in der Regel unmittelbar patientenrelevant. Allerdings lassen sich die verbesserbaren Ereignisse bei Ergebnisindikatoren auch nur in geringerem Maße in tatsächlich verbesserte Ergebnisse umsetzen. Das IQTIG geht daher davon aus, dass Vergleiche der Menge an möglichen Verbesserungen auch zwischen Prozess- und Ergebnisindikatoren angemessen sind, wenn sie anhand der Kennzahlen für das Potenzial zur Verbesserung vorgenommen werden.

### **Beeinflussbarkeit**

Von einer stellungnehmenden Organisation wurden Vorschläge in den „Methodischen Grundlagen“ gewünscht, wie die externe Qualitätssicherung über die Betrachtung einzelner Leistungserbringer hinaus eingesetzt werden könne. Die Beschränkung auf Qualitätsbewertungen, die einzelnen Leistungserbringern zuschreibbar sind, sollte entfallen und es sollten beispielsweise auch Versorgungsnetzwerke und Kooperationen in den Blick genommen werden (PatV, S. 15).

**IQTIG:** Das IQTIG empfiehlt ebenfalls die Entwicklung und den Einsatz leistungserbringerübergreifender Qualitätsmessungen. Insbesondere empfiehlt das IQTIG, den Einsatz von Netzwerkanalysen in der externen Qualitätssicherung zu prüfen, um die Qualität auch bei Themen abzubilden, bei denen regelhaft mehrere Leistungserbringer gemeinsam beteiligt sind. Die bisherige Beschränkung der externen Qualitätssicherung auf Qualitätsmerkmale, deren Ergebnisse primär einem einzelnen Leistungserbringern zuschreibbar sind, ist zu eng und wird der Komplexität der Gesundheitsversorgung nicht gerecht. Anders als in Entwicklungsberichten und Positionspapieren werden in den „Methodischen Grundlagen“ allerdings nur regelhaft vom IQTIG angewendete Methoden dargestellt.



## Brauchbarkeit

Von zwei stellungnehmenden Organisationen wurde der Schilderung in den „Methodischen Grundlagen“ zugestimmt, dass die Beurteilung der Brauchbarkeit von Qualitätsmessungen für anschließende Handlungen und Maßnahmen auch in der Zusammenschau eines Indikatorensets erfolgen sollte (DKG, S. 8; PatV, S. 16). In einer anderen Stellungnahme wurden ergänzende Schilderungen dazu gewünscht (GKV-SV, S. 9).

**IQTIG:** Die Brauchbarkeit eines Indikatorensets für einen Handlungsanschluss beschreibt, ob die Qualitätsmerkmale der Indikatoren so ausgewählt wurden, dass sie die relevanten Inhalte für den vorgesehenen Handlungsanschluss umfassen. Diese Fragestellung entspricht der Inhaltsvalidität des Sets. Weitere Ausführungen dazu finden sich in Kapitel 14 der „Methodischen Grundlagen“.

In einer Stellungnahme wurde angemerkt, das Eignungskriterium „Brauchbarkeit für den Handlungsanschluss“ sei nicht auf alle Versorgungsbereiche anwendbar. In der ambulanten Psychotherapie sei die frühe Verfügbarkeit der Messergebnisse nicht für die Brauchbarkeit maßgeblich. In diesem Bereich seien aus über mehrere Behandlungsfälle aggregierten Ergebnissen keine Erkenntnisse zu Qualitätsverbesserungen ableitbar. Daher sei das Eignungskriterium nicht anwendbar (BPtK, S. 12).

**IQTIG:** Die zeitnahe Verfügbarkeit der Messergebnisse eines Indikators stellt nur einen Aspekt der Brauchbarkeit für den Handlungsanschluss dar (Leitfrage 3) und ist nicht alleine hinreichend für den Nachweis der Brauchbarkeit. Wie in der Stellungnahme zu Recht hervorgehoben, ist zusätzlich auch die Frage relevant, ob sich aus den Messergebnissen Verbesserungsansätze im internen Qualitätsmanagement der Leistungserbringer ableiten lassen. Diese Frage ist daher ebenfalls Teil der Beurteilung des Eignungskriteriums „Brauchbarkeit für den Handlungsanschluss“ (Leitfrage 1). Aus Sicht des IQTIG unterstreicht die Stellungnahme daher, dass die Prüfung des Eignungskriteriums für alle Versorgungsbereiche wichtig ist. Der Auffassung, im Versorgungsbereich der ambulanten Psychotherapie könnten aus der fallübergreifenden Betrachtung von Qualitätsergebnissen keine Erkenntnisse gewonnen werden, die der Qualitätsverbesserung dienen, schließt sich das IQTIG nicht an.

## Objektivität, Reliabilität und Validität der Messung

Von einer stellungnehmenden Organisation wurde als unklar angesehen, wie die Eignungskriterien „Objektivität der Messung“, „Reliabilität der Messung“ sowie „Validität der Messung“ voneinander abgegrenzt würden (GKV-SV, S. 12 f.). Bei Objektivität und Reliabilität beurteile das IQTIG denselben Sachverhalt, nämlich die Klarheit der Definitionen von Datenfeldern, und dieser Sachverhalt werde teilweise auch durch die Validität der Messung adressiert. Die Beurteilerübereinstimmung (Kennzahl 1 zur Beurteilung der Objektivität) sei vom IQTIG bei seinen bisherigen Prüfungen nicht angewendet worden und daher nicht nützlich (GKV-SV, S. 12).

**IQTIG:** Das IQTIG betrachtet anhand der Kriterien Objektivität, Reliabilität und Validität unterschiedliche Aspekte der Operationalisierung eines Qualitätsmerkmals (siehe Kapitel 13 der „Methodische Grundlagen“) und folgt dabei dem in der Wissenschaft etablierten Verständnis dieser Kriterien (Literatur ebd.). Unter der Objektivität einer Messung versteht das IQTIG das Ausmaß, in dem der für einen Sachverhalt erhobene Wert unabhängig von der beobachtenden Person und den Begleitumständen der Messung ist. Bei der Eignungsprüfung von Qualitätsindikatoren auf Basis von QS-Dokumentationsdaten betrachtet das IQTIG dazu zwei Schritte der Messung: die Erhebung des medizinischen Sachverhalts im Rahmen der Versorgung (Befunderhebung) und die Erfassung der erhobenen Sachverhalte für Zwecke der Qualitätssicherung (QS-Dokumentation). Im ersten Schritt können unterschiedliche Beurteilungen des klinischen Sachverhalts resultieren, wenn diese Beurteilung im Wesentlichen auf (unterschiedlichen) Einschätzungen der Beurteilenden beruht (Varianzquelle: Beobachtende) oder wenn Unterschiede in der Durchführung der Befunderhebung bestehen (Varianzquelle: Umsetzung der Messung, z. B. technische Unterschiede der Messgeräte). Im zweiten Schritt kann es zu Unterschieden in den dokumentierten Werten kommen, wenn das Datenfeld von verschiedenen Ausfüllenden unterschiedlich verstanden wird (Varianzquelle: Beobachtende), z. B. bei ungenauer oder missverständlicher Formulierung.

Bei der Reliabilität der Messung untersucht das IQTIG das Ausmaß, zu dem die initiale Befunderhebung frei ist von Zufallsrauschen (= Fehlervarianz). Sie umfasst Messungengenauigkeiten, die auch unter gleichen Untersuchungsbedingungen (dieselbe untersuchende Person, dieselbe technische Umsetzung) zu unterschiedlichen Einschätzungen führen, wenn die Messung bei einer Patientin oder einem Patienten wiederholt würde. Dies kann der (wechselnden) Einschätzung der untersuchenden Person geschuldet sein (Varianzquelle: Beobachtende) oder einer eingeschränkten Präzision der verwendeten Messinstrumente (Varianzquelle: Zufallseffekte).

Unter dem Begriff „Validität“ untersucht das IQTIG, ob die Qualitätsanforderung inhaltlich richtig, z. B. für die richtigen Patientenfälle, operationalisiert wird. In Bezug auf ein bestimmtes Datenfeld wird mit Blick auf die Validität also nicht beurteilt, ob das Datenfeld klar definiert ist, sondern ob es angemessen ist, den vom Datenfeld erfassten Sachverhalt in die Berechnung des Indikators ein- oder auszuschließen.

Unter bestimmten Konstellationen sind gleiche Beurteilungen für Objektivität und Reliabilität der Befunderhebung zu erwarten. Dies ist der Fall, wenn die Beurteilung des Sachverhalts von menschlicher Einschätzung abhängt. Hier wirkt sich die Varianzquelle „Beobachtende“ sowohl auf Einschätzungen durch verschiedenen Personen (Objektivität, beschrieben z. B. mit der sog. Inter-Rater-Reliabilität als Maß für die Beurteilerübereinstimmung) als auch auf Einschätzungen durch dieselbe Person (z. B. mit Maßzahlen wie der Test-Retest-Reliabilität) aus. Falls Hinweise zur Beurteilerübereinstimmung einer Messung vorliegen, sollen diese Informationen berücksichtigt werden. Aus Sicht des IQTIG ist es daher wichtig, für diesen Fall eine Kennzahl für die Beurteilung zu benennen.

Die Zuordnung zum jeweiligen Eignungskriterium hilft, den passenden Ansatzpunkt für entsprechende Verbesserungen zu identifizieren: Beispielsweise würde die Einschätzung, dass die echokardiographische Bestimmung der Pumpfunktion des Herzens auch durch dieselbe Person nur wenig präzise erfolgt (geringe Test-Retest-Reliabilität), nahelegen, die Pumpfunktion für die QS-Dokumentation anders zu erfassen, etwa aus einem radiologischen Befund. Dagegen würde Einschätzung, dass der Zeitpunkt in der Behandlung, zu dem die Pumpfunktion für die QS-Dokumentation erfasst werden soll, von verschiedenen Beurteilern unterschiedlich verstanden wird (geringe Objektivität), nahelegen, das Datenfeld klarer zu formulieren.

Für die Interpretation der Bewertungsergebnisse der Eignungskriterien ist darüber hinaus zu beachten, dass Objektivität und Reliabilität der Messung für einen Indikator jeweils datenfeldübergreifend betrachtet werden. Daher kann sich die Bewertung eines einzelnen Datenfelds bei verschiedenen Indikatoren unterschiedlich auf die Gesamtbewertung des jeweiligen Eignungskriteriums auswirken.

Von einer stellungnehmenden Organisation wurde der Eindruck geschildert, die Bezeichnung des Eignungskriteriums „Objektivität des Messverfahrens“ suggeriere eine detailliertere Prüfung, als tatsächlich vorgenommen werde. Die Bezeichnung lege nahe, es würden auch die Objektivität „der Festlegung der zu messenden Konstrukte“ und die Objektivität der Auswertung beurteilt. Daher solle das Kriterium in „Klarheit der Formulierungen der Datenfelder“ umbenannt werden (GKV-SV, S. 13).

**IQTIG:** Die Klarheit der Formulierung der Datenfelder zu erhöhen ist eine wichtige Maßnahme, um die Objektivität der Indikatoren zu erhöhen. Da die Auswertung der dokumentierten Daten durch die Indikatorrechenregel standardisiert ist und damit Auswertungsobjektivität bereits gegeben ist, kann nach Prüfung der Objektivität der Datenerhebung eine Aussage über die Objektivität der Messung insgesamt getroffen werden. Es bleibt leider unklar, was in der Stellungnahme unter „Objektivität [...] der Festlegung der zu messenden Konstrukte“ verstanden wird. Ob die zu messenden Konstrukte, d. h. die Qualitätsmerkmale, angemessen festgelegt wurden, beschreibt das IQTIG anhand der Eignungskriterien des Qualitätsmerkmals.

### Datenqualität

Von zwei stellungnehmenden Organisationen wurde die Schilderung, dass selbstberichtete Daten mit dem Risiko einer Verzerrung durch gezielte Dokumentation einhergehen, als „Generalverdacht“ (KZBV, S. 7) und „Abwertung“ (DGfN, S. 4) wahrgenommen. Demgegenüber wurden in anderen Stellungnahmen auf „Anreize(...)“, die die Datenqualität beeinflussen können“ (DNVF, S. 3), und mögliche Probleme der Datenqualität durch „nachlässige oder manipulative Dokumentation“ (GKV-SV, S. 10) hingewiesen. Diese Problematik solle ausführlicher geschildert werden (DNVF, S. 3).

**IQTIG:** Die entsprechende Passage in Abschnitt 4.3 dient als Übersicht über die verschiedenen Datenquellen einschließlich einer kurzen Schilderung ihrer Vor- und Nachteile im Vergleich und ist nicht als generelles Votum für die andere oder eine Datenquelle zu verstehen. Mit dem Ziel aussagekräftiger Qualitätsinformationen und auch im Interesse der Fairness unter den Leistungserbringern ist für jede Datenquelle eine möglichst hohe Datenqualität anzustreben. Dies erfordert, mögliche Einschränkungen der Datenquellen zu benennen, damit Maßnahmen zu deren Vermeidung ergriffen werden können, bei QS-Dokumentationsdaten beispielsweise ein Verfahren zur Prüfung der Datenqualität.

In einer Stellungnahme wurde zu bedenken gegeben, dass die Datenqualität keine inhärente Eigenschaft eines Qualitätsindikators sei und eine niedrige Datenqualität daher nicht zur Abschaffung eines Qualitätsindikators führen solle (GKV-SV, S. 13)

**IQTIG:** Die Eignungskriterien von Qualitätsindikatoren sind nicht allein Eigenschaften der zugrunde liegenden Kennzahl, sondern stehen im Bezug zum Kontext, in dem der Indikator angewendet wird. Beispielsweise hängt das Potenzial zur Verbesserung davon ab, in welchem Maß Qualitätsanforderungen durch die Leistungserbringer bereits erreicht sind, und die Objektivität der Messung hängt davon ab, welche Personengruppen den Sachverhalt erheben. Gleiches trifft für die Datenqualität zu, die bei QS-Dokumentationsdaten insbesondere durch die spezifikationskonforme und inhaltlich korrekte Übertragung der Behandlungsinformationen in die QS-Software bestimmt wird. Lässt sich die Ursache für eine niedrige Datenqualität beheben, empfiehlt das IQTIG, die Nutzung der Indikatorergebnisse nur vorübergehend auszusetzen. Lässt sich dagegen keine hinreichende Datenqualität erzielen, können mit dem Qualitätsindikator keine belastbaren Informationen generiert werden und der Qualitätsindikator sollte abgeschafft werden.

### **Angemessenheit der Risikoadjustierung**

Von einer stellungnehmenden Organisation wurde um Erläuterung gebeten, warum das Eignungskriterium „Angemessenheit der Risikoadjustierung“ den Eignungskriterien der Operationalisierung und nicht mehr dem Bewertungskonzept zugeordnet wurde (DGPPN, S. 2).

**IQTIG:** Die Sortierung der Eignungskriterien in drei Gruppen dient einer übersichtlichen und strukturierten Darstellung und hat keine direkte Auswirkung auf deren Anwendung. Eine angemessene Risikoadjustierung ist für aussagekräftige Vergleiche auf Grundlage von Qualitätsmessungen von Bedeutung, nicht nur für Qualitätsindikatoren mit festgelegtem Bewertungskonzept, sondern auch bei Nutzung der Ergebnisse ohne Bewertungskonzept (sog. Qualitätskennzahl). Die Umsortierung des Eignungskriteriums soll dies verdeutlichen.

Von einer stellungnehmenden Organisation wurde eingewandt, eine Risikoadjustierung könne niemals sämtliche Risikofaktoren berücksichtigen. Daher sei die Einschätzung des IQTIG falsch, Unterschiede im Patientenkollektiv könnten stets durch eine angemessene Risikoadjustierung ausgeglichen werden (DKG, S. 8).

**IQTIG:** Angemessenheit der Risikoadjustierung meint nicht, dass sämtliche Risikofaktoren berücksichtigt werden müssen. Gleichwohl geht das IQTIG nicht davon aus, dass jede Risikoadjustierung angemessen ist oder dass für jeden Indikator eine angemessene Risikoadjustierung erreicht werden kann. Gerade deshalb beurteilt das IQTIG die Angemessenheit der Risikoadjustierung mittels des gleichnamigen Eignungskriteriums.

### Praktikabilität der Messung/Aufwandsschätzung

Von mehreren stellungnehmenden Organisationen wurde die Auffassung vertreten, dass auch der Aufwand der Stellungnahmeverfahren nach § 17 Teil 1 DeQS-RL mit Leistungserbringern betrachtet werden solle (DGfN, S. 7; DKG, S. 8; KZBV, S. 12). Der Aufwand für die „Bewertung der Ergebnisse“ eines Indikators solle unter Berücksichtigung der „Spezifität“ in die Aufwandsschätzung einbezogen werden (GKV-SV, S. 11).

**IQTIG:** Bei der Eignungsprüfung von Qualitätsindikatoren bezieht das IQTIG den Aufwand von Stellungnahmeverfahren aus mehreren Gründen nicht in die Aufwand-Nutzen-Abwägung mit ein:

- Der Aufwand des Stellungnahmeverfahrens nach § 17 Teil 1 DeQS-RL hängt in erster Linie von dessen Ausgestaltung ab. Diese erfolgt bisher unterschiedlich zwischen den durchführenden Organisationen (IQTIG 2020: Kapitel 2), was bei empirischen Erhebungen zum Aufwand berücksichtigt werden müsste. Unabhängig davon erscheint vor Abschluss der vom G-BA am 6. März 2024 beauftragten Weiterentwicklung des Verfahrens zur qualitativen Bewertung (G-BA 2024b) eine entsprechende Aufwandserhebung derzeit nicht sinnvoll.
- Zudem kann der Aufwand des Stellungnahmeverfahrens schon bisher über verteilungsbezogene Referenzbereiche gesteuert werden. Soll die Anzahl der Stellungnahmeverfahren aus Aufwandsgründen bei den Landesarbeitsgemeinschaften oder den Leistungserbringern oder beiden reduziert werden, können dazu beispielsweise perzentilbasierte Referenzbereiche eingesetzt werden. Dann werden nur so viele Stellungnahmeverfahren durchgeführt, wie aus Aufwandsgründen für praktikabel gehalten werden.
- Sofern in der Stellungnahme unter „Spezifität“ das Verhältnis von im Stellungnahmeverfahren deklarierten Qualitätsdefiziten (Einstufung als „qualitativ auffällig“) zu sog. „rechnerisch auffälligen“ Ergebnissen gemeint ist, so ist diese nicht nur vom Indikator, sondern auch von der Lage des Referenzbereichs abhängig. Sie eignet sich daher nicht als Kriterium, ob der Indikator verwendet werden sollte. Zudem kann dieses Verhältnis irreführend sein, da auch Qualitätsdefizite vorliegen können, wenn die

Ursache für ein auffälliges Indikatorergebnis im Stellungnahmeverfahren nicht gefunden wurde. Ein Grund dafür ist, dass auch die Datengrundlagen für das Stellungnahmeverfahren, insbesondere für schriftliche Stellungnahmen, sowie dessen Durchführung Limitationen aufweisen. Ein weiterer Grund ist, dass im Stellungnahmeverfahren teilweise andere Sachverhalte beurteilt werden als im Qualitätsindikator und daher eine Korrelation der Ergebnisse von Stellungnahmeverfahren und Qualitätsindikator nicht unbedingt zu erwarten ist.

- Das Stellungnahmeverfahren mit Leistungserbringern in seiner bisherigen Form hat nicht nur den Charakter einer Qualitätsmessung, sondern wird teilweise auch als Instrument zur Qualitätsverbesserung verstanden. Der Aufwand für das Stellungnahmeverfahren ist somit nicht allein der Aufwand-Nutzen-Betrachtung der Qualitätsmessung zuzuordnen, sondern auch einer Aufwand-Nutzen-Betrachtung der qualitätsverbessernden Maßnahmen (IQTIG 2022: 20). Dabei müsste auch die Nutzenseite umfangreicher betrachtet werden, wie in ähnlicher Form auch von einer stellungnehmenden Organisation angemerkt wurde (PatV, S. 17). Dies überschreitet den Rahmen der Eignungsprüfung der Qualitätsmessung und wäre Teil einer umfassenden Evaluation.
- Des Weiteren müssen Qualitätsindikatoren nicht zwingend mit einem Stellungnahmeverfahren verbunden sein, damit ihre Ergebnisse zur Qualitätsverbesserung und zur Herstellung von Qualitätstransparenz genutzt werden können, beispielsweise, wenn sie als Feedback für das interne Qualitätsmanagement der Leistungserbringer oder im Rahmen von Veröffentlichungen für Auswahlentscheidungen der Patientinnen und Patienten genutzt werden.

Zusammenfassend sind die Aufwände im Stellungnahmeverfahren daher nicht maßgeblich für die Eignungsbeurteilung der Indikatoren, sondern sollten Anlass sein, das Stellungnahmeverfahren hinsichtlich seiner Ausgestaltung weiterzuentwickeln.

Von mehreren stellungnehmenden Organisationen wurde angeführt, dass bei der Einschätzung der Praktikabilität der Messung nicht nur der Dokumentationsaufwand, sondern auch weitere Aufwände berücksichtigt werden sollten. Genannt wurde in den Stellungnahmen beispielsweise der Aufwand für Personal, Technik, Software, Materialkosten, Raumkosten, Organisation (KBV, S. 6; KZVB, S. 11). Darüber hinaus solle nicht nur der Aufwand der Leistungserbringer, sondern auch der Aufwand bei den weiteren beteiligten Institutionen auf Bundes- und Landesebene einbezogen werden (DKG, S. 8; KZBV, S. 12), einschließlich des Aufwands für die Entwicklung, Umsetzung und Auswertung der Verfahren (KZBV, S. 12) sowie für die Administration der Verfahren wie „Richtlinienpflege, Spezifikation, Rechenregeln“ (DKG, S. 8).

**IQTIG:** Das IQTIG bezieht Aufwand, der indikatorenübergreifend ist oder der bei anderen Institutionen anfällt, aus folgenden Gründen nicht in die Aufwand-Nutzen-Betrachtung der Qualitätsindikatoren und Qualitätsindikatorenssets mit ein:

- Bisher liegen keine belastbaren Daten zur Höhe des indikatorenübergreifenden Aufwands vor. Erforderlich wäre eine empirische Erhebung, die die verschiedenen Aufwandskomponenten (u. a. Spezifikation, Softwareeinrichtung) bei den verschiedenen Beteiligten (z. B. Leistungserbringer, Landesarbeitsgemeinschaften, auf Bundesebene an der Qualitätssicherung beteiligte Institutionen) und möglicherweise verfahrensspezifische Aspekte berücksichtigt. Eine orientierende Befragung der Landesarbeitsgemeinschaften im Nachgang zum Abschlussbericht „Empfehlungen zur Weiterentwicklung von Verfahren der datengestützten Qualitätssicherung. Indikatorensets der Verfahren *QS PCI*, *QS HSMDEF* und *QS KEP*“ vom 19. Juli 2023 zeigte, dass sich die „weiteren Aufwände“ der Indikatoren und QS-Verfahren nicht mit vertretbarem Aufwand so erfassen lassen, dass aus den Ergebnissen konkrete Konsequenzen gezogen werden könnten.
- Es bleibt unklar, ob die Aussetzung eines oder mehrerer Indikatoren auch den indikatorenübergreifenden Aufwand reduziert. Sofern ein Indikatorenset nicht vollständig ausgesetzt wird, müssen indikatorenübergreifende Aufgaben weiterhin durchgeführt werden. Die Nichtberücksichtigung des indikatorenübergreifenden Aufwands in diesem Kontext vermeidet eine Überschätzung der möglichen Aufwandsreduktion durch Verzicht auf Indikatoren. Indikatorenübergreifender Aufwand wird jedoch bei der Verfahrensprüfung indirekt adressiert: Dieser Aufwand fällt vor allem dann ins Gewicht, wenn andere Aufwände gering sind und ihm nur ein geringer Nutzen gegenübersteht. Das IQTIG spricht daher für Qualitätsindikatoren mit nur „geringem“ Potenzial zur Verbesserung auch dann eine Aussetzungsempfehlung aus, wenn der indicatorspezifische Aufwand der Datenerhebung „gering“ oder „eher gering“ ist.
- Die Umsetzung von Maßnahmen der externen Qualitätssicherung gehört zu den primären Aufgaben der Organisationen auf Landes- und Bundesebene. Der Aufwand für diese Aufgaben geht anders als der Aufwand bei den Leistungserbringern nicht direkt zulasten der Patientenversorgung. Daher erscheint es nicht angemessen, ihn dem patientenrelevanten Nutzen in der Abwägung gegenüberzustellen.

In einer Stellungnahme wurde vorgeschlagen, neben dem Aufwand für die Leistungserbringer auch den Aufwand und die Belastung für befragte Patientinnen und Patienten in die Aufwand-Nutzen-Abwägung einzubeziehen. Dies gelte insbesondere für vulnerable Patientengruppen wie Patientinnen und Patienten mit psychischen Erkrankungen (DGPPN, S. 2).

**IQTIG:** Bei der Entwicklung von Patientenbefragungen werden der Aufwand und die Belastung für die Patientinnen und Patienten vom IQTIG ausdrücklich berücksichtigt (sog. Zumutbarkeit, siehe Abschnitt 7.2), auch in Bezug auf psychische Belastungen, und im Rahmen der Pretestung adressiert. Darüber hinaus handelt es sich für die Patientinnen und Patienten um eine freiwillige und anonyme Befragung.

Von zwei stellungnehmenden Organisationen wurde zu bedenken gegeben, dass auch Patientenbefragungen als Datengrundlage für Qualitätsmessungen mit einem relevanten Aufwand für die Leistungserbringer einhergingen. Der Aufwand für den Versand von Adressdaten sowie von Behandlungsdaten für die Fragebogenauswahl sei nicht „pauschal als niedrig“ anzunehmen (DGfN, S. 7). Auch die engen zeitlichen Fristen für den Versand der Adressdaten sowie die Notwendigkeit von Nachkodierungen oder Falländerungen zeigten, dass Patientenbefragungen mit mehr als „geringem“ Aufwand verbunden seien (DKG, S. 5). Demgegenüber wurde von einer anderen stellungnehmenden Organisation bekräftigt, dass die maßgeblichen Aufwände bei Patientenbefragungen nicht bei den Leistungserbringern entstünden (PatV, S. 16).

**IQTIG:** Der den Leistungserbringern entstehende Aufwand im Rahmen von Patientenbefragungen wird vom IQTIG bei einer Eignungsprüfung nicht pauschal als gering beurteilt, sondern gemäß dem in Anhang A.2.6 geschilderten Vorgehen datenfeldbezogen berücksichtigt. Wie groß dieser Aufwand ist, lässt sich erst nach durchgeführter Prüfung angeben. Es ist wahrscheinlich, dass der Aufwand im Vergleich zu Indikatoren auf Basis von QS-Dokumentationsdaten meist geringer ausfällt, da die Angaben des Leistungserbringers nur einmal je Fragebogen benötigt werden und nicht gesonderte Datenfelder für jeden Indikator dokumentiert werden müssen.

In einer Stellungnahme wurde die Frage aufgeworfen, in welchem Maße die Einschätzung des Dokumentationsaufwands themenspezifische Besonderheiten berücksichtigt. Beispielsweise sollten „Automatisierungs- und Digitalisierungspotenziale“ berücksichtigt werden, wenn in einem Versorgungsbereich viele von der Qualitätssicherung benötigte Daten ohnehin im Rahmen der Behandlung maschinell erfasst und nicht mehr händisch dokumentiert würden (PatV, S. 16 f.).

**IQTIG:** Bei der Beurteilung, welcher Dokumentationsaufwand je Datenfeld entsteht, berücksichtigt das IQTIG die automatische Erfassung von Daten (Schritt 1 der Beurteilung, siehe Anhang A.2.6). Mit dem Ziel einer möglichst realistischen Einschätzung des tatsächlichen Aufwands gehen das IQTIG und die dazu befragten Expertinnen und Experten vom zum Zeitpunkt der Beurteilung üblichen Automatisierungsgrad eines durchschnittlich gut organisierten Leistungserbringers aus. Ergeben sich künftig Hinweise, dass sich für bestimmte Daten der Automatisierungsgrad geändert hat, nimmt das IQTIG dies zum Anlass für eine Anpassung der Beurteilung.

### **Aufwand-Nutzen-Abwägung**

Von mehreren stellungnehmenden Organisationen wurden nähere Erläuterungen gewünscht, wie die strukturierte qualitative Abwägung der Eignungskriterien bei der Gesamtbeurteilung der Eignung eines Indikators erfolge (BPtK, S. 10; DGfN, S. 7; GKV-SV, S. 10; KBV, S. 6; KZBV, S. 11). Es wurde u. a. um Begründung gebeten, warum eine niedrige



Bewertung bei manchen Eignungskriterien wie der Datenqualität vom IQTIG als Ausschlussbedingung gewertet wird und bei anderen Eignungskriterien wie Reliabilität der Messung, Angemessenheit der Risikoadjustierung und „Relevanz“ nicht (GKV-SV, S. 10; KBV, S. 6).

**IQTIG:** Das IQTIG stellt in seinen Berichten zur Eignungsprüfung von Qualitätsindikatoren oder Indikatorensets jeweils dar, wie die Aufwand-Nutzen-Abwägung jeweils konkret vorgenommen wurde (siehe z. B. IQTIG 2023: Abschnitt 2.7). Da das strukturierte Verfahren im Rahmen der Prüfung aller Verfahren nach DeQS-RL erstmalig eingesetzt und intern weiterhin geprüft wird, ist es noch nicht im Detail in den „Methodischen Grundlagen“ festgehalten. Grundsätzlich unterscheidet das IQTIG bei der Abwägung zwei Situationen:

- Bei einem oder mehreren Eignungskriterien liegt eine Ausschlussbedingung vor. In diesem Fall wird ein Qualitätsindikator als derzeit nicht für die Qualitätssicherung geeignet gewertet, unabhängig davon, inwieweit die übrigen Eignungskriterien erfüllt sind, und unabhängig von der Einschätzung des Aufwands. Die Ausschlussbedingungen sind in Anhang A der „Methodischen Grundlagen“ aufgeführt. Lassen sich die Gründe für die fehlende Eignung nicht beheben, empfiehlt das IQTIG, den Indikator nicht mehr einzusetzen (Abschaffung).
- In allen anderen Fällen stellt das IQTIG die Bedeutung für die Patientinnen und Patienten und den Umfang des Potenzials zur Verbesserung dem Aufwand gegenüber. Ebenso wird die Güte der Operationalisierung dem Aufwand gegenübergestellt. Dazu werden die je Eignungskriterium zugeteilten Kategorien anhand vordefinierter Regeln verwendet. Beispielsweise empfiehlt das IQTIG, Indikatoren mit einem „hohen“ oder „eher hohen“ Aufwand nur weiterzuführen, wenn diesem Aufwand in mindestens zwei Eignungskriterien der Operationalisierung Einstufungen als „hoch“ gegenüberstehen. Überwiegt in diesen Gegenüberstellungen der Nutzen, empfiehlt das IQTIG, den betreffenden Indikator weiter einzusetzen.

Nicht bei jedem Eignungskriterium gilt die niedrigste Beurteilungskategorie als Ausschlussbedingung. Dies ist den inhaltlichen Unterschieden in den Eignungskriterien, in den Skalen und im erwarteten Antwortverhalten geschuldet.

- Bei der Bedeutung für die Patientinnen und Patienten wird davon ausgegangen, dass die Kategorie „gering“ zwar für Qualitätsmerkmale vergeben wird, die im Vergleich zu anderen Qualitätsmerkmalen weniger wichtig sind, die aber nicht völlig unbedeutend sind. Anderenfalls wäre das Merkmal in der Entwicklung nicht als Grundlage für einen Indikator gewählt worden.
- Die Reliabilität der Messung wird bei der Prüfung des Eignungskriteriums auf Fallebene beurteilt. Aufgrund der Mehrebenenstruktur der Qualitätsmessung wirkt sich eine „geringe“ Präzision auf Fallebene in Form breiterer Konfidenzintervalle auf

Ebene der Leistungserbringer aus. Dies kann jedoch teilweise dadurch berücksichtigt werden, dass statistische Unsicherheit der Indikatorwerte, z. B. durch die Angabe von Konfidenzintervallen, beachtet wird.

- Die Angemessenheit der Risikoadjustierung ist für Vergleiche relevant, z. B. bei der Veröffentlichung von Indikatorergebnissen. In dieser Verwendung wird eine „nicht angemessene“ Risikoadjustierung als Ausschlussbedingung (für diese Verwendung) gewertet. In anderen Situationen wie etwa der Verwendung für das interne Qualitätsmanagement der Leistungserbringer oder bei der Identifikation von Verbesserungsansätzen im Rahmen der Qualitätsförderung sind direkte Vergleiche dagegen von untergeordneter Bedeutung.

Zu den Gründen für eine Ausschlussbedingung beim Eignungskriterium „Datenqualität“ siehe oben.

In zwei Stellungnahmen wurde die Frage aufgeworfen, warum die Beurteilungskategorien der Eignungskriterien nicht einheitlich formuliert sind und sich je nach Eignungskriterium unterscheiden, etwa in der Anzahl der Kategorien, dem Vorhandensein einer Mittelkategorie und den Benennungen scheinbar gleicher Kategorien (GKV-SV, S. 10, KZBV, S. 15 f.).

**IQTIG:** Der Aufbau der Beurteilungskategorien wurde für jedes Eignungskriterium so gewählt, dass jede Kategorie den dahinterstehenden Erfüllungsgrad des Eignungskriteriums sprachlich möglichst passend beschreibt. Dabei wurden mehrere Aspekte berücksichtigt.

- Die Anzahl der Kategorien ergibt sich daraus, inwieweit eine Differenzierung notwendig und belastbar möglich ist. Beispielsweise ist es nicht sinnvoll, die „Brauchbarkeit für einen Handlungsanschluss“ in mehr Stufen als „ja“ und „nein“ zu unterteilen, und eine Differenzierung der „Bedeutung für die Patientinnen und Patienten“ wäre zwar möglich, würde aber die Frage aufwerfen, anhand welcher Kriterien zwischen feineren Abstufungen unterschieden werden soll. Demgegenüber erlaubt die quantitative Datengrundlage für das „Potenzial zur Verbesserung“ und die „Praktikabilität der Messung“ eine stärkere Differenzierung, die auch für die Aufwand-Nutzen-Abwägung hilfreich ist. Die Vermeidung einer zu großen Anzahl von Kategorien dient somit auch dazu, unnötige Komplexität und eine Scheingenauigkeit zu vermeiden.
- Die Benennung einer Mittelkategorie ist naturgemäß bei dichotomer Kategorienskala („ja“/„nein“) nicht sinnvoll. In allen anderen Fällen ist eine Mittelkategorie vorgesehen. Bei den Kategorien für die Eignungskriterien der Operationalisierung handelt es sich hier - anders als dies häufig bei Items in Befragungen der Fall ist - nicht um äquidistante Kategorien auf einer metrischen Skala, sondern um eine ordinale Skala mit einem Spektrum von „unzureichend“ bis „kaum Verbesserungsbedarf“. Die Mittelkategorie wurde nicht bei allen Eignungskriterien mit der Bezeichnung „mittel“ versehen, um irreführende Konnotationen zu vermeiden. Beispielsweise wäre eine „mittlere Objektivität“ bereits vom Begriff widersprüchlich und würde eine Ausschluss-

bedingung nahelegen, während die Bezeichnung „weitgehend“ die zugrunde liegende Idee einer eingeschränkten, aber noch tolerablen Objektivität besser beschreibt. Das gleiche trifft sinngemäß auf die „eingeschränkt angemessene“ Risikoadjustierung (statt „mittel angemessen“) zu.

Eine eignungskriterienübergreifend gleiche Benennung von Kategorien ist letztendlich auch deswegen nicht erforderlich, weil die Eignungskriterien unterschiedliche Sachverhalte abbilden und kein direkter Vergleich der Kategorien vorgenommen wird.

Von einer stellungnehmenden Organisation wurde kritisiert, dass das Fehlen eines Referenzbereichs vom IQTIG nicht als Ausschlusskriterium für den Einsatz eines Qualitätsindikators gewertet wird (KBV, S. 6, 13).

**IQTIG:** Um über die Versorgungsqualität von Leistungserbringern zu informieren, ist die Angabe eines Referenzbereichs oder die Angabe des Einstufungsergebnisses eines Leistungserbringers in Bezug auf einen Referenzbereich nicht nötig. Das Indikatorergebnis ermöglicht Aussagen zur Erfüllung des Qualitätsmerkmals, und ein Vergleich zwischen Leistungserbringern kann anhand der Indikatorergebnisse erfolgen. Bei bestimmten Verwendungen der Indikatorergebnisse sollten die Entscheidungen auf Grundlage der Indikatorergebnisse nach einheitlichen, objektiven Kriterien getroffen werden, beispielsweise die Entscheidung, eine Qualitätsförderung bestimmter Leistungserbringer anzustoßen. Diese Funktion erfüllen der Referenzbereich und die zugehörige statistische Klassifikationsregel. Für diese Verwendungszwecke ist ein Referenzbereich erforderlich.

Von zwei stellungnehmenden Organisationen wurde kritisiert, es werde nicht hinreichend klar, was das IQTIG unter „Nutzen“ und „Aufwand“ eines Qualitätsindikators verstehe (DGfN, S. 7; KZBV, S. 11).

**IQTIG:** Der Nutzen eines Instruments zur Qualitätsmessung wie beispielsweise eines Qualitätsindikators besteht in der von ihm bereitgestellten Information über die Versorgungsqualität (IQTIG 2022: 20). Dieser Nutzen ergibt sich daraus, dass erstens ein geeignetes Qualitätsmerkmal abgebildet wird und zweitens diese Abbildung sachgerecht operationalisiert ist (siehe Abschnitt 13.2 der „Methodischen Grundlagen“). Der Aufwand eines Qualitätsindikators ist dementsprechend der Aufwand, der durch die Erhebung und Bereitstellung der benötigten Informationen entsteht.

Von einer stellungnehmenden Organisation wurde vorgeschlagen, in der Aufwand-Nutzen-Betrachtung den „Nutzen“ zu verstehen als die „Verbesserung der Werte bei einzelnen Qualitätsindikatoren“ und zu berücksichtigen, inwieweit „die anschließenden Maßnahmen zur Verbesserung der Qualität beigetragen“ haben (KZBV, S. 11 f.).

**IQTIG:** Mit Blick auf die Verbesserung der Qualität ist für die Beurteilung des Nutzens eines Messinstruments maßgeblich, ob erreichbare Qualitätsziele *erreicht sind* (Bestandsaufnahme). Bei erreichten Qualitätszielen kann auf weitere Messungen und Maßnahmen verzichtet werden und umgekehrt bestehen bei nicht erreichten Qualitätszielen weiterhin Möglichkeiten zur Verbesserung, sodass eine Messung weiterhin nützlich sein kann. Das IQTIG beurteilt diese Facette des Nutzens anhand des Eignungskriteriums „Potenzial zur Verbesserung“.

Eine andere Fragestellung ist, ob die eingesetzten Maßnahmen ihre intendierte Funktion erfüllt haben, d. h., ob die Ergebnisse *durch die Qualitätsmessung und die anschließenden Maßnahmen herbeigeführt* wurden (Kausalitätsprüfung) und nicht auf anderen Gründen beruhen. Dazu müsste beurteilt werden, ob der historische Verlauf der Ergebnisse auf die Wirkung der Qualitätssicherungsmaßnahme zurückgeht. Für die Entscheidung über die Fortführung der Qualitätsmessung ist jedoch in erster Linie nicht bedeutsam, ob die bisher eingesetzten Qualitätssicherungsmaßnahmen wirksam waren, sondern ob die Qualitätsziele schon erreicht wurden und ob die Ergebnisse der Messung für mindestens eine potenziell wirksame Maßnahme verwendet werden können. Letzteres wird durch das Eignungskriterium „Brauchbarkeit für den Handlungsanschluss“ untersucht.

Von einer stellungnehmenden Organisation wurde eine Aufwand-Nutzen-Betrachtung der Indikatoren mit gesundheitsökonomischen Methoden gewünscht (KZBV, S. 5 und 11 f.).

**IQTIG:** Eine angemessene Aufwand-Nutzen-Betrachtung von Qualitätsindikatoren erfordert die Berücksichtigung multipler Kriterien, der Eignungskriterien für Qualitätsmessungen. Nicht für alle diese Kriterien liegen allgemein akzeptierte Quantifizierungen vor. Das IQTIG nimmt die Aufwand-Nutzen-Betrachtung daher so vor, dass sowohl qualitative als auch quantitative Aspekte der Eignung berücksichtigt werden können (partielle multikriterielle Entscheidungsanalyse, siehe auch Marsh et al. 2016).

Eine Beschreibung des Aufwands auch über die Parameter Zeitbedarf und Kosten ist aus Sicht des IQTIG grundsätzlich möglich. Diese erfordert jedoch die Erhebung von Zeitbedarf und Kosten mit einer zuverlässigen Methode und in einer repräsentativen Stichprobe von Einrichtungen. Bisher sind keine solchen Studien publiziert und konnten auch vom IQTIG nicht in der erforderlichen Detailtiefe durchgeführt werden. Daher sind derzeit keine belastbaren Aussagen zum Zeitbedarf und zu Kosten möglich.

In einer Stellungnahme wurde vorgeschlagen, unerwünschte Nebeneffekte von Qualitätsmessungen bei der Eignungsprüfung zu berücksichtigen. Beispielsweise könne die Veröffentlichung von Qualitätsinformationen zum Fehlanreiz bei Leistungserbringern führen, bevorzugt Patientengruppen zu behandeln, bei denen gute Behandlungsergebnisse oder eine gute therapeutische Beziehung zu erwarten sind. Auch könnten bestimmte Behandlungsprozesse gezielt zu stark fokussiert werden, „um die Erinnerungsleistung der Patient\*innen beim Ausfüllen der Patientenbefragung zu verbessern“. Des

Weiteren gehe die Veröffentlichung unterdurchschnittlicher Qualitätsergebnisse mit negativen persönlichen Folgen für die Leistungserbringer in Einzelpraxen einher, bei denen die Ergebnisse auf einzelne Personen bezogen werden können (BPtK, S. 11 f.).

**IQTIG:** Falls unerwünschte Effekte wie Fehlanreize auf eine ungeeignete Operationalisierung von Qualitätsindikatoren zurückzuführen sind, schlägt sich dies bei der Eignungsprüfung in den Eignungskriterien der Operationalisierung wieder. Bei einem validen, angemessen risikoadjustierten Qualitätsindikator ist weitgehend sichergestellt, dass schlechte Behandlungsergebnisse oder schwierige Behandlungsprozesse nicht ohne Weiteres zu einer schlechten Qualitätsbewertung führen, sondern dass die Besonderheiten z. B. bestimmter Patientengruppen berücksichtigt werden. So beugt eine angemessene Risikoadjustierung auch dem Fehlanreiz der Risikoselektion, also der bevorzugten Behandlung weniger kranker Patientinnen und Patienten vor.

Das Ziel von QS-Verfahren ist letztlich die Bereitstellung aussagekräftiger Qualitätsinformationen, auch für Leistungserbringer in Einzelpraxen. Aus Sicht des IQTIG kann bei angemessener Wahl der an die Qualitätsmessung anschließenden Maßnahmen ein Qualitätsbewusstsein geschaffen werden, das sowohl für Leistungserbringer als auch für Patientinnen und Patienten hilfreich ist und der Qualitätsverbesserung dient.

## 4 Bewertungskonzepte und Referenzbereiche

Von einer stellungnehmenden Organisation wurde die Aussage in den „Methodischen Grundlagen“ kritisiert, dass kriteriumsbezogene Referenzbereiche hohe und geringe Versorgungsqualität sichtbar machen können. Da sich die Beurteilung auf empirische Daten stütze, müsse darauf aufmerksam gemacht werden, dass die Ergebnisse mit Unsicherheit behaftet seien (DKG, S. 9).

**IQTIG:** Jede Beurteilung auf Grundlage empirischer Erhebungen geht mit Unsicherheit einher. Dies betrifft standardisierte Erhebungen und Auswertungen, z. B. mittels QS-Dokumentationsdaten und Qualitätsindikatoren, ebenso wie weniger standardisierte Erhebungen und Auswerten, z. B. mittels Dialogverfahren und Expertenurteilen. Bei standardisierten Verfahren mit quantitativer Auswertung wird die Unsicherheit explizit berücksichtigt und dargestellt. Dies erfolgt über das statistische Verfahren, mittels dessen Leistungserbringerergebnisse und Referenzbereich miteinander verglichen werden. Daher berücksichtigt das IQTIG bei der Beurteilung des Kriteriums „Klassifikationsgüte“ insbesondere, ob diese Unsicherheit adäquat berücksichtigt wurde, z. B. mittels statistischem Signifikanztest.

In einer Stellungnahme wurde nachgefragt, in welchem Bezug die in Anhang B geschilderte Herleitung des „erreichbaren Indikatorwerts“ zu den Begriffen in Abschnitt 15.3 steht und welche Änderungen sich im Vergleich zu Version 2.0 ergeben haben. Es sei unklar, wie der theoretische Minimal- bzw. Maximalwert ermittelt werden und „ob der hypothetische Wert dem theoretischen Wert entspricht“ oder dem „bestenfalls erreichbaren Indikatorwert“ (KBV, S. 7 f.).

**IQTIG:** Gegenüber der Version 2.0 ist das in Abschnitt 15.3 geschilderte Konzept unverändert und es wurden lediglich redaktionelle Anpassungen vorgenommen. Der Begriff des „erreichbaren Indikatorwerts“ bezeichnet eine Kenngröße zur Beschreibung des fachlich Erreichbaren auf der Indikatorskala. Der erreichbare Indikatorwert wird gemäß der in Anhang B geschilderten statistischen Methodik berechnet und dort als hypothetische Größe bezeichnet, da er nicht direkt gemessen wird, sondern unter den dort geschilderten Annahmen geschätzt wird. Mit dem „theoretischen Minimum bzw. Maximum“ eines Indikatorwerts sind die kleinsten bzw. größten Werte gemeint, die bei einem Indikator gemessen werden können. Beispielsweise entsprechen sie bei einem anteilbasierten Indikator den Skalenenden 0 % und 100 %. Für die Festlegung eines kriteriumsbezogenen Referenzbereichs haben sie keine direkte Bedeutung, da dieser (unter Berücksichtigung einer zusätzlichen Toleranz) vom erreichbaren Indikatorwert abgeleitet wird, und sie müssen daher nicht gesondert ermittelt werden.

Von einer stellungnehmenden Organisation wurde ein Widerspruch zwischen den Aussagen zu verteilungsbezogenen Referenzbereichen und zum erreichbaren Indikatorwert gesehen. Es sei nachvollziehbar, dass ein verteilungsbezogener Referenzbereich keine Aussage über die Einhaltung von vorab definierten Standards treffe (Abschnitt 15.2). Dem widerspreche die Schilderung des erreichbaren Indikatorwerts als Qualität, „die bei Erfüllung aller fachlichen Standards von den Leistungserbringern erreicht werden könnte“ (Anhang B), da der erreichbare Indikatorwert laut der Schilderung auf einem Benchmarking-Ansatz beruhe und somit „der einzelne Leistungserbringer mit einem verteilungsbezogenen Referenzbereich verglichen“ werde (DKG, S. 10).

**IQTIG:** Kennzeichnend für verteilungsbezogene Referenzbereiche ist, dass der Leistungserbringer, dessen Indikatorwert mit dem Referenzbereich verglichen wird, selbst Teil der Vergleichsgruppe ist. Beispielsweise wird ein Leistungserbringer im Erfassungsjahr 2024 mit allen Leistungserbringern im Erfassungsjahr 2024 verglichen. Leistungserbringerergebnis und Vergleichswert stammen also aus demselben Zeitraum. Dadurch ist die Lage des verteilungsbezogenen Referenzbereichs für den Vergleich mit dem Leistungserbringerergebnis nicht vorab festgelegt.

Demgegenüber dient der erreichbare Indikatorwert nicht als Referenzbereich, sondern lediglich als Informationsgrundlage, um den erreichbaren Standard fachlich einzuschätzen und daraus abgeleitet einen angemessenen kriteriumsbezogenen Referenzbereich prospektiv festzusetzen. Beispielsweise kann ein erreichbarer Indikatorwert auf Basis von Daten des Erfassungsjahres 2022 berechnet werden und als Informationsgrundlage für die Festlegung eines kriteriumsbezogenen Referenzbereichs für das Erfassungsjahr 2024 genutzt werden. Der Leistungserbringer wird in diesem Fall nicht mit den Ergebnissen der Leistungserbringer des Erfassungsjahres 2024 verglichen, sondern die Lage des Referenzbereichs für den Vergleich ist vorab festgelegt.

Von einer stellungnehmenden Organisation wurde kritisch kommentiert, dass das IQTIG Interventionsstudien zur Einschätzung des Erreichbaren hinzuzieht. In der Regel handele es sich bei den Interventionen, die in den Studien untersucht würden, um „neue Ansätze“ und es müsse darauf geachtet werden, ob diese im Rahmen der Regelversorgung umsetzbar seien (KBV, S. 7).

**IQTIG:** Im Kontext der Qualitätssicherung handelt es sich bei den hier interessierenden Interventionen nicht um neuartige Untersuchungs- und Behandlungsmethoden, sondern um Prozesse, die als fachliche Standards anzusehen sind. Dabei kann es sich auch um Maßnahmen des internen Qualitätsmanagements handeln, deren Umsetzung zu besseren Indikatorergebnissen führt. Das IQTIG untersucht bei der Bewertung entsprechender Studien, inwieweit diese auf den Kontext der Regelversorgung übertragbar sind.

In einer Stellungnahme wurde vermutet, dass ein Referenzbereich „vom IQTIG praktisch als eine Möglichkeit zur Risikoadjustierung genutzt wird“, da er Einflussfaktoren auf die

Indikatorergebnisse berücksichtige, die in der Berechnungsvorschrift nicht berücksichtigt werden könnten. Mittelwertbasierte Referenzbereiche seien daher nur angemessen, „wenn eine Gleichverteilung der über den Referenzbereich zu „kompensierenden“ Faktoren anzunehmen“ sei (GKV-SV, S. 14).

**IQTIG:** Die Berücksichtigung nicht vom Leistungserbringer zu verantwortender Einflussfaktoren auf die Indikatorergebnisse erfolgt mit unterschiedlichen Methoden, die unterschiedliche Ziele verfolgen. Ein Ziel ist, bei der Darstellung und Verwendung der Indikatorwerte die Unsicherheit kenntlich zu machen, die durch nicht gemessene, zufällig wirkende Einflüsse entsteht, welche zur Abweichung des gemessenen Wertes vom zugrunde liegenden Wert (Kompetenzparameter) führen. Dies erfolgt beispielsweise durch Angabe von Konfidenzintervallen und durch entsprechende statistische Verfahren für den Vergleich zwischen Indikatorwert und Referenzbereich.

Das zweite Ziel besteht darin, möglichst faire Vergleiche zwischen Leistungserbringern zu ermöglichen, deren Fallmenge sich systematisch in bestimmten Einflussfaktoren unterscheidet (z. B. im Casemix). Dazu nimmt das IQTIG eine Standardisierung der betrachteten Fallmengen vor, entweder durch Ausschluss von Einflussfaktoren aus der Indikatorberechnung oder, wenn ein vollständiger Ausschluss nicht sinnvoll ist, durch eine angemessene Risikoadjustierung.

Das dritte Ziel, mit dem das IQTIG nicht vom Leistungserbringer zu verantwortende Einflussfaktoren berücksichtigt, ist eine realistische Schätzung des fachlich Erreichbaren auf der Indikatorkala. Dies ist für kriteriumsbezogene Referenzbereiche wichtig, da sie feste Anforderungen an die Leistungserbringer definieren. Da die betreffenden Einflussfaktoren nicht selbst bekannt sind, sondern nur die Summe ihrer Auswirkungen auf alle Indikatorergebnisse, können Unterschiede zwischen den Leistungserbringern in der Verteilung dieser Einflussfaktoren nicht kompensiert und Referenzbereiche nur einheitlich für alle Leistungserbringer festgelegt werden. Diese Limitation betrifft alle Arten von Referenzbereichen gleichermaßen.

Das IQTIG versteht Referenzbereiche daher ausdrücklich nicht als ein Mittel der Risikoadjustierung, da sie Einflussfaktoren mit dem Ziel realistischer Anforderungen an die Leistungserbringer berücksichtigen, aber nicht die Vergleichbarkeit der Ergebnisse zwischen Leistungserbringern erhöhen.

In zwei Stellungnahmen wurde vermutet, bei Verwendung mittelwertbasierter Referenzbereiche bleibe unklar, worin gute Qualität besteht. Mittelwertbasierte Referenzbereiche stünden im Widerspruch zur Aussage des IQTIG, die Qualitätsrichtung eines Indikators müsse monoton sein, um eine eindeutige Bewertung anhand von Indikatoren vornehmen zu können (KBV, S. 7). Solche Referenzbereiche gäben den „Leistungserbringern keine klare Orientierung“ (KZBV, S. 14 f.). In diesem Zusammenhang wurde von einer stellungnehmenden Organisation um Erläuterung gebeten, welche Vor- und Nachteile mittelwertbasierte Referenzbereiche gegenüber anderen verteilungsbezogenen



Referenzbereichen (z. B. solche, die sich „auf die schlechtesten 5 %“ beziehen) aufweisen und welche Bedeutung der Unterschied für Auswahlentscheidungen von Patientinnen und Patienten habe (PatV, S. 17).

**IQTIG:** Auch bei Verwendung eines mittelwertbasierten Referenzbereichs ist die Qualitätsrichtung des Indikators - abhängig vom zugrunde liegenden Qualitätsmerkmal - eindeutig (monoton) in Richtung höherer Werte oder eindeutig in Richtung niedrigerer Werte. Beispielsweise könnte ein Indikator zur Information und Aufklärung einen Mittelwert von 70 Punkten (von 100 erreichbaren Punkten) aufweisen. In diesem Fall wäre die Qualitätsrichtung eindeutig positiv (je mehr Punkte, desto besser) und für Leistungserbringer mit Ergebnissen signifikant unterhalb des Referenzbereichs würde die Notwendigkeit einer Qualitätsverbesserung angezeigt. Die Bezeichnung „mittelwertbasiert“ bedeutet *nicht*, dass der Mittelwert als anzustrebendes Ziel deklariert wird.

Ein Vorteil mittelwertbasierter Referenzbereiche ist im Gegenteil, dass der Eindruck vermieden wird, mit einem Indikatorergebnis im Referenzbereich seien keine Verbesserungsbestrebungen mehr nötig. Beispielsweise können Leistungserbringer mit Ergebnissen über dem Mittelwert anstreben, noch bessere Ergebnisse zu erzielen und ihre Versorgung so verbessern. Mittelwertbasierte Referenzbereiche sollen Leistungserbringern mit unterschiedlicher Versorgungsqualität Anreize bieten, ihre Qualität zu steigern.

Um Auswahlentscheidungen von Patientinnen und Patienten mittels Indikatorergebnisse zu unterstützen, ist eine allgemeinverständliche und visuelle Aufbereitung und insbesondere Aggregation von Qualitätsergebnissen sinnvoll. Dabei spielen Referenzbereiche eine untergeordnete Bedeutung. Denn um Auswahlentscheidungen zu unterstützen, sollen Patientinnen und Patienten sowie Dritten eigenständige Vergleiche zwischen den Ergebnissen verschiedener Leistungserbringer ermöglicht werden.

Von einer stellungnehmenden Organisation wurde nachgefragt, inwieweit unterschiedliche Fallzahlen der Leistungserbringer bei der Festlegung mittelwertbasierter Referenzbereiche berücksichtigt würden. Es wurde als Risiko wahrgenommen, dass Leistungserbringer mit hoher Fallzahl die Mittelwertbildung überproportional prägen. Außerdem wurde um Klarstellung gebeten, wie die Signifikanzprüfung bei mittelwertbasierten Referenzbereichen erfolge (KBV, S. 7).

**IQTIG:** Unterschiedliche Arten der Mittelwertbildung haben unterschiedliche Vorteile und können für unterschiedliche Indikatoren angemessen sein. In den „Methodischen Grundlagen“ wird keine Festlegung auf eine bestimmte Art der Mittelung getroffen, sondern nur beispielhaft das arithmetische Mittel der Indikatorergebnisse erwähnt. Derzeit finden mittelwertbasierte Referenzbereiche nur bei Indikatoren der Patientenbefragung Anwendung. Wenn mehr Erfahrung mit mittelwertbasierten Referenzbereichen vorliegt, wird diese in zukünftige Versionen der „Methodischen Grundlagen“ einfließen.

Die „Signifikanzprüfung“ bei der Anwendung des Referenzbereichs, d. h. die Berücksichtigung statistischer Unsicherheit beim Vergleich der Indikatorwerte mit dem Referenzbereich, erfolgt nach dem gleichen Vorgehen wie bei anderen Referenzbereichen.

Von einer stellungnehmenden Organisation wurde vorgeschlagen, bei Bewertungskonzepten neben dem Mittelwert auch die Varianz der Datenverteilung zu berücksichtigen (GMDS/IBS-DR, S. 2).

**IQTIG:** Referenzwerte werden mit dem (beobachteten) Indikatorwert verglichen. Der Indikatorwert ist dabei der (geschätzte) Erwartungswert der angenommenen Datenverteilung oder eine Transformation dessen. Grundsätzlich ist es auch möglich, Referenzwerte zu definieren, die mit anderen (geschätzten) Parametern der Datenverteilung verglichen werden. Damit ließen sich insbesondere auch Anforderungen an die Variabilität der Behandlungsqualität der Leistungserbringer stellen. Es werden jedoch bei den allermeisten Indikatoren binäre Ereignisse betrachtet. Als Datenverteilung wird auf Leistungserbringerebene eine Binomialverteilung bzw. generalisierte Binomialverteilung angenommen. Da bei diesen Verteilungen Varianz und Erwartungswert gekoppelt sind, werden keine getrennten Anforderungen an Erwartungswert und Varianz gestellt.

Von einer stellungnehmenden Organisation wurde vorgeschlagen, bei der Festlegung kriteriumsbezogener Referenzbereiche international publizierte Empfehlungen zu berücksichtigen, beispielsweise die Empfehlungen der Weltgesundheitsorganisation (*World Health Organization*, WHO) zur Kaiserschnitttrate (DGHWi, S. 5).

**IQTIG:** Das IQTIG berücksichtigt bei der Festlegung kriteriumsbezogener Referenzbereiche internationale Empfehlungen, indem entsprechende wissenschaftliche Literatur zur Beurteilung des Erreichbaren auf der Indikatorkala hinzugezogen wird. Dabei prüft das IQTIG unter Beratung durch Fachexpertinnen und Fachexperten auch, inwieweit die Literaturergebnisse und -empfehlungen auf den deutschen Versorgungskontext übertragbar sind.

Von einer stellungnehmenden Organisation wurde die große Bedeutung von Referenzbereichen hervorgehoben und vorgeschlagen, die fachliche Einschätzung der Fachkommissionen zu berücksichtigen. (KBV, S. 7).

**IQTIG:** Bei der Beurteilung der Angemessenheit des Referenzbereichs bezieht das IQTIG die Einschätzung von Fachexpertinnen und Fachexperten ein, indem das jeweils zuständige Expertengremium auf Bundesebene konsultiert wird. Die Erfahrungen aus der praktischen Anwendung zeigen allerdings auch, dass die Einschätzung durch die Expertengremien, welcher Indikatorwert fachlich erreichbar ist, mit Unsicherheit verbunden ist, wenn nicht ohnehin belastbare empirische Daten dazu vorliegen. Bei der Eignungsprüfung von Indikatoren im Regelbetrieb erfolgt daher die Schätzung des erreichbaren Indikatorwerts anhand der in Anhang B geschilderten Methodik.

In einer Stellungnahme wurde nachgefragt, wer über die Festlegung von kriteriumsbezogenen Referenzbereichen, einschließlich der Beurteilung des Erreichbaren und der Festsetzung des geforderten Erreichungsgrades, entscheide. Die Entscheidung setze „eine solide Kenntnis der Entwicklung in Forschung, Entwicklung und Praxis voraus“ und benötige daher die Einbindung der Leistungserbringer (DGIM, S. 3).

**IQTIG:** Die *Empfehlung* eines Referenzbereiches an den G-BA ist Aufgabe des IQTIG. Das IQTIG berücksichtigt bei der Entwicklung seiner fachlichen Empfehlungen die Erfahrungen aus „Forschung, Entwicklung und Praxis“ dadurch, dass es die wissenschaftliche Literatur sowie Fachexpertinnen und Fachexperten der jeweiligen Gesundheitsprofessionen einbezieht. Die abschließende *Entscheidung*, welcher Referenzbereich für einen Qualitätsindikator verwendet wird, trifft der G-BA.

Von einer stellungnehmenden Organisation wurde die Bevorzugung kriteriumsbezogener Referenzbereiche kritisiert. Auch mittels verteilungsbezogener Referenzbereiche könne Qualitätsförderung vorgenommen werden. Zudem könnten für einen Indikator auch zwei Aussagen gleichzeitig getroffen werden: „Im Sinne der Transparenz könnte zum Beispiel berichtet werden, dass ein Leistungserbringer den festen Referenzbereich erfüllt, und gleichzeitig in die Gruppe der zum Beispiel 10 % besten Leistungserbringer bei diesem Indikator gehört.“ (DNVF, S. 6) Ähnlich wurde von einer anderen stellungnehmenden Organisation vorgeschlagen, die Benennung fester Referenzbereiche als Orientierung für die Leistungserbringer von der nachgelagerten Steuerung der Handlungsanschlüsse zu trennen (KZBV, S. 14). In einer weiteren Stellungnahme wurde um Klärung gebeten, ob die Verwendung eines verteilungsbezogenen Referenzbereichs, etwa mit dem Ziel der Aufwandsreduktion oder des schrittweisen Erreichens von Standards, dazu führe, dass ein kriteriumsbezogener Referenzbereich ersetzt werde (KBV, S. 7). In einer weiteren Stellungnahme wurde deutlich gemacht, dass das Nichterreichen von Versorgungsstandards unabhängig von Aufwandsüberlegungen transparent gemacht werden müsse und nicht durch die Verwendung verteilungsbezogener Referenzbereiche verdeckt werden dürfe (PatV, S. 3 und 17 f.).

**IQTIG:** Für einen Indikator können grundsätzlich mehrere Referenzbereiche definiert werden, die sich je nach Handlungsanschluss unterscheiden (siehe Kapitel 15 der „Methodischen Grundlagen“). Die Argumente in den Stellungnahmen sind aus Sicht des IQTIG zutreffend und passen zum Konzept mehrerer Referenzbereiche für unterschiedliche Handlungsanschlüsse. Bei Verwendung mehrerer Referenzbereiche sollte klar benannt sein, für welchen Handlungsanschluss der jeweilige Referenzbereich definiert ist. Passend zu den Vorschlägen in den o. g. Stellungnahmen könnten beispielsweise für einen Indikator zwei Referenzbereiche festgelegt werden: ein fester Referenzbereich, der sichtbar macht, ob vom Leistungserbringer ein fachlicher Standard erreicht wurde, unterhalb dessen Maßnahmen zur Qualitätsverbesserung ergriffen werden sollten (Interpretation/Handlungsanschluss: Bedarf an Aktivitäten zur Qualitätsverbesserung),

sowie ein verteilungsbezogener Referenzbereich, der unter Berücksichtigung der verfügbaren Ressourcen festlegt, bei welchen Leistungserbringern externe Maßnahmen zur Qualitätsförderung am dringendsten durchgeführt werden sollten (z. B. die 5 % Leistungserbringer mit den schlechtesten Indikatorwerten).

Das IQTIG sieht es als wichtig an, dass das Erreichen oder Nichterreichen von Standards benannt wird, sofern sich solche Standards angeben lassen. Da in der externen Qualitätssicherung nach DeQS-RL derzeit nur ein Referenzbereich pro Qualitätsindikator vorgesehen ist, empfiehlt das IQTIG bevorzugt kriteriumsbezogene Referenzbereiche. Gleichzeitig kann insbesondere bei hohem Verbesserungsbedarf die Konstellation vorliegen, dass die verfügbaren Ressourcen (z. B. bei den Landesarbeitsgemeinschaften) nicht ausreichen, um bei allen Leistungserbringern, die dies benötigen, qualitätsfördernde Maßnahmen einzusetzen. Welche Ressourcen für solche Maßnahmen aufgewendet werden sollen, liegt in der Entscheidung des G-BA. Gegebenenfalls muss daher eine zusätzliche Priorisierung vorgenommen werden, z. B. anhand einer Rangreihe der Indikatorergebnisse. Letzteres würde praktisch einem zweiten, verteilungsbezogenen Referenzbereich entsprechen.

In einer Stellungnahme wurde die Frage aufgeworfen, warum im Themenbereich der ambulanten Psychotherapie für die Indikatoren zur Patientenbefragung nicht wie bei Indikatoren auf Basis von QS-Dokumentation kriteriumsbezogene Referenzbereiche vom IQTIG bevorzugt wurden. Des Weiteren wurde um Erläuterung gebeten, nach welchen Kriterien die Wahl der Referenzbereichsgrenze (z. B. 90 % versus 95 %) bei den festen Referenzbereichen erfolgte (BPtK, S. 13).

**IQTIG:** In den „Methodischen Grundlagen“ wird das grundsätzliche methodische Vorgehen geschildert, nach dem das IQTIG Referenzbereiche für Qualitätsindikatoren empfiehlt. Für Fragen, Hinweise und Diskussionen zu konkreten Verfahrensentwicklungen und Indikatoren werden gesonderte Beteiligungsverfahren zu den jeweiligen Vorberichten durchgeführt, auf die hier verwiesen sei. Das Vorgehen des IQTIG bei der Empfehlung von Referenzbereichen im Themenbereich der ambulanten Psychotherapie folgt den in Kapitel 15 beschriebenen Grundsätzen. Falls je Indikator nur ein Referenzbereich festgelegt werden soll, bevorzugt das IQTIG kriteriumsbezogene Referenzbereiche. Deren Größe orientiert sich dabei daran, was von den Leistungserbringern fachlich erreichbar ist, sowie daran, welche Abweichung vom Erreichbaren toleriert werden sollte, bevor weitere Maßnahmen initiiert werden.

## 4 Methoden der Datenanalyse

In Kapitel 20 wurde im Entwurf zur Version 2.1 der „Methodischen Grundlagen“ der neue Abschnitt 20.4 zu Volume-Outcome-Analysen eingefügt. Die anderen Abschnitte wurden nur redaktionell angepasst. Hinweise zu den anderen Abschnitten werden daher bei zukünftigen Überarbeitungen der „Methodischen Grundlagen“ berücksichtigt.

### Allgemeine Anmerkungen

Von einer stellungnehmenden Organisation wurde eine Trennung von Prädiktion und kausaler Inferenz in den jeweiligen Abschnitten empfohlen (GMDS/IBS-DR, S. 2).

**IQTIG:** Der kausale Aspekt von statistischen Auswertungen ist im derzeitigen Rahmenkonzept noch nicht verankert. Gleichwohl handelt es sich um einen wichtigen Aspekt, der auch in aktuellen wissenschaftlichen Diskussionen eine große Rolle einnimmt. Das IQTIG nimmt den Hinweis daher dankend zur Kenntnis und wird sich im Rahmen der kommenden Überarbeitungen der „Methodischen Grundlagen“ damit auseinandersetzen.

Von einer stellungnehmenden Organisation wurde vorgeschlagen, die Bedeutung der Berücksichtigung von statistischer Unsicherheit stärker zu betonen sowie die verwendeten statistischen Methoden darzustellen, insbesondere in Hinblick auf Stichprobenziehungen (GMDS/IBS-DR, S. 2).

**IQTIG:** Das IQTIG erkennt die große Bedeutung der Berücksichtigung von Unsicherheit an. Im Abschnitt zur Bewertungsart des Rahmenkonzepts für die statistische Auswertungsmethodik (Abschnitt 20.1.4) wird die Berücksichtigung von Unsicherheit im Kontext von Analysen mit analytischer Zielsetzung sowie mit enumerativer Zielsetzung in Kombination mit Stichprobenziehungen behandelt. Im neuen Abschnitt 20.4 zu Volume-Outcome-Analysen wird zudem auf die Berücksichtigung von Unsicherheit eingegangen, ebenso in den Anhängen C und D. Das IQTIG wird für kommende Überarbeitungen prüfen, wie dieses Thema noch weiter in den Vordergrund gerückt werden kann. Eine propädeutische Darstellung von allgemeinen Methoden zur Darstellung von Unsicherheit ist nicht Ziel der „Methodischen Grundlagen“ des IQTIG.

Von einer stellungnehmenden Organisation wurde der Wunsch nach einer Beschreibung der Auswertungsmethodik bei qualitativ erhobenen Daten geäußert (KBV, S. 8).

**IQTIG:** Kapitel 20 beschäftigt sich mit der quantitativen Datenanalyse. Da der Begriff „Daten“ im Kontext der Qualitätssicherung üblicherweise quantitativ verstanden wird (vgl. die Abkürzung DeQS-RL), sieht das IQTIG hier wenig Verwechslungspotenzial bei der Kapitelüberschrift. Aus dem Bereich der qualitativen Methoden werden am IQTIG vorwiegend Fokusgruppen und Interviews eingesetzt. Diese werden in den „Methodischen Grundlagen“ bereits in anderen Kapiteln beschrieben.

### Rahmenkonzept für die statistische Auswertungsmethodik

Von einer stellungnehmenden Organisation wurde um die Überprüfung des Begriffs „Fall“ gebeten (KBV, S. 8).

**IQTIG:** Die in den „Methodischen Grundlagen“ seit Version 2.0 verwendete Definition ist die in der Qualitätssicherung üblicherweise verwendete. Diese Definition wurde bereits von den Vorgängerinstitutionen des IQTIG verwendet. Sollte es konkrete Hinweise auf häufige Missverständnisse geben, prüft das IQTIG gerne seine Produkte auf Anpassungsbedarf.

Von einer stellungnehmenden Organisation wurde gefragt, ob die Fallzahlabhängigkeit von Ergebnissen regelhaft berücksichtigt werde (KBV, S. 8).

**IQTIG:** Aus der Stellungnahme wird nicht deutlich, was mit der Fallzahlabhängigkeit genau gemeint ist. Das IQTIG interpretiert die Frage so, dass sich die stellungnehmende Organisation dafür interessiert, ob die fallzahlabhängige statistische Unsicherheit der Ergebnisse berücksichtigt wird.

In Abschnitt 20.1.4 zum Rahmenkonzept für die statistische Auswertungsmethodik wird beschrieben, wann statistische Unsicherheit generell berücksichtigt werden soll. Dies bezieht sich auf alle statistischen Analysen des IQTIG.

In Abschnitt 13.6.2 beschreibt das IQTIG die Rolle statistischer Unsicherheit für die Klassifikationsgüte als Eignungskriterium des Bewertungskonzepts.

Bei der rechnerischen Auffälligkeitseinstufung wird aktuell in den Modulen *PCI* und *PPCI* des Verfahrens *QS PCI* statistische Unsicherheit berücksichtigt. Das IQTIG plant, zukünftig bei sämtlichen QS-Verfahren statistische Unsicherheit bei der Ermittlung der Auffälligkeit zu berücksichtigen (siehe dazu auch IQTIG [kein Datum]).

Von zwei stellungnehmenden Organisationen wurde der Begriff „Kompetenzparameter“ als missverständlich (KBV, S. 8) bzw. als unklar (KZBV, S. 13) kritisiert. Das IQTIG wurde um eine Abgrenzung gegenüber dem Begriff „Qualitätsmerkmal“ gebeten.

**IQTIG:** Ein *Qualitätsmerkmal* ist eine Eigenschaft der Versorgung (z. B. Erlangen von Gehfähigkeit, Information über alternative Behandlungsmöglichkeiten), die mit Anforderungen (z. B. ein möglichst hoher Anteil gehfähiger Patientinnen und Patienten nach Operation) verknüpft ist. Demgegenüber handelt es sich bei der *Kompetenz* um eine

Abstraktion von Eigenschaften eines Leistungserbringers, die in Hinblick auf ein gegebenes Qualitätsmerkmal von Bedeutung sind. Der Begriff *Kompetenzparameter* wiederum beschreibt einen numerischen Parameter, in der Regel innerhalb eines statistischen Modells, der diese Kompetenz operationalisiert. Das IQTIG prüft, wie diese Zusammenhänge in zukünftigen Versionen der „Methodischen Grundlagen“ klarer dargestellt werden können.

### Risikoadjustierung

Von zwei stellungnehmenden Organisationen wurden Hinweise auf die Datenerhebung sowie auf Probleme bei der Datenqualität in Bezug auf mögliche Risikofaktoren vermisst (GMDS/IBS-DR, S. 3; DGHWi, S. 5). Laut einer Organisation sei es bei der Erhebung wesentlich, ob ein Risikofaktor Einfluss auf die Behandlung hat (DGHWi, S. 5). Zudem wurde von derselben Organisation auf die Bedeutung der Zusammensetzung der Expertengremien bei der Einschätzung, welche Risikofaktoren von Bedeutung sind, hingewiesen.

**IQTIG:** Die Erhebung und Validierung von Daten für die Risikoadjustierung unterscheidet sich nicht wesentlich von der Erhebung und Validierung anderer Daten, die für die Berechnung von Qualitätsindikatoren benötigt werden. Die Bewertung der Datenqualität sowie auch der Objektivität, Reliabilität und Validität der Erhebung sind wichtige Schritte bei der Eignungsprüfung von Qualitätsindikatoren. Die vom IQTIG verwendete Methodik wird in Abschnitt 13.5 der „Methodischen Grundlagen“ beschrieben.

Bezüglich der Zusammensetzung von Expertengremien verweist das IQTIG auf Kapitel 11 der „Methodischen Grundlagen“.

Von einer stellungnehmenden Organisation wurde eine Diskussion vermisst, wie geprüft wird, ob eine vorhandene Risikoadjustierung in der Lage ist, Unterschiede zwischen Patientenkollektiven angemessen auszugleichen (GMDS/IBS-DR, S. 3). Von einer anderen stellungnehmenden Organisation wurde auf generelle Bedenken in der Literatur bezüglich der Angemessenheit der Risikoadjustierungsmodelle der externen Qualitätssicherung verwiesen (DNVF, S. 8).

**IQTIG:** Die Angemessenheit der Risikoadjustierung ist ein wichtiges Eignungskriterium von Qualitätsindikatoren. Die Güte der Risikoadjustierung beurteilt das IQTIG nach der in Kapitel 13 beschriebenen Methodik. Für die Beurteilung der Angemessenheit der Risikoadjustierung sind zwei Aspekte wichtig: 1. Liegen in den Daten genügend Informationen über die Unterschiede zwischen Patientenkollektiven vor, d. h., sind alle relevanten Risikofaktoren erfasst? Dies wird in der Regel anhand von Hinweisen aus der Literatur oder von Expertinnen und Experten geprüft. Außerdem fließen die Erfahrungen aus den QS-Verfahren ein. 2. Berücksichtigt das Risikoadjustierungsmodell die vorliegenden Informationen in angemessener Weise? Diese Frage wird derzeit in den „Methodischen Grundlagen“ nicht ausführlich behandelt. Das IQTIG erwägt, hier in künftigen Versionen ausführlicher zu werden.

Das IQTIG widerspricht deutlich der Einschätzung aus der zitierten Arbeit (Vorbeck et al. 2022) und der darin geäußerten Generalkritik an der in der externen Qualitätssicherung verwendeten Risikoadjustierung. Das IQTIG verweist auf seinen Leserbrief zu dieser Arbeit, in welchem methodische Mängel der Arbeit dargelegt werden. Insbesondere wurde in der zitierten Arbeit die Güte der Risikoadjustierung gar nicht untersucht. Stattdessen schlussfolgern Vorbeck et al. indirekt, aufgrund unerwarteter Ergebnisse, dass diese durch eine ungenügende Risikoadjustierung zustande gekommen sein müssen. In seinem Leserbrief dazu weist das IQTIG auf eine statistische Erklärung für die für Vorbeck et al. unerwarteten Ergebnisse hin: Die Ergebnisse beruhen zumindest teilweise auf einem statistischen Artefakt der Analysen von Vorbeck et al. Eine substantielle Kritik an der Güte der Risikoadjustierung würde erfordern zu zeigen, dass bekannte und relevante patientenseitige Einflüsse in den Risikoadjustierungsmodellen nicht oder nicht korrekt berücksichtigt sind.

Von einer stellungnehmenden Organisation wurde ein Kriterium vermisst, wann welche Methodik der Risikoadjustierung zur Anwendung kommt (DNVF, S. 8).

**IQTIG:** Bei den allermeisten Indikatoren führt das IQTIG die Risikoadjustierung mittels statistischer Modellierung nach Abschnitt 20.2.3 durch, weil dadurch eine große Zahl von verschiedenen Risikofaktoren flexibel berücksichtigt werden kann. Das IQTIG plant, dies in künftigen Versionen der „Methodischen Grundlagen“ klarer darzustellen.

Von zwei stellungnehmenden Organisationen wurde eine genauere Beschreibung der Methodik zur Modellentwicklung gewünscht (GMDS/IBS-DR, S. 3; DNVF, S. 8). Außerdem sollen neuere Entwicklungen in der statistischen Modellierung berücksichtigt werden, insbesondere Propensity Score Matching (PSM) (GMDS/IBS-DR, S. 3).

**IQTIG:** Eine genauere Darstellung der inhaltlichen und statistischen Überlegungen bei der Selektion der Variablen und bei der Modellierung des Einflusses stetiger Risikofaktoren ist wichtig und soll daher mit einer zukünftigen Version der „Methodischen Grundlagen“ erfolgen.

Das IQTIG beobachtet neuere Entwicklungen in der statistischen Modellierung und passt daraufhin seine Methodik an. Beispielsweise ist das IQTIG dazu übergegangen, bei der Risikoadjustierung stetige Variablen nicht mehr zu diskretisieren sowie die Clusterstruktur der Daten mittels zufälliger Effekte zu berücksichtigen. Diese Fortschritte wurden jeweils mit den Rechenregeln veröffentlicht, finden sich aber noch nicht in den „Methodischen Grundlagen“. Bei diesen Beispielen handelt es sich um generelle statistische Prinzipien, die nicht nur im Kontext der Risikoadjustierung eine Rolle spielen, sondern auch bei anderen Analysen, wie beispielsweise Volume-Outcome-Analysen. Im Abschnitt dazu werden Clusterung und stetige Effekte bereits erwähnt. Das IQTIG wird in zukünftigen Überarbeitung der „Methodischen Grundlagen“ diese übergreifenden statistischen Prinzipien zentraler darstellen.



Aus Sicht des IQTIG ist die Adjustierung für patientenseitige Einflussvariablen durch multivariable Regression auf das Outcome weiterhin sehr gut für die Risikoadjustierung geeignet. Als Standardisierungsmethode verwendet das IQTIG indirekte Standardisierung. Herkömmlich werden propensity-score-basierte Methoden in Beobachtungsstudien verwendet, um Behandlungs- und Kontrollgruppe vergleichbarer zu machen. Mittlerweile gibt es Ansätze, diese Methodik auf mehr als zwei Gruppen zu erweitern. Die gleichzeitige Standardisierung der Ergebnisse einer Vielzahl an Leistungserbringern ist jedoch noch keine Standardanwendung von Propensity Scores. Zudem entsprechen gängige propensity-score-basierte Methoden konzeptuell eher einer direkten Standardisierung, da versucht wird, verschiedene Populationen durch Gewichtung oder Matching einander ähnlich zu machen.

Von einer stellungnehmenden Organisation wurde die Darstellung der Bewertung der Kalibrierung von Risikoadjustierungsmodellen gefordert (GMDS/IBS-DR, S. 3).

**IQTIG:** Das IQTIG bedankt sich für die Anregung. Derzeit wird die Prüfung der statistischen Eigenschaften, wie die Kalibrierung, von Regressionsmodellen, wie Risikoadjustierungsmodellen, in den „Methodischen Grundlagen“ nicht thematisiert. Dieses Thema ist jedoch eine mögliche Ergänzung für künftige Versionen.

Von zwei stellungnehmenden Organisationen wurde die Komplexität der Konstruktion von additiven Scores betont (DGIM, S. 4; DGHWi, S. 6). Von einer Organisation wurde zudem nach der Methodik für additive Scores gefragt (DGIM, S. 4).

**IQTIG:** Additive Scores spielen bei der Risikoadjustierung am IQTIG eine untergeordnete Rolle. Wie in Abschnitt 20.2.2 erläutert, werden Scores in der Regel aus der Literatur entnommen und nicht selbst entwickelt. Als Beispiel wird im Abschnitt 20.2.2 auf den CRB-65-Score verwiesen, der im Verfahren *Ambulant erworbene Pneumonie (QS CAP)* verwendet wird.

### Auswertung von Follow-up-Indikatoren

Von einer stellungnehmenden Organisation wurden die Kriterien vermisst, nach denen verschiedene Verfahren der Follow-up-Analyse bei verschiedenen Indikatoren zur Anwendung kommen (DNVF, S. 8).

**IQTIG:** Das IQTIG bedankt sich für den Hinweis. In der Tat fehlt diese Information derzeit in den „Methodischen Grundlagen“. Das IQTIG prüft, ob diese Informationen in einer zukünftigen Überarbeitung der „Methodischen Grundlagen“ Eingang finden können.

### Volume-Outcome-Analysen

In einigen Stellungnahmen wurde die Ergänzung des Abschnitts zu Volume-Outcome-Analysen begrüßt (EbM-Netzwerk, S. 1.; GMDS/IBS-DR, S. 3; AWMF, S. 4). Zwei stellungnehmende Organisationen regten an, im Kontext der Empfehlung von Mindestmengen

die Ermittlungsmethoden von Bruchpunkten aus Volume-Outcome-Analysen genauer zu diskutieren (GMDS/IBS-DR, S. 3; GKV-SV, S. 18). Eine Organisation wünschte sich in diesem Zusammenhang die Auseinandersetzungen mit neuerer Literatur (GKV-SV, S. 18).

**IQTIG:** Das IQTIG bedankt sich für den Hinweis und wird in einer zukünftigen Version der „Methodischen Grundlagen“ entsprechende Ergänzungen vornehmen. Der Fokus dieser ersten Version des Abschnitts zu Volume-Outcome-Analysen liegt auf der Modellierung des Zusammenhangs selbst. Bruchpunktanalysen und zugehörige methodische Literatur werden dementsprechend nicht behandelt. Der Absatz zu regulatorischen Maßnahmen wurde trotzdem bereits dahingehend ergänzt, dass zur Ableitung von Mindestmengen geschätzte Bruchpunkte im Volume-Outcome-Zusammenhang in umfassendere Kosten-Nutzen-Abwägungen eingebunden werden sollten.

Eine stellungnehmende Organisation verwies auf die Möglichkeit, quasi-experimentelle Studiendesigns, wie das IQTIG sie bereits in Kapitel 9 der „Methodischen Grundlagen“ darlegt, auch in Volume-Outcome-Analysen zu nutzen (dggö, S. 2).

**IQTIG:** Das IQTIG bedankt sich für den Vorschlag und hat den Abschnitt 20.4.2 um eine entsprechende Fußnote ergänzt. Die Nutzung solcher Designs sieht das IQTIG generell nicht als Ersatz, sondern als potenzielle Erweiterung der dargelegten Regressionsanalysen an.

Eine stellungnehmende Organisation merkte an, dass Strukturmerkmale abgefragt werden müssten (DGIM, S. 4). Sie sprach sich dafür aus, den Anteil des Personals zu erheben, welcher ausreichend Prozeduren durchgeführt habe.

**IQTIG:** Aus der Stellungnahme wird weder deutlich, worauf sich die Forderung, Strukturmerkmale abzufragen, genau bezieht, noch, wie eine ausreichende Anzahl an Prozeduren zu definieren sei. Das IQTIG interpretiert die Stellungnahme so, dass sich die DGIM für personenbezogene Volume-Outcome-Analysen ausspricht. Das Ergebnis dieser Volume-Outcome-Analysen könnte dann Anhaltspunkte dafür liefern, was eine ausreichende Anzahl an Prozeduren pro Person sein kann. Darauf basierend ließe sich erheben, welcher Anteil des Personals ausreichend Prozeduren durchgeführt hat. Die Feststellung, dass nicht nur standortbezogene Fallzahlen in Volume-Outcome-Analysen betrachtet werden können, deckt sich mit der Aussage in Abschnitt 20.4.1, in dem das IQTIG bereits knapp auf den Zusammenhang von Volume und Outcome auf Arzzebene eingeht. Da dem IQTIG keine arztbezogenen Daten vorliegen, hat es sich bislang nicht näher mit Volume-Outcome-Analysen auf Arzzebene auseinandergesetzt. Die in den „Methodischen Grundlagen“ skizzierten zugrunde liegenden Annahmen und statistischen Grundsätze von Volume-Outcome-Analysen hängen nicht davon ab, ob standortbezogene oder arztbezogene Analysen durchgeführt werden. Ob eine arztbezogene

oder eine standortbezogene Analyse sinnvoller ist, hängt letztlich von der betrachteten Prozedur und der Fragestellung ab.

Eine stellungnehmende Organisation vermisste eine Diskussion von Volume-Outcome-Analysen bei Patientenbefragungen (KBV, S. 8). Sie stellte zudem fest, dass sich das Kapitel nur auf die stationäre Leistungserbringung konzentrierte.

**IQTIG:** Das IQTIG bedankt sich für den Hinweis und stimmt zu, dass Patientenbefragungen ebenso als Datenquelle für Volume-Outcome-Analysen infrage kommen. Dies wurde im einleitenden Teil des Abschnitts zur *Stichprobenart* bei Volume-Outcome-Analysen nun ergänzt. Bisher spielten Patientenbefragungen als Datenquelle für Volume-Outcome-Analysen des IQTIG keine Rolle. Dies kann sich jedoch zukünftig ändern. Der Abschnitt zu Volume-Outcome-Analysen stellt die generellen statistischen Grundsätze dar und leitet diese her. Diese Grundsätze sind unabhängig von der Datenquelle und gelten somit auch für Volume-Outcome-Analysen basierend auf Patientenbefragungsdaten. Gleiches gilt für Leistungserbringungen aus dem ambulanten Bereich. Der Abschnitt zu Volume-Outcome-Analysen ist nicht auf die stationäre Leistungserbringung beschränkt.

Zwei stellungnehmende Organisationen gaben Rückmeldungen zum Aspekt der Risikoadjustierung als Teil der Volume-Outcome-Analyse. Dabei wurde angemerkt, dass die Methodik zur Adjustierung ausführlicher dargestellt werden sollte (DKG, S. 9; GMDS/IBS-DR, S. 3) und dass die Datenverfügbarkeit zu diskutieren sei (GMDS/IBS-DR, S. 3).

**IQTIG:** Das IQTIG bedankt sich für die Hinweise. Die Adjustierung für patientenseitige Merkmale spielt in der Tat eine wichtige Rolle in Volume-Outcome-Analysen und wird daher im Kapitel 20 mehrfach erwähnt. Dabei unterscheiden sich die inhaltlichen und methodischen Abwägungen allerdings nicht grundlegend von denen, die für die Risikoadjustierung im Zuge der externen Qualitätssicherung durch das IQTIG getroffen werden. Diese werden bereits in Abschnitt 20.2 ausführlich beschrieben, einschließlich Aspekten wie Auswahl von Einflussvariablen, Messung, Datenerfassung und Limitationen. Ein entsprechender Verweis auf den vorherigen Abschnitt befindet sich bereits in Abschnitt 20.4.1 und wird vom IQTIG derzeit als ausreichend erachtet. In Berichten zu Volume-Outcome-Analysen geht das IQTIG deshalb auf die Angemessenheit der Risikoadjustierung im jeweils betrachteten klinischen Kontext genauer ein. Um die üblichen Datengrundlagen für Volume-Outcome-Analysen, und damit auch für die Risikoadjustierung, deutlicher darzulegen, wurde zudem ergänzt, dass die Vollerhebungen i. d. R. „auf QS-Daten und/oder Sozialdaten bei den Krankenkassen“ beruhen.

Eine stellungnehmende Organisation verwies auf die Arbeiten zum Thema Mindestmengen-Regelungen der Abteilung „Fachberatung Medizin“ der G-BA-Geschäftsstelle (KZBV, S. 13).

**IQTIG:** Das IQTIG bedankt sich für den Hinweis. Der Beschluss des G-BA über eine Beauftragung des IQTIG mit der Analyse von Volume-Outcome-Zusammenhängen enthielt bereits Verweise auf etwaige Vorarbeiten der Abteilung „Fachberatung Medizin“ der G-BA-Geschäftsstelle (G-BA 2019). In diesem Fall setzte sich das IQTIG bei der Bearbeitung des Auftrags mit den Erkenntnissen der Abteilung „Fachberatung Medizin“ auseinander. Unabhängig davon befasst sich das IQTIG bei jeder neuen Beauftragung mit der entsprechenden Literatur.

Eine stellungnehmende Organisation wies darauf hin, dass die Beschreibung der Annahmen der funktionalen Form des modellierten Zusammenhangs in Abschnitt „Fallzahl als stetige Größe mit stetigem Effekt“ missverständlich sein könnte (GKV-SV, S. 18).

**IQTIG:** Das IQTIG stimmt zu, dass ein Fallzahleffekt nicht vorab als linear angenommen werden sollte. Ein nicht linearer Fallzahleffekt ist ein Spezialfall eines stetigen Fallzahleffekts, welcher vom IQTIG im angesprochenen Textabschnitt bereits als zentrale Annahme beschrieben wird. Ein stetiger Fallzahleffekt steht im Gegensatz zu „springenden“ Fallzahleffekten. Solche eher unplausiblen Effektverläufe würden resultieren, wenn die Fallzahlvariable anhand von Kategoriengrenzen diskretisiert würde. Dies ist im vorliegenden Textabschnitt bereits erläutert. Um trotzdem noch klarer zu werden, hat das IQTIG nun im angesprochenen Textabschnitt ergänzt, dass seine flexiblen Modellierungsansätze „auch komplexe, nicht lineare Zusammenhänge abbilden können“.

Eine stellungnehmende Organisation kritisierte die Formulierung, dass die Fallzahl eines Leistungserbringers nur indirekt seine Behandlungsergebnisse beeinflusse (GKV-SV, S. 17). Sie bestritt, dass es sich bei der Fallzahl um ein Surrogat handle. Es wurde betont, dass mit einer Mindestmenge einzig die Erfahrung des Personals, nicht aber die technische Ausstattung des Standortes adressiert würde, und eine Klarstellung gefordert, dass Mindestmengenregelungen nicht durch Strukturvorgaben ersetzt werden könnten, wobei sich dabei auf ein Urteil des Bundessozialgerichts berufen wurde.

**IQTIG:** Der missverständliche Satz zum indirekten Einfluss wurde gestrichen. Der Begriff „indirekt“ sollte zum Ausdruck bringen, dass die Fallzahl selbst keinen Einfluss auf die Behandlungsergebnisse hat. Vielmehr bewirkt eine höhere Fallzahl, dass das Personal über mehr Erfahrung und Routine verfügt. Diese Erfahrung und Routine ließen sich prinzipiell auch anders messen als durch die Fallzahl. Daher handelt es sich bei der Fallzahl um ein Surrogat.

Während eine höhere Fallzahl i. d. R. unmittelbar dazu führt, dass das angestellte Personal über mehr Erfahrung verfügt, sind andere Konsequenzen (z. B. Investitionen in technische oder personelle Ausstattung) eher mittelbar und unterliegen der Entscheidung des Leistungserbringers. Sie können auch unabhängig von der Fallzahl erfolgen. Um dies klarer zu machen, hat das IQTIG eine Ergänzung in Abschnitt 20.4.1 vorgenommen.

Das IQTIG weist darauf hin, dass der durch eine Volume-Outcome-Analyse geschätzte Effekt der Fallzahl nur dann den *reinen* Effekt der Erfahrung und Routine beinhalten kann, wenn für alle anderen Variablen, die mit der Fallzahl und der Behandlungsqualität potenziell korrelieren, im Modell korrekt adjustiert würde (z. B. Strukturmerkmale). Da dies in der Praxis nie der Fall sein wird, kann der Fallzahleffekt nicht als reiner Effekt der Erfahrung des Personals interpretiert werden, auch wenn dies wünschenswert wäre. Vor dem Hintergrund des erwähnten Urteils des Bundessozialgerichts weist das IQTIG darauf hin, dass zu unterscheiden ist zwischen (a) der gesetzlichen Feststellung, dass eine Maßnahme nicht durch eine andere ersetzt werden kann, und (b) der Tatsache, dass in einer empirischen Untersuchung die in den Daten vorliegenden gemessenen Größen u. U. nicht ausreichen, um bei der Interpretation eindeutige Kausalschlüsse zu ziehen. Während es sich bei (a) um eine Rechtsnorm handelt, handelt es sich bei (b) um eine methodische Einschränkung, die das IQTIG transparent darlegt.

## Literatur

- AWMF [Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften],  
Ständige Kommission Leitlinien (2023): Das AWMF-Regelwerk Leitlinien. Version 2.1. [Stand:]  
05.09.2023. Marburg: AWMF. URL: [https://www.awmf.org/fileadmin/user\\_upload/dateien/downloads\\_regelwerk/20230905\\_AWMF-Regelwerk\\_2023\\_V2.1\\_final.pdf](https://www.awmf.org/fileadmin/user_upload/dateien/downloads_regelwerk/20230905_AWMF-Regelwerk_2023_V2.1_final.pdf) (abgerufen am:  
28.06.2024).
- Bader, H; Abdulelah, M; Maghnam, R; Chin, D (2021): Clinical peer Review; A mandatory process  
with potential inherent bias in desperate need of reform. *Journal of Community Hospital  
Internal Medicine Perspectives* 11(6): 817-820. DOI: 10.1080/20009666.2021.1965704.
- BÄK [Bundesärztekammer]; Hrsg. (2013): Curriculum. Ärztliches Peer Review. 2. Auflage. (Texte  
und Materialien der Bundesärztekammer zur Fortbildung und Weiterbildung, Band 30).  
Berlin: BÄK. URL: [http://www.bundesaerztekammer.de/fileadmin/user\\_upload/downloads/  
CurrAerztlPeerReview2013.pdf](http://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/CurrAerztlPeerReview2013.pdf) (abgerufen am: 03.02.2023).
- Bernal, JL; Cummins, S; Gasparrini, A (2017): Interrupted time series regression for the evaluation  
of public health interventions: a tutorial. *International Journal of Epidemiology* 46(1): 348-355.  
DOI: 10.1093/ije/dyw098.
- Bion, J; Alderman, JE (2023): Peer review of quality of care: methods and metrics. *BMJ: Quality &  
Safety* 32: 1-5. DOI: 10.1136/bmjqs-2022-014985.
- Boyd, A; Addicott, R; Robertson, R; Ross, S; Walshe, K (2017): Are inspectors' assessments  
reliable? Ratings of NHS acute hospital trust services in England. *Journal of Health Services  
Research & Policy* 22(1): 28-36. DOI: 10.1177/1355819616669736.
- DeGEval [Gesellschaft für Evaluation] (2016): Standards für Evaluation. Erste Revision auf Basis  
der Fassung 2002. Langfassung. [Stand:] 21.09.2016. [Mainz]: DeGEval. URL: [https://www.  
degeval.org/fileadmin/Publikationen/DeGEval-Standards\\_fuer\\_Evaluation.pdf](https://www.degeval.org/fileadmin/Publikationen/DeGEval-Standards_fuer_Evaluation.pdf) (abgerufen  
am: 09.10.2023).
- Döring, N; Bortz, J (2016): Forschungsmethoden und Evaluation in den Sozial- und  
Humanwissenschaften. 5. vollständig überarbeitete, aktualisierte und erweiterte Auflage.  
Berlin: Springer. ISBN: 978-3-642-41088-8.
- Flodgren, G; Gonçalves-Bradley, DC; Pomey, MP (2016): External inspection of compliance with  
standards for improved healthcare outcomes (Review). *Cochrane Database of Systematic  
Reviews* 12: CD008992. DOI: 10.1002/14651858.CD008992.pub3.

- G-BA [Gemeinsamer Bundesausschuss] (2019): Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des IQTIG mit der Auswertung von esQS-Daten hinsichtlich Volume-Outcome-Beziehungen bei Revisionseingriffen bei Knie-Endoprothesen. [Stand:] 16.05.2019. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/39-261-3790/2019-05-16\\_IQTIG-Beauftragung\\_Auswertung-esQS-Daten-Knie-Endoprothesen.pdf](https://www.g-ba.de/downloads/39-261-3790/2019-05-16_IQTIG-Beauftragung_Auswertung-esQS-Daten-Knie-Endoprothesen.pdf) (abgerufen am: 18.09.2024).
- G-BA [Gemeinsamer Bundesausschuss] (2022a): Beschluss des Gemeinsamen Bundesausschusses über Eckpunkte zur Weiterentwicklung der datengestützten gesetzlichen Qualitätssicherung. [Stand:] 21.04.2022. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/39-261-5386/2022-04-21\\_DeQS-RL\\_Eckpunkte\\_Weiterentwicklung.pdf](https://www.g-ba.de/downloads/39-261-5386/2022-04-21_DeQS-RL_Eckpunkte_Weiterentwicklung.pdf) (abgerufen am: 02.12.2022).
- G-BA [Gemeinsamer Bundesausschuss] (2022b): Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des IQTIG mit der Prüfung und Entwicklung von Empfehlungen zur Weiterentwicklung von Verfahren der datengestützten gesetzlichen Qualitätssicherung. [Stand:] 19.05.2022. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/39-261-5456/2022-05-19\\_IQTIG-Beauftragung\\_Empfehlungen-Weiterentwicklung-DeQS.pdf](https://www.g-ba.de/downloads/39-261-5456/2022-05-19_IQTIG-Beauftragung_Empfehlungen-Weiterentwicklung-DeQS.pdf) (abgerufen am: 04.09.2024).
- G-BA [Gemeinsamer Bundesausschuss] (2023a): Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des IQTIG mit der Entwicklung und Anwendung eines Konzepts zur kontinuierlichen und systematischen Identifizierung von relevanten Qualitätsdefiziten und Verbesserungspotenzialen. [Stand:] 12.05.2023. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/39-261-6017/2023-05-12\\_IQTIG-Beauftragung\\_Identifizierung-Qualitätsdefizite-Verbesserungspotenziale\\_DeQS-RL.pdf](https://www.g-ba.de/downloads/39-261-6017/2023-05-12_IQTIG-Beauftragung_Identifizierung-Qualitätsdefizite-Verbesserungspotenziale_DeQS-RL.pdf) (abgerufen am: 18.09.2024).
- G-BA [Gemeinsamer Bundesausschuss] (2023b): Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des IQTIG mit der Überarbeitung aller bestehenden Verfahren der datengestützten Qualitätssicherung auf Basis der Ergebnisse aus den drei Modellverfahren. [Stand:] 19.01.2023. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/39-261-5841/2023-01-19\\_IQTIG-Beauftragung\\_Ueberarbeitung-aller-QS-Verfahren-DeQS-RL.pdf](https://www.g-ba.de/downloads/39-261-5841/2023-01-19_IQTIG-Beauftragung_Ueberarbeitung-aller-QS-Verfahren-DeQS-RL.pdf) (abgerufen am: 04.09.2024).
- G-BA [Gemeinsamer Bundesausschuss] (2024a): Beschluss des Gemeinsamen Bundesausschusses über die Freigabe des Abschlussberichts der Medizinischen Hochschule Brandenburg CAMPUS GmbH und der revFLect GmbH zur Evaluation der Richtlinie zum Zweitmeinungsverfahren zur Veröffentlichung. [Stand:] 15.02.2024. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/39-261-6487/2024-02-15\\_Zm-RL\\_MHB\\_Evaluationsbericht-2024.pdf](https://www.g-ba.de/downloads/39-261-6487/2024-02-15_Zm-RL_MHB_Evaluationsbericht-2024.pdf) (abgerufen am: 28.08.2024).

- G-BA [Gemeinsamer Bundesausschuss] (2024b): Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des IQTIG mit der Weiterentwicklung des Verfahrens der qualitativen Beurteilung. [Stand:] 06.03.2024. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/39-261-6502/2024-03-06\\_IQTIG-Beauftragung\\_Weiterentwicklung-Verfahren-qual-Beurteilung\\_DeQS-RL.pdf](https://www.g-ba.de/downloads/39-261-6502/2024-03-06_IQTIG-Beauftragung_Weiterentwicklung-Verfahren-qual-Beurteilung_DeQS-RL.pdf) (abgerufen am: 24.05.2024).
- Gigerenzer, G; Edwards, A (2003): Simple tools for understanding risks: from innumeracy to insight. *BMJ* 327(7417): 741-744. DOI: 10.1136/bmj.327.7417.741.
- Grol, R (1994): Quality improvement by peer review in primary care: a practical guide. *Quality in Health Care* 3(3): 147-152. DOI: 10.1136/qshc.3.3.147.
- Hategeka, C; Ruton, H; Karamouzian, M; Lynd, LD; Law, MR (2020): Use of interrupted time series methods in the evaluation of health system quality improvement interventions: a methodological systematic review. *BMJ: Global Health* 5(10). DOI: 10.1136/bmjgh-2020-003567.
- Heller, RF; Dobson, AJ (2000): Disease impact number and population impact number: population perspectives to measures of risk and benefit. *BMJ* 321(7266): 950-953. DOI: 10.1136/bmj.321.7266.950.
- Hut-Mossel, L; Ahaus, K; Welker, G; Gans, R (2021): Understanding how and why audits work in improving the quality of hospital care: A systematic realist review. *PLoS One* 16(3): e0248677. DOI: 10.1371/journal.pone.0248677.
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2020): Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern. Abschlussbericht zu Stufe 1 und Stufe 2. Stand: 31.01.2020. Berlin: IQTIG. [unveröffentlicht].
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2022): Kriterien für die Aussetzung und Aufhebung von Qualitätsindikatoren oder QS-Verfahren. Methodisches Vorgehen des IQTIG bei der Entwicklung von Empfehlungen an den G-BA. Abschlussbericht. Stand: 14.07.2022. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2022/IQTIG\\_Aussetzungskriterien-QI-Abschlussbericht\\_2022-07-14-1.pdf](https://iqtig.org/downloads/berichte/2022/IQTIG_Aussetzungskriterien-QI-Abschlussbericht_2022-07-14-1.pdf) (abgerufen am: 04.09.2024).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2023): Empfehlungen zur Weiterentwicklung von Verfahren der datengestützten gesetzlichen Qualitätssicherung. Indikatorensets der Verfahren *QS PCI*, *QS HSMDEF* und *QS KEP*. Abschlussbericht. [Stand:] 19.07.2023. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2023/IQTIG\\_Weiterentwicklung-datengestuetzte-QS\\_Verfahren-PCI-HSMDEF-KEP\\_2023-07-19-barrierefrei.pdf](https://iqtig.org/downloads/berichte/2023/IQTIG_Weiterentwicklung-datengestuetzte-QS_Verfahren-PCI-HSMDEF-KEP_2023-07-19-barrierefrei.pdf) (abgerufen am: 28.05.2024).



- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] ([kein Datum]): Biometrische Methodik zur Auffälligkeitseinstufung. Berlin: IQTIG. URL: <https://iqtig.org/das-iqtig/wie-wir-arbeiten/grundlagen/biometrische-grundlagen/biometrische-methodik-zur-auffaelligkeitseinstufung/> (abgerufen am: 18.09.2024).
- IQWiG [Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen] (2023): Allgemeine Methoden. Version 7.0. [Stand:] 19.09.2023. Köln: IQWiG. ISBN: 978-3-9815265-5-4. URL: [https://www.iqwig.de/methoden/allgemeine-methoden\\_version-7-0.pdf](https://www.iqwig.de/methoden/allgemeine-methoden_version-7-0.pdf) (abgerufen am: 18.07.2024).
- Klein, DO; Rennenberg, RJMW; Koopmans, RP; Prins, MH (2018): Adverse event detection by medical record review is reproducible, but the assessment of their preventability is not. *PLoS One* 13(11): e0208087. DOI: 10.1371/journal.pone.0208087.
- Lilford, R; Edwards, A; Girling, A; Hofer, T; Di Tanna, GL; Petty, J; et al. (2007): Inter-rater reliability of case-note audit: a systematic review. *Journal of Health Services Research & Policy* 12(3): 173-180. DOI: 10.1258/135581907781543012.
- Maats, P (2007): Einführung in das Datenmanagement und die Datenauswertung. Kapitel 9. In: Stockmann, R: *Handbuch zur Evaluation. Eine praktische Handlungsanleitung.* (Sozialwissenschaftliche Evaluationsforschung, Band 6). Münster [u. a.]: Waxmann, 278-313. ISBN: 978-3-8309-1766-3.
- Marsh, K; M, IJ; Thokala, P; Baltussen, R; Boysen, M; Kaló, Z; et al. (2016): Multiple Criteria Decision Analysis for Health Care Decision Making—Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value in Health* 19(2): 125-137. DOI: 10.1016/j.jval.2015.12.016.
- Meyer, W (2004): Indikatorenentwicklung: Eine praxisorientierte Einführung. 2. Auflage. (CEval-Arbeitspapier, Nr. 10). Saarbrücken: Universität des Saarlandes, Fakultät 05 Empirische Humanwissenschaften, CEval [Centrum für Evaluation]. URL: <http://www.ssoar.info/ssoar/bitstream/handle/document/11124/ssoar-2004-meyer-indikatorenentwicklung.pdf> (abgerufen am: 05.07.2022).
- Meyer, W (2007): Datenerhebung: Befragungen – Beobachtungen – Nicht-reaktive Verfahren. Kapitel 8. In: Stockmann, R: *Handbuch zur Evaluation. Eine praktische Handlungsanleitung.* Münster [u. a.]: Waxmann, 223-277. ISBN: 978-3-8309-1766-3.
- NQF [National Quality Forum] (2015): Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement. [Stand:] April 2015. Washington, US-DC: NQF. URL: <http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=79434> (abgerufen am: 05.07.2022).
- Nyumba, TO; Wilson, K; Derrick, CJ; Mukherjee, N (2018): The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution* 9(1): 20-32. DOI: 10.1111/2041-210X.12860.

- Page, MJ; McKenzie, JE; Bossuyt, PM; Boutron, I; Hoffmann, TC; Mulrow, CD; et al. (2021): The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372: n71. DOI: 10.1136/bmj.n71.
- Schermelleh-Engel, K; Werner, CS (2012): Methoden der Reliabilitätsbestimmung. Kapitel 6. In: Moosbrugger, H; Kelava, A; Hrsg.: *Testtheorie und Fragebogenkonstruktion*. 2., aktualisierte und überarbeitete Auflage. Berlin [u. a.]: Springer, 119-143. ISBN: 3-642-20071-0.
- Scheuren, F (2004): What is a Survey? Alexandria, US-VA: ASA [American Statistical Association]. URL: [https://www.whatisasurvey.info/downloads/pamphlet\\_current.pdf](https://www.whatisasurvey.info/downloads/pamphlet_current.pdf) (abgerufen am: 13.09.2024).
- Schnell, R; Hill, PB; Esser, E (2013): Methoden der empirischen Sozialforschung. 10., überarbeitete Auflage. München: Oldenbourg. ISBN: 978-3-486-72899-6.
- Sens, B; Fischer, B; Bastek, A; Eckardt, J; Kaczmarek, D; Paschen, U; et al. (2007): Begriffe und Konzepte des Qualitätsmanagements – 3. Auflage. *GMS Medizinische Informatik, Biometrie und Epidemiologie* 3(1): Doc05. URL: <http://www.egms.de/tools/download.jsp?path=journals/mibe/2007-3/mibe000053.1a.pdf&mime=application/pdf&name=GMS-QM-Glossar-Auflage3.pdf> (abgerufen am: 13.09.2024).
- Sens, B; Pietsch, B; Fischer, B; Hart, D; Kahla-Witzsch, H; von Friedrichs, V; et al. (2018): Begriffe und Konzepte des Qualitätsmanagements – 4. Auflage. *GMS Medizinische Informatik, Biometrie und Epidemiologie* 14(1). Doc04. DOI: 10.3205/mibe000182.
- Thai, T; Loudon, DKN; Adamson, R; Dominitz, JA; Doll, JA (2022): Peer evaluation and feedback for invasive medical procedures: a systematic review. *BMC: Medical Education* 22:581. DOI: 10.1186/s12909-022-03652-9.
- Veit, C; Lüken, F; Bungard, S; Trümner, A; Tewes, C; Hertle, D (2013): Rahmenkonzept Evaluation bezogen auf Evaluationen nach § 137b SGB V. Version 1.1, Entwurf vom 17.07.2013. Düsseldorf: BQS [Institut für Qualität und Patientensicherheit]. [unveröffentlicht, auf Nachfrage vom G-BA zugesandt].
- Vorbeck, L; Naumoska, D; Geraedts, M (2022): Assoziation von Strukturvariablen mit der Versorgungsqualität der Krankenhäuser in Deutschland. *Das Gesundheitswesen* 84(3): 242-249. DOI: 10.1055/a-1341-1246.
- Weissman, NW; Allison, JJ; Kiefe, CI; Farmer, RM; Weaver, MT; Williams, OD; et al. (1999): Achievable benchmarks of care: the ABC<sup>TM</sup>s of benchmarking. *Journal of Evaluation in Clinical Practice* 5(3): 269-281. DOI: 10.1046/j.1365-2753.1999.00203.x.
- Widmer, T (2012): Unabhängigkeit in der Evaluation. *LeGes – Gesetzgebung & Evaluation* 23(2): 129-147. URL: [https://leges.weblaw.ch/dam/jcr:942bcb64-bda7-4929-b78b-eae2d8fd8bfc/LeGes\\_2012\\_2\\_129-147.pdf](https://leges.weblaw.ch/dam/jcr:942bcb64-bda7-4929-b78b-eae2d8fd8bfc/LeGes_2012_2_129-147.pdf) (abgerufen am: 28.08.2024).

# Impressum

## HERAUSGEBER

---

IQTIG – Institut für Qualitätssicherung  
und Transparenz im Gesundheitswesen  
Katharina-Heinroth-Ufer 1  
10787 Berlin

Telefon: (030) 58 58 26-0

[info@iqtig.org](mailto:info@iqtig.org)

[iqtig.org](http://iqtig.org)