



Institut für Qualitätssicherung und  
Transparenz im Gesundheitswesen

# Methodische Grundlagen V1.0

Stand: 15. September 2017

---

# Impressum

**Thema:**

Methodische Grundlagen V1.0

**Ansprechpartner:**

Dr. Dennis Boywitt, Dr. Sven Bungard

**Stand:**

15. September 2017

**Zitierhinweis:**

IQTIG (2017): Methodische Grundlagen V1.0. Stand: 15. September 2017. Berlin: Institut für Qualitätssicherung und Transparenz im Gesundheitswesen.

URL: [https://iqtig.org/downloads/berichte/2017/IQTIG\\_Methodische-Grundlagen-V1.0.pdf](https://iqtig.org/downloads/berichte/2017/IQTIG_Methodische-Grundlagen-V1.0.pdf)

**Herausgeber:**

IQTIG – Institut für Qualitätssicherung  
und Transparenz im Gesundheitswesen

Katharina-Heinroth-Ufer 1  
10787 Berlin

Telefon: (030) 58 58 26-0  
Telefax: (030) 58 58 26-999

[methoden@iqtig.org](mailto:methoden@iqtig.org)  
[info@iqtig.org](mailto:info@iqtig.org)  
<https://www.iqtig.org>

# Übersicht

Tabellenverzeichnis.....	8
Abbildungsverzeichnis.....	9
Abkürzungsverzeichnis.....	10
Vorwort.....	12
<b>Teil A: Grundlagen der gesetzlichen Qualitätssicherung und -förderung im Gesundheitswesen.....</b>	<b>13</b>
1 Einleitung.....	14
2 Qualität und Qualitätssicherung im Gesundheitswesen.....	17
3 Elemente der Verfahrensentwicklung.....	36
<b>Teil B: Entwicklung und Durchführung von Qualitätssicherungsverfahren.....</b>	<b>39</b>
4 Neuentwicklung von Qualitätssicherungsverfahren.....	40
5 Durchführung, Pflege und Weiterentwicklung von Qualitätssicherungsverfahren.....	77
<b>Teil C: Methodische Elemente .....</b>	<b>86</b>
6 Grundlagen der systematischen Literatur- und Informationsbeschaffung.....	87
7 Einbeziehung externen Wissens und externer Erfahrung.....	97
8 Eignung von Qualitätsindikatoren.....	109
9 Veröffentlichung von Indikatorergebnissen.....	126
10 Bewertungskonzepte und Referenzbereiche .....	129
11 Grundgesamtheit, Stichprobenziehung und Vollerhebung.....	140
12 Datenvalidierung von QS-Dokumentationsdaten und Strukturierter Dialog.....	145
13 Methoden der Datenanalyse.....	152
Glossar .....	165
Literatur.....	170

# Inhaltsverzeichnis

Tabellenverzeichnis.....	8
Abbildungsverzeichnis.....	9
Abkürzungsverzeichnis.....	10
Vorwort.....	12
<b>Teil A: Grundlagen der gesetzlichen Qualitätssicherung und -förderung im Gesundheitswesen.....</b>	<b>13</b>
1 Einleitung.....	14
2 Qualität und Qualitätssicherung im Gesundheitswesen.....	17
2.1 Qualität der Gesundheitsversorgung.....	17
2.1.1 Qualitätsbegriff in der Gesundheitsversorgung.....	17
2.1.2 Qualitätsdimensionen und Systematiken von Qualitätskriterien.....	17
2.1.3 Patientenorientierung.....	21
2.2 Qualitätssicherung und Qualitätsmanagement.....	22
2.2.1 Historischer Ursprung der externen Qualitätsvergleiche.....	22
2.2.2 Prinzip der externen Vergleiche.....	23
2.2.3 Qualitätssicherung.....	24
2.2.4 Qualitätsmanagement.....	25
2.2.5 Transparenz und externe Intervention.....	26
2.3 Legitime Anforderungen als Grundlage der Qualitätsmessung.....	26
2.4 Verfahrenszwecke der Qualitätssicherung.....	29
2.4.1 Qualitätsförderung.....	29
2.4.2 Information.....	29
2.4.3 Regulierung durch Planung der Versorgungsstrukturen.....	30
2.4.4 Regulierung durch Selektivverträge mit Anreizen.....	31
2.4.5 Regulierung durch Qualitätszu- und -abschläge.....	31
2.4.6 Zusammenfassung: Verfahren mit unterschiedlichen Verfahrenszwecken.....	32
3 Elemente der Verfahrensentwicklung.....	36
3.1 Qualitätsaspekte der Gesundheitsversorgung.....	36
3.2 Qualitätsmodelle als Basis der QS-Verfahren.....	36
3.3 Qualitätsmerkmale als erster Schritt der Konkretisierung.....	37

3.4	Qualitätsindikatoren .....	38
<b>Teil B: Entwicklung und Durchführung von Qualitätssicherungsverfahren.....</b>		<b>39</b>
4	Neuentwicklung von Qualitätssicherungsverfahren .....	40
4.1	Konzeptstudie für Qualitätssicherungsverfahren .....	43
4.1.1	Themenerschließung.....	45
4.1.2	Identifikation von Qualitätsaspekten.....	46
4.1.3	Auswahl von Qualitätsaspekten für das Qualitätsmodell .....	47
4.2	Dokumentations- und sozialdatenbasierte Neuentwicklung von Qualitätssicherungsverfahren.....	48
4.2.1	Entwicklungsschritte .....	48
4.2.2	Entwicklungsergebnisse .....	54
4.3	Machbarkeitsprüfung für Qualitätssicherungsverfahren .....	55
4.3.1	Ziel der Machbarkeitsprüfung.....	55
4.3.2	Teilnehmende Einrichtungen und Datenbasis .....	56
4.3.3	Strukturierte Interviews mit den Leistungserbringern .....	57
4.3.4	Auswertung der Daten .....	57
4.3.5	Ergebnis der Machbarkeitsprüfung, Musterauswertung und Rückmeldeberichte .....	58
4.3.6	Empfehlungen zur Spezifikation.....	59
4.4	Neuentwicklung von Patientenbefragungen im Rahmen von Qualitätssicherungsverfahren .....	60
4.4.1	Generelle Anforderungen an die Patientenbefragungen .....	60
4.4.2	Faktenorientierte Befragungsansätze .....	61
4.4.3	Methodische Schritte der Fragebogenentwicklung.....	63
4.4.4	Messtheoretische Gütekriterien der Patientenbefragung.....	71
4.4.5	Risikoadjustierung von Qualitätsindikatoren der Patientenbefragung .....	75
5	Durchführung, Pflege und Weiterentwicklung von Qualitätssicherungsverfahren .....	77
5.1	Verfahrensdurchführung.....	77
5.2	Verfahrenspflege.....	78
5.2.1	Informationsquellen.....	79
5.2.2	Einbindung von Expertengruppen.....	80
5.2.3	Handlungsoptionen .....	80
5.3	Verfahrensweiterentwicklung.....	83

5.3.1	Anlässe für Verfahrensweiterentwicklungen.....	84
5.3.2	Durchführung von Verfahrensweiterentwicklungen .....	84
<b>Teil C: Methodische Elemente .....</b>		<b>86</b>
<b>6</b>	<b>Grundlagen der systematischen Literatur- und Informationsbeschaffung.....</b>	<b>87</b>
6.1	Systematische Recherche .....	88
6.1.1	Präzisierung der Fragestellung.....	88
6.1.2	Informationsquellen der systematischen Recherche nach wissenschaftlicher Literatur und Leitlinien .....	89
6.1.3	Weitere Recherchequellen.....	90
6.1.4	Ableitung der Suchwörter, Suchstrategie und Dokumentation.....	91
6.1.5	Auswahlprozess.....	91
6.2	Kritische Bewertung relevanter Publikationen .....	92
6.3	Zusammenfassende Qualitätsbewertung der Gesamtevidenz.....	93
6.3.1	Primärstudien.....	94
6.3.2	Systematische Übersichtsarbeiten.....	96
6.3.3	Leitlinien.....	96
6.3.4	Qualitative Studien.....	96
<b>7</b>	<b>Einbeziehung externen Wissens und externer Erfahrung.....</b>	<b>97</b>
7.1	Ziele der Einbindung externen Wissens.....	97
7.2	Beratungszwecke bei der Einbindung externen Wissens .....	97
7.3	Methoden der Einbindung externen Wissens und externer Kompetenzen .....	99
7.3.1	Leitfadengestützte Einzelinterviews .....	99
7.3.2	Fokusgruppen.....	100
7.3.3	Einbindung externer Expertengruppen.....	101
<b>8</b>	<b>Eignung von Qualitätsindikatoren.....</b>	<b>109</b>
8.1	Konzepte zur Eignungsbeurteilung in der Literatur .....	109
8.2	Grundsätze zur Eignungsbeurteilung.....	109
8.3	Eignungskriterien des Qualitätsziels des Indikators.....	112
8.4	Eignungskriterien des Messverfahrens .....	117
8.5	Eignungskriterien des Bewertungskonzepts .....	122
8.6	Bezug zum Verfahrenszweck .....	124
<b>9</b>	<b>Veröffentlichung von Indikatorergebnissen.....</b>	<b>126</b>
<b>10</b>	<b>Bewertungskonzepte und Referenzbereiche .....</b>	<b>129</b>

10.1	Begriffsbestimmung und Funktion von Bewertungskonzepten .....	129
10.2	Entwicklung von Bewertungskonzepten.....	131
10.3	Festlegung von Referenzbereichen.....	132
10.3.1	Informationsbewertung für die Festlegung von Referenzbereichen.....	133
10.3.2	Spezifikation des Referenzbereichs.....	134
10.4	Festlegung der Zuordnungsregel .....	136
10.5	Sentinel Events.....	139
11	Grundgesamtheit, Stichprobenziehung und Vollerhebung.....	140
11.1	Ebenen der Stichprobenziehung und Vollerhebung .....	140
11.2	Kriterien für Stichprobenziehung und Vollerhebungen .....	142
11.3	Vorgehen bei Stichprobenziehungen.....	143
12	Datenvalidierung von QS-Dokumentationsdaten und Strukturierter Dialog.....	145
12.1	Technische Prüfung bei Dateneingabe und -annahme .....	146
12.2	Statistische Prüfung nach Datenannahme.....	147
12.3	Datenabgleich mit Referenzquelle.....	148
12.4	Strukturierter Dialog / Stellungnahmeverfahren.....	150
13	Methoden der Datenanalyse.....	152
13.1	Grundlegende Annahmen.....	152
13.2	Risikoadjustierung.....	154
13.2.1	Stratifizierung.....	155
13.2.2	Regressionsmodelle .....	157
13.2.3	Zusammenfassung und Ausblick .....	158
13.3	Aggregation mehrerer Qualitätsmessungen zum Zweck der Qualitätsförderung..	159
13.3.1	Hierarchisierung von Ergebnisgrößen .....	160
13.3.2	Indizes in der Qualitätsmessung .....	161
	Glossar .....	165
	Literatur.....	170

## Tabellenverzeichnis

Tabelle 1: OECD-Modell der Leistungsfähigkeit von Gesundheitssystemen (Arah et al. 2006) .	19
Tabelle 2: Qualitätskriterien des Institute of Medicine (IOM 2001), mit Übersetzung der Kriterien durch die Autoren der „Methodischen Grundlagen“ .....	20
Tabelle 3: Schema zur Evidenzbewertung in Bezug auf das Kriterium Quantität.....	94
Tabelle 4: Schema zur Evidenzbewertung in Bezug auf das Kriterium Qualität .....	94
Tabelle 5: Schema zur Evidenzbewertung in Bezug auf das Kriterium Konsistenz.....	95
Tabelle 6: Beratungszwecke der Einbindung externen Wissens mit Beispielen für die Einsatzbereiche und für die verwendeten Methoden.....	98
Tabelle 7: Beispiele für unmittelbar und mittelbar patientenrelevante Indikatorziele.....	115
Tabelle 8: Darstellung von Sensitivität und Spezifität bei der Feststellung von Abweichungen vom Referenzbereich .....	137
Tabelle 9: Im Krankenhaus verstorbene Patientinnen und Patienten mit isoliertem koronarchirurgischem Eingriff geschichtet nach Dringlichkeit des Eingriffs im Jahr 2015..	155

# Abbildungsverzeichnis

Abbildung 1: Differenz zwischen erreichbarer und erreichter Qualität bei einem Leistungserbringer als Ansatzpunkt des internen Qualitätsmanagements.....	24
Abbildung 2: Integration des Qualitätsbegriffs (Erfüllung von Anforderungen) in die Prozessabläufe der externen Qualitätssicherung zur Definition und Bewertung von Versorgungsqualität.....	28
Abbildung 3: Phasen der Neuentwicklung von QS-Verfahren .....	42
Abbildung 4: Ablaufschema der Entwicklung des Qualitätsmodells im Rahmen einer Konzeptstudie für ein neues QS-Verfahren .....	44
Abbildung 5: Entwicklungsschritte vom Qualitätsmodell bis zum Qualitätsindikatorenset.....	50
Abbildung 6: Modell der Patientenbefragung im Rahmen der externen Qualitätssicherung....	63
Abbildung 7: Schematische Darstellung der Entwicklungsschritte einer Patientenbefragung ..	64
Abbildung 8: Methodische Schritte (dunkelgrün) und Zwischenprodukte (hellgrün) der Fragebogenentwicklung.....	65
Abbildung 9: Überblick über die Eignungskriterien für Qualitätsindikatoren .....	111
Abbildung 10: Schematische Darstellung des Ausmaßes der leistungserbringerbedingten und nicht leistungserbringerbedingten Einflüsse auf den Indikatorwert anhand von Beispielen fallbezogener Qualitätsindikatoren.....	117
Abbildung 11: Schritte und zugehörige Eignungskriterien bei der Ermittlung und Aufarbeitung von Informationen für die Berechnung von Indikatorwerten .....	118
Abbildung 12: Funktion und Bestandteile des Bewertungskonzepts von Qualitätsindikatoren .....	130
Abbildung 13: Beispielhafte Darstellung der zugrunde liegenden Indikatorwerte (y-Achse), die mit einer Spezifität von 85 % in Abhängigkeit von Fallzahl (x-Achse) und Sensitivität (80 % vs. 70 %) mit einem einseitigen exakten Binomialtest als signifikant verschieden von einem Wert von 90 % klassifiziert werden .....	138

## Abkürzungsverzeichnis

Abkürzung	Erläuterung
AQUA-Institut	AQUA – Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
BQS	Bundesgeschäftsstelle Qualitätssicherung
DIMDI	Deutsches Institut für Medizinische Information und Dokumentation
G-BA	Gemeinsamer Bundesausschuss
GKV-FQWG	Gesetz zur Weiterentwicklung der Finanzstruktur und der Qualität in der gesetzlichen Krankenversicherung (GKV-Finanzstruktur- und Qualitätsweiterentwicklungsgesetz)
HTA	Health Technology Assessments
ICD	International Statistical Classification of Diseases and Related Health Problems. Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme
InEK	Institut für das Entgeltsystem im Krankenhaus
IQTIG	Institut für Qualitätssicherung und Transparenz im Gesundheitswesen
KHEntgG	Krankenhausentgeltgesetz
KHG	Krankenhausfinanzierungsgesetz
KHSG	Krankenhausstrukturgesetz
LE	Leistungserbringer
OPS	Operationen- und Prozedurenschlüssel
PREM	Patient-Reported Experience Measures
PROM	Patient-Reported Outcome Measures
Qb-R	Regelungen zum Qualitätsbericht der Krankenhäuser
Qesü-RL	Richtlinie zur einrichtungs- und sektorenübergreifenden Qualitätssicherung
QI	Qualitätsindikator
QS	Qualitätssicherung
QSKH-RL	Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern
RCT	Randomisierte kontrollierte Studien

<b>Abkürzung</b>	<b>Erläuterung</b>
SGB	Sozialgesetzbuch
SQS	Servicestelle Qualitätssicherung

## Vorwort

Zum 1. Januar 2016 hat das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG, im folgenden Text auch „Institut“) auf der Basis des § 137a SGB V vom Gemeinsamen Bundesausschuss (G-BA) den Auftrag erhalten, die Verfahren der externen Qualitätssicherung nach §§ 136 ff. SGB V fortzuführen, weiterzuentwickeln, sich an deren Durchführung maßgeblich zu beteiligen und deren Ergebnisse zu veröffentlichen. Diese Tätigkeit basiert auf den Vorarbeiten der Servicestelle Qualitätssicherung (SQS, 1996–2000), der Bundesgeschäftsstelle Qualitätssicherung (BQS, 2001–2009) sowie des AQUA-Instituts (2010–2015).

Das IQTIG ist eine Einrichtung der Stiftung für Qualitätssicherung und Transparenz im Gesundheitswesen, die vom G-BA auf der Basis des § 137a SGB V am 9. Januar 2015 gegründet wurde. Es ist sowohl nach dem Gesetz als auch nach seiner Satzung ein unabhängiges, wissenschaftliches Institut und hat zu gewährleisten, dass es seine Aufgaben auf Basis der maßgeblichen, international anerkannten Standards der Wissenschaften auf transparente Weise erfüllt.

Hierzu veröffentlicht das IQTIG die vorliegenden „Methodischen Grundlagen“ in ihrer ersten Version. In diesem Dokument werden die methodischen Arbeitsgrundlagen des IQTIG als fachlich unabhängiges, wissenschaftliches Institut nach § 137a SGB V festgehalten und es umfasst Methoden der Entwicklung, Weiterentwicklung und Durchführung von Qualitätssicherungsverfahren im Auftrag des G-BA. Die nachvollziehbare Anwendung dieser Methoden legitimiert in der Verfahrensentwicklung die Empfehlungen an den G-BA und legitimiert in der Durchführung der Qualitätssicherungsverfahren deren Ergebnisse.

Die „Methodischen Grundlagen“ des IQTIG werden kontinuierlich weiterentwickelt, öffentlichen Stellungnahmeverfahren unterzogen und in Folgeversionen veröffentlicht. Die nächste Version soll bis Ende des Jahres 2018 erscheinen.

# **Teil A: Grundlagen der gesetzlichen Qualitätssicherung und -förderung im Gesundheitswesen**

# 1 Einleitung

Die „Methodischen Grundlagen“ des Instituts für Qualitätssicherung und Transparenz im Gesundheitswesen stellen die Methoden dar, die der Entwicklung, Weiterentwicklung und Durchführung von Qualitätssicherungsverfahren (QS-Verfahren) des Instituts zugrunde liegen. Damit schafft das Institut Transparenz über seine Arbeit sowie über das Zustandekommen seiner Arbeitsergebnisse.

## **Inhalte und Struktur**

Die vorliegende Fassung beschränkt sich dabei auf diejenigen Methoden, die vom Institut zum Zeitpunkt der Erstellung dieses Dokuments angewendet werden. Das Dokument ist somit weder ein Lehrbuch für Qualitätssicherung noch beinhaltet es eine Abhandlung der Geschichte der medizinischen Qualitätssicherung.

Erfordern Entwicklungsaufträge neue Methoden, werden diese im Rahmen der Beauftragung entwickelt und im entsprechenden Abschlussbericht dargestellt. Eingang in die „Methodischen Grundlagen“ des Instituts finden diese Methoden dann bei der nächsten Aktualisierung des Dokuments.

Die „Methodischen Grundlagen“ gliedern sich in drei inhaltliche Teile:

- In Teil A werden neben der Einführung in das Dokument grundlegende Begriffe und Vorgehensweisen bei der Entwicklung, Weiterentwicklung und Durchführung von QS-Verfahren erläutert.
- In Teil B werden diese Prozesse im Detail dargestellt.
- In Teil C werden methodische Elemente dargestellt, die an verschiedenen Stellen der Entwicklungs- und Weiterentwicklungsprozesse eine Rolle spielen, wie etwa Fokusgruppen und Auswertungsmethoden.

## **Stellungnahmeverfahren**

Die „Methodischen Grundlagen“ und deren künftige Aktualisierungen wurden ab 1. Februar 2017 auf der Website des Instituts als Download bereitgestellt und einem öffentlichen Stellungnahmeverfahren unterzogen. Zur Stellungnahme aufgefordert wurden insbesondere:

- die Vertreterinnen und Vertreter der nach § 137a Abs. 7 SGB V zu beteiligenden Institutionen:
  - die Kassenärztlichen Bundesvereinigungen
  - die Deutsche Krankenhausgesellschaft
  - der Spitzenverband Bund der Krankenkassen
  - der Verband der Privaten Krankenversicherung
  - die Bundesärztekammer, die Bundeszahnärztekammer und die Bundespsychotherapeutenkammer
  - die Berufsorganisationen der Krankenpflegeberufe

- die wissenschaftlichen medizinischen Fachgesellschaften
  - das Deutsche Netzwerk Versorgungsforschung
  - die für die Wahrnehmung der Interessen der Patientinnen und Patienten und der Selbsthilfe chronisch kranker und behinderter Menschen maßgeblichen Organisationen auf Bundesebene
  - der oder die Beauftragte der Bundesregierung für die Belange der Patientinnen und Patienten
  - zwei von der Gesundheitsministerkonferenz der Länder zu bestimmende Vertreter
  - die Bundesoberbehörden im Geschäftsbereich des Bundesministeriums für Gesundheit, soweit ihre Aufgabenbereiche berührt sind
- die Mitglieder des wissenschaftlichen Beirats des Instituts
  - die Mitglieder des Kuratoriums des Instituts
  - die Mitglieder der Expertengruppen auf Bundesebene

Darüber hinaus wurde auch der interessierten Fachöffentlichkeit die Möglichkeit zur Stellungnahme eingeräumt. Stellungnahmen wurden bis 31. März 2017 in elektronischer Form entgegengenommen. Die Argumente aller Stellungnahmen werden einschließlich entsprechender Würdigungen durch das IQTIG zusammen mit den „Methodischen Grundlagen V1.0“ auf der Website des IQTIG veröffentlicht.

### **„Methodische Grundlagen“, Verfahrenshandbuch und Diskussionspapiere**

Die wissenschaftlichen Methoden und die praktischen Abläufe von QS-Verfahren werden in separaten Dokumenten dargestellt. Die wissenschaftlichen Methoden, die den Entwicklungen und Weiterentwicklungen von QS-Verfahren zugrunde liegen, werden in den „Methodischen Grundlagen“ dargelegt. Das Verfahrenshandbuch<sup>1</sup> beschreibt dagegen die konkreten Abläufe der Verfahren der sektorspezifischen und der sektorenübergreifenden Qualitätssicherung ohne ausführliche Erläuterungen der Grundlagen ihrer Instrumente. Es fasst die Einzelheiten konkreter Verfahren zusammen und soll u. a. denjenigen als Leitfaden dienen, die für die Durchführung der Verfahren verantwortlich sind. Bei Änderungen in der praktischen Durchführung der Verfahren wird die jeweilige Beschreibung kurzfristig angepasst. Das Dokument „Methodische Grundlagen“ bleibt dagegen von solchen Änderungen unberührt und wird nur aktualisiert, sofern sich die methodischen Grundlagen der Entwicklung und Weiterentwicklung der Verfahren ändern.

Änderungen von Methoden werden verfahrensübergreifend berücksichtigt, sobald dies auf sinnvolle Weise möglich ist. Methodische Änderungen in Projekten des G-BA werden diesem rechtzeitig angezeigt und ggf. wird ein entsprechender Beschluss erwirkt, damit die methodischen Änderungen auch eine entsprechende Rechtsgrundlage haben. Neue, autorisierte Methoden können auch dann schon angewandt werden, wenn sie noch nicht in den „Methodischen Grundlagen“ veröffentlicht sind. An entsprechender Stelle ist dann auf die veränderte methodische Grundlage hinzuweisen.

---

<sup>1</sup> <https://iqtig.org/das-iqtig/grundlagen/verfahrensgrundlagen>

Die hier beschriebenen Methoden beziehen sich in großen Teilen auf die QS-Verfahren des G-BA, die auf Richtlinien beruhen. Vorschläge zur Weiterentwicklung der Verfahren sind nicht Bestandteil der „Methodischen Grundlagen 1.0“, sondern werden künftig in Diskussionspapieren des IQTIG formuliert und den zuständigen Fachexpertinnen und Fachexperten und Gremien zur Beratung gegeben.

Die wissenschaftlichen Methoden des IQTIG werden fortlaufend ergänzt und weiterentwickelt. Außerdem weist die derzeitige Version der „Methodische Grundlagen“ noch Lücken auf, die geschlossen werden sollen. Dementsprechend wird bereits jetzt deren Folgeversion vorbereitet. Einige der Themen, für die Weiterentwicklungsbedarf gesehen wird und mit denen sich das IQTIG als nächstes schwerpunktmäßig befassen wird, sind:

- die grundlegenden methodischen Konzepte für den Einsatz von Qualitätsmessungen in der Krankenhausplanung und für Qualitätszu- und -abschläge der Vergütung
- die Auswirkungen solcher, über die Qualitätsförderung hinausgehender Verfahrenszwecke auf die Beurteilung der Eignung von Qualitätsaspekten, Qualitätsindikatoren und Indikatorensets für die Qualitätsdarstellung
- Konzepte für die Zuschreibbarkeit und Nutzung leistungserbringerübergreifender Qualitätsmessungen, z. B. bei sektorenübergreifenden Langzeit-Follow-up-Indikatoren oder bei regionalen Indikatoren
- Grundsätze für die Bearbeitung von Evaluationsaufträgen des G-BA
- die Weiterentwicklung der Datenvalidierung für die in die Qualitätssicherung einbezogenen Datenquellen
- Grundsätze für die Entwicklung von Risikoadjustierungsmodellen für QS-Verfahren
- die Methodik zur Beurteilung der Eignung von Qualitätsindikatoren laufender QS-Verfahren

Ein Teil dieser Entwicklungen wird voraussichtlich in die nächste Version der „Methodischen Grundlagen“ Eingang finden, die Ende 2018 erscheinen soll.

## 2 Qualität und Qualitätssicherung im Gesundheitswesen

### 2.1 Qualität der Gesundheitsversorgung

#### 2.1.1 Qualitätsbegriff in der Gesundheitsversorgung

Unser Leben wird in ganz besonderer Weise von unserer Gesundheit geprägt. Die Maßnahmen, die unsere Gesundheit im Krankheitsfall wiederherstellen oder die es uns ermöglichen, mit einer Erkrankung bestmöglich zu leben, sind für alle Menschen von zentraler Bedeutung. Die Verbesserung oder zumindest Stabilisierung der Gesundheitssituation von Menschen ist das zentrale Ziel. Gesundheitliche Versorgung wird primär an dieser Zielerreichung gemessen. Dies entspricht der Definition der DIN-Norm „Qualitätsmanagementsysteme – Grundlagen und Begriffe“ (DIN EN ISO 9000), wonach – in Analogie dazu – Qualität den Grad beschreibt, mit dem „ein Satz inhärenter Merkmale“ der Versorgung die an sie gestellten „Anforderungen erfüllt“ (DIN EN ISO 9000, zit. nach Sens et al. 2007: 3).

Qualität ist ein relativer Begriff. Im konkreten Kontext hängt seine Bedeutung von den Anforderungen ab, deren Erreichung als Qualitätsmaßstab gilt. Die Qualität der Patientenversorgung wird von Patientinnen und Patienten, von Hausärztinnen und -ärzten, von der Krankenhausadministration, von Politikerinnen und Politikern, von Krankenkassenvertreterinnen und -vertretern, von Pflegekräften, von Angehörigen etc. meist mit ganz unterschiedlichen Inhalten und Schwerpunkten gesehen. Wer von Qualität spricht, muss immer auch sagen, welche Anforderungen er damit meint, weil ansonsten der Begriff der Qualität inhaltslos bleibt.

Qualität wird entweder der Versorgung selbst, also den Prozessen, oder aber auch den agierenden Einrichtungen und Personen zugeschrieben: beschrieben wird alltagssprachlich z. B. die Qualität einer Behandlung, die Qualität einer Ärztin oder eines Arztes, die Qualität einer Praxis oder eines Krankenhauses. Bei den Personen und den Einrichtungen ist die Kompetenz der Akteure gemeint, gute Versorgungsqualität zu gewährleisten. Diese Unterscheidung spielt eine Rolle, wenn die Qualitätsaussage sich auf die konkret geleistete Versorgungsqualität bei den Patientinnen und Patienten eines Jahres beziehen soll, oder wenn stattdessen die generelle Qualitätskompetenz einer Einrichtung dargestellt werden soll, abgeleitet vom Ergebnis der Behandlungen von Patientinnen und Patienten eines Jahres, die gleichsam als Stichprobe des Könnens der Praxis oder der Einrichtung angesehen wird. In letzterem Fall ist die Zufälligkeit „der Stichprobe“ auch statistisch zu berücksichtigen. Beides stellt entsprechende methodische Anforderungen an eine belastbare Messung und Darstellung der Versorgungsqualität.

#### 2.1.2 Qualitätsdimensionen und Systematiken von Qualitätskriterien

Nun kann nach Donabedian (1966) unterschieden werden, ob die Struktur-, Prozess- oder Ergebnisqualität betrachtet werden soll, die bekannten Qualitätsdimensionen nach Donabedian. Es gibt immer wieder Grenzfälle, an denen die Zugehörigkeit von Qualitätsmerkmalen zu einer der Dimensionen diskutiert wird. Dabei wird in diesem Methodenpapier die Zugehörigkeit zur

Prozessqualität durch die direkte Beeinflussbarkeit des Akteurs bestimmt. In diesem Zusammenhang wird oft die Indikationsqualität diskutiert, die von einigen als Ergebnisqualität eingestuft wird. Dem folgt dieses Methodenpapier nicht. Die Indikationsstellung ist zwar für das Ergebnis einer Behandlung von großer Bedeutung, aber die Indikationsstellung selbst ist durch die behandelnden Ärztinnen oder Ärzte gestaltbar, also Prozessqualität. Die ungeplante Wiederaufnahme ist ein weiteres diskutiertes Beispiel. Die Wiederaufnahme in ein Krankenhaus nach Primärbehandlung wird als Ergebnisqualität eingestuft, weil es nicht darum geht, dass eine Wiederaufnahme erfolgt, sondern dass eine Patientin oder ein Patient eine so schwere Komplikation erlitten hat, dass stationäre Behandlung erforderlich ist. Es geht nicht um den Prozess der Wiederaufnahme, sondern um die erhebliche Behandlungsbedürftigkeit der Patientin oder des Patienten als Komplikation nach Primärbehandlung. Besonders zu betrachten sind auch Merkmale, die die Behandlungsprozesse beschreiben, z. B. OP-Dauer oder Strahlenbelastung bei Eingriffen mit Durchleuchtung. Es ist nicht eindeutig herzuleiten, wie diese Merkmale in der Systematik einzuordnen sind. Dieses Methodenpapier folgt der Setzung, dass diese prozessbezogenen Parameter eben den Behandlungsprozess beschreiben, mit dem Patientinnen und Patienten versorgt werden, und die daher der Prozessqualität zugeordnet werden.

Neben der Qualität der Versorgung der einzelnen Leistungserbringer kommt es auch auf deren effektive Kooperation bei der Behandlung von Patientinnen und Patienten an. In Gesundheitssystemen spielen auch Aspekte des Zugangs zur Versorgung (*access to care*), Aspekte der Versorgungsgerechtigkeit (*equity*) und solche des Übergangs der Patientinnen und Patienten von der Versorgung in ihre alltägliche Lebenswelt eine Rolle. Dies sind Systemqualitäten, die bei der Beschreibung von Versorgungsqualität oberhalb der Ebene der Leistungserbringer zu berücksichtigen sind.

Mehrere Modelle fassen solche prinzipiellen Aspekte, die bei der Bewertung von Gesundheitsversorgung und Gesundheitswesen zu berücksichtigen sind, zu Systematiken zusammen. Diese sind mal enger, mal weiter gefasst, je nach Intention der Institutionen, die diese Modelle publiziert haben und anwenden: z. B. als Grundlage für die Entwicklung und das Monitoring von Qualitätssicherung und Qualitätsmanagement im Gesundheitswesen (Committee on Quality of Health Care in America/Institute of Medicine 2001, Commonwealth Fund 2006, Davis 2005), zum internationalen Vergleich von Gesundheitswesen und ihren Qualitätskonzepten (Arah et al. 2006) oder als Richtschnur für die Akkreditierung von Gesundheitseinrichtungen (z. B. Joint Commission 2016).

Das OECD-Modell (Arah et al. 2006) fasst die Kriterien zur Bewertung der Leistungsfähigkeit eines Gesundheitswesens in einer komprimierten Liste bzw. Tabelle zusammen (siehe Tabelle 1) und umfasst folgende Aspekte unter dem Hauptsegment Qualität:

- Qualität
  - Effektivität (Effectiveness)
  - Sicherheit der Versorgung (Safety)
  - Patientenzentrierung (Responsiveness / patient-centered).

Weitere Hauptsegmente sind:

- Zugang zur Versorgung (Access)
- Ausgaben/Kosten (Expenditure).

Während die unter dem Oberbegriff „Quality“ zusammengefassten Aspekte traditioneller Teil der extern vergleichenden Qualitätssicherung in Deutschland sind, ist der „Zugang zur Versorgung“ ein Systemaspekt, der erst schrittweise auch in den Fokus der QS-Betrachtungen rückt.

Tabelle 1: OECD-Modell der Leistungsfähigkeit von Gesundheitssystemen (Arah et al. 2006)

Health Care System Performance					
	Quality			Access	Cost / Expenditure
Healthcare Needs	Effectiveness	Safety	Responsiveness / Patient-centeredness	Accessibility	
Staying healthy					
Getting better					
Living with illness or disability					
Coping with end-of-life					

Die Ausgaben bzw. Kosten für Gesundheitsversorgung sowie andere ökonomische Parameter, die im OECD-Modell ein eigenes Segment bilden, werden in der Methodik der externen Qualitätssicherung in Deutschland primär nicht betrachtet. Es geht stattdessen um eine reine Betrachtung des Qualitätsaspekts. Erst sekundär wird ggf. reflektiert, inwieweit finanzielle Rahmenbedingungen Anreize oder Fehlanreize setzen, die sich auf die Versorgungsqualität auswirken. Diese ökonomischen Parameter werden aber durch gänzlich andere Regelkreise bestimmt. Erst Steuerungsinstrumente, die mit finanziellen Anreizen arbeiten (z. B. Qualitätszu- und -abschläge), verschränken diese Wirkkreise. Die Wirtschaftlichkeit der versorgenden Einrichtungen ist zwar ein weiterer, wichtiger Aspekt für ein Gesundheitswesen, bleibt aber bei der Qualitätsbewertung unberücksichtigt. Unzureichende Versorgungsqualität bei unwirtschaftlich arbeitenden Einrichtungen zu tolerieren wäre unethisch.

Als weiteres Beispiel möglicher Qualitätskriterien werden die des Institute of Medicine (Committee on Quality of Health Care in America/Institute of Medicine 2001: 5 f.) aufgelistet (siehe Tabelle 2).

Tabelle 2: Qualitätskriterien des Institute of Medicine (IOM 2001), mit Übersetzung der Kriterien durch die Autoren der „Methodischen Grundlagen“

Qualitätskriterium	Übersetzung
safe	sicher
effective	wirksam in der Versorgungspraxis
patient-centered	patientenorientiert
timely	rechtzeitig, zeitnah
efficient	wirtschaftlich
equitable	gerecht

Die Liste der Kriterien des IOM ist sehr heterogen zusammengesetzt: „Effectiveness“ betrifft die Wirksamkeit der medizinischen Maßnahmen, „Efficiency“ beschreibt die Wirtschaftlichkeit. „Timeliness“ kommt in den QS-Verfahren des IQTIG z. B. als Entschluss-Entbindungs-Zeit bei Notfallkaiserschnitten (Prozesszeit) vor, während in anderen Zusammenhängen unter „Timeliness“ die Kapazität eines (Gesundheits-)Systems gemeint ist, einer Patientin oder einem Patienten schnell eine Behandlung zukommen zu lassen, sobald der Bedarf hierzu offensichtlich wird (Systemzeit, AHRQ 2011). Sterblichkeit und vermeidbare schwerwiegende Komplikationen beschreiben die Versorgungssicherheit (Safety). Die Versorgungsgerechtigkeit wiederum ist eine Systemeigenschaft.

Es werden in den verschiedenen Modellen, die alle untereinander Ähnlichkeiten aufweisen, aber auch eigene Charakteristika haben, sehr unterschiedliche Aspekte und Verantwortlichkeiten in der Gesundheitsversorgung angesprochen. Sie sind für die jeweilige Zielsetzung der Qualitätsbewertung zusammengestellt (Arah et al. 2006), wobei insbesondere Aspekte zu unterscheiden sind, die sich auf das Gesundheitssystem beziehen, und solche, die die Versorgungsqualität einzelner Leistungserbringer thematisieren.

Für die externe Qualitätssicherung nach §§ 136 ff. SGB V scheint das OECD-Modell das geeignetste zu sein. Es ist auf wenige Aspekte fokussiert und stimmt am ehesten mit den Fragestellungen der extern vergleichenden Qualitätssicherung überein. Sie fokussiert im Segment „Quality“ vornehmlich auf Aspekte, die im Vergleich von Leistungserbringern und im Rahmen der Patientenorientierung von Bedeutung sind. Zusätzlich ist der Zugang zu Versorgungsangeboten (Accessibility) durchaus ein wichtiger Qualitätsaspekt, der sowohl das System (Erreichbarkeit ausreichender Versorgungsangebote) als auch den einzelnen Leistungserbringern (niedrigschwelliger Zugang zur Versorgung dieses Leistungserbringers) betrifft. Insofern wird das Segment „Quality“ der OECD-Systematik durch das der „Accessibility“ bei der externen Qualitätssicherung sinnvoll ergänzt.

Vorhersehbar ist, dass auch die externe Qualitätssicherung nach §§ 136 ff. SGB V künftig Qualität nicht nur leistungserbringerbezogen, sondern auch systembezogen thematisieren wird. Dies wird künftig ermöglicht durch die Verfahren der sektorenübergreifenden Qualitätssicherung sowie die Einbeziehung von Follow-up-Erhebungen, Patientenbefragungen und Sozialdaten als

weiteren Datenquellen neben der traditionellen QS-Dokumentation. Zum Beispiel wird die Versorgungsqualität von Patientinnen und Patienten mit chronischen Erkrankungen sowohl im Zeitverlauf als auch in der parallelen und abwechselnden Versorgung durch unterschiedliche Einrichtungen abbildbar. Für die Auswertung und Bewertung dieser Versorgungsqualität auf Systemebene sind neue Instrumente zu entwickeln. Insbesondere wird bei einer Versorgung, die mehrere Leistungserbringer über einen gewissen Zeitraum einschließt, die Frage nach der Verantwortlichkeit und dem Umgang mit Verbesserungsbedarfen zu stellen sein. Hier könnten z. B. Area-Indikatoren regionale Versorgungsprobleme aufzeigen, für die dann eine Organisationsform für die Bearbeitung der Probleme gefunden werden muss. Areaindikatoren dürfen nicht zu Ergebnissen führen, für die kein Adressat sich verantwortlich fühlt.

Das OECD-Modell kann jedenfalls auch Aspekte auf Systemebene abbilden. Wie bereits angesprochen, können seine Kriterien der

- „Effectiveness“ (Wirksamkeit der Versorgungspraxis),
- „Safety“ (Sicherheit der Versorgungspraxis),
- „Patient-centeredness“ / „Responsiveness“ (Patientenorientierung) und
- „Accessibility“ (Zugänglichkeit von Versorgungsangeboten)

für die extern vergleichende Qualitätssicherung als Grundraster genutzt werden. Gleichzeitig spricht das OECD-Modell den primären Gesundheitszustand von betroffenen Patientinnen und Patienten an und öffnet hiermit eine weitere Dimension patientenorientierter Beschreibung von Gesundheitsversorgung.

Bei der Entwicklung von Qualitätsindikatoren hat sich gezeigt, dass die Systematik der OECD nur grobe Kategorien bildet, die für die Entwicklung eines QS-Verfahrens nur bedingt hilfreich sind. Der Nutzen der Systematik liegt darin, dass sie die Qualitätsaspekte einer Versorgungssituation übersichtlich gruppiert, sodass anhand der Verteilung der adressierten Themen auf die einzelnen Aspekte die Ausgewogenheit der Zusammenstellung kontrolliert werden kann. Die Systematik spielt somit bei der Qualitätskontrolle von Indikatorensets eine Rolle sowie als Übersicht bei der Darstellung derselben.

### **2.1.3 Patientenorientierung**

Das Ziel der Gesundheitsversorgung ist die optimale Versorgung der Patientinnen und Patienten. Sie sind nicht nur Objekt der Gesundheitsversorgung, sondern Betroffene, deren Bedarfe von besonderer Bedeutung sind. Patientenorientierung ist ein zentraler Aspekt der Bewertung von Gesundheitsversorgung.

Klemperer (2000) beschreibt Patientenorientierung wie folgt:

*Patientenorientierung (P.) bezeichnet die Ausrichtung von Strukturen, Prozessen und Ergebnissen des Systems der gesundheitlichen Versorgung auf die Interessen, Bedürfnisse und Wünsche des individuellen Patienten.*

*P[atientenorientierung]<sup>2</sup> äußert sich u. a. darin, dass der Patient im Gesundheitssystem mit seinen Interessen, Bedürfnissen und Wünschen wahrgenommen und respektiert wird, ihm mit Empathie und Takt begegnet wird, er die Leistungen erhält, die nutzbringend und von ihm erwünscht sind und er über verbrieft Rechte und Pflichten verfügt.*

Diese Definition beschreibt sehr kompakt

- die Ausrichtung der professionellen Strukturen und Prozesse am Nutzen für die Patienten,
- die Ausrichtung der relevanten Ergebnisparameter am Patientennutzen,
- die Wahrnehmung und Berücksichtigung der Individualität der Patientinnen und Patienten,
- die soziale und persönliche Komponente der Interaktion der Behandelnden und Pflegenden mit Patientinnen und Patienten.

Am Anfang der Entwicklung der externen QS-Verfahren in Deutschland definierten Ärztinnen und Ärzte sowie Pflegekräfte die professionellen Standards ihres Fachgebiets als Grundlage der Qualitätsmessungen: einzuhaltende Prozesse und erwartbare Ergebnisse. Die explizite Patientenorientierung kam später hinzu und prüft u. a., ob die Prioritäten der professionellen Betrachtung wirklich primär am Patientennutzen orientiert sind. Sie erweitert außerdem die Aspekte um die Berücksichtigung jener Bedarfe, die Patientinnen und Patienten in ihrer besonderen Lebenssituation haben.

Dieser emanzipatorische Ansatz wurde nicht zuletzt in den Beteiligungsrechten der Patientenvertretung in den Gremien des G-BA und des IQTIG, der Herstellung öffentlicher Transparenz der Ergebnisse der Qualitätssicherung sowie der Etablierung der Patientenbefragung als Instrument der externen Qualitätssicherung gesetzlich verankert.

## **2.2 Qualitätssicherung und Qualitätsmanagement**

### **2.2.1 Historischer Ursprung der externen Qualitätsvergleiche**

Mit der Münchner Perinatalstudie (1975 bis 1977) wurde erstmals der Schritt von der Versorgungsforschung zur systematischen externen Qualitätssicherung unter der späteren Leitung von Prof. H. K. Selbmann vollzogen (Conrad 1977). Anlass war eine als zu hoch empfundene neonatale Sterblichkeit im Raum München in den Jahren 1970 bis 1972. Durch eine einheitliche Dokumentation der beteiligten Kliniken und die krankenhausbefundene, vergleichende Auswertung der Studie sollte

- ein Bild von der neonatologischen Versorgungsqualität in der Region München ermöglicht werden,
- „erste Schritte in Richtung einer selbstverantwortlichen Eigenkontrolle der beteiligten Kliniken gegangen“ werden und
- „Statistische Unterlagen zur Beantwortung gemeinsamer perinatologischer Fragestellungen bereitgestellt werden“ (Conrad et al, 1977).

---

<sup>2</sup> Ergänzung durch die Autoren der „Methodischen Grundlagen“.

Dies gilt als die Geburtsstunde der externen Qualitätssicherung in Deutschland, der bald weitere Projekte folgten, z. B. in der Chirurgie, Neonatologie, Anästhesiologie u. a. m.

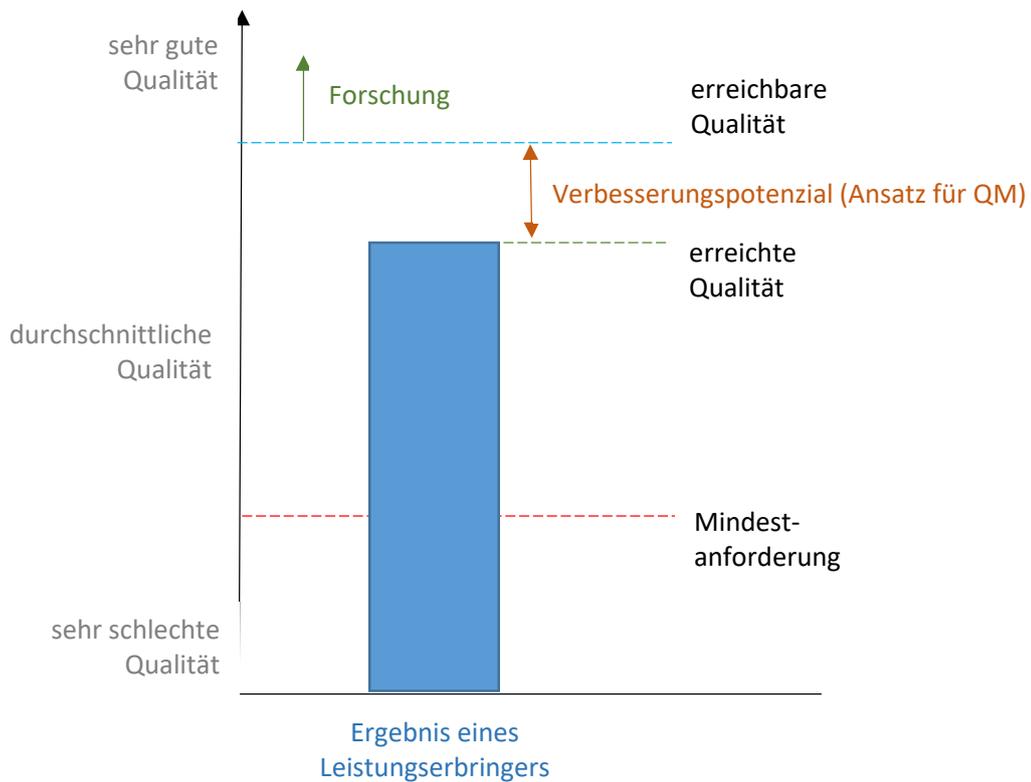
### **2.2.2 Prinzip der externen Vergleiche**

Das Prinzip der externen Vergleiche beruhte bereits in diesen ersten Verfahren auf der Festlegung relevanter Qualitätsmerkmale (z. B. neonatale Sterblichkeit), der Durchführung einer einheitlichen Dokumentation, der statistischen Auswertung der Daten und deren fachlicher Bewertung im Hinblick auf Versorgungsqualität. Dies bildet bereits die Grundstruktur von Qualitätsindikatoren, nämlich die Definition des Qualitätsziels, die Qualitätsmessung sowie die Qualitätsbewertung.

Durch die vergleichende Messung von Versorgungsqualität kann festgestellt werden, ob ein Leistungserbringer erreichbare Qualität erreicht (siehe Abbildung 1). Die erreichbare Qualität wird empirisch aus den besten Ergebnissen eines solchen Qualitätsvergleichs oder aus entsprechenden Studien abgeleitet. Darüber hinaus können bestimmte Qualitätsforderungen auf Basis sozialer Normen festgelegt werden. Die erreichte Qualität einer Einrichtung wird meist ermittelt als statistisches Ergebnis der Falldokumentationen und der Prüfung auffälliger Ergebnisse im fachlichen Dialog. Aufgabe der Qualitätssicherung und des einrichtungsinternen Qualitätsmanagements ist es, den Abstand zwischen erreichter und eigentlich erreichbarer Versorgungsqualität zu minimieren. Medizinischer Fortschritt und Weiterentwicklung der Versorgung in all ihren Aspekten sorgen dafür, dass sich die Grenze der erreichbaren Qualität kontinuierlich weiterentwickelt, z. B. durch neue Verfahren, Medikamente, Versorgungsformen etc. Fachliche Standards definieren darüber hinaus den Mindeststandard an Versorgungsqualität, der nicht unterschritten werden darf. Damit sind bereits die Grundstrukturen externer Qualitätsvergleiche und externer Qualitätssicherung kurz skizziert.

Externe Vergleiche wurden schon früh zur medizinischen Qualitätssicherung eingesetzt, da sie ermöglichen, anhand empirischer Daten Qualitätsstandards zu formulieren und Verbesserungspotenziale (Erreichbares noch nicht erreicht) bei einzelnen Einrichtungen zu identifizieren. Dabei werden Qualitätsanforderungen bereits dadurch legitimiert, dass einige Einrichtungen eine entsprechende Versorgungsqualität bereits erreicht haben. Es wird nur Machbares gefordert. Das interne Qualitätsmanagement der Einrichtungen versucht dann, das Zurückbleiben der erreichten Qualität hinter der erreichbaren Qualität zu minimieren (siehe Abbildung 1).

Abbildung 1: Differenz zwischen erreichbarer und erreichter Qualität bei einem Leistungserbringer als Ansatzpunkt des internen Qualitätsmanagements



### 2.2.3 Qualitätssicherung

Die Orientierung an der Aufdeckung von Defiziten anhand von statistischen Qualitätsvergleichen erklärt die frühe Prägung des Begriffs der „Qualitätssicherung“, die ein gefordertes Maß an Qualität sicherstellen soll. Passend dazu formuliert die derzeitige Normendefinition von Qualitätssicherung, dass diese „auf das Erzeugen von Vertrauen [...] gerichtet ist, dass Qualitätsanforderungen erfüllt werden.“ (DIN EN ISO 9000:2005, zitiert nach Sens et al. 2007).

Von dieser Definition weicht das IQTIG in seinem Methodenpapier jedoch ab. Dabei muss unterschieden werden, ob die Qualitätssicherung im Gesundheitswesen allgemein oder die spezifische Funktion der Verfahren nach §§ 136 ff. SGB V gemeint sind.

Das Verständnis von Qualitätssicherung im Gesundheitswesen ist historisch gewachsen. Natürlich hat die Etablierung von Qualitätssicherungsmaßnahmen im Gesundheitswesen auf gesetzlicher Grundlage auch das Ziel, durch Instrumente der Messung, der Kontrolle, des Dialogs, der kontinuierlichen Verbesserung und vor allem auch durch Transparenz Vertrauen der Bevölkerung in die Qualität der Gesundheitsversorgung zu schaffen. Die Qualitätssicherung setzt sich zusammen aus verschiedensten Aktivitäten der Akteure im Gesundheitswesen. Aufgabe des IQTIG als Einrichtung der Selbstverwaltung ist dabei vor allem die unabhängige Messung von Versorgungsqualität in Kooperation mit den verschiedenen Partnern im Gesundheitswesen sowie die transparente Darstellung der Ergebnisse. Vertrauen können die Berichte des IQTIG dann

erzeugen, wenn das IQTIG unabhängig und kritisch berichtet und sekundär dann die dargestellten Ergebnisse das Vertrauen rechtfertigen. Gerade weil das IQTIG nicht primär beauftragt ist, Vertrauen zu schaffen, sondern kritisch zu prüfen, kann es dem vertrauensbildenden Ziel der Qualitätssicherungsmaßnahmen insgesamt dienlich sein. Nicht das IQTIG soll Vertrauen schaffen, sondern die Ergebnisse der Qualitätssicherungsmaßnahmen im Gesundheitswesen sollen das bewirken. Somit passt das Begriffsverständnis der DIN, dass Qualitätssicherung auf die Erzeugung von Vertrauen gerichtet ist, auf das Gesundheitssystem als Ganzes, in dem Einrichtungen wie das IQTIG eine unabhängige Funktion erfüllen.

In der Praxis der externen Qualitätssicherung im deutschen Gesundheitswesen hat sich das Verständnis der Verfahren dahingehend erweitert, dass nicht nur die Qualitätsmessung und die Behebung von Qualitätsdefiziten das Ziel ist, sondern auch die Förderung der Weiterentwicklung von Versorgungsqualität. So sieht das Krankenhausstrukturgesetz nicht nur Abschläge bei unzureichender, sondern auch Zuschläge bei außerordentlich guter Qualität vor, und mit den Qualitätsverträgen nach § 110a SGB V soll durch Selektivverträge ein überdurchschnittliches Qualitätsniveau in bestimmten Leistungsbereichen gefördert werden. Damit unterscheidet sich der Qualitätssicherungsbegriff der externen Qualitätssicherung von dem der DIN EN ISO 9000.

Dies führt zu dem (scheinbaren) Dilemma, dass zum gleichen Wort „Qualitätssicherung“ zwei Bedeutungen vorgefunden werden, nämlich die der Norm und diejenige, die dem gewachsenen Verständnis der externen Qualitätssicherung im Gesundheitswesen entspricht. Bezogen auf das Gesundheitswesen ist die Normendefinition die zutreffende. Gleichzeitig tritt aber auf der Ebene der konkreten Projekte nach §§ 136 ff. SGB V der Vertrauensaspekt in den Hintergrund, da er den Auftrag, kritisch zu prüfen und zu berichten, stören würde, wenn das Schaffen von Vertrauen primäres Ziel wäre. Im konkreten Handlungskontext des IQTIG wird Qualitätssicherung daher definiert als die „Prozesse und Maßnahmen, durch die die Qualität der Versorgung gewährleistet oder verbessert werden soll“, ohne dass der übergeordnete Rahmen der Vertrauensbildung dadurch ausgeblendet würde (siehe Glossar auf S. 167).

#### **2.2.4 Qualitätsmanagement**

Versorgungsqualität kann primär nur vom Leistungserbringer selbst verbessert werden, da er der Akteur ist. Dabei wird Qualitätsmanagement nach DIN EN ISO definiert als „Aufeinander abgestimmte Tätigkeiten zum Leiten und Lenken einer Organisation bezüglich Qualität“ (zitiert nach Sens et al. 2007). Dieses Management bedient sich einer Reihe von Instrumenten, wie z. B. dem Deming- oder PDCA Zyklus, nach dem Planung, Umsetzung, Prüfung und „Re-Aktion“ auf das Prüfergebnis in einem fortlaufenden Verbesserungszyklus zyklisch durchlaufen werden.

Das Konzept der externen Qualitätssicherung ist nun, dem internen Qualitätsmanagement der Einrichtungen valide Messergebnisse aus adjustierten Vergleichen mit anderen Einrichtungen als Feedback zur Verfügung zu stellen, um ihr internes Qualitätsmanagement zu unterstützen (Altenhofen et al. 2009: 5). So können sie ihre Stärken und Verbesserungspotenziale besser einschätzen und die Erfolge ihrer kontinuierlichen Verbesserungsprojekte kontrollieren.

Die primäre Verantwortlichkeit für die kontinuierliche Verbesserung der medizinischen Versorgungsqualität liegt beim internen Qualitätsmanagement der Einrichtungen. Diese wird durch die

externe Qualitätssicherung unterstützt und bisweilen auch zu Aktivitäten aufgefordert. Letztere hat daher stets zu prüfen, ob sie ihre unterstützende Funktion in geeigneter Weise erfüllt.

### **2.2.5    Transparenz und externe Intervention**

Mit der Etablierung der externen Qualitätssicherung auf gesetzlicher Grundlage (aktuell §§ 136 ff. SGB V) wurde auch die Verpflichtung zur Transparenz festgelegt, d. h. die öffentliche Zugänglichkeit zu einrichtungsbezogenen Qualitätsdaten aus den Qualitätsvergleichen für den stationären Bereich.

Weiterhin hat sich mit den Jahren gezeigt, dass alleine das Feedback der externen Qualitätssicherung an das interne Qualitätsmanagement der Einrichtungen nicht ausreicht, um unzureichende Versorgungsqualität bei vereinzelt Leistungserbringern zu beheben. Da nun verschiedene Anreize im Gesundheitswesen das Leistungsgeschehen beeinflussen, lag es nahe, über eine Anpassung der Anreizsysteme Fehlanreize für Überversorgung, Fehlversorgung oder Unterversorgung zu beheben und verstärkt Anreize für Qualitätsverbesserungen zu setzen – oder bei einzelnen Abteilungen eine anhaltend unzureichende Versorgung zu unterbinden. Solche Regulierungsansätze finden sich z. B. im Krankenhausstrukturgesetz mit planungsrelevanten Qualitätsindikatoren oder Qualitätszu- und -abschlägen. Hier wird Neuland betreten, entsprechende Instrumente werden für verschiedene, neue Verfahrenszwecke entwickelt (siehe Abschnitt 2.4) und dann in die Routine eingeführt.

Da diese neuen Verfahren derzeit noch in der Entwicklung oder Erprobung sind, erfolgt ihre methodische Beschreibung zunächst in den verfahrensspezifischen Konzeptpapieren. Nach Konsolidierung der Verfahrensroutine sollen diese spezifischen methodischen Grundlagen in das allgemeine Methodenpapier aufgenommen werden.

## **2.3        Legitime Anforderungen als Grundlage der Qualitätsmessung**

Es gibt vielfältige Arten, die Qualität der Gesundheitsversorgung zu bewerten. Wenn man Patientinnen oder Patienten nach der Qualität ihrer Versorgung fragt, so gibt es aufgrund der individuellen Erwartungen und des unterschiedlichen persönlichen Erlebens eine Vielzahl von Facetten, die bei der Bewertung durch die einzelne Person zusammenspielen. Dieses umfassende, lebensweltliche Qualitätsurteil von Patientinnen und Patienten ist der Ausgangspunkt der Qualitätsvorstellung in der gesundheitlichen Versorgung.

Die genannte individuelle, von Patientinnen und Patienten empfundene Qualität in all ihren persönlichen Erlebnisaspekten kann jedoch von keinem noch so umfassenden Zahlenwerk dargestellt werden. Eine objektive, Vergleiche ermöglichende Darstellung von Versorgungsqualität setzt voraus, dass eine begrenzte Anzahl legitimer Anforderungen an die Versorgung identifiziert und definiert werden. Messungen und (vergleichende) Bewertungen der Qualität der Versorgung können dann, in Einklang mit der genannten Definition von Qualität, das Ausmaß beschreiben, in dem diese Anforderungen erfüllt werden. Da diese Anforderungen der externen Qualitätssicherung Grundlage für Verbesserungsinitiativen und Steuerungsverfahren im Gesundheitswesen sein sollen, müssen sie in besonderer Weise legitimiert sein. Diese Legitimation ergibt sich durch die Erfüllung folgender Kriterien:

- **Patientenorientierung der Ziele:** Die Anforderungen an die medizinische Versorgung müssen sich primär am Nutzen für Patientinnen und Patienten orientieren. Dies können zum einen Anforderungen sein, die z. B. als Ergebnisqualität den Patientennutzen direkt ausdrücken. Zum anderen können auch Anforderungen an die Prozess- und Strukturqualität gestellt werden, sofern Evidenz dafür vorliegt, dass deren Erfüllen die Ergebnisqualität für die Patientinnen und Patienten in relevanter Weise fördert. Darüber hinaus können aber auch gesellschaftliche Werte und Übereinkünfte patientenzentrierte Anforderungen vorgeben, wie etwa die Erwartung partizipativer Entscheidungsfindung in der Versorgung.
- **Beeinflussbarkeit durch die Leistungserbringer:** Es muss den Leistungserbringern prinzipiell möglich sein, die patientenbezogenen Ziele durch eine entsprechende Gestaltung ihrer Versorgung zu erreichen.
- **Unbedenklichkeit:** Die gestellten Anforderungen dürfen darüber hinaus keine für die Patientinnen und Patienten schädlichen Nebenwirkungen entfalten, die den eigentlich angestrebten Nutzen durch begleitende Nebenwirkungen konterkarieren. Qualitätsindikatoren und ggf. daran anknüpfende Anreize sind so zu gestalten, dass Über-, Unter- oder Fehlversorgung nicht gefördert werden. Mögliche Reaktionen der Leistungserbringer auf neue Regularien sind zu beachten, wobei stets eine am Patientenwohl orientierte Versorgung der Leistungserbringer vorauszusetzen ist.

Die Fokussierung auf definierte, legitime Anforderungen ist eines der zentralen Charakteristika der externen Qualitätssicherung. Ein zweites ist die primär statistische Herangehensweise an die Messung und Bewertung der Versorgungsqualität der einzelnen Leistungserbringer im Vergleich zueinander. Die statistische Analyse von Qualitätssicherungsdaten ist eine über den Einzelfall hinausgehende Betrachtungsweise und kann dadurch Aspekte der Qualität quantitativ darstellen, die in Einzelfallanalysen nicht sichtbar werden. Die Bewertung der Versorgungsqualität eines Leistungserbringers erfolgt anhand statistischer Daten der Versorgung von Patientinnen und Patienten, die dieser in einem bestimmten Zeitraum (meist während eines Jahres) behandelte. Die externe Qualitätssicherung hat damit einen anderen Schwerpunkt als etliche Instrumente der internen Qualitätssicherung, die mehr auf kasuistischen Einzelfallbewertungen vor Ort beruhen (z. B. Mortalitäts- und Morbiditätskonferenzen, Peer-Review-Verfahren anhand einzelner Patientenakten und Vor-Ort-Begutachtung). Lediglich bei seltenen Ereignissen, besonderen Einzelkonstellationen oder sehr kleinen Fallzahlen greift auch die externe Qualitätssicherung auf kasuistische Analysemethoden zurück, z. B. bei der Untersuchung von Sentinel Events. Statistische und kasuistische Instrumente der Qualitätsbewertung haben ihre jeweils spezifischen Sichtweisen und ihre spezifischen Limitationen. Qualitätssicherung kann dann besonders gut funktionieren, wenn die statistischen Erkenntnisse der externen Qualitätsvergleiche und die fachlichen Rückmeldungen des Strukturierten Dialogs für die interne Qualitätssicherung nützlich sind und auch bestmöglich genutzt werden. Hierzu zählt zum Beispiel, dass zwischen der externen und der internen Qualitätssicherung zeitnahe, inhaltlich relevante Schnittstellen existieren.

Zusammenfassend lässt sich in Anlehnung an die bereits erwähnte Qualitätsmanagement-Norm DIN EN ISO 9000 konkretisieren:

**Versorgungsqualität** im Rahmen der extern vergleichenden Qualitätssicherung ist definiert als der Grad der Erfüllung der legitimen Anforderungen, die durch das jeweilige QS-Verfahren normativ vorgegeben sind.

Mit dieser Präzisierung wird die Suche nach relevanten Kriterien für gute Versorgungsqualität zu einer Frage der Identifikation legitimer Anforderungen an die medizinische Versorgung. Die „Methodischen Grundlagen“ legen hier in Analogie zur DIN-EN-ISO-Norm fest, wie die Versorgungsqualität durch entsprechende Anforderungen definiert wird (siehe Abbildung 2). Qualität gilt als Erfüllen der Anforderungen, in der Abbildung dargestellt als Pfeil zwischen der Versorgung und den Anforderungen. Um diese zu definieren werden im Handlungskontext des IQTIG zunächst Qualitätsaspekte für ein neues QS-Verfahren identifiziert und daraus ein sogenanntes Qualitätsmodell zusammengestellt (siehe Abschnitt 3.2), das dem G-BA als Ergebnis einer Konzeptstudie zur Beschlussfassung vorgelegt wird (siehe Abschnitt 4.1). Aus diesem Qualitätsmodell werden die Qualitätsanforderungen abgeleitet. Mit der Entwicklung von Indikatoren werden die Anforderungen präzisiert und die Qualitätsindikatoren dann zur Messung und Bewertung der Erfüllung der Anforderungen durch die einzelne konkrete eingesetzt. Die oben dargestellte Konkretisierung der Normen-Definition von Qualität nach DIN EN ISO 9000:2015 als Versorgungsqualität verankert somit den Qualitätsbegriff im konkreten Verfahrensablauf der gesetzlichen Qualitätssicherung nach §§ 136 ff. SGB V.

Abbildung 2: Integration des Qualitätsbegriffs (Erfüllung von Anforderungen) in die Prozessabläufe der externen Qualitätssicherung zur Definition und Bewertung von Versorgungsqualität



In den folgenden Abschnitten wird erläutert, wie im Rahmen der Verfahrensentwicklung solche Anforderungen definiert, legitimiert, konkretisiert, deren Erreichen mess- und bewertbar gemacht und schließlich in einem QS-Verfahren angewandt werden.

## 2.4 Verfahrenszwecke der Qualitätssicherung

Qualitätssicherungsverfahren können – wie oben bereits angesprochen – unterschiedliche Zwecke haben, die sich aus den jeweiligen Beauftragungen durch den G-BA ergeben.

- Qualitätsförderung (z. B. traditionelle Verfahren der externen stationären Qualitätssicherung)
- Information (z. B. Patienteninformation über Versorgungsqualität einzelner Leistungserbringer)
- Regulierung durch Planung der Versorgungsstrukturen (planungsrelevante Qualitätsindikatoren)
- Regulierung durch Selektivverträge mit Anreizen (Qualitätsverträge nach § 110a SGB V)
- Regulierung durch kollektivvertragliche Qualitätszu- und -abschläge

Entsprechend den unterschiedlichen Zwecken müssen die Mess- und Bewertungsinstrumente sehr unterschiedlichen Anforderungen genügen. Dies hat Auswirkungen auf die Entwicklung und Ausgestaltung der Verfahren.

### 2.4.1 Qualitätsförderung

Die Qualitätsförderung, wie sie in den QS-Verfahren der externen stationären Qualitätssicherung von den Landesgeschäftsstellen (und künftig – auch sektorenübergreifend – durch die Landesarbeitsgemeinschaften) in Zusammenarbeit mit den Einrichtungen praktiziert wird, dient dem gezielten Feedback der Qualitätsergebnisse und der Unterstützung des internen Qualitätsmanagements der einzelnen Einrichtungen. Diese Maßnahmen entsprechen konzeptuell einer gezielten Förderung der Qualitätsentwicklung bei den Leistungserbringern mit entsprechenden Möglichkeiten zum kollegialen Dialog, zum Erfahrungsaustausch, zum Lernen von den Besten, aber auch zur Zielvereinbarung. Qualitätsförderung ist stets Verfahrenszweck, insbesondere bei fallbezogenen Indikatoren, bei denen es um Prozess- und Ergebnisqualität geht. Alle Qualitätsindikatoren sind zur Qualitätsförderung geeignet, dieser Verfahrenszweck ist nicht gesondert zu beauftragen.

### 2.4.2 Information

Transparenz ist grundlegende Erfordernis bei den QS-Verfahren und kein besonderer Verfahrenszweck. Sie steht für eine transparente Verfügbarkeit der Qualitätsergebnisse im Rahmen der gesetzlichen Regelungen. Allerdings hat sich gezeigt, dass die Fülle dieser Einzelergebnisse so groß ist, dass sie kaum von der Öffentlichkeit genutzt werden kann. Unüberschaubarkeit führt durch die Menge an Informationen zu einer gewissen Intransparenz.

Von der Transparenz wird daher die gezielte Information im Sinne eines „Public Reporting“ abgegrenzt. Informationen werden aus vorhandenen Daten so abgeleitet, dass ihre Inhalte in bestimmten Informations- und Entscheidungssituationen hilfreiche Antworten und für Entscheidungen eine handhabbare Grundlage bieten. Hierzu ist zu definieren, welche Informationen Personen in typischen Auskunft- und Entscheidungssituationen benötigen und in welcher Form sie am besten nutzbar sind. Als Beispiele seien genannt die Einrichtung einer laienverständlichen Internetseite, anhand derer Patientinnen und Patienten die Versorgungsqualität verschiedener Einrichtungen vergleichen können, um sich für eine Behandlung die geeignetste zu suchen, oder

die Nutzung von Qualitätssiegeln und Zertifikaten zur Information der Patientinnen und Patienten über besonders qualifizierte Versorgungseinrichtungen.

Transparenz und Informationsangebote sollten stets gleichzeitig verfügbar sein. Informationen sind geeignete Zusammenfassungen einer Fülle transparenter Einzeldaten, die dadurch handhabbar werden. Andererseits legitimieren sich die Informationsangebote dadurch, dass ihre Herleitung aus verfügbaren Einzeldaten transparent überprüfbar ist.

Die Schaffung von Informationen greift auf vorhandene Einzeldaten zurück und prüft diese auf Nutzbarkeit, ggf. durch geeignete Transformation, Aggregation und Erläuterung. Es kann durchaus sein, dass Informationsangebote auch Daten verwenden, die nicht als Qualitätsergebnisse vorliegen, aber dennoch z. B. für Patientinnen und Patienten bei der Auswahl von Gesundheitseinrichtungen für ihre Behandlung von Nutzen sein können. Darunter fallen z. B. Informationen zum Versorgungsangebot oder zur Erreichbarkeit einer Einrichtung. Prinzipiell sind alle Qualitätsindikatoren zur gezielten Information nutzbar. Es ist allerdings zu prüfen, ob die Fragestellungen und Ergebnisse der Indikatoren in ausreichender Weise Antworten auf die Belange der Patientinnen und Patienten geben können.

### **2.4.3 Regulierung durch Planung der Versorgungsstrukturen**

Nach § 6 Abs. 1 Krankenhausfinanzierungsgesetz (KHG) stellen die Länder Krankenhauspläne auf, um nach § 1 KHG „eine qualitativ hochwertige, patienten- und bedarfsgerechte Versorgung der Bevölkerung mit leistungsfähigen, qualitativ hochwertig und eigenverantwortlich wirtschaftenden Krankenhäusern zu gewährleisten“. Um die Versorgungsqualität der Einrichtungen bei der Krankenhausplanung berücksichtigen zu können, entwickelt das IQTIG im Auftrag des G-BA nach § 136c Abs. 2 Satz 1 SGB V entsprechende Qualitätsindikatoren sowie Maßstäbe und Bewertungskriterien für die Qualitätsergebnisse. Einrichtungen, die „nicht nur vorübergehend eine in einem erheblichen Maß unzureichende Qualität aufweisen, dürfen insoweit ganz oder teilweise nicht in den Krankenhausplan aufgenommen werden.“ (§ 8 Abs. 1a KHG). „Plankrankenhäuser, die nach den in Absatz 1a Satz 1 genannten Vorgaben nicht nur vorübergehend eine in einem erheblichen Maß unzureichende Qualität aufweisen [...], sind insoweit durch Aufhebung des Feststellungsbescheides ganz oder teilweise aus dem Krankenhausplan herauszunehmen“ (§ 8 Abs. 1a KHG).

Die für diesen Verfahrenszweck zu identifizierenden oder neu zu entwickelnden Qualitätsindikatoren müssen sich auf planungsrelevante Versorgungsaspekte beziehen und in der Lage sein, „in einem erheblichen Maß unzureichende Qualität“ nachweisen zu können. Dabei sind sowohl die Planungsrelevanz als auch die Kriterien für das vom Gesetz vorgegebene Qualitätsniveau zu definieren. Hierzu hat der G-BA beim IQTIG ein umfassendes Konzept in Auftrag gegeben, dessen Fertigstellung für Ende April 2018 geplant ist. Seit 1. Januar 2017 ist jedoch bereits ein erstes Pilotprojekt der planungsrelevanten Qualitätsindikatoren in den Leistungsbereichen Gynäkologie, Geburtshilfe und Mammachirurgie in der Routine. Dieses greift zurück auf Qualitätsindikatoren, die im Rahmen der QSKH-RL bereits genutzt werden. Im ersten Konzept für den Verfahrenszweck Krankenhausplanung (IQTIG 2016) wurden erstmals Kriterien für die Auswahl der Leistungsbereiche, für die Definition von Planungsrelevanz, für die Feststellung unzureichender

Qualität sowie für die Auswahl geeigneter Qualitätsindikatoren festgelegt. Da diese bis zum Abschluss des Gesamtkonzepts im Frühjahr 2018 jedoch nur einen Zwischenstand abbilden, wird hier auf den entsprechenden Abschlussbericht des IQTIG vom August 2016 verwiesen, ohne dessen Inhalte bereits in die „Methodischen Grundlagen“ zu integrieren.

Da die Qualitätsindikatoren dieses ersten Konzepts sich nur auf solche der QSKH-RL beziehen, handelte es sich um fallbezogene Qualitätsindikatoren, die etwas zur Versorgungspraxis aussagen. Für die Krankenhausplanung sind allerdings auch strukturbezogene Qualitätsaspekte von Bedeutung, die in der weiteren Entwicklung der planungsrelevanten Qualitätsindikatoren berücksichtigt werden.

#### **2.4.4 Regulierung durch Selektivverträge mit Anreizen**

Bei den Qualitätsverträgen nach § 110a SGB V handelt es sich um Selektivverträge, die von Krankenkassen mit ausgewählten Krankenhäusern in vier ausgewählten Versorgungsbereichen geschlossen werden können. „Ziel der Qualitätsverträge ist die Erprobung, inwieweit sich eine weitere Verbesserung der Versorgung mit stationären Behandlungsleistungen, insbesondere durch die Vereinbarung von Anreizen sowie höherwertigen Qualitätsanforderungen erreichen lässt.“ (§ 110a SGB V). Die Wahl und die Bemessung der Anreize obliegen den vertragsschließenden Krankenkassen und Krankenhäusern. Der G-BA wählte durch Plenumsbeschluss am 18. Mai 2017 folgende vier Versorgungsbereiche für die Erprobung von Qualitätsverträgen aus:

- Endoprothetische Gelenkversorgung
- Prävention des postoperativen Delirs bei der Versorgung von älteren Patientinnen und Patienten
- Respiratorentwöhnung von langzeitbeatmeten Patientinnen und Patienten
- Versorgung von Menschen mit geistiger Behinderung oder schweren Mehrfachbehinderungen im Krankenhaus.

Die Fertigstellung des Verfahrenskonzepts ist für Dezember 2017 vorgesehen, sodass erst dann Aussagen zur Methodik dieses besonderen Verfahrensbereichs gemacht werden können.

#### **2.4.5 Regulierung durch Qualitätszu- und -abschläge**

§ 136b SGB V regelt, dass „der Gemeinsame Bundesausschuss [...] ein Verfahren“ regelt, „das den Krankenkassen und den Krankenhäusern ermöglicht, [...] Qualitätszuschläge für außerordentlich gute und Qualitätsabschläge für unzureichende Leistungen zu vereinbaren. Hierfür hat er insbesondere jährlich Bewertungskriterien für außerordentlich gute und unzureichende Qualität zu veröffentlichen“. Durch Beschluss des G-BA vom 20. Oktober 2016 soll „das zugrundeliegende Konzept [...] auch eine indikatorenübergreifende Bewertung ermöglichen“.

Weiter ist im § 5 Abs. 3a Krankenhausentgeltgesetz (KHEntG) geregelt:

*Qualitätszu- oder -abschläge sind bezogen auf die betreffenden Leistungen oder Leistungsbereiche jeweils für Aufnahmen ab dem ersten Tag des Folge-monats der Vereinbarung anzuwenden. Sie sind für Aufnahmen bis zum letzten Tag des Monats anzuwenden, in dem die Vertragsparteien nach Satz 1*

*feststellen, dass die Voraussetzungen für die weitere Erhebung von Qualitätszu- oder -abschlägen nicht mehr vorliegen. Sofern die Vertragsparteien nach Satz 1 unzureichende Qualität feststellen, hat die Vereinbarung auch zu beinhalten, dass die Qualitätsmängel innerhalb eines Jahres ab dem Vereinbarungszeitpunkt zu beseitigen sind; in dieser Zeit sind keine Qualitätsabschläge zu erheben. Werden die Qualitätsmängel nicht innerhalb eines Jahres beseitigt, ist der vereinbarte Qualitätsabschlag zu erheben, sofern die Vertragsparteien feststellen, dass die Voraussetzungen für die Erhebung weiterhin vorliegen; dabei ist der Qualitätsabschlag für die Dauer von zwölf Kalendermonaten in doppelter Höhe zu erheben. Die zeitliche Begrenzung für die Erhebung von Qualitätsabschlägen nach § 8 Absatz 4 Satz 2 Nummer 2 auf höchstens drei Jahre ist zu beachten.*

Nach § 8 Abs. 1b KHG sind

*Plankrankenhäuser, die nach den in Absatz 1a Satz 1 genannten Vorgaben nicht nur vorübergehend eine in einem erheblichen Maß unzureichende Qualität aufweisen oder für die in höchstens drei aufeinanderfolgenden Jahren Qualitätsabschläge nach § 5 Absatz 3a des Krankenhausentgeltgesetzes erhoben wurden, [...] insoweit durch Aufhebung des Feststellungsbescheides ganz oder teilweise aus dem Krankenhausplan herauszunehmen.*

Vorbericht und Abschlussbericht zum ersten Entwicklungsschritt wurden zeitgerecht am 30. April 2017 bzw. am 30. Juni 2017 beim G-BA vorgelegt. Auf dessen Wunsch wird derzeit der Abschlussbericht erweitert mit aktualisiertem Abgabetermin am 15. Oktober 2017. Daher kann erst danach über die Methodik bei Qualitätszu- und -abschlägen berichtet werden. Insbesondere ist im dritten Schritt des G-BA-Auftrags bis Ende 2018 die Erstellung eines Gesamtkonzepts zur Methodik der Qualitätszu- und -abschläge vorgesehen, das die gesamte Thematik methodisch betrachten wird. Das im ersten Bericht am 15. Oktober 2017 vorzuschlagende Verfahren wird daher ein erstes Pilotprojekt sein.

#### **2.4.6 Zusammenfassung: Verfahren mit unterschiedlichen Verfahrenszwecken**

Die Nutzung von Qualitätsmessungen und deren Ergebnissen bezog sich bislang auf die Qualitätsförderung und die Veröffentlichung dieser Ergebnisse in den Qualitätsberichten der Krankenhäuser. Der Gesetzgeber hat die oben genannten drei neuen Verfahrenszwecke in der gesetzlichen Qualitätssicherung eingeführt:

- **Planungsrelevante Qualitätsindikatoren**  
Verfahrenszweck: **Strukturplanung** im Gesundheitswesen
- **Qualitätszu- und -abschläge**  
Verfahrenszweck: **finanzielle Anreize** für eine Verbesserung der Versorgungsqualität, andererseits Einschränkung des Feststellungsbescheids und der Finanzierung bestimmter Leistungen bei anhaltend schlechter Versorgungsqualität

#### ▪ **Qualitätsverträge nach § 110a SGB V**

Verfahrenszweck: **selektivvertragliche Anreize** (ggf. unterschiedlicher Art) für außerordentlich gute Versorgungsqualität

Das, was sie wesentlich unterscheidet, sind die unterschiedlichen Handlungsanschlüsse, die die Qualitätsergebnisse zur weiteren Steuerung des Gesundheitswesens auf je eigene Weise nutzen. Die Qualitätsmessungen werden nicht durch ein unterschiedliches Verständnis von Versorgungsqualität geprägt, sondern dadurch, welche Leitdifferenzen in der Qualitätsbewertung benötigt werden, um entsprechende Handlungen daran anknüpfen zu können: wann ist Versorgungsqualität planungsrelevant, wann abschlagspflichtig, wann zuschlagswürdig, wann förderungswürdig? Welche Aspekte der Versorgung eignen sich, um planungsrelevant zu sein, welche eignen sich für Förderung? Hier sind für jeden der genannten drei neuen Verfahrenszwecke spezifische Kriterien zu entwickeln. Dies ist methodisch aber erst mit den beauftragten Gesamtkonzepten zu erreichen, die noch in Entwicklung sind. Es können daher an dieser Stelle nur erste Hinweise gegeben werden.

#### **Differenzierung der Bewertungskategorien**

Die klassische Qualitätsförderung unterscheidet bislang lediglich qualitativ unauffällige und qualitativ auffällige Ergebnisse, also gute und nicht ausreichende Qualität. Das KHSG identifiziert dagegen beispielsweise im Rahmen von qualitätsbezogenen Zu- und Abschlägen außerordentlich gute und am anderen Ende der Skala unzureichende Qualität und damit indirekt auch die dazwischen liegende, durchschnittlich gute Qualität. Damit wird deutlich, dass zukünftig weitere, über die Unterscheidung von guter und nicht ausreichender Qualität hinausgehende Differenzierungen notwendig werden.

Eine solche feiner granulいた Einstufung kann zum Beispiel bei der Publikation von Qualitätsergebnissen von Bedeutung sein und wird bereits in verschiedenen Formen in Krankenhausportalen angewandt (z. B. Qualitätskliniken.de, AOK-Krankenhausnavigator).

Eine Differenzierung im Bereich guter bis überdurchschnittlicher Qualität verlagert dabei den Schwerpunkt der Qualitätsbetrachtung. Es geht nicht mehr einzig darum, die Leistungserbringer mit unzureichender Qualität zu identifizieren und deren Mängel zu beheben, sondern auch darum, die Entwicklung von hervorragender Qualität sichtbar zu machen und zu fördern. Damit geht die Qualitätssicherung über die Sicherung der Einhaltung von Standards hinaus. Auch im Rahmen planungsrelevanter Qualitätsindikatoren findet sich eine feinere Differenzierung der Versorgungsqualität: Minimalanforderungen bei Planungsentscheidungen, Ausschlusskriterien bei Interventionsentscheidungen und Exzellenzaspekte bei Vergabeentscheidungen. Im KHG wird für Interventionen der Planungsbehörden von einer „nicht nur vorübergehend in einem erheblichen Maß unzureichende[n] Qualität“ (§ 8 Abs. 1a) gesprochen. Hier wird Versorgungsqualität sowohl nach Schweregrad als auch nach Zeitdauer differenziert.

Die Kategorien der Bewertung von Qualität hängen also vom jeweiligen Verfahrenszweck und den damit verbundenen Handlungsoptionen ab. Die Bewertungskategorien, die u. a. integraler Bestandteil von Qualitätsindikatoren sind, stellen die Schnittstelle zu den Handlungsoptionen

dar, für die die Qualitätsmessung durchgeführt wird. Wer planen will, muss wissen, welche planungsrelevanten Ergebnisse vorliegen, und wer fördern will, muss wissen, was förderungswürdig ist. Aufgrund der noch ausstehenden umfassenden Konzeptentwicklungen können die methodischen Grundlagen erst in der nächsten Version dieses Dokuments dargestellt werden. Bis dahin geben die Einzelkonzepte der vorgeschalteten Pilotprojekte Auskunft über die bislang entwickelten Methoden und Kriterien auf diesem Gebiet.

### **Eignung von Verfahren und Versorgungsbereichen für spezifische Verfahrenszwecke**

Auch wenn die umfassenden Konzepte für die verschiedenen Verfahrenszwecke derzeit noch in Entwicklung sind, lassen sich aus den bisherigen Kriterien zur Auswahl der Pilotprojekte bereits erste relevante Fragestellungen für die Auswahl von Verfahrenszwecken, Versorgungsbereichen und Qualitätsaspekten nennen, ohne Anspruch auf systematische Abgeschlossenheit oder Vollständigkeit:

- **Erwartbarer Nutzen**
  - das Potenzial einer Qualitätsverbesserung in diesem Versorgungsbereich
  - die Dringlichkeit einer Qualitätsverbesserung in diesem Versorgungsbereich
  - die Eignung des gewählten Verfahrens, eine solche Qualitätsverbesserung zu bewirken
- **Verfügbarkeit geeigneter Indikatoren**
  - thematisch passende Indikatoren für den jeweiligen Verfahrenszweck (z. B. für die Krankenhausplanung oder zu fördernde Exzellenzqualität)
  - Differenzierungsmöglichkeit im relevanten Ergebnisbereich der Indikatoren
- **Umsetzungsaspekte**
  - Anzahl betroffener Leistungserbringer (auch erwartbarer Nutzen)
  - Fallzahlen bei den betroffenen Leistungserbringern (auch erwartbarer Nutzen)
  - Verfügbarkeit relevanter Daten
  - zeitliche Aspekte der Verfügbarkeit von Daten und Ergebnissen (z. B. bei Follow-up-Zeiten (z. B. auch Eignungskriterium Brauchbarkeit bei den Indikatoren)
  - Kosten-Nutzen-Abwägungen und Umsetzbarkeitsaspekte
  - Besonderheiten bei sektorenübergreifenden Verfahren, insbesondere wenn es um einrichtungsübergreifende Systemqualität geht

Die Komplexität der Anforderungen, die an eine Auswahl von Verfahrenszweck, Versorgungsbereich und Qualitätsaspekten zu stellen sind, machen es erforderlich, dass noch vor der Festlegung des Qualitätsmodells (siehe Abschnitt 3.2) eines neuen QS-Verfahrens, das ja Grundlage der Indikatorenentwicklung ist, der Verfahrenszweck durch den G-BA festzulegen ist. Ergibt sich dies noch nicht aus dem Kriterienkatalog des Themenfindungs- und Priorisierungsverfahrens, steht künftig das IQTIG dem G-BA vor der Beauftragung einer Verfahrensentwicklung auf der Basis der Gesamtkonzepte für die einzelnen Verfahrenszwecke beratend zur Verfügung.

Während aber bislang immer die Frage bestand, welche Versorgungsbereiche sich für die Entwicklung eines Verfahrens eignen, wird künftig gefragt sein, welche Steuerungsinstrumente und

somit Verfahrenszwecke für welche Qualitätsdefizite am geeignetsten erscheinen, nachhaltige Verbesserungen oder Entwicklungen zu initiieren. Da sich insbesondere die regulierenden Verfahren noch in der Entwicklung befinden, ist auch hier ein gewisser Lernprozess zu erwarten.

## 3 Elemente der Verfahrensentwicklung

### 3.1 Qualitätsaspekte der Gesundheitsversorgung

Die Bewertung von Versorgungsqualität beginnt mit einer Beschreibung der Gesundheitsversorgung in einem ausgewählten Versorgungsbereich im Hinblick auf Aspekte der Qualität. Bereits unter Abschnitt 2.1.2 waren Systematiken solcher Aspekte vorgestellt worden.

Diese Aspekte, mit denen die Beschreibung eines Versorgungsbereichs beginnt, reichen von der Erwartung der professionellen Einhaltung leitlinienbasierter Standards der Diagnostik und Therapie über die Hoffnung der Patientinnen und Patienten auf schnelle Gesundung bis hin zur Bilanz von Kosten und Erlösen der versorgenden Einrichtungen oder dem funktionierenden Zusammenspiel der Versorgungsstrukturen einer Region. Die möglichen Blickweisen zu diesem Thema sind vielfältig.

Die Qualitätssicherung der Gesundheitsversorgung befasst sich primär mit den Aspekten, die die patientenorientierte Ergebnisqualität abbilden, denn gute Ergebnisse sind das Ziel jeder Behandlung im Rahmen der gegebenen Möglichkeiten bei Patientinnen und Patienten. Andere Qualitäten wie die Prozess- und Strukturqualität werden zusätzlich einbezogen, wenn sie mittelbar für das Erzielen guter, patientenorientierter Ergebnisse dienlich sind. Die Identifikation von patientenzentrierten Qualitätsaspekten stellt den Ausgangspunkt der Entwicklung von Qualitätssicherungsverfahren dar. Qualitätsaspekte sind konkrete Themen der Versorgungspraxis, wie z. B. Indikationsstellung zur Herzkatheteruntersuchung, Komplikationen bei Hüftoperation, anhand derer sich die Qualität der Gesundheitsversorgung beschreiben lässt und die durch konkrete Merkmale dann näher zu operationalisieren sind (siehe auch Abschnitte 4.1.2 ff.). Wie ein Aspekt in der täglichen Versorgung umgesetzt wird (z. B. Entlassungsmanagement mit Überleitung in die häusliche Versorgung), kann durch mehrere, sich ergänzende Qualitätsindikatoren gemessen und bewertet werden. Aspekte sollten sich idealerweise nicht überlappen und sind nur sinnvoll, wenn es in ihrem Themenfeld Verbesserungsbedarf und Verbesserungsmöglichkeiten gibt. Qualitätsaspekte, die für die Verfahrensentwicklung ausgesucht werden, sollten noch nicht so detailliert sein, dass sie bereits konkrete Anforderungen formulieren oder bereits Konkretisierungen der Messung beinhalten.

### 3.2 Qualitätsmodelle als Basis der QS-Verfahren

Im Rahmen der Verfahrensentwicklung werden zunächst für den ausgewählten Versorgungsbereich Qualitätsaspekte identifiziert. Anschließend wird anhand spezifischer Kriterien geprüft, ob diese sich für den beauftragten Verfahrenszweck eignen (Die Entwicklung dieser Kriterien erfolgt noch im Rahmen der Gesamtkonzepte insbesondere für planungsrelevante Qualitätsindikatoren und Qualitätszu- und -abschläge).

Die für einen Versorgungsbereich identifizierten und für den Verfahrenszweck geeigneten Aspekte werden in ihrer Gesamtheit als Qualitätsmodell bezeichnet. Ein Qualitätsmodell ergibt ein

am Patientennutzen orientiertes Gesamtbild der Versorgung im Rahmen des geplanten Qualitätssicherungsverfahrens mit besonderem Fokus auf Verbesserungspotenziale. Auch wenn sich in der späteren Entwicklung für einen Aspekt kein praktikabler Indikator finden lässt, bleibt der Aspekt Teil des Qualitätsmodells. Damit wird sichtbar, welche Aspekte in der Verfahrensroutine durch die Qualitätsindikatoren abgebildet werden und welche nicht berücksichtigt werden können. Dies ist für das Gesamtverständnis der Qualitätsergebnisse, ihre Aussagekraft und ihre Limitationen von Bedeutung.

Wird für einen Versorgungsbereich ein weiteres QS-Verfahren mit einem anderen Verfahrenszweck beauftragt, so sind die zu Anfang gefundenen Aspekte erneut dahingehend zu prüfen, ob sie für diesen neuen Zweck relevant sind. Für die Konstellation von bekanntem Versorgungsbereich und neuem Verfahrenszweck wird ein entsprechend neues, spezifisches Qualitätsmodell festgelegt.

### **3.3 Qualitätsmerkmale als erster Schritt der Konkretisierung**

Qualitätsmerkmale stellen einen Zwischenschritt in der Entwicklung von Qualitätsindikatoren auf der Basis von Qualitätsaspekten dar. Die Qualitätsmerkmale sind Eigenschaften der Versorgung, die sich auf Anforderungen beziehen. In der Norm DIN EN ISO 9000:2005 (und auch später) wird dies so ausgedrückt: ein Qualitätsmerkmal ist ein „inhärentes Merkmal eines Produkts, Prozesses oder Systems, das sich auf eine Anforderung bezieht.“ (zitiert nach Sens et al. 2007). Merkmal und Anforderung entstehen im Entwicklungsprozess meist gekoppelt. Die Anforderung, dass zeitnah nach Entschluss für einen Notkaiserschnitt die Entbindung erfolgen sollte, hat gleich auch die entsprechende Anforderung an die sogenannte E-E-Zeit und ihre Befristung im Blick. Ab dann steht das Merkmal E-E-Zeit für die Anforderung nach schneller Entbindung in einer Notfallsituation mit Entschluss zur Notsectio. Die Verwendung des Begriffs ist daher konform mit der DIN-EN-ISO-Norm.

Für einen Qualitätsaspekt lassen sich meist mehrere, sich nicht überlappende Qualitätsmerkmale finden. Beispielsweise stellen „Information über den Grund der Untersuchung“ und „Information über Behandlungsalternativen“ konkretisierte Merkmale des Qualitätsaspekts „Patienteninformation“ dar. Qualitätsmerkmale weisen, im Gegensatz zu den Qualitätsaspekten, spezifische Qualitätsziele auf (z. B. „Patienten sollen über Behandlungsalternativen aufgeklärt sein“). Hier wird deutlich, warum die Suche nach Optionen der Konkretisierung nicht schon auf der Ebene der Aspekte zu sehr eingeschränkt werden darf. Erst die Qualitätsmerkmale formulieren konkrete, legitime Anforderungen an die medizinische Versorgung für einen spezifischen Aspekt, anhand derer die Versorgungsqualität beurteilt werden kann. Auf Basis dieser Qualitätsmerkmale erfolgt die Entwicklung von Qualitätsindikatoren.

### 3.4 Qualitätsindikatoren

Die Messung der Qualität medizinischer Versorgung erfolgt grundsätzlich mittels Qualitätsindikatoren. Qualitätsindikatoren sind Konstrukte, die aus Versorgungsdaten nachvollziehbare Bewertungen der Versorgungsqualität ableiten. Sie bestehen immer aus drei Komponenten,

- einem konkreten Ziel für die Versorgungsqualität (z. B. „die Häufigkeit von Komplikationen soll möglichst gering sein“),
- einem spezifischen Dokumentations- und Messverfahren für das Qualitätsmerkmal (z. B. Spezifikation, Dokumentation, Rechenregel) sowie
- einem Bewertungskonzept, das die Zielerreichung bewertet und einen Handlungsanschluss bietet (z. B. Referenzbereich einschließlich Klassifikationsverfahren).

Die Ermittlung rechnerischer Ergebnisse aus Versorgungsdaten führt zu Kennzahlen. Erst das Hinzukommen sowohl einer klaren Definition des Qualitätsziels als auch eines definierten Verfahrens, das eine sinnvolle Wertung der Messergebnisse ermöglicht, erweitert die Kennzahl zum Qualitätsindikator.

Durch die indikatorengeliteten Bewertungsergebnisse wird dann der Anschluss zu möglichen Handlungskonsequenzen geschaffen, die letztlich die Verbesserung und Weiterentwicklung der Gesundheitsversorgung bewirken sollen.

# **Teil B: Entwicklung und Durchführung von Qualitätssicherungsverfahren**

## 4 Neuentwicklung von Qualitätssicherungsverfahren

Die Entwicklung neuer Verfahren für die Qualitätssicherung nach §§ 136 ff. SGB V zur Messung und Darstellung der Versorgungsqualität mittels Qualitätsindikatoren ist eine Kernaufgabe des IQTIG. In diesem Kapitel wird die grundlegende Methodik der Entwicklung solcher Qualitätssicherungsverfahren (QS-Verfahren) dargestellt. Die Methoden der diesen Entwicklungen zugrunde liegenden Elemente (Methoden der Stichprobenziehung, der Informationsrecherche, Auswertungsmethoden etc.) werden in Teil C separat dargestellt.

Die Qualitätssicherung gliedert sich in drei Hauptabschnitte:

- **Die Entwicklung der Instrumente:** Ausgehend von der Festlegung der relevanten Qualitätsaspekte eines Versorgungsbereichs wird ein konkretes Set an Qualitätsindikatoren entwickelt.
- **Die Qualitätsbewertung der Patientenversorgung:** Aus Versorgungsdaten wird mithilfe der Qualitätsindikatoren eine qualitative Bewertung der Versorgungsqualität abgeleitet.
- **Die qualitätsbezogenen Handlungskonsequenzen:** An die Bewertungen der Versorgungsqualität schließen sich die informierenden, die fördernden und die regulierenden Instrumente der Qualitätssicherung und -entwicklung an.

Zwischen dem ersten, konzeptionellen Teil und den folgenden Teilen der Durchführung der Qualitätssicherung bilden Versorgungsdaten den Verknüpfungspunkt. Sie sind Teil des Konzepts und zugleich Grundlage aller Praxis.

Für Neuentwicklungen von QS-Verfahren stehen derzeit drei verschiedene Standarddatenquellen zur Verfügung:

- **Dokumentationen der Leistungserbringer:** Bislang stützen sich QS-Verfahren vor allem auf die Dokumentation von Behandlungsfällen zu Zwecken der Qualitätssicherung durch die Leistungserbringer (fallbezogene QS-Dokumentation). Ebenfalls unmittelbar durch die Leistungserbringer dokumentiert werden Angaben auf Einrichtungsebene, die als einrichtungsbezogene QS-Dokumentation einmal jährlich durchgeführt werden. Darunter fallen z. B. Strukturangaben, interne Regelungen etc.
- **Sozialdaten bei den Krankenkassen:** Auf Grundlage des § 299 SGB V dürfen die Sozialdaten bei den Krankenkassen für die Zwecke der Qualitätssicherung genutzt werden. Hierzu zählen insbesondere Versichertendaten nach § 284 SGB V, die Daten zur Abrechnung ärztlicher Leistungen nach § 295 SGB V, Daten zur Verordnung von Arzneimitteln nach § 300 SGB V, Daten zu Krankenhausbehandlungen nach § 301 SGB V sowie Daten zu Heil- und Hilfsmitteln nach § 302 SGB V. Die Sozialdaten bei den gesetzlichen Krankenkassen umfassen nur Daten zu den GKV-Versicherten. Sie stehen erst mit einer Verzögerung von 3 Monaten (Krankenhausdaten) bis 9 Monaten (Daten der Vertragsärzte, die zuvor über die Kassenärztlichen Vereinigungen an die Krankenkassen weitergeleitet werden müssen) für die QS-Verfahren nach §§ 136 ff. zur Verfügung.

- Befragung von Patientinnen und Patienten: Nach §§ 136 ff. und § 299 SGB V dürfen Patientenbefragungen für die Zwecke der gesetzlich verbindlichen, externen Qualitätssicherung genutzt werden.

Mit diesen Datenquellen steht eine breite Informationsbasis zur Verfügung, um die qualitätsrelevanten Aspekte der Versorgung bewerten zu können. Dabei decken sie unterschiedliche, aber auch sich überlappende oder komplementäre Aspekte der Versorgung ab. Daten, die durch die Leistungserbringer dokumentiert werden, machen vor allem durchgeführte Prozeduren und klinische Detailparameter wie z. B. medizinische Diagnose- und Ergebnisaspekte abbildbar. Die Sozialdaten bei den Krankenkassen erlauben darüber hinaus, den weiteren Behandlungsverlauf einschließlich verschiedener Outcome-Parameter sowie Diagnosen auch im Rahmen von Follow-up-Erhebungen in die Qualitätsbewertungen einzubeziehen. Auf der anderen Seite stellen die Patientinnen und Patienten selbst eine weitere, sehr wichtige Informationsquelle für die Qualität der medizinischen Versorgung dar, die durch Patientenbefragungen zugänglich gemacht werden kann. So können Patientinnen und Patienten beispielsweise Auskunft über Behandlungsabläufe, Behandlungserfolg und Rahmenbedingungen der Behandlung geben.

Neben den Informationen, die die verschiedenen Datenquellen abdecken, unterscheiden sie sich auch im Hinblick auf den Erhebungsaufwand. So ist der Dokumentationsaufwand für die Leistungserbringer unter Umständen hoch, wenn in einem Verfahren viele fallbezogene Daten nochmals für die Qualitätssicherung erfasst werden sollen. Geringer bleibt der Aufwand bei einer einmal jährlich durchgeführten einrichtungsbezogenen QS-Dokumentation. Diese eignet sich jedoch nur zur Erfassung von Strukturen, prinzipiellen Verfahrensregelungen und vor Ort bereits vorliegenden aggregierten Ergebnisdaten (z. B. Datensets in Bezug auf Hygiene und Infektionen). Für die Nutzung der Sozialdaten bei den Krankenkassen fällt bei den Leistungserbringern dagegen kein zusätzlicher Aufwand an, da diese routinemäßig erfasst und an die Krankenkassen weitergeleitet werden. Auch für Patientenbefragungen fällt bei den Leistungserbringern nur ein geringer Aufwand an, da die Befragungen nicht durch die Leistungserbringer durchgeführt werden sollen. Dieser Aufwand fällt dann lediglich bei den Befragten in Form des Ausfüllens eines Fragebogens an und bei den beauftragten Einrichtungen, die diese Fragebögen versenden und jenen, die sie auswerten.

Die verschiedenen Aspekte der Datenquellen, wie z. B. zu erwartender Informationsgehalt und Erhebungsaufwand, werden bereits im ersten Schritt der Neuentwicklung von QS-Verfahren, der Konzeptstudie, berücksichtigt, in deren Rahmen eine Empfehlung bezüglich der geeignetsten Datenquellen für das QS-Verfahren erfolgt. Bei Bedarf können in der Verfahrensentwicklung aber auch weitere, neue Datenquellen im Rahmen der gesetzlichen Vorgaben erschlossen werden.

Die Verfahrensentwicklung im Bereich der Qualitätsförderung wird in Abbildung 3 schematisch dargestellt. Sie gliedert sich in eine konzeptionelle und eine empirische Phase.

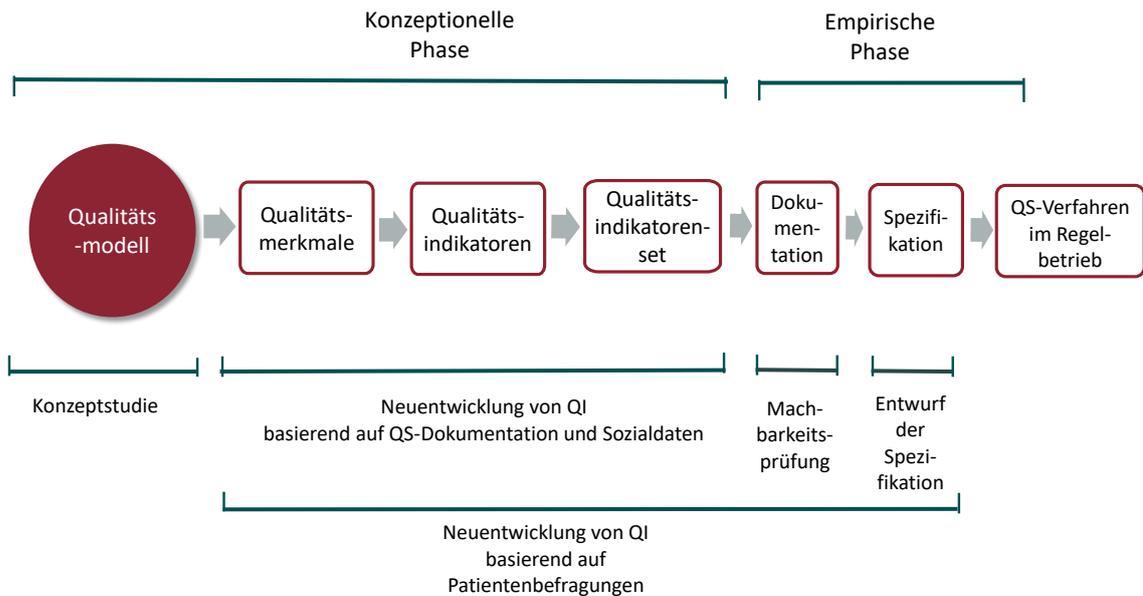


Abbildung 3: Phasen der Neuentwicklung von QS-Verfahren

Die konzeptionelle Phase beginnt mit der Konzeptstudie, an deren Ende das verfahrensspezifische und für alle folgenden Entwicklungsschritte verbindliche Qualitätsmodell steht. Darauf basierend erfolgt die Entwicklung des eigentlichen QS-Verfahrens. Ausgehend vom Qualitätsmodell, welches die selektierten Qualitätsaspekte umfasst, werden zunächst die Qualitätsmerkmale abgeleitet und anschließend die Qualitätsindikatoren entwickelt. Ziel der konzeptionellen Phase ist es, ein empfohlenes Qualitätsindikatorenset zu erarbeiten, das zentraler Bestandteil des QS-Verfahrens ist. Aufgrund der methodischen Unterschiede in der Entwicklung von Fragebögen zur Patientenbefragung einerseits und der Entwicklung von Qualitätsindikatoren auf Basis von Dokumentations- und Sozialdaten andererseits sind jeweils eigene, den Erfordernissen gerecht werdende Entwicklungsmethoden vorgesehen, die in den folgenden Abschnitten dargestellt werden. Diese Darstellungen beziehen sich dabei ausschließlich auf den Verfahrenszweck der Qualitätsförderung (siehe Abschnitt 2.4). Beschreibungen für andere Verfahrenszwecke werden zunächst in ihren spezifischen Entwicklungsberichten beschrieben und später in dieses übergeordnete Dokument integriert.

In der anschließenden empirischen Phase folgt eine Machbarkeitsprüfung, die insbesondere die Erfassungsinstrumente der Falldokumentation in der probenhaften Anwendung durch freiwillig teilnehmende Leistungserbringer für den Regelbetrieb optimieren soll. Die Entwicklung einer Patientenbefragung vereint immer den konzeptionellen Teil der Fragebogenentwicklung und den empirischen Teil der Fragebogenvalidierung. Die empirische Phase kann für beide Entwicklungsprodukte in das QS-Verfahren hineinreichen, sofern der G-BA in seiner Richtlinie eine Erprobungsphase des QS-Verfahrens im Regelbetrieb vorsieht. Diese ist dann jedoch nicht mehr Teil der Neuentwicklung, sondern wird im Rahmen des Regelbetriebs (siehe Abschnitt 5.2) durchgeführt. Am Ende aller in den folgenden Abschnitten beschriebenen Entwicklungen stehen Berichte an den Auftraggeber, den G-BA. In der dokumentations- und sozialdatenbasierte Neuentwicklung von Qualitätsindikatoren und in der Entwicklung von Patientenbefragungen wird

ein Stellungnahmeverfahren mit den zu beteiligenden Institutionen nach § 137a Abs. 7 SGB V durchgeführt.

Alle Verfahrensentwicklungen orientieren sich an folgenden grundsätzlichen Anforderungen:

- Einrichtungsvergleiche sollen ermöglicht werden.
- Das Verfahren soll möglichst sektorenübergreifend, interdisziplinär und interprofessionell sein.
- Das Verfahren soll auf einen bestimmten Verfahrenszweck (siehe Abschnitt 2.4) hin konzipiert werden.
- Die patientenrelevante Versorgungsqualität einschließlich der Ergebnisqualität der Behandlung soll einbezogen werden.
- Die Patientenperspektive soll in die Entwicklung einbezogen werden.
- Das Verfahren soll Verbesserungsbedarfe sichtbar machen.
- Durch die Indikatorenergebnisse sollen Handlungen bzw. Entscheidungen angestoßen werden.
- Das Verfahren soll möglichst datensparsam sein.
- Das Verfahren soll möglichst aufwandsarm für die Leistungserbringer sein.
- Das Verfahren soll transparent sein.

#### **4.1 Konzeptstudie für Qualitätssicherungsverfahren**

Die Entwicklungsarbeit für ein neues QS-Verfahren beginnt mit einer Konzeptstudie. Diese wird durch den G-BA beauftragt und baut ggf. auf einem Kriterienkatalog des G-BA zur Themenfindung und Priorisierung auf. Entscheidend ist, dass möglichst bei der ersten Beauftragung durch den G-BA der Verfahrenszweck (in den Kategorien Qualitätsförderung, Information oder Regulierung, siehe Abschnitt 2.4) für das spätere QS-Verfahren durch den G-BA schon verbindlich vorgegeben wird. Ist dies nicht a priori möglich, weil noch nicht ausreichende Informationen zur Verfügung stehen, soll während der Erstellung des Kriterienkatalogs durch den G-BA eine Entscheidung anhand der bis dahin gewonnenen Erkenntnisse getroffen werden. Das IQTIG steht dem G-BA dabei beratend zur Verfügung. Nur mit einer Festlegung des Verfahrenszwecks kann das zu entwickelnde Konzept und in der Folge die Neuentwicklung des QS-Verfahrens optimal auf diesen Verfahrenszweck hin ausgerichtet werden (siehe Abschnitt 2.4). Unterschiedliche Verfahrenszwecke führen zu verfahrensspezifischen Qualitätsmodellen und erfordern darüber hinaus ein spezifisches methodisches Vorgehen (siehe Abschnitt 3.2).

Die Funktion einer Konzeptstudie ist die inhaltliche Skizzierung eines neuen QS-Verfahrens durch:

- die Entwicklung eines Qualitätsmodells, das spezifisch auf das vom G-BA vorgegebene Thema und den Verfahrenszweck ausgerichtet ist und der weiteren Verfahrensentwicklung verbindlich zugrunde gelegt werden soll
- die Identifizierung von Datenquellen und Instrumenten, die hierfür genutzt bzw. neu entwickelt werden sollten
- die Beurteilung, ob das geplante Verfahren in dieser Form grundsätzlich umsetzbar erscheint
- die Darstellung des voraussichtlich zu erwartenden Aufwand-Nutzen-Verhältnisses

Dadurch soll sichergestellt werden, dass ein zur weiteren Entwicklung empfohlenes QS-Verfahren einerseits ausreichend auf relevante Verbesserungspotenziale fokussiert, andererseits mögliche Umsetzungshürden frühzeitig erkannt und bewertet werden können. Eine Verfolgung von unangemessen aufwendigen oder vor dem Hintergrund der Versorgungsrealität nicht durchführbaren Verfahrensansätzen soll auf diese Weise vermieden werden. Die Konzeptstudie schafft damit die notwendigen systematischen Grundlagen für eine anschließende Neuentwicklung von Qualitätsindikatoren.

Die Entwicklung eines Qualitätsmodells gliedert sich in drei Entwicklungsschritte: Die Themenererschließung der Versorgungspraxis, die Identifikation der Qualitätsaspekte und die Selektion der Qualitätsaspekte für das Qualitätsmodell (siehe Abbildung 4).

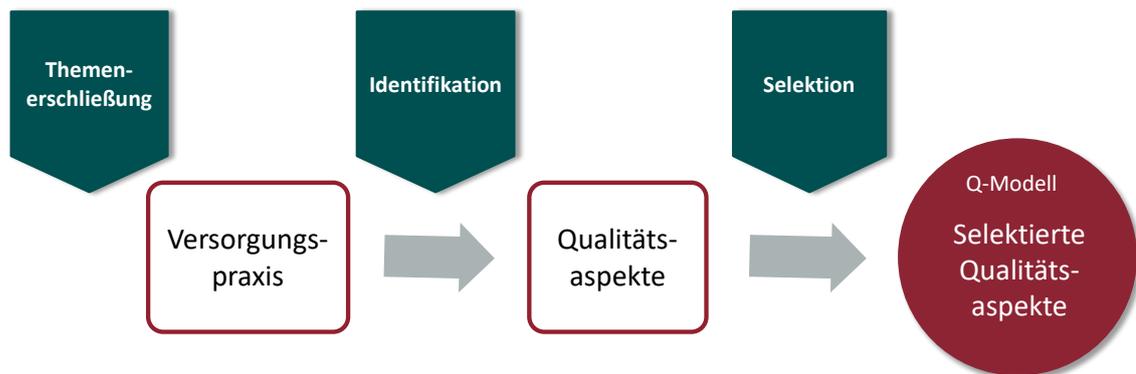


Abbildung 4: Ablaufschema der Entwicklung des Qualitätsmodells im Rahmen einer Konzeptstudie für ein neues QS-Verfahren

Nach der Themenererschließung – einer umfassenden Analyse und Bewertung der Versorgungspraxis des adressierten Versorgungsbereichs – beginnt die Entwicklung des verfahrensspezifischen Qualitätsmodells zunächst mit der Festlegung der Zielpopulationen hinsichtlich der Patientinnen und Patienten und der Festlegung der voraussichtlich vom Verfahren betroffenen Leistungserbringer. Daraufhin werden patientenrelevante Qualitätsaspekte aus der Analyse der ärztlichen, pflegerischen und weitere Professionen betreffenden Versorgungspraxis abgeleitet. Dabei wird von Anfang an die Perspektive von Patientinnen und Patienten in die konzeptionellen Überlegungen einbezogen. Die zu definierenden Qualitätsaspekte sollen sich stets auf die Ergebnisqualität und Prozessqualität beziehen und können – soweit in der Beauftragung des G-BA nicht anders formuliert – auch die Strukturqualität einschließen. Für das Qualitätsmodell werden jene Qualitätsaspekte ausgewählt, für die ein Qualitätsverbesserungsbedarf und die grundsätzliche Beeinflussbarkeit durch die Leistungserbringer dargelegt werden können. Den ausgewählten Qualitätsaspekten werden schließlich die im Hinblick auf Umsetzbarkeit und Aufwand jeweils am besten geeigneten Datenquellen und Erfassungsinstrumente zugeordnet. Dabei können neben den etablierten Instrumenten (fallbezogene QS-Dokumentation, einrichtungsbezogene QS-Dokumentation, Sozialdaten bei den Krankenkassen, Patientenbefragung) auch neu zu entwickelnde Erfassungsinstrumente vorgeschlagen werden. Aus diesem Vorgehen heraus

ergibt sich abschließend eine sorgfältige Abschätzung der grundsätzlichen verfahrenstechnischen Umsetzbarkeit und des Aufwand-Nutzen-Verhältnisses, die zu einer Empfehlung an den G-BA führt, die Entwicklung des Verfahrens weiterzuverfolgen oder nicht.

In den folgenden Abschnitten folgt die genaue Darstellung der drei genannten Entwicklungsschritte.

#### **4.1.1 Themenerschließung**

Zunächst wird eine umfängliche Themenerschließung zur qualitativen und quantitativen Analyse der Versorgungspraxis durchgeführt. Diese basiert auf einer Literatur- und Leitlinienrecherche, einer Recherche der aktuellen normativen Rahmenbedingungen (Gesetze, Richtlinien etc.), einer explorativen Sozialdatenanalyse sowie leitfadengestützten Einzelgesprächen mit Expertinnen und Experten. Die Versorgungspraxis wird darauf aufbauend u. a. anhand folgender Punkte dargestellt:

- Ausführungen zu Epidemiologie, Ätiologie und Risikofaktoren der Erkrankung, sowie zu Diagnosestellung und Komorbiditäten
- Ausführungen zu Indikation, Therapie, Häufigkeiten des Leistungsgeschehens
- Ausführung zu Behandlungsergebnissen und nachfolgender Lebenserwartung und Lebensqualität

Als Hilfsmittel für die Identifikation der Qualitätsaspekte wird im Zuge der Themenerschließung ein Versorgungspfad abgeleitet unter Zuhilfenahme der Analyse

- der verschiedenen Versorgungssettings (stationär, ambulant, teilstationär und rehabilitativ) und ihrer Patienten- und Leistungszahlen,
- der Versorgung mit Medikamenten, Heil- und Hilfsmitteln,
- ggf. der bestehenden Versorgungsnetze, der integrierten Versorgung, der ambulanten spezialfachärztlichen Versorgung und von Disease-Management-Programmen,
- der beteiligten Berufsgruppen (z. B. Ärztinnen und Ärzte, Pflege und andere Gesundheitsprofessionen) sowie
- der zugrunde liegenden Anreizsysteme (z. B. Vergütung).

Der Versorgungspfad ermöglicht nicht nur eine Visualisierung und Strukturierung des gesamten Versorgungsgeschehens, sondern hilft auch, sowohl die identifizierten als auch die selektierten Qualitätsaspekte zuzuordnen und nach Sektoren differenziert darzustellen.

Im Rahmen der Themenerschließung erfolgt auch eine Suche nach größeren regionalen bzw. überregionalen Qualitätsinitiativen und Versorgungsmodellen.

Für die Themenerschließung ist eine umfangreiche Recherche nach aggregierter Evidenz des zu untersuchenden Versorgungsbereichs notwendig (vgl. Kapitel 6). Dabei werden zunächst alle verwertbaren Quellen systematisch gesichtet. Da der Zusammenhang zwischen dem projektierten Verfahren und der vorhandenen Evidenz zu Beginn der Untersuchungen nicht eingeschätzt werden kann, erfolgt die Zusammenstellung wertfrei, d. h. es erfolgt in der synoptischen Sichtung möglicher Quellen keine Grenzziehung bezüglich der Höhe eines erforderlichen Evidenzgrades. Um sicherzustellen, dass alle relevanten Aspekte zur Entwicklung eines Qualitätsmodells

betrachtet werden, wird neben der systematischen Recherche von aggregierter Evidenz anhand von systematischen Übersichtsarbeiten, Health Technology Assessments (HTA) und Leitlinien auch eine ergänzende Handsuche anhand von Suchalgorithmen in den zusammengetragenen Evidenzquellen sowie auf Websites relevanter Organisationen vorgenommen.

Die explorative Sozialdatenanalyse wird mithilfe von anonymisierten Datensätzen durchgeführt, die von einer kooperierenden, gesetzlichen Krankenkasse im Rahmen der Vorgaben von § 75 SGB X angefordert werden. Die genutzten Daten umfassen alle nach § 299 SGB V für ein QS-Verfahren im Regelbetrieb nutzbaren Datenarten, sofern sie themenspezifisch relevant sind.

#### **4.1.2 Identifikation von Qualitätsaspekten**

Für die Identifikation der patientenrelevanten Qualitätsaspekte ist es bedeutsam, die themenspezifischen Behandlungsziele sowohl aus der wissenschaftlichen Perspektive (Leitlinien der wissenschaftlichen Fachgesellschaften, Nationale Versorgungsleitlinien, Pflegestandards etc.) als auch aus der Patientenperspektive aufzugreifen und diese mit patientenrelevanten Aspekten aus der Literatur abzugleichen. Schlüsselstrukturen und -prozesse sowie Schnittstellen und Beteiligte werden im Versorgungspfad abgebildet. Weitere Informationsquellen für die Identifizierung von relevanten Qualitätsaspekten sind Fokusgruppen (Abschnitt 7.3.2) und leitfadengestützte Interviews (Abschnitt 7.3.1) mit Beteiligten aus allen Bereichen des Versorgungsgeschehens. Durch die Einbindung der Patientenperspektive über Fokusgruppen können qualitätsrelevante Themen oder wichtige Merkmale und Ereignisse einer Behandlung eruiert werden, die aus Sicht der Patientinnen und Patienten von Bedeutung sind. Auf diese Weise wird die Patientensicht von Beginn an eingebunden. Sollte der gruppendedynamische Prozess einer Fokusgruppe in Einzelfällen als nicht zielführend eingeschätzt werden, so können an ihre Stelle auch leitfadengestützte Interviews mit Patientinnen und Patienten treten.

Um auch das Erfahrungswissen aus der Perspektive der Leistungserbringer für die Identifikation von Qualitätsaspekten nutzen zu können, werden auch mit Gesundheitsfachkräften Fokusgruppen durchgeführt. Während die oben angesprochenen Expertinnen und Experten in der Themerschließung vor allem eine übergeordnete (systembezogene) Perspektive vermitteln sollen, sollen in den Fokusgruppen nun ausschließlich (aktive) Praktikerinnen und Praktiker des alltäglichen Behandlungsgeschehens angesprochen werden. Diese sollen aus allen beteiligten Sektoren der Versorgung und aus allen relevanten Berufsgruppen stammen. Um eine Überlagerung durch möglicherweise gegensätzliche berufsständische Interessen zu vermeiden, können die Fokusgruppen der Gesundheitsfachkräfte auch professionsbezogen durchgeführt oder im Einzelfall durch leitfadengestützte Interviews ersetzt oder ergänzt werden. Dies kann auch bei Rekrutierungsschwierigkeiten bei Gesundheitsfachkräften infrage kommen. Die Einordnung der identifizierten Qualitätsaspekte in den Versorgungspfad ergibt schließlich ein Bild ihrer Verteilung über den gesamten Versorgungsbereich.

### 4.1.3 Auswahl von Qualitätsaspekten für das Qualitätsmodell

In diesem Entwicklungsschritt werden aus den aus der Literatur und dem Erfahrungswissen der Expertinnen und Experten abgeleiteten, patientenrelevanten Qualitätsaspekten einzelne Qualitätsaspekte für das verfahrensspezifische Qualitätsmodell ausgewählt. Hierfür sind vier Kriterien maßgeblich:

- Übereinstimmung des Qualitätsaspekts mit dem gegebenen Zweck des QS-Verfahrens und den spezifischen Vorgaben des G-BA-Auftrags (Ausschlüsse)
- Verbesserungsbedarf für die Patientinnen und Patienten
- Beeinflussbarkeit durch die Leistungserbringer
- Grundsätzliche Erfassbarkeit durch die Qualitätssicherung

Das erste Auswahlkriterium stellt die Übereinstimmung des Qualitätsaspekts mit dem Zweck und den Zielen des geplanten QS-Verfahrens dar (in der Beauftragung durch den G-BA festgelegt). Darüber hinaus müssen die ausgewählten Qualitätsaspekte einen belegbaren Verbesserungsbedarf für die Patientinnen und Patienten und ein Verbesserungspotenzial durch die Leistungserbringer aufweisen. Hieraus ergibt sich das auf den jeweiligen Aspekt bezogene Qualitätsziel. Der Einschätzung der Qualitätsaspekte anhand dieser beiden Kriterien werden sowohl veröffentlichte Informationen als auch vom IQTIG berechnete Daten zugrunde gelegt. Ausgewertet werden hierfür Interventionsstudien, Längsschnittvergleiche und Vergleiche von Querschnittsergebnissen aus Studien oder Qualitätsinitiativen. Die untersuchten Versorgungssettings müssen dabei dem deutschen Qualitätsstandard prinzipiell vergleichbar sein; so werden beispielsweise bei internationalen Vergleichen nur solche zwischen Industrieländern einbezogen. Auch aus den Ergebnissen großer repräsentativer Erhebungen, z. B. aus Registern, können über die Spannweite der Ergebnisse eines Jahres Verbesserungspotenziale erkannt werden. Schließlich werden auch Befragungsergebnisse, insbesondere zum Umsetzungsgrad von Prozessschritten, in die Analyse einbezogen. Verbesserungsbedarfe und -potenziale können auch aus der explorativen Sozialdatenanalyse oder aus Aussagen mehrerer Expertinnen und Experten (Patientinnen und Patienten, Leistungserbringer) abgeleitet werden. In die Bewertung des Verbesserungsbedarfs werden zudem exemplarische Darstellungen größerer einrichtungsübergreifender Qualitätsinitiativen und -projekte einbezogen, die der Literatur bzw. Expertenangaben entnommen wurden (s. o.).

Als weiteres Auswahlkriterium müssen die Qualitätsaspekte von den Leistungserbringern grundsätzlich beeinflussbar sein. Die Frage der Zuordnung einer einrichtungsbezogenen Verantwortlichkeit stellt sich dagegen erst auf der Ebene der Qualitätsmerkmale bzw. Qualitätsindikatoren. Die solchermaßen selektierten Qualitätsaspekte werden zum Qualitätsmodell zusammengefasst.

Anschließend werden die Aspekte des Qualitätsmodells einer Prüfung der grundsätzlichen Erfassbarkeit durch die Qualitätssicherung unterzogen. Die grundsätzliche Erfassbarkeit ist gegeben, wenn der Qualitätsaspekt einer zugänglichen Datenquelle und einem realisierbaren Erfassungsinstrument zugeordnet werden kann. Letzteres bedeutet, dass die Inhalte des Qualitätsaspekts grundsätzlich mit diesem Instrument erfassbar (z. B. Abbildung in Sozialdaten) erschei-

nen, der Leistungserbringer damit technisch erreicht wird (bei Auslösung einer fall- oder einrichtungsbezogenen QS-Dokumentation) und der Aufwand als angemessen gelten kann. Darüber hinaus sollte – soweit auf Basis öffentlich zugänglicher Dokumente möglich – abgeschätzt werden, in welchem Umfang Patientinnen und Patienten in entsprechenden Selektivverträgen zum jeweiligen Themengebiet versorgt werden und welche Erfassungsprobleme sich ggf. hieraus ergeben. Sollten neue Erfassungsinstrumente vorgeschlagen werden, so sind diese soweit zu skizzieren, dass die oben genannten Fragen adressiert werden können. Im Ergebnis der Prüfung kann sich herausstellen, dass nicht alle Qualitätsaspekte des Qualitätsmodells unter den derzeitigen Rahmenbedingungen umsetzbar sind. Dies berührt jedoch inhaltlich nicht das Qualitätsmodell, sondern die konkreten Empfehlungen für eine eventuelle Verfahrensentwicklung.

Auf dieser Grundlage werden am Ende der Konzeptstudie Empfehlungen gegeben, welche Qualitätsaspekte im zu entwickelnden Verfahren durch welche Instrumente erfasst werden sollen. Dabei kann einem Aspekt durchaus auch mehr als ein Erfassungsinstrument zugeordnet werden, wenn damit unterschiedliche Teile des Aspekts erfasst werden sollen. Doppelerhebungen sind im Sinne der Aufwandsangemessenheit jedoch zu vermeiden, bei alternativen Erfassungsmöglichkeiten ist dem aufwandsärmeren Instrument der Vorzug zu geben.

Die besondere Kennzeichnung der empfohlenen Qualitätsaspekte im Versorgungspfad dient der Veranschaulichung einer angemessenen Fokussierung innerhalb des abgebildeten Versorgungsbereichs.

## 4.2 Dokumentations- und sozialdatenbasierte Neuentwicklung von Qualitätssicherungsverfahren

Die Grundlage für die Neuentwicklung von QS-Verfahren ist das in einer vorausgegangenen Konzeptstudie formulierte verfahrensspezifische Qualitätsmodell. Auf dieser Basis erfolgen die Entwicklung von Qualitätsindikatoren und die Erarbeitung von Empfehlungen zur Umsetzung eines künftigen QS-Verfahrens. Dabei zielt die gesamte Entwicklung auf den vom G-BA festgelegten Verfahrenszweck.

In diesem Abschnitt werden die einzelnen Schritte zur Entwicklung von QS-Verfahren beschrieben, die auf Dokumentationen der Leistungserbringer und/oder auf Sozialdaten bei den Krankenkassen beruhen. Die Entwicklung von Patientenbefragungen folgt einer eigenen Methode, die in Abschnitt 4.4 beschrieben wird. Bei paralleler Entwicklung von dokumentations- und sozialdatenbasierten Indikatoren sowie von Indikatoren einer Patientenbefragung für das gleiche QS-Verfahren erfolgt eine enge, wiederholte Abstimmung zwischen den Projektteams im IQTIG. Abschließend werden die Ergebnisse genannt, die am Ende des Entwicklungsprozesses stehen.

### 4.2.1 Entwicklungsschritte

Ausgehend vom Qualitätsmodell erfolgt die Entwicklung von Qualitätsindikatoren in folgenden Schritten:

- **Konkretisierung**, welche Qualitätsmerkmale die selektierten Qualitätsaspekte am besten repräsentieren können

- die **Operationalisierung** der Qualitätsmessung auf Basis der ausgewählten Qualitätsmerkmale zu Qualitätsindikatorentwürfen (QI-Entwürfen)
- die **Kondensierung** der Liste möglicher Qualitätsindikatoren zu fokussierten Qualitätsindikatorensets

Am Ende des Prozesses steht ein für die Realisierung des entwickelten QS-Verfahrens handhabbares und überschaubares Set an Qualitätsindikatoren zur Verfügung.

Bei der Neuentwicklung von Qualitätsindikatoren wird regelhaft ein beratendes Expertengremium einbezogen. Die Einbeziehung der Expertenmeinungen erfolgt dabei nach Verfahren, die dem jeweiligen Zweck angemessen sind. Im Folgenden wird das am häufigsten angewandte Vorgehen im Sinne eines modifizierten Delphi-Verfahrens (siehe Abschnitt 7.3.3 ab S. 107) beschrieben, ohne auszuschließen, dass bei bestimmten Aufgabestellungen, bei denen Expertinnen und Experten beratend hinzugezogen werden, auch andere Vorgehensweisen methodisch gerechtfertigt sein können. Beim modifizierten Delphi-Verfahren werden die Expertinnen und Experten bei jedem Entwicklungsschritt zuerst um eine schriftliche Rückmeldung zu den spezifischen merkmals- bzw. indikatorbezogenen Fragestellungen gebeten. Daran anschließend findet jeweils ein persönliches Treffen statt, in dem ein Austausch in der Gruppe ermöglicht wird.

Im Prozess der Entwicklung von dokumentations- und sozialdatenbasierten Qualitätsindikatoren sind insgesamt vier Expertentreffen vorgesehen. Beim ersten Treffen werden den Mitgliedern des Expertengremiums erste Informationen über das zu entwickelnde QS-Verfahren vermittelt (Auftrag des G-BA, Rahmenbedingungen) und es wird das zugrunde liegende Qualitätsmodell vorgestellt. Darüber hinaus wird der Ablauf des Entwicklungsprozesses dargelegt und die Aufgaben der Expertinnen und Experten werden näher erläutert. Die weiteren drei Treffen des Expertengremiums finden immer im Anschluss an die jeweiligen genannten Entwicklungsschritte des IQTIG statt. Die Aufgaben der Expertinnen und Experten sind dabei:

- die **Priorisierung** der Qualitätsmerkmale, die zur Abbildung von Qualitätsaspekten infrage kommen
- die **Kommentierung** der operationalisierten QI-Entwürfe
- die **Reflexion** des gesamten Sets an Qualitätsindikatoren

In Abbildung 5 ist der gesamte sechsstufige Entwicklungsprozess vom Qualitätsmodell bis hin zum Qualitätsindikatorenset schematisch dargestellt. In ihm verbinden sich die Entwicklungsschritte des IQTIG (dunkelgrün) einerseits – nämlich die Konkretisierung zu Qualitätsmerkmalen, die Operationalisierung der QI-Entwürfe und die Kondensierung der möglichen Qualitätsindikatoren zum fokussierten Qualitätsindikatorenset – und die Aufgaben der Expertinnen und Experten (grau) andererseits – und zwar die Priorisierung der Qualitätsmerkmale, die Kommentierung der QI-Entwürfe und die Reflexion des Qualitätsindikatorensets.

Die einzelnen Entwicklungsschritte werden im Folgenden näher erläutert.

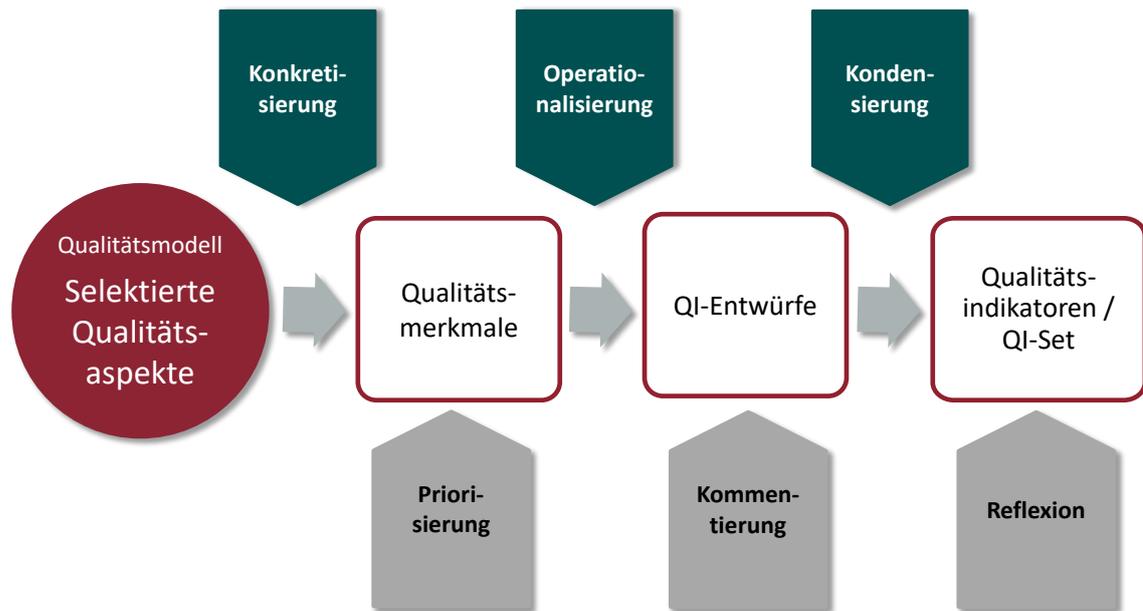


Abbildung 5: Entwicklungsschritte vom Qualitätsmodell bis zum Qualitätsindikatorenset

### Konkretisierung zu Qualitätsmerkmalen

Für die selektierten Qualitätsaspekte des Qualitätsmodells werden im ersten Entwicklungsschritt Merkmale der Versorgung identifiziert, die den jeweiligen Aspekt konkret abbilden. Dabei werden auf Basis der identifizierten Verbesserungsbedarfe und -potenziale merkmalsbezogene Qualitätsziele formuliert. Ist dies gegeben, qualifizieren sich diese Merkmale als Qualitätsmerkmale. Für jeden Qualitätsaspekt können meist mehrere Qualitätsmerkmale gefunden werden, die zumindest theoretisch zur Messung herangezogen werden können. So lassen sich beispielsweise für den Qualitätsaspekt „Komplikationen nach dem Eingriff“ eine Reihe möglicher Komplikationen als konkretisierte Merkmale dieses Aspekts definieren. Qualitätsaspekte können generell durch ein oder mehrere Qualitätsmerkmale abgebildet werden; inhaltliche Überlappungen der Merkmale werden jedoch vermieden.

Um relevante Abläufe oder Ereignisse der Patientenversorgung zu identifizieren, die für einen Qualitätsaspekt als Qualitätsmerkmal genutzt werden könnten, findet eine erneute, fokussierte Literaturrecherche statt (siehe Kapitel 6). Stehen evidenzbasierte Leitlinien (S3, S2e) oder Pflegestandards wissenschaftlicher Fachgesellschaften zur Verfügung, so bilden diese den Ausgangspunkt der Merkmalsableitung. Bei ausländischen Leitlinien ist zusätzlich zu prüfen, ob die Empfehlungen auf das deutsche Gesundheitssystem übertragbar sind. Außerdem wird für Qualitätsmerkmale, die nicht per se patientenrelevant sind, wie z. B. Surrogatparameter (siehe Kapitel 8), ermittelt, ob sie in einem relevanten, mittelbaren Zusammenhang mit dem Qualitätsaspekt stehen. Zudem findet eine ergänzende themenspezifische Recherche nach bereits national oder international vorhandenen Qualitätsindikatoren statt. Diese gehen jedoch nicht als gesetzte Vorlagen in spätere Phasen des Entwicklungsprozesses ein, sondern dienen lediglich als gedankliche Inputs für die verfahrensspezifische Neuentwicklung.

Für die Konkretisierung der Qualitätsaspekte zu Qualitätsmerkmalen wird des Weiteren festgelegt, bei welchen Patienten- und auch Leistungserbringergruppen das jeweilige Qualitätsmerkmal abgefragt werden soll. Zudem wird ermittelt, ob Datenquellen (siehe S. 40) zur Erfassung der Merkmale nutzbar sind und wenn ja, welche von diesen zur Abbildung des jeweiligen Qualitätsmerkmals am besten geeignet sind. Die Wahl der geeigneten Datenquelle für ein Qualitätsmerkmal orientiert sich dabei einerseits daran, inwieweit aus einer potenziellen Datenquelle valide und vollständige Daten zu erwarten sind und andererseits daran, ob der Aufwand zur Datenerhebung im Verhältnis zum Nutzen gerechtfertigt ist. Als weiterer Schritt der Konkretisierung wird geklärt, ob das gewählte Ereignis- oder Prozessmerkmal durch die Leistungserbringer beeinflussbar ist und ob diesen die Verantwortung dafür zugeschrieben werden kann.

Am Ende des Schritts der Konkretisierung liegt eine Liste selektierter Qualitätsmerkmale vor, die für die Weiterentwicklung zu Qualitätsindikatoren prinzipiell tauglich sind. Diese werden zu einem ersten Katalog von QI-Entwürfen zusammengestellt. Gegebenenfalls wird eine übermäßig große Anzahl von QI-Entwürfen schon in diesem Schritt auf eine handhabbare Größe reduziert. Dabei wird jedoch sichergestellt, dass für jeden Aspekt des Qualitätsmodells mindestens ein QI-Entwurf entwickelt wird.

### **Priorisierung der Qualitätsmerkmale**

Nach der Konkretisierung der Qualitätsmerkmale werden diese erstmals durch das externe Expertengremium hinsichtlich folgender Eignungskriterien bewertet (siehe Kapitel 8):

- Bedeutung des Qualitätsmerkmals für die patientenrelevanten Qualitätsaspekte (Relevanz, Verbesserungsbedarf/-potenzial und das daraus abgeleitete konkrete Qualitätsziel)
- Verantwortlichkeit der Leistungserbringer für die Ausprägung des Qualitätsmerkmals

Die Erfassung der Expertenmeinung erfolgt dabei in Anlehnung an die RAND/UCLA-Appropriateness-Method (RAM). Hierbei geben die Expertinnen und Experten sowohl in den schriftlichen Vorabbewertungen als auch in den persönlichen Treffen ihre Einschätzung merkmals- bzw. indikatorbezogen auf einer 9-stufigen Skala ab, wobei die Eignung ab einem Punktwert von 7 gegeben ist (Fitch et al. 2001). Konsens der Gruppe hinsichtlich der Bewertung wird dann angenommen, wenn mehr als 75 % der Expertinnen und Experten ein Kriterium als gegeben bewerten (ähnlich der Schwelle, die die AWMF (2012) bei Abstimmungen über Leitlinienempfehlungen verwendet).

Außerdem sollen die Expertinnen und Experten dazu Stellung nehmen, inwieweit mögliche Nebenwirkungen gegen eine Verwendung der QI-Entwürfe als Indikatoren sprechen (siehe Abbildung 5).

Ziel der Priorisierung ist es, von den Expertinnen und Experten eine Einschätzung der am besten geeigneten QI-Entwürfe für den betrachteten Versorgungsbereich zu erhalten, um eine zielgerichtete Reduzierung des Katalogs an QI-Entwürfen zu ermöglichen. Hierbei wird sichergestellt, dass mindestens ein QI-Entwurf je Qualitätsaspekt in den nächsten Entwicklungsschritt der Operationalisierung überführt wird.

### Operationalisierung der QI-Entwürfe

Im Zuge der Operationalisierung der geeigneten QI-Entwürfe werden die jeweils relevanten Grundgesamtheiten der Patientinnen und Patienten sowie die Merkmalsausprägungen konkretisiert, die für die Qualitätsmessung am geeignetsten erscheinen. In den QI-Datenblättern wird die Rationale mit der zugrunde liegenden, bewerteten wissenschaftlichen Evidenz, dem Verbesserungsbedarf und -potenzial sowie dem sich daraus ergebenden, konkreten Qualitätsziel niedergeschrieben. Des Weiteren wird festgelegt, mit welchem Erfassungsinstrument (fall- bzw. einrichtungsbezogene QS-Dokumentation, Sozialdaten bei den Krankenkassen) die Daten für die künftigen Indikatoren erhoben werden sollen. Lassen sich die erforderlichen Daten mit bisher etablierten Instrumenten nicht erheben, so müssen ggf. neue Erfassungsinstrumente konzipiert werden. Dem ausgewählten Erfassungsinstrument entsprechend werden die für die Berechnung des Indikators notwendigen Datenfelder und Kodierungen festgelegt. Darüber hinaus wird anhand der Literatur geprüft, welche patientenbezogene Risikofaktoren das Indikatorergebnis beeinflussen können und in einem späteren Risikoadjustierungsmodell für den Indikator berücksichtigt werden sollten. Im Rahmen des Auswertungskonzepts soll auch bereits ein erster Vorschlag für einen Referenzbereich (siehe Kapitel 10) formuliert werden. Hierbei kann es sich um einen fixen Referenzbereich auf Basis der Literatur oder um einen Perzentil-Referenzbereich (üblicherweise 5./95. Perzentil) handeln.

Wenn Qualitätsindikatoren mit Sozialdaten berechnet werden sollen, wird die Abbildbarkeit der Qualitätsindikatoren im Rahmen einer empirischen Prüfung mit Testdaten bereits vorab getestet. Analysiert wird unter anderem,

- ob die Grundgesamtheit der für das QS-Verfahren relevanten Behandlungsfälle in den Daten korrekt identifiziert werden kann,
- ob die erwarteten Diagnosen/Prozeduren in den Sozialdaten vorhanden und geeignet sind, die Grundgesamtheiten und die für die Qualität relevanten Merkmalsausprägungen abzubilden,
- ob Unterschiede im Kodierverhalten innerhalb und/oder zwischen den verschiedenen Sektoren bestehen.

Die Vorabberechnungen auf Grundlage der Sozialdaten ermöglichen eine erste Einschätzung, ob eine geeignete Operationalisierung der angedachten Qualitätsindikatoren möglich ist. Gleichzeitig können schon während des Entwicklungsprozesses fortlaufende Anpassungen zur Optimierung der QI-Entwürfe erfolgen (z. B. Anpassung der verwendeten Codes, Veränderungen an der Rechenregel des Qualitätsindikators). Darüber hinaus können bereits erste Probeauswertungen zu folgenden Fragestellungen erfolgen:

- zu den Indikatorergebnissen
- zur Häufigkeit des Eintretens von Ausschlusskriterien für die Grundgesamtheit der Patientinnen und Patienten des QS-Verfahrens bzw. für dessen Indikatoren
- zur Anzahl der einbezogenen Leistungserbringer
- zur Anzahl und Verteilung der Fälle pro Leistungserbringer
- zur Verteilung der Indikatorergebnisse der Leistungserbringer
- zur Anzahl von auffälligen Leistungserbringern

Weiterhin werden in diesem Schritt gegebenenfalls vorgesehene Indizes konkret ausgestaltet (siehe Kapitel 13).

### **Kommentierung der QI-Entwürfe**

Im Anschluss an die Operationalisierung der QI-Entwürfe werden diese dem Expertengremium zur Kommentierung vorgelegt und gemeinsam mit diesem weiter optimiert. Hierbei wird nochmals die Definition der Grundgesamtheit und der relevanten Merkmalsausprägungen beraten sowie die Korrektheit der vorgeschlagenen Kodierungen (z. B. ICD-/OPS-/ATC-Kodes oder GOP) und der ausgewählten Datenfelder abgeklärt. Des Weiteren wird die Verständlichkeit der vorgesehenen Datenfelder sowie die Notwendigkeit von erläuternden Ausfüllhinweisen eruiert. Darüber hinaus werden die bei der Risikoadjustierung zu berücksichtigenden Faktoren thematisiert und die möglichen Bewertungskategorien und Referenzbereiche mit den Expertinnen und Experten diskutiert. Bei unüberwindbar erscheinenden Problemen kann an dieser Stelle ein QI-Entwurf begründet aus dem Katalog herausgenommen werden. Bei der Verwendung von Sozialdaten können neben den Expertinnen und Experten des Gremiums ggf. noch externe Kodierexpertinnen und -experten zur Beantwortung spezieller Kodier- und Abrechnungsfragen einbezogen werden.

Am Ende der Operationalisierung liegt damit eine Liste möglicher Qualitätsindikatoren für das künftige QS-Verfahren vor.

### **Kondensierung der Liste möglicher Qualitätsindikatoren zum fokussierten Qualitätsindikatorenset**

Im letzten Entwicklungsschritt werden die möglichen Qualitätsindikatoren für ein valides, ausgewogenes, fokussiertes und praktikables Set an Qualitätsindikatoren ausgewählt. Dabei wird vor allem reflektiert, ob das Qualitätsindikatorenset den ursprünglichen Zweck des Verfahrens (im hier beschriebenen Teil: Qualitätsförderung) erfüllt und das Qualitätsmodell angemessen abbildet. Dabei kann es zu einer weiteren Reduktion der Qualitätsindikatoren im Set kommen, wenn bestimmte Aspekte mit der Zahl ihrer Indikatoren als überrepräsentiert angesehen werden. Qualitätsindikatoren können zudem in Indizes (vgl. Abschnitt 13.3.2) zusammengeführt werden.

### **Reflexion des Qualitätsindikatorensets**

Am Ende des Entwicklungsprozesses wird dem Expertengremium das fokussierte Set von Qualitätsindikatoren für eine finale Reflexion vorgelegt. Gemeinsam mit den Expertinnen und Experten werden die Ausgewogenheit und die Angemessenheit des Qualitätsindikatorensets bezogen auf den zu betrachtenden Versorgungsbereich und den Verfahrenszweck thematisiert. Konkret sind folgende Fragen zu prüfen:

- Liegen Überschneidungen zwischen den Indikatoren vor? (Mehrfachbewertungen der gleichen Strukturen/Prozesse/Ergebnisse sind zu vermeiden)
- Entspricht die Zahl der Indikatoren der Bedeutung des Qualitätsaspekts im Vergleich zu den anderen Qualitätsaspekten und deren Zahl an Indikatoren?

- Kann das Indikatorenset als ein ausgewogenes Bündel angesehen werden, dessen Ziele/Anreize miteinander im Gleichklang stehen oder gibt es hierin Widersprüche?
- Handelt es sich um ein in der Erhebung datensparsames, im Aufwand für die Leistungserbringer angemessenes und auf die wichtigsten Verbesserungsbedarfe fokussiertes Indikatorenset?

Darüber hinaus werden im Expertentreffen weitere Inhalte des Auswertungskonzepts (z. B. die Bildung von Indizes) diskutiert. Ziel ist es, am Ende des Entwicklungsprozesses das vorgesehene QS-Verfahren mit dem entwickelten Qualitätsindikatorenset im Hinblick auf die praktische Umsetzung, mögliche Umsetzungshürden, die Akzeptanz und die erwartbaren Ergebnisse zu reflektieren. Die Einschätzungen des Expertengremiums werden im Abschlussbericht dokumentiert.

#### 4.2.2 Entwicklungsergebnisse

Die Ergebnisse der Neuentwicklung eines QS-Verfahrens werden einem externen Stellungnahmeverfahren gemäß § 137a Abs. 7 SGB V unterzogen, optimiert und anschließend in einem Abschlussbericht zusammengefasst. Dieser beschreibt ausführlich den Entwicklungsprozess und das vorgesehene Qualitätsindikatorenset (Version 1.0). Zusätzlich enthält der Abschlussbericht weitere Ausführungen zur Umsetzung des vorgesehenen QS-Verfahrens:

- zu verwendende Erfassungsinstrumente<sup>3</sup>:
  - Dokumentationsbögen für die fall- und/oder einrichtungsbezogene QS-Dokumentation, inkl. Ausfüllhinweisen
  - Spezifikation für den Patientenfilter sowie für den Leistungs- und Medikationsfilter für die Sozialdaten bei den Krankenkassen
  - ggf. weitere, neue Erfassungsinstrumente
- Auslösung der verwendeten Erfassungsinstrumente:
  - QS-Filter für die fall-/einrichtungsbezogene QS-Dokumentation
  - Patientenfilter für die Sozialdaten bei den Krankenkassen
- Erhebungsformen:
  - Datenflüsse
  - Erfassungszeiträume (Follow-up)
  - Verfügbarkeit der Daten (Lieferfristen, Datenexporte)
  - Überprüfung der Dokumentationspflicht (Sollstatistik)
  - Anforderungen an eine Datenvalidierung
  - Begründung eines Stichprobenverfahrens oder einer Vollerhebung
- differenziertes Auswertungskonzept:
  - Berichtswesen:
    - Berichtszeitpunkte

---

<sup>3</sup> Das Instrument der Patientenbefragung wird in Abschnitt 4.4 gesondert dargestellt.

- Inhalte für die Rückmeldeberichte an die beteiligten Leistungserbringer (Qualitätsindikatoren, Kennzahlen)
- Skizzierung der angestrebten Risikoadjustierung und mögliche Risikofaktoren auf Basis der Literatur
- Vorschlag eines Referenzbereichs (fix oder Perzentil)
- möglicher Zeitplan bis zur Umsetzung des QS-Verfahrens in den Regelbetrieb
- weitere Schritte bis zum Regelbetrieb

### 4.3 Machbarkeitsprüfung für Qualitätssicherungsverfahren

Mit der Entwicklung eines dokumentations- und/oder sozialdatenbasierten Qualitätsindikatorensets ist die konzeptionelle Phase der Verfahrensentwicklung abgeschlossen. Vor der Umsetzung des Verfahrens, d. h. der Aufnahme des Regelbetriebs, sollen die fallbezogenen und einrichtungsbezogenen Dokumentationen regelhaft in einer vom G-BA beauftragten Machbarkeitsprüfung empirisch getestet werden: Die Dokumentation und die Qualitätsindikatoren werden in Kooperation mit freiwillig teilnehmenden Einrichtungen erprobt und optimiert, bevor die Empfehlungen zur Spezifikation (siehe Abschnitt 4.3.6) erarbeitet werden.

Um sicherzustellen, dass eine erfolversprechende Aufnahme des Regelbetriebs möglich ist, wird eine systematische Erprobung, Überprüfung und Anpassung der genannten Verfahrenskomponenten mit Fällen aus der Praxis der Leistungserbringer durchgeführt. Eine ggf. erforderliche Anpassung lässt Ziele und konzeptuelle Merkmale der neuen Qualitätsindikatoren unverändert; dem bestehenden Indikatorenset werden auch keine zusätzlichen Indikatoren hinzugefügt. Falls sich jedoch einzelne Indikatoren als nicht umsetzbar erweisen, kann deren Streichung empfohlen werden. Nach der Machbarkeitsprüfung ergibt sich somit das Indikatorenset in der Version 2.0.

Unter dem Begriff Machbarkeitsprüfung wird im Kontext der Entwicklung von QS-Verfahren die Überprüfung der fall- und/oder einrichtungsbezogenen QS-Dokumentation und der aus dieser Datenquelle basierenden Qualitätsindikatoren verstanden. Abzugrenzen ist sie von der Umsetzungsprüfung einer Patientenbefragung (Fragebogenvalidierung, siehe Abschnitt 4.4.3) und der explorativen/empirischen Prüfung von Sozialdaten, die bereits im Rahmen der Konzeptstudie bzw. der Neuentwicklung der Qualitätsindikatoren erfolgen kann. Auch die technische Prüfung der Datenflüsse gehört nicht zum Gegenstandsbereich der Machbarkeitsprüfung. Die Datenflüsse eines neu entwickelten Verfahrens werden stattdessen in einer Probephase zu Beginn des Regelbetriebs überprüft.

#### 4.3.1 Ziel der Machbarkeitsprüfung

In den Regelbetrieb soll ein optimiertes QS-Verfahren gehen, dessen Wartungsaufwand gering ist – der Ressourceneinsatz für die nachfolgende Verfahrenspflege sollte deutlich begrenzt werden. Die frühzeitige Abstimmung mit den Leistungserbringern sollte zudem zu einer höheren Akzeptanz des neuen QS-Verfahrens und seiner notwendigen Dokumentation beitragen. Und nicht zuletzt sollen optimierte und trennscharfe Indikatoren es ermöglichen, zielgerichtete und

effiziente qualitätsfördernde Maßnahmen (Qesü-RL) und Strukturierte Dialoge (QSKH-RL) durchzuführen.

#### 4.3.2 Teilnehmende Einrichtungen und Datenbasis

Schwerpunkt der Machbarkeitsprüfung ist der empirische Test der fallbezogenen und/oder der einrichtungsbezogenen QS-Dokumentation:

- Es ist vorgesehen, dass auf freiwilliger Basis etwa 15 bis 30 Einrichtungen an einer Machbarkeitsprüfung beteiligt werden.
- In jeder teilnehmenden Einrichtung werden ca. 15 bis 25 Falldokumentationen vorgenommen und ggf. eine einrichtungsbezogene Dokumentation durchgeführt.
- Im Anschluss an die fall- oder einrichtungsbezogenen Dokumentationen findet eine ausführliche Befragung der an Behandlung und Dokumentation Beteiligten statt.

Bei freiwilliger Teilnahme und geringen Fallzahlen kann keine repräsentative Auswahl der an der Machbarkeitsstudie teilnehmenden Einrichtungen (Krankenhäuser, Vertragsärztinnen und -ärzte) sowie der Patientinnen und Patienten erfolgen. Daher wird bei der Auswahl von Einrichtungen und/oder Behandlungsfällen nach dem Prinzip einer Maximierung der Heterogenität verfahren, um die Bandbreite der Leistungserbringer und Behandlungsfälle abzubilden.

Für die Auswahl teilnehmender Einrichtungen hat dies zur Folge, dass zentrale Unterscheidungsmerkmale berücksichtigt und in ihrer Variationsbreite abgedeckt werden müssen, z. B.:

- Sektoren der medizinischen Versorgung: ambulant und stationär, ggf. auch spezielle Versorgungsformen
- Größe der Einrichtungen
- Versorgungsstufe der Krankenhäuser, belegärztliche Versorgung
- Organisationsart der ambulanten Einrichtungen (Krankenhausambulanz, Einzelpraxis, Berufsausübungsgemeinschaft/Gemeinschaftspraxis, Medizinisches Versorgungszentrum [MVZ])
- Region (ganzes Bundesgebiet, Land–Stadt)

Maßgeblich für die Auswahl von Behandlungsfällen sind die Ein- und Ausschlusskriterien des neuen QS-Verfahrens. Da jedoch noch keine Filtersoftware zur Verfügung steht und je Leistungserbringer nur relativ wenige Fälle ausgewählt werden, kann es bei umfangreichen QS-Filtern notwendig werden, aus dem Original-Filter prototypische Auslösefälle (Kodes) gezielt herauszugreifen. Auch hierbei gilt wiederum das Prinzip einer möglichst großen Heterogenität. Bei weniger umfangreichen QS-Filtern können dagegen konsekutive Fälle berücksichtigt werden. Entsprechend den übermittelten Auslösekodes suchen die teilnehmenden Einrichtungen aktuelle Fälle über ihr Krankenhausinformationssystem (KIS) oder ihr Praxis-Verwaltungssystem (PVS der Vertragsärztinnen und -ärzte) heraus und dokumentieren jeweils die gesamten Datensätze der QS-Dokumentation auf einem Papierbogen oder in einem elektronischen Dokument. In dieser Phase steht den Einrichtungen eine Ansprechpartnerin oder ein Ansprechpartner des IQTIG zur Verfügung, der telefonisch oder per E-Mail kontaktiert werden kann, um die dokumentierenden Mitarbeiterinnen oder Mitarbeiter der Einrichtungen bei etwaigen Unklarheiten zu beraten oder Fragen zu beantworten.

### 4.3.3 Strukturierte Interviews mit den Leistungserbringern

Nach Abschluss der Dokumentation wird vor Ort in der Einrichtung ein strukturiertes Interview zu den Merkmalen der Datenfelder und zu ihrer Handhabbarkeit bei der Erfassung der jeweils abzubildenden Einrichtungs- oder Patientendaten durchgeführt. Interviewer sind Mitarbeiterinnen und Mitarbeiter des IQTIG, Befragte sind die an Behandlung und Dokumentation in der Einrichtung Beteiligten.

Die Befragung erfolgt sukzessive zu allen Datenfeldern der Einrichtungs- resp. Falldokumentation und betrifft folgende Themen:

- **Funktion:** Den Befragten wird die Funktion des Datenfeldes erläutert. Es wird erklärt, für welchen Qualitätsaspekt das erhobene Datum von Bedeutung ist und wie das Datum für die Berechnung des zugehörigen Qualitätsindikators ausgewertet wird. Die Befragten werden aufgefordert einzuschätzen, ob die intendierte Funktion erfüllt wird; ggf. können Alternativen vorgeschlagen werden.
- **Klarheit/Verständlichkeit:** Dieses Thema betrifft die Beschriftungen im Datenfeld einschließlich der Schlüsselbezeichnungen und Ausfüllhinweise zum Datenfeld.
- **Dokumentationsaufwand:** Generell unterliegt die Dokumentation zu Zwecken der Qualitätssicherung dem Gebot der Datensparsamkeit. Dokumentationsaufwand entsteht jedoch nicht ausschließlich durch die Menge der Datenfelder. Von erheblichem Einfluss ist auch die Verfügbarkeit der Daten innerhalb der Einrichtung. Im Interview zur Datenerhebung wird daher nachgefragt, ob die zur Bearbeitung der Datenfelder erforderlichen Informationen den Patientenakten oder dem EDV-System des Hauses unmittelbar entnommen werden können. Es wird eine grobe Abschätzung des Zeitaufwands für die Informationsbereitstellung und die eigentliche Dokumentation durch die einzelnen Leistungserbringer erfragt; zu beachten ist jedoch, dass der Zeitaufwand im späteren Regelbetrieb abweichen kann – sobald die Dokumentation routinemäßig erfolgt, reduziert dies in der Regel den zeitlichen Aufwand.
- **Klinische Einschätzung:** Für die fallbezogene QS-Dokumentation wird nachgefragt, ob die dokumentierten Datensätze die Fälle medizinisch adäquat abbilden, um eine angemessene Bewertung durch Qualitätsindikatoren zu ermöglichen. Zusätzlich wird diskutiert, ob besondere Fallkonstellationen ggf. zu einem Ausschluss aus dem Indikator führen sollten.

### 4.3.4 Auswertung der Daten

In der Auswertung der erhobenen Einrichtungs- oder Falldaten werden die Rechenregeln des neuen QS-Verfahrens erstmalig angewendet. Hierbei können jedoch nur solche Indikatoren ausgewertet werden, die keine Verknüpfung verschiedener Datenquellen (z. B. mit Sozialdaten) erfordern, da dies erst im Regelbetrieb möglich wird. Wegen der begrenzten Fallzahl erlauben die Auswertungsergebnisse noch keine Bewertung der Versorgungsqualität. Es handelt sich vielmehr um eine exemplarische Auswertung, die demonstriert, dass eine Auswertung der Daten mit den für das Indikatorenset formulierten Rechenregeln möglich ist. Außerdem werden elementare Voraussetzungen der Indikatorvalidität geprüft: Unterschiede zwischen den dokumentierten Behandlungsfällen müssen sich auf der Ebene der berechneten Indikatorwerte in entsprechenden und plausiblen Unterschieden widerspiegeln.

Nach der Datenerhebung und der Berechnung folgen Analysen zu zwei Bereichen:

- **Struktur und Effizienz der Algorithmen:** Die Algorithmen (Rechenregeln) der Qualitätsindikatoren sollten eine möglichst einfache Struktur aufweisen. Eine einfache Struktur bzw. eine Untergliederung umfangreicher und komplexer Regeln in einfach strukturierte Komponenten erleichtert die Fehlerkontrolle und die spätere Verfahrenspflege. Ein Ergebnis der Struktur- und Effizienzanalyse kann ggf. sein, dass die Anzahl von Schlüsselwerten in Datenfeldern angepasst/reduziert oder dass Datenfelder komplett aus der Spezifikation entfernt werden.
- **Voraussetzungen von Validität<sup>4</sup>:** Valide Qualitätsindikatoren machen Unterschiede in der Versorgungsqualität zwischen Einrichtungen sichtbar. Voraussetzung der Indikatorvalidität ist eine zutreffende Zuordnung oder Nichtzuordnung der dokumentierten Behandlungsfälle zum Zähler des jeweiligen Indikators. Bei dieser Prüfung spielen „kritische Fälle“, d. h. Behandlungsfälle, bei denen komplizierte Verläufe oder Komplikationen aufgetreten sind, eine wichtige Rolle.

Auch die Befragungsdaten, die dokumentiert wurden, werden deskriptiv-statistisch ausgewertet. Die Ergebnisse werden zur Anpassung der Spezifikation und ggf. auch zur Anpassung der Qualitätsindikatoren und ihrer Rechenregeln genutzt. Diese Änderungen werden abschließend im Umlaufverfahren mit den Mitgliedern des Expertengremiums der QI-Neuentwicklung abgestimmt.

Änderungen der Spezifikation und der Rechenregeln der Qualitätsindikatoren einschließlich der Streichung von Datenfeldern und Indikatoren erfolgen in der Verantwortung des IQTIG und sind zu begründen. Das IQTIG kann bei der Optimierung der Spezifikation und der Rechenregeln Expertinnen und Experten konsultieren, z. B. Mitglieder des Expertenpanels, das an der Verfahrensentwicklung mitgewirkt hat.

#### **4.3.5 Ergebnis der Machbarkeitsprüfung, Musterauswertung und Rückmeldeberichte**

Nach Abschluss der Machbarkeitsprüfung wird dem G-BA ein Abschlussbericht übermittelt, der Vorgehen und Ergebnisse der Machbarkeitsprüfung beschreibt sowie als Anhang überarbeitete und optimierte Versionen der Dokumentation und der Rechenregeln enthält. Daraufhin kann die Beauftragung zur Erstellung einer Spezifikation für das neue QS-Verfahren erfolgen.

Vor Beginn des Regelbetriebs erhält der G-BA eine Musterauswertung; die Einrichtungen, welche an der Machbarkeitsprüfung teilgenommen haben, erhalten Rückmeldeberichte mit einer Auswertung zu den jeweils dokumentierten Fällen.

---

<sup>4</sup> Eine umfassende Prüfung der Validität der Qualitätsindikatoren kann in der Machbarkeitsprüfung nicht erfolgen, weil die Repräsentativität freiwillig teilnehmender Einrichtungen und ihrer Patientinnen und Patienten nicht unterstellt werden kann. Allerdings sind grundlegende Voraussetzungen der Validität überprüfbar.

#### **4.3.6 Empfehlungen zur Spezifikation**

Nach der eigentlichen Machbarkeitsprüfung ist gesondert beauftragt auch ein erster Abschlussbericht zur Spezifikation der Datenerfassungen der fallbezogenen bzw. einrichtungsbezogenen QS-Dokumentation und/oder der Sozialdaten bei den Krankenkassen zu erstellen. Dieser wird nach Diskussion im G-BA überarbeitet in die richtlinienbezogenen Gesamtempfehlungen zur Spezifikation eingefügt. Nach Beschluss der Gesamtempfehlungen zur Spezifikation und der themenspezifischen Bestimmungen der Richtlinie erfolgt die Veröffentlichung der technischen Spezifikation.

## 4.4 Neuentwicklung von Patientenbefragungen im Rahmen von Qualitätssicherungsverfahren

Patientenbefragungen haben sich mittlerweile als wesentliches Bewertungsinstrument medizinischer Behandlungen etabliert und auch Einzug in die Qualitätssicherung gehalten (siehe z. B. National Health Service in Großbritannien, Centers for Medicare & Medicaid Services in den USA). Die Qualität einer Behandlung, die für Patientinnen und Patienten durchgeführt wird, soll nicht ohne die Stimme der Patientinnen und Patienten selbst beurteilt werden.

Die Patientinnen und Patienten können als Adressaten der Behandlung und als diejenigen, die das Behandlungsgeschehen beobachten, nicht nur Auskunft über ihren Behandlungserfolg geben (Ergebnisqualität), sie sind gleichermaßen in der Lage, über die Abläufe und Rahmenbedingungen der Behandlung zu berichten (Struktur- und Prozessqualität). Die Befragung von Patientinnen und Patienten bietet damit neben den Sozialdaten bei den Krankenkassen und den Dokumentationen der Leistungserbringer eine weitere Datenquelle für Qualitätsindikatoren.

Neben Patientenbefragungen können als weitere Befragungsinstrumente im Rahmen der gesetzlichen Qualitätssicherung Angehörigenbefragungen oder Befragungen von Leistungserbringern, wie beispielsweise Einweiserbefragungen, eine wichtige Rolle spielen. Bei Patientinnen und Patienten, die nicht selbst Fragen beantworten können, sind auch Proxy-Befragungen denkbar, bei denen Dritte, z. B. Angehörige, Auskunft aus Sicht der Patientin / des Patienten geben. Da diese Befragungen aktuell noch nicht Beauftragungsgegenstand sind, fokussieren die folgenden Ausführungen ausschließlich auf die Befragung von Patientinnen und Patienten.

### 4.4.1 Generelle Anforderungen an die Patientenbefragungen

Für die Entwicklung einer Patientenbefragung im Kontext eines Qualitätssicherungsverfahrens bestehen normative Regelungen in Form von gesetzlichen Grundlagen (SGB V) sowie der Richtlinie zur einrichtungs- und sektorenübergreifenden Qualitätssicherung (Qesü-RL), die bestimmte Anforderungen an die Entwicklung von Patientenbefragungen stellen. Dazu zählen:

- verpflichtende Teilnahme aller infrage kommenden Leistungserbringer
- Durchführbarkeit von Einrichtungsvergleichen
- Ermöglichung einrichtungsbezogener Auswertungen
- sektorenübergreifende Anwendung (stationär und ambulant)
- Veröffentlichungsfähigkeit der Ergebnisse im Qualitätsbericht der Krankenhäuser
- Fragebogenlogistik und Datenschutz entsprechend den Regelungen nach § 299 SGB V

Darüber hinaus stellen die methodisch-wissenschaftlichen Grundprinzipien des IQTIG konkrete Anforderungen an die Entwicklung der Patientenbefragungen (siehe Abschnitt 2.3):

- Fokussierung auf Qualitätsaspekte, die Verbesserungsbedarf und -potenzial aufweisen
- Beeinflussbarkeit der Qualitätsaspekte durch die Leistungserbringer und Zuschreibbarkeit (Handlungsrelevanz)
- Abbildung von Qualitätsaspekten und Qualitätsmerkmalen, die von Patientinnen und Patienten gleichermaßen beobachtbar und bewertbar sind

- Patientenrelevanz der Qualitätsaspekte und Qualitätsmerkmale
- Abstimmung und Verzahnung mit anderen Instrumenten und Datenquellen, die im jeweiligen QS-Verfahren eingesetzt werden
- Eignung für den Einsatz im Regelbetrieb

Diese Anforderungen haben einen wesentlichen Einfluss auf die Methoden der Fragebogenentwicklung, die im Folgenden beschrieben werden.

#### 4.4.2 Faktenorientierte Befragungsansätze

Befragungskonzepte können unterteilt werden in die Messung der Patientenzufriedenheit, Patient-Reported Outcome Measures (PROM) und Patient-Reported Experience Measures (PREM). Patientenzufriedenheit wird oft als Indikator der Behandlungsqualität verstanden. Das Ausmaß der Zufriedenheit wird häufig als Maß für die Übereinstimmung von vorab getroffenen Erwartungen und den tatsächlichen Erfahrungen verstanden (Jacob und Bengel 2000, Lingenfelder und Schneider 1990, Pascoe 1983). Die Annahme, dass Zufriedenheit ein Maß für Qualität ist, wird jedoch vielfach infrage gestellt (Linder-Pelz 1982, Williams 1994). Denn ungeachtet tatsächlich gemessener Qualitätsunterschiede zeigt die Praxis gleichermaßen hohe Zufriedenheitsraten im Sinne von Deckeneffekten (Heidegger et al. 2008, Jacob und Bengel 2000). Auf welcher Basis die Patientinnen und Patienten ihre Urteile bilden und ob es tatsächlich zu einer Gegenüberstellung von Erwartungen und Erfahrungen kommt, kann nicht vollständig geklärt werden. Die Ergebnisse der Zufriedenheitsbefragungen sind folglich von einer hohen Subjektivität geprägt, die einen großen Interpretationsspielraum lassen, aber wenig konkrete Handlungsansätze für Verbesserungsmaßnahmen bieten.

Im Rahmen der externen Qualitätssicherung müssen die Indikatoren der Patientenbefragung qualitätsrelevante Aspekte adressieren, die von allen Patientinnen und Patienten gleichermaßen und weitgehend objektiv beantwortet werden können, d. h. der individuelle Interpretationsspielraum der Patientinnen und Patienten muss so gering wie möglich gehalten werden. Vor diesem Hintergrund wird deutlich, dass Patientenzufriedenheitsmessungen für die externe Qualitätssicherung ungeeignet sind, da zu starke Limitationen hinsichtlich der Objektivität und Eindeutigkeit der Aussagen bzw. der dahinterliegenden Konstrukte bestehen.

Deshalb werden bei der Entwicklung von Patientenbefragungen faktenorientierte Ansätze zugrunde gelegt, wie sie im Rahmen der PREM und PROM verfolgt werden. Als PREM werden Befragungskonzepte bezeichnet, die die Erfahrungen der Patientinnen und Patienten mit dem Versorgungsprozess erfassen. Der Fokus von PREM liegt auf patientenrelevanten Aspekten der Behandlung und den diesbezüglichen Erfahrungsberichten der Patientinnen und Patienten (Hodson et al. 2013). Eine Form von PREM, die einen faktenorientierten Ansatz verfolgt, ist die ereignisorientierte Patientenbefragung (Cleary 1999, Gerteis et al. 2017, Klein 2004). Hier stehen konkrete Erfahrungen oder Beobachtungen bzw. Ereignisse im Vordergrund, über deren Eintreten oder Nichteintreten die Patientinnen und Patienten berichten (z. B. „Wurde Ihnen erklärt, weshalb Sie die Ihnen verschriebenen Medikamente einnehmen sollen?“). Die Ereignisse sollen Schlüsselsituationen in der Versorgung aus Sicht der Patientinnen und Patienten darstellen, die einen Aufschluss über die Qualität der Versorgung geben können. Zudem können über

diesen Ansatz kritische Aspekte im Behandlungsverlauf bzw. Ereignisse, die Hinweise auf mögliche Qualitätsdefizite geben, erfragt werden.

PROM werden als Oberbegriff für verschiedene Ansätze zur Erfassung von Therapieeffekten verwendet, die von Patientinnen und Patienten berichtet werden (z. B. „Können Sie schmerzfrei Treppen steigen?“). So wird als PROM jeder Bericht über den Gesundheitszustand einer Patientin oder eines Patienten bezeichnet, der direkt von ihr oder ihm ohne Interpretation einer anderen Person (z. B. durch eine Ärztin / einen Arzt oder Angehörige) abgegeben wird (HHS.gov et al. 2009). PROM werden genutzt, da zum einen manche Gesundheitsaspekte nur durch die Betroffenen selbst bewertet werden können (z. B. Schmerzen) und zum anderen auf diese Weise der Fokus auf Aspekte fällt, die für die Patientinnen und Patienten von besonderer Relevanz sind.

Zusammenfassend zielen PROM und die ereignisorientierte Befragung im Rahmen von PREM als faktenorientierte Ansätze auf konkrete Aspekte der Behandlung ab, ohne nach individuellen Werturteilen der Patientinnen und Patienten zu fragen. Auf diese Weise werden Beobachtungen des Behandlungsgeschehens und individuelle Bewertungen voneinander abgekoppelt. So wird beispielsweise nach den konkreten Wartezeiten (Erfahrung) gefragt und nicht, wie diese von den Patientinnen und Patienten empfunden wurden (Bewertung). Damit wird die Interpretation der Ergebnisse eindeutiger, wodurch konkrete Hinweise für mögliche Handlungsansätze gegeben werden können.

Wie in Abbildung 6 dargestellt, lassen sich PREM und PROM den Qualitätsdimensionen der Struktur-, Prozess- und Ergebnisqualität zuordnen (Donabedian 1966). Die Ergebnisqualität wird im Rahmen von PROM über Fragen zur Effektivität einer Behandlung und Fragen zu Aspekten der Patientensicherheit abgebildet (z. B. gesundheitsbezogene Lebensqualität, Symptome und Funktionalität). Struktur- und Prozessqualität werden im Rahmen von PREM mittels Fragen zu Aspekten der Patientenpartizipation und Behandlungsabläufen adressiert (z. B. Abfragen konkreter Beobachtungen, Ereignisse und Erfahrungen und objektiver Problemlagen).

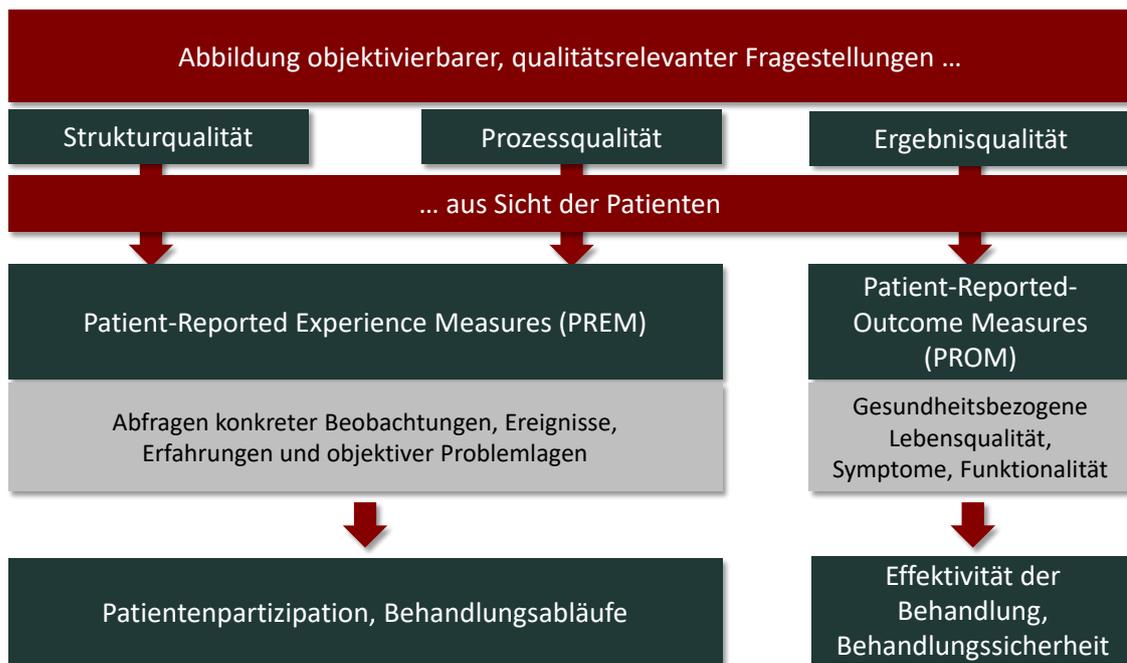


Abbildung 6: Modell der Patientenbefragung im Rahmen der externen Qualitätssicherung

In Anlehnung an das OECD-Rahmenkonzept für Qualitätsindikatoren (Arah et al. 2006) lassen sich damit den beiden Befragungsansätzen die übergeordneten, allgemeinen Zielsetzungen zuordnen, die über Patientenbefragungen adressiert und verfahrensspezifisch ausgestaltet werden: die Sicherstellung des Behandlungserfolges, die Vermeidung von Risiken, die Einbindung der Patientinnen und Patienten und die Sicherstellung von Qualitätsstandards im Rahmen der Behandlungsabläufe (Arah et al. 2006).

#### 4.4.3 Methodische Schritte der Fragebogenentwicklung

Die Fragebogenentwicklung geht immer von dem im Rahmen einer vorangegangenen Konzeptstudie (siehe Abschnitt 4.1) erarbeiteten Qualitätsmodell aus. Analog zur verfahrensspezifischen Entwicklung von Qualitätsmodellen erfolgt auch die Entwicklung einer Patientenbefragung verfahrensspezifisch. Das bedeutet, dass die Befragungsinhalte für jedes Verfahren entsprechend den Zielen des QS-Verfahrens individuell entwickelt werden. Zudem müssen Fragebogenlogistik und Datenflüsse für jede Entwicklung individuell geklärt werden, da je nach Verfahren unterschiedliche Befragungszeiträume und Auslösekriterien relevant sein können. Die Qualitätsaspekte des Qualitätsmodells, die durch eine Patientenbefragung adressiert werden sollen, bestimmen dabei die Fragebogenentwicklung (siehe Abbildung 7).

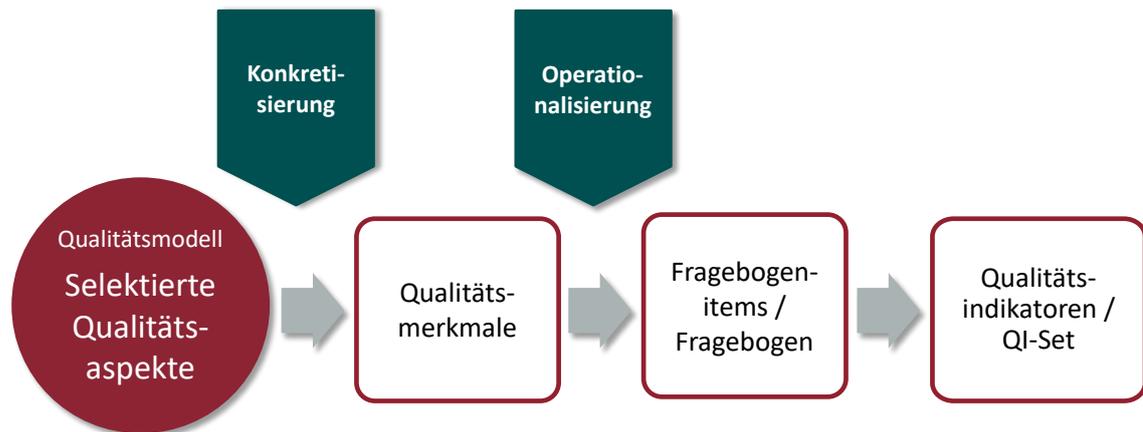


Abbildung 7: Schematische Darstellung der Entwicklungsschritte einer Patientenbefragung

Eine Patientenbefragung kann auch als Ergänzung beauftragt werden, wenn ein QS-Verfahren bereits entwickelt wurde oder sich bereits im Regelbetrieb befindet. In diesem Fall werden zunächst die Qualitätsaspekte erarbeitet, die mittels Patientenbefragung adressierbar sind.

Die im Regelfall auf eine Konzeptstudie folgenden methodischen Schritte der Fragebogenentwicklung (Abbildung 8) beginnen zunächst mit der explorativen Phase der Entwicklung der Qualitätsmerkmale, wozu die Literaturrecherche und die Durchführung von Fokusgruppen und ggf. leitfadengestützte Interviews zählen. Die auf dieser Basis entwickelten Qualitätsmerkmale werden im nächsten Schritt mit einem Expertengremium kritisch diskutiert und gegebenenfalls überarbeitet. Die Qualitätsmerkmale bilden als Konkretisierungen der Qualitätsaspekte die Basis für die Fragebogenentwicklung – sie setzen die Themen, die mit dem Fragebogen adressiert werden. Daher wird nach der Entwicklung der Qualitätsmerkmale ein Stellungnahmeverfahren durchgeführt, in dem die nach § 137a Abs. 7 SGB V zu beteiligenden Organisationen die Entwicklungen bis zur Themensetzung der Fragebogenentwicklung kommentieren können.

Nach dem Stellungnahmeverfahren beginnt die Itementwicklung, die in einem ersten Fragebogenentwurf mündet. Dieser Fragebogenentwurf wird in der folgenden Phase der Fragebogenoptimierung einem kognitiven Pretest unterzogen und auf der Basis der Ergebnisse überarbeitet. In einem anschließenden Standard-Pretest wird die Befragung einschließlich der Eignung des Fragebogens nochmals getestet. Am Ende des Entwicklungsprozesses steht ein validierter Fragebogen. Abschließend wird definiert, welche Fragebogenitems bzw. Indizes als Qualitätsindikatoren zur Beurteilung der Behandlungsqualität gelten sollen. Diese methodischen Schritte werden in den folgenden Abschnitten jeweils genauer erläutert.

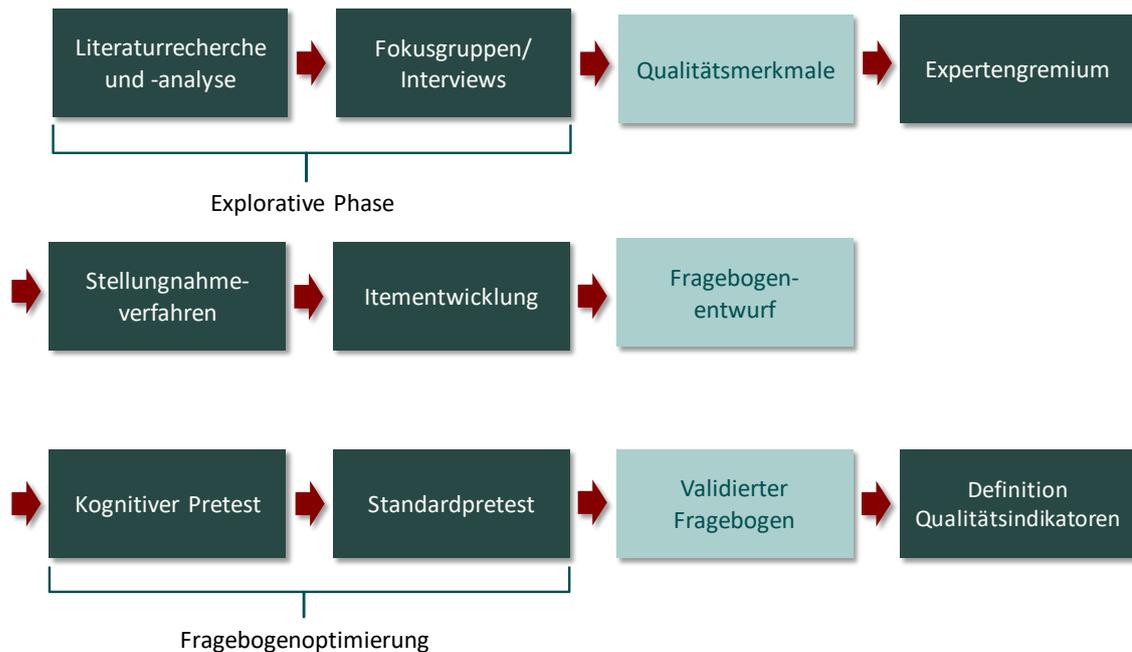


Abbildung 8: Methodische Schritte (dunkelgrün) und Zwischenprodukte (hellgrün) der Fragebogenentwicklung

### Literaturrecherche und -analyse

Die Entwicklung eines Fragebogens für eine Patientenbefragung beginnt mit einer Literaturrecherche und -analyse (siehe Kapitel 6). Die Ergebnisse der Literaturrecherche dienen zum einen der Generierung von Themen für die Diskussionsleitfäden der Fokusgruppen und ggf. für die Interviewleitfäden, zum anderen geben sie Hinweise für die spätere Generierung von Fragebogenitems. Auf Basis der Erkenntnisse aus den Fokusgruppen und Interviews sowie der Literaturrecherche werden die Qualitätsmerkmale entwickelt.

Die Literaturrecherche umfasst neben dem Recherche teil auch immer eine Bewertung der gesichteten Literatur. Im Kontext der Patientenbefragung bedeutet Literaturbewertung nicht, dass Evidenz für Zusammenhänge zwischen patientenorientierten Prozessen, wie etwa partizipativer Entscheidungsfindung, und klinischen Ergebnissen vorliegen muss. Gerade für die Patientenorientierung einer Behandlung, die sich primär auf Prozesse im Behandlungsgeschehen bezieht, spielen Daten zu deren klinischer Effektivität eine untergeordnete Rolle. Die patientenorientierte Behandlung folgt einer Rollenänderung der Patientinnen und Patienten von passiven Behandlungsempfängern zu aktiven Beteiligten des Behandlungsgeschehens. Die damit einhergehende Anerkennung der Patientenbedürfnisse, die Einbindung in Entscheidungsprozesse und die Informiertheit über die Behandlung sind zunächst Ausdruck dieser gesellschaftlichen Einstellungs- und Rollenänderung und weniger eine Reaktion auf unzureichende Behandlungsergebnisse im klinischen Sinn. Folglich begründet sich Patientenorientierung weniger aus einem evidenzbasierten Wirkungszusammenhang, sondern vielmehr aus gesellschaftlichen Werten, deren Auswirkungen – beispielweise auf die Therapietreue, Lebensqualität oder Ängste – in einzelnen Bereichen untersucht wurden (z. B. Härter et al. 2005, Haynes et al. 1996, Scheibler et al.

2003). Diesem Umstand wird bei der Literaturbewertung im Kontext der Fragebogenentwicklung Rechnung getragen. Folglich leitet bei der Entwicklung einer Patientenbefragung vor allem das Eignungskriterium „Bedeutung für die Patientinnen und Patienten“ die Literaturrecherche (siehe Kapitel 8).

Durch die Literaturbewertung wird Transparenz darüber hergestellt, mit welcher Deutlichkeit die recherchierten Quellen Hinweise auf konkrete patientenrelevante Qualitätsmerkmale geben, damit vor diesem Hintergrund bewusste Entscheidungen zu deren Verwendung bei der Fragebogenentwicklung getroffen werden können.

Darüber hinaus beinhaltet die Literaturrecherche immer auch eine Recherche nach etablierten Fragebögen für den entsprechenden Themenbereich (siehe Abschnitt 6.1.3). Ziel ist hier nicht die Übernahme bestimmter Skalen und Items. Dies ist zum einen aufgrund des spezifischen Kontextes, in dem die Patientenfragebögen als ein in sich geschlossenes Instrument entwickelt und validiert werden, selten bis kaum möglich, zum anderen kommen meist lizenzrechtliche Fragen hinzu. Vielmehr soll die Fragebogenrecherche als Teil der Literaturrecherche eine weitere Informationsquelle erschließen, anhand derer konkrete Qualitätsmerkmale entwickelt werden.

### **Fokusgruppen und Interviews**

Den Fokusgruppen (siehe Abschnitt 7.3.2) wird ein besonders hoher Stellenwert eingeräumt, da durch die Diskussion in Gruppen vertiefte Einblicke in die Bedürfnisse und Schwerpunkte der Patientinnen und Patienten im Rahmen ihrer Behandlung gewonnen werden (Flick 2016, Lamnek 2010). Das Ziel der Fokusgruppen liegt dementsprechend in der Vertiefung und Konkretisierung der patientenspezifischen Qualitätsaspekte, sodass die Themensetzungen des späteren Fragebogens genauer bestimmt und einzelne Fragen (Items) und Indikatoren abgeleitet werden können.

Um dies zu erreichen, werden im Rahmen der Fragebogenentwicklung in der Regel zwei Arten von Fokusgruppen eingesetzt:

- Fokusgruppen mit Patientinnen und Patienten
- Fokusgruppen mit Gesundheitsfachkräften, die maßgeblich am Versorgungsgeschehen beteiligt sind (z. B. Ärztinnen und Ärzte, Pflegekräfte)

Ziel der Fokusgruppen mit Patientinnen und Patienten ist, die in der Literatur und im Rahmen der Konzeptskizze benannten qualitätsrelevanten Themen, die als Qualitätsaspekte in das Qualitätsmodell eingingen, anhand ihrer Erfahrungen und Erlebnisse zu vertiefen. Dabei sollen wichtige Merkmale der Behandlung benannt werden, die

- aus ihrer Sicht von hoher Relevanz für die Qualität der thematisierten Versorgung sind,
- sich auf die konkreten Versorgungsabläufe beziehen und
- von den Patientinnen und Patienten in gleicher Weise beobachtbar und beurteilbar sind und damit keine Einzelmeinung darstellen.

Die Fokusgruppen mit den Gesundheitsfachkräften sollen diese Merkmale aus ihrem Erfahrungshorizont heraus ergänzen.

Im Falle von Rekrutierungsschwierigkeiten (z. B. Schwierigkeiten bei der Findung eines gemeinsamen Termins) werden alternativ persönliche leitfadengestützte Interviews durchgeführt.

Die Analyse der Ergebnisse der Fokusgruppen und Interviews wird in den Abschnitten 7.3.2 und 7.3.1 beschrieben.

### **Expertenberatung zu den Qualitätsmerkmalen**

Auf Basis der Ergebnisse der Literaturrecherche und der Fokusgruppen bzw. Interviews werden die Qualitätsmerkmale definiert. Die Qualitätsmerkmale stellen Konkretisierungen der jeweiligen Qualitätsaspekte dar und stellen im Kontext der Patientenbefragung die Basis für die Itemgenerierung dar. Folglich haben die Qualitätsmerkmale eine wegweisende Funktion bei der Fragebogenentwicklung und damit auch für die Definition der Qualitätsindikatoren.

Bevor anhand der Qualitätsmerkmale Fragebogenitems entwickelt werden, wird sichergestellt, dass die Qualitätsmerkmale die folgenden Eignungskriterien (siehe Kapitel 8) erfüllen:

- Patientenbezogene Bedeutung des Qualitätsmerkmals für den Qualitätsaspekt
- Zuschreibbarkeit der Ausprägung des Qualitätsmerkmals
- Ausgewogenheit des vorgeschlagenen Sets an Qualitätsmerkmalen

Für eine Diskussion der Qualitätsmerkmale hinsichtlich dieser Kriterien werden externe Fachexpertinnen und -experten in Form eines Expertengremiums einbezogen. Ziel der Expertenkonsultation ist zunächst die Darlegung der möglicherweise unterschiedlichen Ansichten und Einschätzungen der Expertinnen und Experten hinsichtlich der o. g. Kriterien (Krick 2013, Niederberger und Wassermann 2015). Auf Basis dieser Einschätzungen wird entschieden, ob die Qualitätsmerkmale in Form von Fragebogenitems in den Fragebogen übernommen werden sollen.

Um ein möglichst weites Spektrum an Fachexpertise zu erhalten und die Vielfalt der Einschätzungen und Erkenntnisse abzubilden, sollen möglichst alle für das jeweilige QS-Verfahren relevanten Gruppen eingebunden werden. In der Regel sind folgende Gruppen vertreten:

- medizinische Fachexpertinnen und -experten aus den beteiligten Versorgungsbereichen (Fachexpertise der Gesundheitsprofessionen)
- Patientenvertretung (Fachexpertise der Patientinnen und Patienten)
- Fachexpertinnen und -experten aus dem übergeordneten Versorgungskontext, z. B. Epidemiologie, Versorgungsforschung (kontextbezogene Fachexpertise)

Für eine zielorientierte Arbeitsweise sollen insgesamt mindestens 9 und nicht mehr als 15 Expertinnen und Experten vertreten sein. Dabei wird eine Ausgewogenheit vor allem der gesundheitsprofessionsspezifischen Fachexpertise und der Fachexpertise der Patientinnen und Patienten sichergestellt (Niederberger und Wassermann 2015). Der formalisierte Auswahlprozess der Expertinnen und Experten ist in Abschnitt 7.3.3 dargelegt.

Die Expertenberatung besteht aus zwei Phasen. In der ersten Phase werden die ausgewählten Qualitätsmerkmale mit Zuordnung zum Qualitätsaspekt und Darstellung ihrer Ableitung (Fokusgruppen, Literatur) den Expertinnen und Experten auf postalischem Weg zur Bewertung vorgelegt. Die Ergebnisse dieser postalischen Vorabbefragung dienen als Moderationsinstrument für

die zweite Phase, in der die aufbereiteten Antworten vor Ort diskutiert werden. Die protokollierten Aussagen der Expertinnen und Experten werden zusammenfassend aufbereitet und in den weiteren Entwicklungsprozess der Qualitätsmerkmale und Fragebogenitems einbezogen.

Sowohl die zusammengestellten Expertenbewertungen als auch der Umgang mit diesen werden als Teil des Abschlussberichts zur Entwicklung einer Patientenbefragung transparent gemacht.

### **Itementwicklung**

Damit die Ausprägung der Qualitätsmerkmale von den Patientinnen und Patienten beurteilt werden kann, werden sie mithilfe von Fragen und Antwortkategorien operationalisiert, die als Items in einem Fragebogen zusammengefasst werden. Mit dem Begriff „Item“ ist die Fragestellung gemeint, zu der auch immer eine Antwortskala gehört.<sup>5</sup> Die Qualitätsmerkmale stellen damit die Grundlage für die Itementwicklung dar.

Jeder Fragebogen enthält eine systematisch begründete Auswahl von Fragen, welche das dahinterliegende Konstrukt wiedergeben sollen. Dadurch ist ein Fragebogen das Verbindungsstück zwischen theoretischem Konzept und praktischer Analyse (Porst 1996). Bezogen auf die Qualitätssicherung beinhaltet der Fragebogen Fragen zu den systematisch und begründet ausgewählten Qualitätsmerkmalen. Der Fragebogen ist damit ein in sich geschlossenes Konzept, das die vorab definierten Qualitätsaspekte des Qualitätsmodells abbildbar macht.

An die Itementwicklung werden folgende Anforderungen gestellt:

- Alle für die Patientenbefragung relevanten Qualitätsmerkmale sollen im Sinn der Inhaltsvalidität im Fragebogen durch die Fragebogenitems abgebildet werden.
- Alle Fragen und Antwortkategorien sollen die zur Messung der Ausprägung des Qualitätsmerkmals relevanten Informationen zuverlässig abbilden. Es muss eine inhaltlich angemessene Operationalisierung der Qualitätsmerkmale vorliegen.
- Antworttendenzen, d. h. ein Antwortverhalten, das unabhängig von den Inhalten der Fragen ist, soll ausgeschlossen oder mindestens kontrolliert werden.

Da das Ziel der Patientenbefragung die vergleichende Darstellung der Qualität von Leistungserbringern ist, setzen sich die Fragebögen größtenteils aus geschlossenen Fragen mit vorgegebenen Antwortkategorien zusammen. Die Wahl der Antwortskalen wird durch die Art und den Inhalt der Frage bestimmt, sie können somit nominal-, ordinal- oder intervallskaliert sein.

Die Verbalisierung oder Bezifferung der Antwortskalen erfolgt – wenn inhaltlich gerechtfertigt – möglichst einheitlich, d. h., es wird verzichtet auf einen Wechsel der Antwortpole, eine unterschiedliche Anzahl der Skalenpunkte (z. B. einmal 5-stufig, einmal 7-stufig) oder unterschiedliche Benennungen der Skalenabstufungen, die nicht inhaltsbegründet sind. Solche unerwarteten Wechsel steigern nicht die Aufmerksamkeit der Befragten, sondern führen vielmehr zu nicht plausiblen Unterschieden im Antwortverhalten (Hippler et al. 1991, Porst 2014).

---

<sup>5</sup> Eine Ausnahme stellen offene Fragen dar, die im vorliegenden Kontext allerdings eine untergeordnete Rolle spielen.

Neben den Items zu den Qualitätsmerkmalen wird ein Fragebogen immer auswertungsrelevante Zusatzfragen beinhalten, die in keinem direkten Zusammenhang mit den einzelnen Qualitätsmerkmalen stehen (z. B. Items für Zwecke der Risikoadjustierung oder Stratifizierung, wie soziodemografische Angaben oder Angaben zum Gesundheitszustand oder dem Gesundheitsverhalten).

### **Pretestung**

Um Hinweise über die Funktionsfähigkeit oder eventuelle Mängel von Fragen oder des Fragebogens insgesamt zu erhalten, unterläuft jeder Fragebogenentwurf einer Testung. Pretests gelten als unabdingbarer Bestandteil jeder Fragebogenentwicklung und als eine Voraussetzung für hohe Datenqualität (Bradburn et al. 2004, Lenzner et al. 2015, Porst 2014, Prüfer und Rexroth 1996). Nach Porst (2014) werden mit der Pretestung folgende Ziele verfolgt:

- Prüfung der Verständlichkeit der Fragen
- Prüfung der Reihenfolge der Fragen
- Untersuchung von Kontexteffekten
- Exploration von Problemen der Befragten mit ihrer Aufgabe
- Exploration des Zustandekommens der Antworten
- Exploration von Interesse und Aufmerksamkeit der Befragten
- Einschätzungen zum Wohlbefinden der Befragten bei der Befragung
- Ermittlung der Zeitdauer der Beantwortung des Fragebogens
- Feststellung technischer Probleme
- Berechnung statistischer Kennwerte

Es werden im Allgemeinen zwei Arten von Pretests unterschieden: kognitive Pretests und Standard-Pretests<sup>6</sup>, die jeweils beide auch bei der Fragebogenentwicklung im Rahmen der QS-Verfahren eingesetzt werden.

### **Kognitiver Pretest**

Kognitive Pretests geben Einblicke in die kognitiven Prozesse, die bei den Befragten bei der Beantwortung der Fragen ablaufen. So werden die Verständlichkeit und Interpretation der Fragen sowie das Zustandekommen der Antworten, d. h. die zugrunde liegenden Entscheidungsprozesse, Erinnerungsleistungen und Zuordnungen der Antwort zur formalen Antwortkategorie, aktiv bei den Befragten erfasst. Ziel des kognitiven Pretests ist es, mögliche Probleme bei der Beantwortung des Fragebogens zu ermitteln, deren Ursachen aufzudecken und Verbesserungen zu erarbeiten. Kognitive Pretests beziehen sich auf die Prüfung von Einzelitems und werden zu Beginn der Fragebogenentwicklung eingesetzt (Lenzner et al. 2015, Porst 2014, Prüfer und Rexroth 2005, Willis 2004, Willis 2005).

Zu den am häufigsten verwendeten Techniken zählen:

- Nachfragetechniken (*Probing*), bei denen Begriffe, Fragentexte oder gegebene Antworten mittels einer oder mehrerer Zusatzfragen (*probes*) hinterfragt werden

---

<sup>6</sup> In der Literatur werden für „Standard-Pretest“ häufig auch die Bezeichnungen „Standardbeobachtungspretest“ oder „klassischer Pretest“ verwendet.

- Bewertung der Verlässlichkeit der Antwort (*Confidence Rating*), bei der die Befragten den Grad der Verlässlichkeit ihrer Antworten bewerten
- Paraphrasieren (*Paraphrasing*), bei dem die befragte Person die beantwortete Frage in eigenen Worten wiedergibt
- Sortiertechniken (*Card Sorting*), bei denen die Befragten Begriffe oder Situationen bestimmten Kategorien zuordnen
- Technik des lauten Denkens (*Think Aloud*), bei der die Befragten sämtliche Gedankengänge bei der Beantwortung des Fragebogens bzw. der Fragen laut verbalisieren

Die Wahl der Technik richtet sich nach dem dahinterliegenden Ziel der Pretestung, aber auch nach den Möglichkeiten ihres Einsatzes. Grundsätzlich nehmen Nachfragetechniken in der kognitiven Pretestung des IQTIG eine zentrale Rolle ein. Sie gelten generell als sehr effektiv, da die Befragten sehr konkret Aussagen zum Verständnis einer Frage oder einzelner Begriffe geben. Je nach Zielsetzung werden diese um weitere Techniken ergänzt (z. B. um das Paraphrasieren oder die Technik des lauten Denkens), sodass auch Mischformen der kognitiven Pretestung zum Einsatz kommen (Prüfer und Rexroth 2005).

Die Testung erfolgt in Form von Einzelinterviews mit Patientinnen und Patienten, die sich aus Repräsentanten der Zielgruppe des QS-Verfahrens (d. h. Patientinnen und Patienten, die vom Versorgungsbereich des QS-Verfahrens betroffen sind) zusammensetzen. Die Anzahl der Testpersonen liegt je nach Fragestellung zwischen 10 und 30 Patientinnen und Patienten. Vor Durchführung des Pretests wird festgelegt, welche der Fragebogenitems im kognitiven Pretest getestet werden sollen. Die Auswahl richtet sich nach der Bedeutung der Fragen, aber auch danach, ob bereits bestimmte Probleme erwartet werden bzw. Unsicherheiten in der Fragenformulierung bestehen.

Grundlage des kognitiven Pretests ist ein halbstandardisierter Fragebogen, in dem die zu testenden Fragen zusammen mit der entsprechenden Nachfrage aufgeführt sind. Im offenen Teil werden die Kommentare der Befragten und die Erkenntnisse der Testleiterin bzw. des Testleiters eingetragen. Zusätzlich werden die Interviews mit Einverständnis der Befragten aufgezeichnet und im Anschluss transkribiert und analysiert. Auf Basis der Ergebnisse des kognitiven Pretests wird der erste Fragebogenentwurf modifiziert.

### **Standard-Pretest**

Zur Prüfung der messtheoretischen Eigenschaften des Fragebogens wird ein Standard-Pretest durchgeführt. Dieser wird am Ende der Fragebogenentwicklung eingesetzt, nachdem der kognitive Pretest durchgeführt wurde und dadurch grundlegende Verständnisprobleme bereits beseitigt wurden. Für den Standard-Pretest wird der Fragebogen unter annähernd ähnlichen Bedingungen, wie sie im Regelbetrieb des QS-Verfahrens erwartet werden, in einer Patientenstichprobe eingesetzt. Die Rekrutierung der Teilnehmerinnen und Teilnehmer erfolgt über die Leistungserbringer. Ambulante und stationäre Einrichtungen, die auch im späteren QS-Verfahren einbezogen werden, ermöglichen durch ihre freiwillige Teilnahme die Befragung einer Patientenstichprobe.

Auf Basis der im Standard-Pretest gewonnenen Daten findet die Validierung des Fragebogens statt. Hierzu werden Verfahren zur Evaluation der Fragenqualität, wie der Analyse und Darstellung der

Antwortverteilungen, und Verfahren zur Überprüfung der Gütekriterien der Messung angewendet. Außerdem wird auf Basis inhaltlicher Überlegungen geprüft, welche Items zu Indizes zusammengefasst werden sollen.

Zusammenfassend erlaubt die Kombination von kognitivem Pretest und Standard-Pretest eine umfassende Prüfung des Fragebogens, die zu dessen Optimierung beiträgt. Nach erfolgreicher Validierung liegt ein Fragebogen vor, der regelhaft in einem QS-Verfahren eingesetzt werden kann.

### **Definition von Qualitätsindikatoren**

Im letzten Schritt der Entwicklung wird vorgeschlagen, welche Fragebogenitems und/oder zu Indizes zusammengefasste Items als Qualitätsindikatoren gelten sollen. Als Qualitätsindikatoren gelten grundsätzlich leistungserbringerbezogene Qualitätsmessungen, die ein konkretes Qualitätsziel verfolgen und eine Qualitätsbewertung (siehe Abschnitt 3.4) ermöglichen. Dies sind bei der Patientenbefragung die Fragebogenitems, die aufgrund der Fragebogenkonstruktion häufig zu Indizes zusammengefasst werden. Schon bei der Entwicklung der Qualitätsmerkmale, die die Basis für die Itementwicklung darstellen, kommen nur solche Qualitätsmerkmale in Betracht, die sich dieser Definition folgend für die Beurteilung der Behandlungsqualität eignen. Folglich sollen alle auf Basis der Merkmale entwickelten Items bzw. Indizes grundsätzlich geeignet sein, als Qualitätsindikatoren zu fungieren. Durch Festlegung eines Referenzbereichs wird ein Item bzw. ein Index zum Qualitätsindikator.

Neben den Items, die die Kriterien als Qualitätsindikatoren erfüllen, enthalten Fragebögen zur Patientenbefragung weitere Items, die nicht als Qualitätsindikatoren fungieren, wie etwa Items zu soziodemografischen Angaben oder Items für die Risikoadjustierung, die für die Bewertung der Qualitätsindikatoren aber von Bedeutung sind. Auch bei Befragungen gilt das Gebot der Datensparsamkeit.

#### **4.4.4 Messtheoretische Gütekriterien der Patientenbefragung**

Die Eignung der entwickelten Fragebögen für Patientenbefragungen wird nach den allgemeinen Eignungskriterien für Qualitätsindikatoren der medizinischen Versorgung (siehe Kapitel 8) beurteilt. Für Fragebogeninstrumente sind insbesondere die messtheoretischen Kriterien Objektivität, Reliabilität und Validität zu berücksichtigen. Dies betrifft sowohl die besonderen Anforderungen an die Objektivität als auch die spezifischen Anforderungen, die der Einsatz von PREM und PROM an Methoden der Reliabilitäts- und Validitätsbeurteilung im Kontext der Qualitätssicherung stellt.

### **Objektivität**

Als Objektivität wird das Maß der Unabhängigkeit eines Messergebnisses von den Untersuchungsumständen bezeichnet (Bühner 2011, Moosbrugger und Kelava 2012). Die Durchführungsobjektivität von Patientenbefragungen wird durch standardisierte Befragung mittels Fragebogen sichergestellt – in Abgrenzung zu beispielsweise mündlichen Befragungen. Auf Basis von § 299 SGB V werden grundsätzlich alle Fragebögen an die Privatadresse der Patientinnen

und Patienten durch eine externe Stelle versendet. Damit wird sichergestellt, dass Effekte beim Ausfüllen in den Räumlichkeiten des Leistungserbringers, wie etwa soziale Erwünschtheit oder Aufforderungscharakter der Situation, keinen Einfluss auf die Durchführungsobjektivität haben. Weiterhin wird die Auswertungs- und Interpretationsobjektivität durch standardisierte Auswertungs- und Rückmeldekonzeppte sichergestellt.

### Reliabilität

Reliabilität stellt ein Maß der klassischen Testtheorie für die Präzision einer Messung dar (z. B. Bühner 2011, Moosbrugger und Kelava 2012).<sup>7</sup> Der klassischen Testtheorie liegt die Annahme zugrunde, dass sich Messwerte in einen wahren, nicht direkt beobachtbaren Wert und einen Messfehler zerlegen lassen. Beobachtbar ist allein die Summe aus beiden: der tatsächlich beobachtete Messwert. Inwieweit ein Messwert dem wahren Wert nahekommt, kann jedoch nicht ohne Weiteres bestimmt werden. Dies impliziert, dass nicht alle Unterschiede (= Varianz) zwischen den Messwerten einer Messreihe (z. B. Unterschiede zwischen den Angaben mehrerer Befragter) auf wahre Unterschiede im gemessenen Merkmal zwischen den Befragten zurückgehen müssen, sondern auch durch Messfehler bedingt sein können.

Reliabilität ist formal definiert als der Anteil an der Varianz ( $\sigma^2$ ) einer Messreihe, der auf tatsächliche („wahre“) Unterschiede zwischen den Messobjekten zurückgeht (Bühner 2011). Reliabilitätskoeffizienten können dementsprechend auch als ein „Signal-zu-Rauschen-Verhältnis“ verstanden werden, wobei das „Signal“ den wahren Unterschieden zwischen den Messobjekten entspricht:

$$\text{Reliabilität} = \frac{\sigma_{\text{Signal}}^2}{\sigma_{\text{Signal}}^2 + \sigma_{\text{Rauschen}}^2}$$

Je größer der Varianzanteil ist, der auf wahre Unterschiede zwischen den Messobjekten zurückgeht, desto höher ist die Reliabilität und desto sicherer kann davon ausgegangen werden, dass Unterschiede zwischen zwei gemessenen Werten auch tatsächlich wahre Unterschiede widerspiegeln. Werden beispielsweise Patientinnen und Patienten zu partizipativer Entscheidungsfindung befragt, ist die Reliabilität ein Maß dafür, inwieweit Unterschiede in den Angaben wahre Unterschiede in dem Ausmaß partizipativer Entscheidungsfindung zwischen den Patientinnen und Patienten widerspiegeln. Fehlervarianz (= „unechte“ Unterschiede) kommt beispielsweise durch Fehlinterpretationen von Fragen, soziale Erwünschtheit im Antwortverhalten etc. zustande.

Methoden, um die Reliabilität von Messungen auf der Ebene der Patientinnen und Patienten zu schätzen, sind beispielsweise die Test-Retest-Methode, die Paralleltest-Methode und Methoden der internen Konsistenz (z. B. Cronbachs  $\alpha$ ). Diese Methoden stellen unterschiedliche Anforderungen an den Befragungskontext und werden daher nach Angemessenheit für die konkrete Befragungssituation eingesetzt. So setzen Methoden der internen Konsistenz beispielsweise ein-

<sup>7</sup> Da die klassische Testtheorie Intervallskalenniveau voraussetzt, beschränkt sich diese Betrachtung auf intervallskalierte Messungen.

dimensionale Skalen voraus und die Test-Retest-Methode unterstellt, dass zeitlich stabile Merkmale gemessen werden, keine Reifung und keine Erinnerung an die Antworten der ersten Befragung vorliegt (Bühner 2011).

Im Kontext der Qualitätssicherung ist das Ziel der Patientenbefragungen meist ein Vergleich zwischen den Ergebnissen der Leistungserbringer untereinander oder ein Vergleich der Ergebnisse der Leistungserbringer mit einem fixen Wert (z. B. ein Referenzwert oder der Mittelwert aller Leistungserbringer; siehe Kapitel 10). Dazu werden in der Regel die Angaben der befragten Patientinnen und Patienten eines Leistungserbringers gemittelt, um einen Indikatorwert pro Leistungserbringer zu berechnen. Auch auf dieser Ebene kann die Frage gestellt werden, inwieweit Unterschieden in den Indikatorwerten zwischen Leistungserbringern wahre Unterschiede zwischen den Leistungserbringern zugrunde liegen (O'Brien 1990). Diese Betrachtungsweise ergänzt die notwendige Betrachtung der Reliabilität auf Ebene der Patientinnen und Patienten um einen für die Qualitätssicherung wichtigen Aspekt, da Qualitätsindikatoren immer auf Ebene der Leistungserbringer betrachtet werden.

Unter der Annahme, dass auch für Leistungserbringer ein wahrer (zugrunde liegender), nicht direkt beobachtbarer Indikatorwert existiert, kann die Varianz in den Leistungserbringerwerten auch in wahre Unterschiede zwischen den Leistungserbringern und Fehlervarianz zerlegt werden. Analog zur obigen Definition kann Reliabilität in diesem Fall verstanden werden als das Verhältnis von wahrer Varianz zwischen den Leistungserbringern zu der Varianz des gemessenen Indikators (O'Brien 1990). Diese Betrachtungsweise wird durch Kennzahlen wie die Intraklassenkorrelation oder verwandte Reliabilitätskoeffizienten aufgegriffen (Krippendorff 1970, O'Brien 1990, Shrout und Fleiss 1979). Sofern angemessen werden diese Kennzahlen berechnet, um auch die Reliabilität auf Ebene der Leistungserbringerwerte zu schätzen.

### **Reliabilität und Stichprobengröße im Regelbetrieb**

Im Regelbetrieb wirkt sich mangelnde Reliabilität von Messungen auf Patientenebene auf die Breite der Konfidenzintervalle um die Ergebnisse der Leistungserbringer (z. B. Mittelwert) aus. Die Breite dieses Konfidenzintervalls wird bei konstantem  $\alpha$ -Niveau (Irrtumswahrscheinlichkeit für einen Fehler erster Klasse) allein durch die Größe des Standardfehlers (auch Stichprobenfehler genannt) bestimmt. Der Standardfehler hängt seinerseits maßgeblich von der Varianz der Messung auf Patientenebene ab: Je größer die Varianz der Patientenangaben innerhalb der Gruppe behandelter Patientinnen und Patienten eines Leistungserbringers ist, desto größer wird der Standardfehler und desto breiter wird damit das Konfidenzintervall. Die Varianz der Patientenangaben innerhalb eines Leistungserbringers hängt ihrerseits von der Reliabilität der Messung auf Patientenebene ab: Je geringer diese ist, desto größer wird die Varianz der Patientenangaben eines Leistungserbringers und desto größer wird damit der Standardfehler und in letzter Konsequenz wird das Konfidenzintervall breiter. Die Größe des Konfidenzintervalls ist demnach ein Maß für die Präzision der Schätzung des wahren (zugrunde liegenden) Indikatorwerts des Leistungserbringers, in die die Reliabilität auf Ebene der Patientinnen und Patienten eingeht.

Da allerdings auch die Größe der Patientenstichproben des jeweiligen Leistungserbringers in die Berechnung des Standardfehlers und damit des Konfidenzintervalls eingeht, werden beide mit steigender Stichprobengröße kleiner bzw. schmaler. Daher steigt die Präzision der Schätzung des Indikatorwerts eines Leistungserbringers allein schon durch die Berücksichtigung möglichst vieler Patientinnen und Patienten pro Leistungserbringer (Moosbrugger und Kelava 2012). Aufgrund dieses Aggregationsvorteils sind, im Gegensatz zum individualdiagnostischen Einsatz, sinnvolle Aussagen auf Ebene der Leistungserbringer auch bei niedrigerer Reliabilität auf Patientenebene möglich (s. a. Frost et al. 2007).

### **Validität**

Als valide im psychometrischen Sinn wird eine Messung bezeichnet, die tatsächlich misst, was sie messen soll (Hartig et al. 2012). Während das bei vielen medizinischen Messungen selbstevident ist (z. B. Messung der Körpergröße), ist dies bei der Messung nicht direkt beobachtbarer, also latenter Konstrukte (z. B. Persönlichkeit) nicht der Fall. Bei der Prüfung der Validität einer Messung unterscheidet man daher zwischen Messungen latenter Konstrukte im psychometrischen Sinn und Fragen nach konkreten Sachverhalten. Letzteres ist beispielsweise per Definition bei PREM der Fall (z. B. „Wurde Ihnen erklärt, weshalb Sie die Ihnen verschriebenen Medikamente einnehmen sollen?“), kann aber auch auf PROM zutreffen (z. B. „Wie viele Treppenstufen können Sie nach der Behandlung wieder problemlos steigen?“). Im Gegensatz zu Fragen nach konkreten Sachverhalten zielen Fragen zur Erfassung latenter Konstrukte nicht auf die unmittelbar erfragten Sachverhalte ab, sondern auf ein zugrunde liegendes Konstrukt. Beispielsweise zielt die Frage „Stehen Sie gerne im Mittelpunkt?“ als Teil eines Persönlichkeitsfragebogens nicht darauf ab herauszufinden, ob die befragte Person gerne im Mittelpunkt steht, sondern soll der Messung des latenten Konstrukts „Extraversion“ dienen. Ob allerdings diese Frage tatsächlich dieses latente Konstrukt abbildet, muss belegt werden, da es nicht selbstevident ist – der Zweck der Validitätsprüfung im psychometrischen Sinn.

Da das Institut einen faktenorientierten Befragungsansatz verfolgt, werden Fragebögen sowohl Items enthalten, die latente Konstrukte erfassen sollen, als auch Items, die konkrete und damit nicht latente Sachverhalte erfassen. Vor allem bei der Erfragung konkreter Sachverhalte wie etwa im Kontext von PREM muss sichergestellt werden, dass die ereignisbezogenen Fragen alle für den konkreten Qualitätsaspekt relevanten Ereignisse berücksichtigen und keine darüber hinaus. Dieser Sachverhalt wird als Inhaltsvalidität bezeichnet. Grundsätzlich lässt sich Inhaltsvalidität nicht numerisch abbilden, sondern kann nur aufgrund sachlogischer Überlegungen als mehr oder weniger gegeben eingestuft werden (Moosbrugger und Kelava 2012). Um die Inhaltsvalidität der Fragebögen zur Patientenbefragung sicherzustellen, werden wie weiter oben beschrieben folgende Instrumente eingesetzt (siehe Abschnitt 4.4.3):

- Literaturrecherche und -analyse
- Fokusgruppen
- Konsultationen mit externen Expertinnen und Experten (Expertengremium)

All diese Instrumente dienen dazu, auf Basis der relevanten Qualitätsaspekte und -merkmale einen repräsentativen und damit inhaltsvaliden Itempool zu generieren.

### **Nebengütekriterien**

Neben diesen Hauptgütekriterien spielen für die Patientenbefragung auch die Nebengütekriterien Ökonomie, Nützlichkeit und Zumutbarkeit eine entscheidende Rolle. Es sollen Fragebögen entwickelt werden, die in angemessener Zeit ausgefüllt werden können, möglichst wenig Material benötigen und einfach durchzuführen sind (Ökonomie). Weiterhin sollen die entwickelten Fragebögen für den Verfahrenszweck der Qualitätssicherung nützlich sein und keine darüber hinausgehenden Aspekte beinhalten (Nützlichkeit). Nicht zuletzt wird großen Wert gelegt auf die Zumutbarkeit der Fragebögen für die Patientinnen und Patienten in zeitlicher, psychischer und physischer Hinsicht. In der Summe bedeutet dies, dass Fragebögen entwickelt werden, bei denen die Anzahl zu befragender Patientinnen und Patienten und deren zeitliche und ggf. auch psychische Belastung durch die Befragung im Verhältnis zum Nutzen für die Qualitätssicherung stehen. Die einzelnen methodischen Entwicklungsschritte werden durch diese Kriterien geleitet.

#### **4.4.5 Risikoadjustierung von Qualitätsindikatoren der Patientenbefragung**

Der Zweck der Risikoadjustierung ist es, einen fairen Vergleich zwischen verschiedenen Leistungserbringern herzustellen, die sich in der Zusammensetzung ihrer Patientengruppen unterscheiden. Das grundsätzliche Vorgehen bei der Entwicklung von Risikoadjustierungsmodellen in der Qualitätssicherung wird in Abschnitt 13.2 dargestellt. Im Folgenden wird daher nur auf Besonderheiten von Risikoadjustierungskonzepten bei Patientenbefragungen eingegangen.

Für die Risikoadjustierung von Qualitätsindikatoren der Patientenbefragung kann aus datenflusslogistischen Gründen zunächst keine Verknüpfung auf Fallebene zwischen den Befragungsdaten und den anderen Daten eines QS-Verfahrens (z. B. QS-Dokumentation, Sozialdaten bei den Krankenkassen) hergestellt werden. Daher können bisher nur über den Fragebogen erfassbare Risikovariablen für die Adjustierung von Qualitätsindikatoren der Patientenbefragung herangezogen werden. Die Erfassung von Risikovariablen über die Befragung der Patientinnen und Patienten selbst eröffnet allerdings auch die Möglichkeit, Risikovariablen zu berücksichtigen, über die nur die Patientinnen und Patienten selbst Auskunft geben können (z. B. körperliche Belastbarkeit, psychische Verfassung). Dabei wird berücksichtigt, dass die zumutbare Anzahl an Items zur Risikoadjustierung in einem Fragebogen begrenzt ist und dass bestimmte Fragen nicht zuverlässig von allen Patientinnen und Patienten beantwortet werden können (z. B. konkrete Medikationen, Laborparameter).

Bei der Entwicklung von Konzepten für die Risikoadjustierung von Qualitätsindikatoren der Patientenbefragung werden zunächst innerhalb der initialen Literaturrecherche grundsätzliche, patientenseitige Risikofaktoren identifiziert. Im Schritt der Itementwicklung werden neben den Items für die eigentliche Patientenbefragung auch Items zur Erfassung der identifizierten Risikofaktoren entwickelt. Im Rahmen des Standard-Pretests werden explorative Analysen durchgeführt, um die empirische Relevanz der operationalisierten Risikofaktoren festzustellen. In den finalen Fragebogen werden nur solche Items zur Risikoadjustierung aufgenommen, die im Standard-Pretest einen statistisch signifikanten und substanziellen Effekt auf die Qualitätsindikatoren aufweisen. Die Auswertung der Befragungsdaten wird grundsätzlich mit den Standardinstrumenten der Statistik, Risikoadjustierung, Indexbildung und Bewertung vorgenommen. So

können Qualitätsaspekte auch durch ein Zusammenführen von Ergebnissen der QS-Dokumentation und der Befragung valide abgebildet werden.

## 5 Durchführung, Pflege und Weiterentwicklung von Qualitätssicherungsverfahren

Die Aufgaben des IQTIG schließen neben der Wahrnehmung von Entwicklungsaufgaben auch die Beteiligung an der Umsetzung der vom G-BA beschlossenen Qualitätssicherungsverfahren ein. Vor dem Hintergrund der Weiterentwicklung der medizinischen Wissenschaft ist es in diesem Zusammenhang notwendig, die QS-Verfahren im Regelbetrieb nicht nur richtlinienkonform durchzuführen, sondern sie auch kontinuierlich auf Anpassungsbedarf zu prüfen und ggf. notwendigen Änderungsbedarf zeitnah zur Umsetzung vorzuschlagen. Letzteres geschieht im Rahmen der Verfahrenspflege, die eng mit den Prozessen der Verfahrensdurchführung verbunden ist. Verfahrensweiterentwicklung beschreibt dagegen einen weitergehenden Prozess, in dem losgelöst von der Durchführung ein umsetzbares Konzept zur Umsetzung umfangreicherer Änderungen (z. B. Erweiterungen eines QS-Verfahrens um eine neue Datenquelle) erarbeitet wird, das dann vom G-BA beschlossen werden muss, der das IQTIG mit der Umsetzung beauftragt.

Im Folgenden werden vor allem die Verfahrenspflege und -weiterentwicklung und deren Teilprozesse näher dargestellt. Es wird dabei auch darauf eingegangen, wie die methodischen Instrumente des IQTIG im Rahmen dieser beiden Vorgehensweisen eingesetzt werden. Da sich die externe Qualitätssicherung des G-BA derzeit in einer Umbruchsituation befindet, in der mehrere Richtlinien zugleich Regelungskraft entfalten, wird auf Besonderheiten an entsprechender Stelle hingewiesen.

### 5.1 Verfahrensdurchführung

Der Begriff „Verfahrensdurchführung“ beschreibt die Gesamtheit aller Prozesse, die vonseiten des IQTIG notwendig sind, um die vom G-BA beschlossenen QS-Verfahren gemäß seiner Richtlinien durchzuführen. Dabei werden direkte (bundesbezogene) und indirekte (landesbezogene) QS-Verfahren unterschieden. Direkte QS-Verfahren zeichnen sich dadurch aus, dass das IQTIG alle Aufgaben der Durchführung übernimmt, die mit der Umsetzung des Regelbetriebs verbunden sind (insbesondere Datenannahme, Auswertung, Rückmeldeberichte, Klärung von rechnerischen Auffälligkeiten mit den Leistungserbringern, Vereinbarung und Durchführung weiterer Maßnahmen). Dies betrifft derzeit insbesondere QS-Verfahren mit relativ wenigen beteiligten Leistungserbringern (z. B. Organtransplantationen). Bei diesen QS-Verfahren wäre auf Ebene der Bundesländer der Aufbau einer Fachgruppe zur Klärung von rechnerischen Auffälligkeiten aufgrund der wenigen innerhalb eines Bundeslandes verfügbaren Expertinnen und Experten nicht möglich. Darüber hinaus könnten die Beratungen bei sehr wenigen Leistungserbringern je Bundesland faktisch nicht pseudonymisiert durchgeführt werden. Bei indirekten QS-Verfahren werden im Gegensatz dazu wesentliche Anteile von den jeweils zuständigen Stellen auf Ebene der Bundesländer durchgeführt. Dies beinhaltet insbesondere die Klärung von rechnerischen Auffälligkeiten im Strukturierten Dialog mit den jeweils betroffenen Leistungserbringern.

Im Unterschied zur Verfahrenspflege und zur Verfahrensweiterentwicklung werden bei der Verfahrensdurchführung per Definition keine inhaltlichen Änderungen an den QS-Verfahren im Regelbetrieb vorgenommen. Sie betrifft insbesondere die Datenannahme, die Anwendung der finalisierten Rechenregeln, die Erstellung der Auswertungen, die Klärung von rechnerischen Auffälligkeiten mit den Leistungserbringern sowie im Bedarfsfall auch die Vereinbarung von weitergehenden Maßnahmen mit den betroffenen Leistungserbringern. Auch die Verfahrensdurchführung wendet allerdings regelhaft methodische Instrumente an. Beispiele hierfür sind die Bewertung der Bundesergebnisse unter Beteiligung der Bundesfachgruppen (Expertengruppen des IQTIG) oder die Bewertung der Qualitätsindikatoren hinsichtlich ihrer Eignung zur einrichtungsbezogenen Veröffentlichung ihrer Ergebnisse. Eine detaillierte Beschreibung der Prozesse im Rahmen der Verfahrensdurchführung unter Berücksichtigung der jeweils zugrunde liegenden Richtlinien des G-BA erfolgt im Verfahrenshandbuch<sup>8</sup> des IQTIG, das regelmäßig aktualisiert wird. Auch die Anpassung der QS-Verfahren an geänderte Vorgaben in den Richtlinien des G-BA ist Gegenstand der Verfahrensdurchführung.

Alle inhaltlichen Änderungen, die das IQTIG an den QS-Verfahren im Regelbetrieb vornimmt, werden entweder im Rahmen der Verfahrenspflege oder der Verfahrensweiterentwicklung vorgenommen.

## 5.2 Verfahrenspflege

Der Begriff „Verfahrenspflege“ bezeichnet die Gesamtheit aller Prozesse, die das IQTIG in regelmäßigen Abständen mit dem Ziel durchführt, inhaltlichen Anpassungsbedarf der QS-Verfahren im Regelbetrieb festzustellen und die nötigen Änderungen vorzuschlagen. Anpassungsbedarf entsteht insbesondere durch die Weiterentwicklung der Erkenntnisse der medizinischen Wissenschaften, z. B. ausgedrückt durch Weiterentwicklung der Klassifikationen des Deutschen Instituts für Medizinische Dokumentation und Information (DIMDI), aber auch durch Überarbeitung von Leitlinien oder durch die Veröffentlichung von einschlägigen Forschungsergebnissen. Darüber hinaus werden im Rahmen der Verfahrenspflege Fehler korrigiert, Formulierungen präzisiert sowie einzelne Verfahrensbestandteile optimiert. Die Prozesse der Verfahrenspflege werden in der Regel im Jahresrhythmus durchgeführt.

Die Verfahrenspflege soll gewährleisten, dass die QS-Verfahren im Regelbetrieb zuverlässig auf dem höchsten möglichen wissenschaftlich-fachlichen Niveau durchgeführt werden. Dies beinhaltet die Prüfung und im Bedarfsfall das Hinwirken auf Anpassungen aller Verfahrensbestandteile hinsichtlich

- der Eignung der eingesetzten Qualitätsindikatoren zur Abbildung der ihnen zugrunde liegenden Qualitätsmerkmale,
- des Verhältnisses von Aufwand und Nutzen,
- Fehlervermeidung, Fehlersuche und Fehlerbehebung sowie
- der Sicherstellung einer kontinuierlichen Optimierung.

---

<sup>8</sup> <https://iqtig.org/das-iqtig/grundlagen/verfahrensgrundlagen>

Anders als bei der Verfahrensdurchführung schlägt das IQTIG bei der Verfahrenspflege inhaltliche Anpassungen an den im Regelbetrieb durchgeführten QS-Verfahren vor. Das dem jeweiligen QS-Verfahren (mindestens implizit) zugrunde liegende Qualitätsmodell soll durch die Umsetzung dieser Vorschläge jedoch nicht verändert werden, d. h., im Rahmen der Verfahrenspflege wird das IQTIG keine Qualitätsindikatoren zu Qualitätsaspekten vorschlagen, die in dem bestehenden QS-Verfahren nicht betrachtet werden. Sofern es aus medizinisch-fachlicher Sicht notwendig ist, können jedoch Anpassungen an bestehenden Qualitätsindikatoren oder Streichungen vorgeschlagen werden. Möglich ist auch die Neuentwicklung von Qualitätsindikatoren zu den bereits im QS-Verfahren abgebildeten Qualitätsaspekten. Alle Änderungsvorschläge werden mit dem Gemeinsamen Bundesausschuss abgestimmt, und soweit notwendig, von diesem beschlossen.

Im Gegensatz zur Verfahrensweiterentwicklung beruht die Verfahrenspflege nicht auf einem verfahrensspezifisch formulierten Auftrag des G-BA, sondern auf der gesetzlichen Grundlage des § 137a Abs. 3 Satz 2 Nr. 3 SGB V in Verbindung mit dem Beschluss des G-BA über die Beauftragung des IQTIG zur Beteiligung an der Durchführung der einrichtungsübergreifenden Qualitätssicherung und zur Veröffentlichung der Qualitätssicherungsmaßnahmen vom 27. November 2015 (G-BA 2015). Der dort verwendete Begriff der „Systempflege“ wurde vonseiten des IQTIG zugunsten des Begriffs „Verfahrenspflege“ verlassen, da sich die Prozesse der Verfahrenspflege nicht auf ein abstraktes „System“, sondern auf konkrete QS-Verfahren beziehen.

### 5.2.1 Informationsquellen

Die Verfahrenspflege nutzt eine Reihe unterschiedlicher Informationsquellen:

- Hinweise (z. B. von Leistungserbringern, QS-Stellen auf der Landesebene, Fachgesellschaften, Verbänden)
- die jährliche Bundesauswertung sowie ggf. Quartalsauswertungen (einschließlich der Ergebnisse aus Vorjahren, sofern diese bezogen auf den einzelnen Qualitätsindikator vorliegen und vergleichend bewertet werden können)
- vorgenommene Änderungen an den Qualitätsindikatoren
- vorgenommene Änderungen an der Datenerfassung einschließlich der eingesetzten QS-Filter
- Ergebnisse der Strukturierten Dialoge (einschließlich der Ergebnisse aus Vorjahren, sofern diese bezogen auf den einzelnen Qualitätsindikator vorliegen und vergleichend bewertet werden können)
- Ergebnisse der Verfahren zur Datenvalidierung
- Ergebnisse aus Aktualisierungen der Evidenzgrundlagen (insbesondere Einbeziehung von Leitlinien-Aktualisierungen sowie neuen Forschungsergebnissen)
- Ergebnisse von Sonderauswertungen
- Ergebnisse der Prüfung der Qualitätsindikatoren auf Erfüllung der Eignungskriterien

### 5.2.2 Einbindung von Expertengruppen

Jedes QS-Verfahren im Regelbetrieb wird auf Bundesebene von einer Expertengruppe medizinisch-wissenschaftlich begleitet; eine Expertengruppe kann auch für mehrere QS-Verfahren zuständig sein. Diese Expertengruppen übernehmen im Rahmen der Verfahrenspflege Aufgaben, die mit den Aufgaben der Entwicklungspanels im Rahmen der Verfahrensneuentwicklung vergleichbar sind, nämlich der Bewertung von Informationen aus Expertensicht. Sie werden vom IQTIG zu allen aus Sicht des Instituts wesentlichen im Verfahrensverlauf gesammelten und erarbeiteten Informationen beratend hinzugezogen und sollen auf der Grundlage ihrer fachlichen Expertise Empfehlungen zum weiteren Vorgehen in den jeweiligen QS-Verfahren abgeben.

Für die Zusammensetzung dieser Expertengruppen sind die hierzu festgelegten Grundsätze des IQTIG (siehe Kapitel 7) maßgeblich. Erreicht werden soll eine angemessene Repräsentation der von den jeweiligen QS-Verfahren betroffenen medizinischen Fachgebiete (z. B. Kardiologie) und der Fachvertreter der jeweils einbezogenen Bereiche der Versorgung (z. B. stationäre Versorgung, vertragsärztliche Versorgung). Bei der Zusammensetzung wird darüber hinaus die Expertise von Patientinnen und Patienten sowie von Vertreterinnen und Vertretern der Pflege und falls notwendig auch weiterer beteiligter Gesundheitsfachberufe angemessen berücksichtigt. Durch die Einbindung eines breiten Wissens- und Erfahrungsspektrums soll eine fundierte medizinisch-fachliche Beratung des IQTIG für die Verfahrensdurchführung und die Verfahrenspflege sichergestellt werden. Damit wird zugleich eine möglichst hohe Akzeptanz für die QS-Verfahren bei den Leistungserbringern angestrebt. Sofern die Richtlinien des G-BA Vorgaben zur Zusammensetzung von Expertengruppen machen, behält sich das IQTIG vor, zur Sicherstellung einer unabhängigen medizinisch-wissenschaftlichen Beratung weitere Expertinnen und Experten hinzuzuziehen. Das Nähere zur Zusammensetzung der Expertengruppen auf Bundesebene wird richtlinienspezifisch und – soweit unterschiedliche Regelungen bestehen – verfahrensspezifisch im Verfahrenshandbuch des IQTIG dargelegt. Die Zusammensetzung der einzelnen Expertengruppen wird auf der Website des Instituts veröffentlicht.

### 5.2.3 Handlungsoptionen

Sofern im Rahmen der Verfahrenspflege Änderungsbedarf festgestellt wird, so werden die notwendigen Anpassungen im Einklang mit den Vorgaben des G-BA und unter Beachtung der Vorgaben der „Methodischen Grundlagen“ des IQTIG erarbeitet und in Abstimmung mit dem G-BA umgesetzt. Grundsätzlich können hiervon alle Bestandteile eines QS-Verfahrens betroffen sein. In der Regel führt die Verfahrenspflege zu Vorschlägen hinsichtlich folgender Verfahrensbestandteile:

- Qualitätsindikatoren (Beschreibungstexte, Rechenregeln, Risikoadjustierung, Referenzbereiche, Literaturangaben, Vergleichbarkeit zum Vorjahr, Streichen von Qualitätsindikatoren, Entwicklung neuer Qualitätsindikatoren),

- Anpassung der Instrumente zur Datenerfassung (QS-Filter, Datenfeldbezeichnungen, Ausgestaltung einzelner Datenfelder [z. B. Zahl und Art der Ausprägungen eines Merkmals], Ausfüllhinweise, Plausibilitätsregeln, Streichen von Datenfeldern, Einführung neuer Datenfelder), sowie
- Anpassung der Maßnahmen zur Datenvalidierung.

Möglich sind auch weitergehende Empfehlungen, insbesondere

- Empfehlung von Themen für Verfahrensweiterentwicklungen,
- Vorschläge für Sonderauswertungen, sowie ggf.
- Empfehlungen zu Lösungsmöglichkeiten für Konflikte mit anderen verpflichtenden oder freiwilligen Datenerhebungen.

### **Anpassung der Qualitätsindikatoren**

Die in der Qualitätsindikatordatenbank des IQTIG dokumentierten Qualitätsindikatoren und Kennzahlen werden jährlich überprüft und, wenn nötig, in Abstimmung mit dem G-BA angepasst. Maßgebliche Grundlagen für Anpassungen sind die Ergebnisse aus den zukünftigen Prüfungen zur Einhaltung der Eignungskriterien für Qualitätsindikatoren (siehe Kapitel 8), die Behebung von Fehlern, die im Rahmen der Auswertungen identifiziert wurden, und des Optimierungsbedarfs aus Biometrie-Sicht (insbesondere Eindeutigkeit, Verständlichkeit und Übersichtlichkeit der Rechenregeln), sowie ggf. festgestellter Anpassungsbedarf für die angewendeten Risikoadjustierungsmodelle.

Für alle im Rahmen des Regelbetriebs erhobenen Qualitätsindikatoren erfolgt zukünftig regelmäßig eine Aktualisierung der Literaturrecherche. Der Zeitpunkt für die nächste Überarbeitung wird nach jeder durchgeführten Literaturrecherche neu festgelegt. Sollte durch aktuelle Erkenntnisse eine frühzeitigere Aktualisierung notwendig sein, z. B. durch Änderungen der Regelungen der betrachteten Interventionen, Änderungen der Evidenz zu Nutzen und Schaden, Änderungen von Leitlinien(-empfehlungen), wird diese vorgezogen. Die Durchführung erfolgt nach der in Kapitel 6 beschriebenen Methodik, beschränkt auf den Zeitraum seit der letzten systematischen Literaturrecherche. Die am 1. Januar 2016 übernommenen QS-Verfahren werden schrittweise aktualisiert und im Bedarfsfall durch systematische Literaturrecherchen und -bewertungen ergänzt.

### **Streichen von Qualitätsindikatoren**

Sofern die Prüfung eines Qualitätsindikators hinsichtlich der Eignungskriterien nach Kapitel 8 ergibt, dass er auch unter Berücksichtigung möglicher Anpassungen nicht mehr für die Messung des jeweiligen Qualitätsmerkmals geeignet ist, wird der Indikator dem G-BA zur Streichung empfohlen. In einem solchen Fall wird durch das Institut geprüft, ob das von diesem Indikator abgebildete Qualitätsmerkmal bereits hinreichend durch einen anderen Qualitätsindikator abgedeckt wird oder ob die Neuentwicklung eines Qualitätsindikators nötig ist.

### Entwicklung neuer Qualitätsindikatoren

Als Ergebnis der Prüfung der QS-Verfahren im Regelbetrieb kann sich im Rahmen der Verfahrenspflege die Notwendigkeit ergeben, für bereits bestehende Qualitätsaspekte bzw. Qualitätsmerkmale neue Indikatoren zu entwickeln. Beispielsweise kann ein neu zu entwickelnder Qualitätsindikator Lücken im Qualitätsindikatorenset schließen, die durch Streichungen entstanden sind oder die neu identifiziert wurden, wenn z. B. ein bereits im QS-Verfahren enthaltener Qualitätsaspekt nicht ausreichend durch die bestehenden Indikatoren abgedeckt ist.

Der Entwicklungsprozess folgt in diesen Fällen grundsätzlich dem in Kapitel 4 beschriebenen Vorgehen zur Entwicklung von Qualitätsindikatoren. Die Umsetzung dieser Methodik wird an die Abläufe der Verfahrenspflege angepasst, hierzu befindet sich derzeit ein Konzept in der Entwicklung, dass nach Fertigstellung und Erprobung in die „Methodischen Grundlagen“ aufgenommen werden soll. Die Einbeziehung der medizinisch-fachlichen Expertise erfolgt durch die Einbindung der jeweils zuständigen Expertengruppe des IQTIG. Sollte im Rahmen der Neuentwicklung eines Indikators Fachexpertise erforderlich sein, die in der bestehenden Expertengruppe nicht oder nicht ausreichend vorhanden ist, so können auf den Entwicklungszweck begrenzt weitere Fachexpertinnen und -experten einbezogen werden.

### Anpassungen der Instrumente zur Datenerfassung

Die grundsätzlich derzeit in der externen gesetzlich verpflichtenden Qualitätssicherung verfügbaren Instrumente zur Datenerfassung sind:

- die QS-Dokumentation durch die Leistungserbringer
- die Risikostatistik für das QS-Verfahren *Pflege: Dekubitusprophylaxe* in den Krankenhäusern
- die Einrichtungsbefragung im QS-Verfahren *Vermeidung nosokomialer Infektionen – postoperative Wundinfektionen*
- die Nutzung der Sozialdaten bei den Krankenkassen

In absehbarer Zukunft wird als neues Erfassungsinstrument die Patientenbefragung hinzukommen. Perspektivisch bestehen darüber hinaus Möglichkeiten zur Nutzung der Daten aus der gesetzlichen Krebsregistrierung sowie des Transplantationsregisters.

Die Datenerfassung in allen derzeit operationalisierten Verfahren zur Datenerfassung beruht auf vom IQTIG erstellten Spezifikationen, die in der Regel durch Softwarehersteller in ein IT-Produkt umgesetzt und von den Leistungserbringern eingesetzt werden. Anpassungen dieser Spezifikationen für die Instrumente zur Datenerfassung können zwei Gründe haben. Zum einen können Anpassungen durch Änderungen der zugrunde gelegten Klassifikationen notwendig werden, z. B. durch die Änderung von Diagnose- und/oder Prozedurenkodens in den hierfür relevanten Klassifikationen des DIMDI. Zum anderen kann eine Anpassung der Datenfelder der Datenerfassungsinstrumente, ihrer Ausfüllhinweise oder ggf. der zugehörigen Plausibilitätsregeln sinnvoll sein, um die Messeigenschaften einzelner Qualitätsindikatoren zu optimieren. Anpassungen der Datenerfassungsinstrumente können auch Datenfelder einschließen, die nicht unmittelbar für die Berechnung eines Qualitätsindikators, sondern für andere Prozesse des QS-Verfahrens wie

z. B. die Datenvalidierung benötigt werden. Alle Spezifikationen zur Datenerfassung werden jährlich mit dem Ziel der Datensparsamkeit hinsichtlich Aufwand und Nutzen geprüft.

### **Maßnahmen zur Datenvalidierung**

Für alle im Regelbetrieb durchgeführten QS-Verfahren sind Maßnahmen zur Validierung der erhobenen Daten vorgesehen. Zu Art und Umfang der Datenvalidierung hat der G-BA detaillierte Vorgaben in seinen Richtlinien beschlossen. Die methodische Sicht des IQTIG ist in Kapitel 12 dargelegt. Darüber hinaus bestehen bereits jetzt sowie perspektivisch detaillierte Vorgaben in den Richtlinien des G-BA. Das konkrete Vorgehen zur Umsetzung der G-BA-Vorgaben unter Berücksichtigung der methodischen Grundlagen des IQTIG ist im Verfahrenshandbuch dargelegt.

### **Regelmäßige Überprüfung hinsichtlich möglicher Themen für die Verfahrensweiterentwicklung**

Für die Identifizierung von Themen für mögliche Weiterentwicklungsprojekte werden neben eigenen Erkenntnissen des Instituts aus der Verfahrensdurchführung und der Verfahrenspflege der QS-Verfahren insbesondere die Informationen aus Anfragen und Hinweisen der Landesebene und der Leistungserbringer herangezogen. Die gesammelten Themen werden jährlich mit den Expertengruppen beraten und an den G-BA berichtet.

## **5.3 Verfahrensweiterentwicklung**

In Abgrenzung zur Verfahrenspflege stellt die Verfahrensweiterentwicklung einen klar definierten Weiterentwicklungsauftrag dar, der qualitativ und/oder quantitativ über die Möglichkeiten der Verfahrenspflege hinausgeht. Gegenstand können sowohl grundlegendere inhaltliche Veränderungen (z. B. die Einbeziehung von Sozialdaten bei den Krankenkassen als neues Erfassungsinstrument) als auch weitergehende organisatorisch-strukturelle Veränderungen eines QS-Verfahrens im Regelbetrieb sein (z. B. Verknüpfung von zwei bestehenden QS-Verfahren). Verfahrensweiterentwicklungen beruhen immer auf einer Beauftragung des G-BA. Das Ziel der Verfahrensweiterentwicklung ist, dass die QS-Verfahren im Regelbetrieb inhaltlich dem Stand der medizinischen Wissenschaft und den Anforderungen der Gesellschaft an die Qualität der medizinischen Leistungserbringung entsprechen. Es ist damit identisch mit dem Ziel der Verfahrenspflege. Der Unterschied liegt in der Größenordnung der angestrebten Änderungen sowie in der Abarbeitung des Auftrags in einem zeitlich begrenzten Projekt.

Die Verfahrensweiterentwicklung umfasst die Erarbeitung von umsetzbaren Änderungsvorschlägen alle Prozesse, die notwendig sind, um den jeweiligen Auftrag umzusetzen. Das kann beispielsweise – im Unterschied zur Verfahrenspflege – die Hinzunahme weiterer Qualitätsaspekte in das QS-Verfahren (z. B. die Indikationsstellung) oder auch die Erschließung neuer Datenquellen bedeuten, um die Messeigenschaften von Qualitätsindikatoren zu bestehenden Qualitätsaspekten zu optimieren oder zu erweitern. Im Gegensatz zur Verfahrenspflege kann im Rahmen einer Verfahrensweiterentwicklung damit auch das Qualitätsmodell des zugrunde liegenden QS-Verfahrens in Teilen geändert werden. Im Unterschied zur Neuentwicklung eines QS-Verfahrens wird das Qualitätsmodell jedoch lediglich angepasst und nicht von Grund auf neu entworfen.

Verfahrensweiterentwicklungen können auch mit einem neuen oder erheblich veränderten Aufwand für die Verfahrensteilnehmer begründet werden (z. B. bei einer erheblichen Zunahme der zu dokumentierenden Daten). Bereits im Vorfeld der Beauftragung zur Verfahrensweiterentwicklung wird daher geklärt, ob die geplanten Veränderungen noch unter Beibehaltung des ursprünglichen Qualitätsmodells realisiert werden können. Die Verfahrensweiterentwicklung erfolgt stets als Projekt, das mit einem Abschlussbericht und/oder einem anderweitig definierten Weiterentwicklungsergebnis abgeschlossen wird.

Die Verfahrensweiterentwicklung soll gewährleisten, dass tiefgreifende Veränderungen an bestehenden QS-Verfahren, die den Umfang der Verfahrenspflege überschreiten, strukturiert umgesetzt werden.

### **5.3.1 Anlässe für Verfahrensweiterentwicklungen**

Weiterentwicklungen für bestehende QS-Verfahren werden in aller Regel vom IQTIG angestoßen. Grundlage hierfür können Anpassungsbedarfe aus der Verfahrenspflege oder andere institutsinterne Überlegungen sein. Prinzipiell besteht auch die Möglichkeit, dass Themen für Weiterentwicklungen von außen an das IQTIG herangetragen werden.

Zunächst wird geprüft, ob ein bestimmter Anpassungsbedarf im Rahmen der Verfahrenspflege umgesetzt werden kann. Kommt diese Prüfung zu dem Ergebnis, dass die angedachte Anpassung des bestehenden QS-Verfahrens eine Neuentwicklung des zugrunde liegenden Qualitätsmodells erfordern würde, so kann diese Anpassung nur im Rahmen einer Verfahrensneuentwicklung umgesetzt werden.

Sofern die Prüfung ergibt, dass eine Umsetzung im Rahmen der Verfahrenspflege nicht möglich ist, das Qualitätsmodell aber grundsätzlich beibehalten werden kann, wird der Vorschlag als Verfahrensweiterentwicklung eingestuft. In diesem Fall werden zunächst Empfehlungen zu Eckpunkten der angestrebten Umsetzung erarbeitet. Diese Empfehlungen legen insbesondere dar, welche Entwicklungsschritte im Rahmen der Umsetzung voraussichtlich notwendig sind und welche Konsequenzen die Umsetzung für die Richtlinien des G-BA haben könnten. Nach Abschluss dieser Arbeiten informiert das IQTIG den G-BA über den Sachstand, der dann zu entscheiden hat, ob er einen inhaltlich differenzierten Auftrag zur Neuentwicklung erteilen will.

### **5.3.2 Durchführung von Verfahrensweiterentwicklungen**

Das Vorgehen entspricht grundsätzlich den Vorgaben für Verfahrensneuentwicklungen (siehe Kapitel 4).

Die wissenschaftlich-fachliche Expertise wird über die Einbindung der jeweils zuständigen Expertengruppe auf Bundesebene in die Entwicklungsarbeit gewährleistet. Die damit verbundenen Aufgaben entsprechen dem Expertenpanel bei der Verfahrensentwicklung. Um eine möglichst praxisnahe Entwicklungsarbeit zu gewährleisten, ist darauf zu achten, dass die Mitglieder dieser Expertengruppe über die hierfür notwendigen Erfahrungen verfügen. Bei Bedarf können weitere Fachexpertinnen und -experten hinzugezogen werden.

Auf Grundlage von Art und Umfang der vorgesehenen Änderungen am jeweiligen QS-Verfahren wird stets geprüft, ob eine gesonderte Machbarkeitsprüfung (siehe Abschnitt 4.3) notwendig ist.

# Teil C: Methodische Elemente

## 6 Grundlagen der systematischen Literatur- und Informationsbeschaffung

Ziel der systematischen Informationsgewinnung aus wissenschaftlicher Literatur ist es, bestmögliche wissenschaftliche Erkenntnisse für Projekte der Neu- und Weiterentwicklung sowie der Durchführung von Qualitätssicherungsverfahren (QS-Verfahren) zu identifizieren. Als Beispiele seien hier die Literaturrecherchen im Rahmen der Konzeptstudie bzw. der Entwicklung des verfahrensspezifischen Qualitätsmodells, im Rahmen einer Fragebogenentwicklung sowie bei der Verfahrenspflege genannt.

In diesem Kapitel werden die Kernelemente der Informationsgewinnung aus wissenschaftlicher Literatur und ergänzenden Quellen dargestellt, die in der Arbeit des Instituts angewendet werden:

- die systematische Recherche
- die kritische Bewertung der relevanten Publikationen
- die zusammenfassende Qualitätsbewertung der Gesamtevidenz

Wichtige Prinzipien sind dabei die Transparenz und Nachvollziehbarkeit der Methodik und des Prozesses der Informationsgewinnung. Dies schließt insbesondere die vollständige und reproduzierbare Dokumentation des gesamten Rechercheprozesses ein (Droste und Lühmann 2014).

In Bezug auf die Zielsetzung und Fragestellungen der Literaturrecherchen werden in der Arbeit des Instituts methodisch zwei Formen der systematischen Evidenzaufbereitung unterschieden:

- Systematische Recherche und Aufbereitung der vorhandenen Evidenz orientiert an der Methodik eines Scoping Review:

Nach Colquhoun et al. (2014) ist ein Scoping Review eine Art der Darlegung des Wissenstands, bei der zu einer explorativen Forschungsfrage durch systematische Suche, Auswahl und Darstellung des vorhandenen Wissens die Schlüsselkonzepte, Arten der Evidenz und Forschungslücken bezogen auf einen definierten Bereich oder ein definiertes Feld aufgezeigt werden.

Diese Art der Evidenzaufbereitung hat zum Ziel, einen systematischen Überblick zu einem spezifischen Thema im Rahmen der Neu- oder Weiterentwicklung von QS-Verfahren zu erstellen. Zu einer (oft eher breiten) Fragestellung erfolgt eine systematische Literaturrecherche (entsprechend Abschnitt 6.1) und Darstellung der Erkenntnisse, bei der auch nicht vorab festgelegte, sondern sich schrittweise annähernde Prozessschritte vorhanden sein können, z. B. bei der Entwicklung oder Anpassung von Ein- und Ausschlusskriterien für die Literaturauswahl. Eine kritische Bewertung der Literatur erfolgt hier regelhaft nicht (Schmucker et al. 2013).

Beispiele hierfür sind epidemiologische Fragestellungen oder Fragen nach Versorgungsprozessen im Rahmen der Themerschließung von Entwicklungsprojekten.

- Systematische Recherche und Bewertung der identifizierten Evidenz und der Empfehlungen:  
Ziel ist die systematische Identifizierung, Bewertung und Zusammenfassung aller verfügbaren und bekannten Publikationen, die zur Beantwortung der Fragestellung von Bedeutung sind. Wissenschaftliche Literatur kann nach Primärliteratur (meist dem Studiendesign entsprechend, z. B. randomisierte kontrollierte Studien) und Sekundärliteratur (z. B. systematische Übersichtsarbeiten, Health Technology Assessments (HTA), Leitlinien) unterschieden werden. Welche Publikationen von Bedeutung sind, hängt von der zu beantwortenden Fragestellung ab (Droste und Lühmann 2014, Perleth et al. 2014) und wird vorab durch die Ein- und Ausschlusskriterien definiert.

Es erfolgt eine systematische Literaturrecherche in den ausgewählten Informationsquellen, mit a priori definierten Ein- und Ausschlusskriterien zur Auswahl der identifizierten Publikationen, und regelhaft eine kritische Bewertung der im Volltext eingeschlossenen Literatur.

Beispiele hierfür sind Literaturrecherchen im Rahmen der Entwicklung des verfahrensspezifischen Qualitätsmodells bei Neu- und Weiterentwicklungen von QS-Verfahren.

## 6.1 Systematische Recherche

Der Begriff „systematisch“ bezieht sich bei der systematischen Recherche auf die strukturierte Vorgehensweise nach einem a priori definierten Konzept auf der Suche nach den Erkenntnissen oder Informationen, die für die Beantwortung der Eingangsfrage relevant sind. Der Ablauf einer systematischen Recherche umfasst die Ableitung der Zielsetzungen und Fragestellungen der Literaturrecherche aus den Zielen und Fragestellungen des Projektes, die weitere Präzisierung der Fragestellung, die themenspezifische Auswahl der Informationsquellen, die Entwicklung der Suchstrategie inkl. Durchführung der Suche, die Auswahl relevanter Publikationen sowie die Dokumentation des gesamten Rechercheablaufs.

Vorab werden passend zur Art der geplanten Evidenzaufbereitung

- die inhaltlichen Ein- und Ausschlusskriterien (z. B. Population, Indikation, Intervention, Setting, Endpunkte) sowie
  - die methodisch-formalen Ein- und Ausschlusskriterien (z. B. Studien- bzw. Leitlinientyp, erforderliche Studienqualität)
- definiert (Liberati et al. 2009).

Hierfür kann ergänzend eine nicht systematische Vorabrecherche notwendig sein, z. B. zur Aufbereitung des Hintergrunds (Droste und Lühmann 2014, Perleth et al. 2014).

### 6.1.1 Präzisierung der Fragestellung

Zur Vorbereitung der Recherche ist es notwendig, eine meist ungenau bzw. allgemein formulierte Fragestellung für die Literaturrecherche in eine strukturierte, präzise Frage umzuwandeln. Hiermit wird die Voraussetzung dafür geschaffen, dass die für die Fragestellung relevanten Publikationen auch in den ausgewählten Informationsquellen identifiziert werden können. Zur Präzisierung wird die Fragestellung in verschiedene Komponenten aufgeschlüsselt, wie z. B. die Population von Interesse, eine definierte Intervention, die relevanten Endpunkte sowie das

Studiendesign (z. B. systematische Reviews, randomisierte kontrollierte Studien, Kohortenstudien), welches zur Beantwortung der Frage geeignet erscheint (Khan et al. 2004). Verschiedene etablierte Schemata werden in Abhängigkeit von der Zielsetzung der Literaturrecherche angewendet. Als Beispiele seien hier das PICO-Schema (Problem, Intervention, Comparison, Outcome) oder alternative Schemata wie z. B. ECLIPSE (Expectation, Client, Group, Location, Impact, Professionals, Service) oder PIPOH (Population, Interventions, Professionals/Patients, Outcome, Health care setting) genannt (Kloda und Bartlett 2013).

### **6.1.2 Informationsquellen der systematischen Recherche nach wissenschaftlicher Literatur und Leitlinien**

Die Auswahl der Informationsquellen orientiert sich an der Fragestellung und den hierfür vorab getroffenen Kriterien.

#### **Recherche nach Primärliteratur**

Die Recherche nach Primärliteratur erfolgt immer in den bibliografischen Datenbanken MEDLINE (via Ovid und/oder via PubMed) und Embase (via Elsevier). Abhängig von der Fragestellung werden Recherchen in weiteren Datenbanken wie z. B. CINAHL (via EBSCO) für pflegespezifische Fragestellungen und PsycINFO (via Ovid) für den Bereich der Psychologie und Psychiatrie durchgeführt. Sofern erforderlich und vorab definiert erfolgt ergänzend eine Suche nach unpublizierten Daten (z. B. Studienregister, Autorenanfragen) und weitere Handsuche, z. B. in den Referenzlisten ausgewählter Schlüsselartikel bzw. in ausgewählten Fachzeitschriften.

#### **Recherche nach systematischen Übersichtsarbeiten / Health Technology Assessments (HTA)**

Bei der Suche nach systematischen Übersichtsarbeiten und HTA-Berichten wird immer in der Cochrane Library recherchiert, hier insbesondere in den Datenbanken der „Cochrane Database of Systematic Reviews (CDSR)“, „Database of Abstracts of Reviews of Effects (DARE)“ sowie in der „Health Technology Assessment Database (HTAD)“. Gegebenenfalls erfolgt themenspezifisch ergänzend hierzu eine Suche in den Datenbanken über die Website des Centre for Reviews and Dissemination (CRD). Ebenso erfolgt eine Suche in den Literaturdatenbanken MEDLINE (via Ovid und/oder via PubMed) und Embase (via Elsevier) und abhängig von der Fragestellung ergänzend in weiteren Datenbanken wie z. B. CINAHL (via EBSCO) und PsycINFO (via Ovid). Ergänzend wird das International prospective register of systematic reviews (PROSPERO)<sup>9</sup> nach aktueller Forschung durchsucht.

Über die Deutsche Agentur für Health Technology Assessment (DAHTA, via DIMDI) und das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) wird ergänzend nach nationalen HTA-Berichten recherchiert. Über das Hinzuziehen weiterer Informationsquellen (z. B. Recherche auf Websites) wird themenspezifisch entschieden.

---

<sup>9</sup> <http://www.crd.york.ac.uk/PROSPERO/>

## Recherche von Leitlinien

Für die Leitliniensuche werden Leitliniendatenbanken wie das „Portal der wissenschaftlichen Medizin“ der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)<sup>10</sup>, die „International Guideline Library“ des Guideline International Network (G-I-N)<sup>11</sup> sowie das „National Guideline Clearinghouse“ (NGC)<sup>12</sup> der Agency for Healthcare Research and Quality (AHRQ) genutzt.

Zudem findet – abhängig von der Thematik – eine Suche in weiteren Datenbanken von nationalen und/oder internationalen Leitlinienerstellern, Websites der Fachgesellschaften, sowie in Literaturdatenbanken (wie MEDLINE oder Embase) statt.

### 6.1.3 Weitere Recherchequellen

In Abhängigkeit der Fragestellung ist es – neben der systematischen Suche nach Informationen aus wissenschaftlichen Studien oder Leitlinien – zur Entwicklung der Projekte notwendig, z. B. nach bereits vorhandenen Fragebögen, Qualitätsindikatoren und Versorgungsdaten zu recherchieren. Die Quellen hierfür werden jeweils themenspezifisch ausgewählt.

#### Recherche nach Fragebögen

Für die Identifikation von Fragebögen im Rahmen der Entwicklung von Patientenbefragungen werden Datenbanken wie z. B. PROQOLID (Patient-Reported Outcome and Quality of Life Instruments Database) und PROMIS (Patient-Reported Outcomes Measurement Information System) genutzt.

Ob weitere Informationsquellen (z. B. Recherche auf spezifischen Websites) für die Recherche hinzugezogen werden, wird themenspezifisch entschieden.

#### Recherche nach Qualitätsindikatoren

Für die Recherche nach national und international bereits vorhandenen Qualitätsindikatoren wird themenspezifisch auf den Websites von Institutionen, Organisationen und Fachgesellschaften des Gesundheitswesens recherchiert. Dazu gehören beispielsweise die Agency for Health Research and Quality (AHRQ), OECD Health Care Quality Indicators, Health and Social Care Information Centre des NHS und Quineth, der Qualitätsindikatorenthesaurus des GKV-Spitzenverbandes. Zudem findet – abhängig von der Thematik – eine ergänzende Suche in Literaturdatenbanken (wie z. B. MEDLINE und Embase) statt.

---

<sup>10</sup> <http://www.awmf.org/leitlinien/aktuelle-leitlinien.html>

<sup>11</sup> <http://www.g-i-n.net/>

<sup>12</sup> <http://www.guideline.gov/>

## Recherche nach Versorgungsdaten

Themenspezifisch, z. B. bei Fragen zum Versorgungsprozess, erfolgen Recherchen z. B. in Morbiditätsregistern, Arzneimittelregistern, Auswertungen von Routinedaten, Berichten des Statistischen Bundesamts, der Gesundheitsberichterstattung des Bundes etc. Diese werden jeweils projektbezogen vorab definiert.

### 6.1.4 Ableitung der Suchwörter, Suchstrategie und Dokumentation

Die Suchstrategie wird entsprechend den Bedingungen und Vorgaben der Informationsquelle entwickelt. Sie ist u. a. davon abhängig, ob eine Suchmaske die Eingabe von Suchwörtern erlaubt oder ob z. B. relevante Websites nach Veröffentlichungen bzw. gesamte Listen von Veröffentlichungen durchsucht werden müssen.

Bei der Literatursuche in bibliografischen Datenbanken wird die Suchstrategie zunächst für die Datenbank MEDLINE entwickelt und dann entsprechend den spezifischen Suchmöglichkeiten der außerdem genutzten Datenbanken Embase, CINAHL etc. angepasst. Ausgehend von der strukturierten Fragestellung (siehe Abschnitt 6.1.1) und ggf. vorab vorliegenden relevanten Publikationen werden die Suchbegriffe abgeleitet. Wenn möglich, wird anhand eines Testsets – aus bereits bekannten relevanten Publikationen – überprüft, ob diese Studien mit der entwickelten Suchstrategie gefunden werden. Anhand der „Checkliste für die Bewertung von Suchstrategien“ (Gechter et al. 2013)<sup>13</sup> wird die Suche nochmals überprüft und angepasst. Bei der Leitlinienrecherche ist die Formulierung der Suchstrategie abhängig von der jeweiligen Leitliniendatenbank bzw. den Websites der relevanten Fachgesellschaften. Teilweise ist es möglich, nach bestimmten Suchwörtern zu recherchieren, häufig ist jedoch auf den Websites der Fachgesellschaften ein Screening der gesamten Leitlinien durchzuführen, um die themenrelevanten zu identifizieren. Wird eine Leitlinienrecherche ergänzend in Literaturdatenbanken (wie MEDLINE oder Embase) durchgeführt, folgt diese dem oben beschriebenen Vorgehen. Die themenspezifische Recherche in weiteren Informationsquellen erfolgt entsprechend den Anforderungen der Informationsquelle.

Die Suchstrategie wird für jede Datenbank bzw. Informationsquelle schriftlich tabellarisch dokumentiert, inkl. der Suchoberfläche, des Suchzeitraums, der Suchbegriffe, der Trefferzahl sowie sprachlichen und anderen Einschränkungen bzw. Filtern.

### 6.1.5 Auswahlprozess

Zwei Personen sichten unabhängig voneinander Titel und Kurzfassung („Abstract“) bezogen auf die a priori festgelegten Ein- und Ausschlusskriterien; im Anschluss vergleichen sie ihre Ergebnisse. Uneinheitliche Bewertungen werden diskutiert und es wird für den Fall einer fehlenden Einigung eine dritte Person hinzugezogen.

Die Volltexte der als potenziell relevant eingeschlossenen Referenzen werden anschließend von zwei Personen unabhängig voneinander im Hinblick auf die a priori festgelegten Ein- und Ausschlusskriterien inhaltlich überprüft. Uneinheitliche Bewertungen werden diskutiert und für den

---

<sup>13</sup> Diese basiert auf McGowan et al. (2010).

Fall einer fehlenden Einigung wird eine dritte Person hinzugezogen. Die Ausschlussgründe aller im Volltext ausgeschlossenen Studien werden dokumentiert.

Bei der Leitlinienrecherche werden alle Titel der identifizierten Leitlinien dokumentiert und von zwei Personen in Bezug auf ihre inhaltliche Relevanz überprüft. Inhaltlich irrelevante Titel werden ausgeschlossen. Anschließend werden die Volltexte der ausgewählten Leitlinien von zwei Personen unabhängig voneinander dahingehend überprüft, ob die a priori festgelegten Einschlusskriterien zutreffen. Uneinheitliche Bewertungen werden diskutiert und für den Fall einer fehlenden Einigung wird eine dritte Person hinzugezogen. Die Ausschlussgründe aller im Volltext ausgeschlossenen Leitlinien werden dokumentiert.

Der gesamte Literatursuchprozess wird in einem Flussdiagramm dokumentiert (Liberati et al. 2009).

## **6.2 Kritische Bewertung relevanter Publikationen**

In Abhängigkeit von der Fragestellung und Art der Evidenzaufbereitung erfolgt eine kritische Bewertung der methodischen Qualität der identifizierten relevanten Publikationen. Systematische Fehler (Verzerrungen/Bias) in der Studienmethodik können zu inkorrekten Ergebnissen führen, woraus ggf. Über- oder Unterschätzungen der Effekte resultieren können. Hierbei spielt nicht nur das Studiendesign selbst eine Rolle, sondern auch die Studiendurchführung. Die kritische Bewertung der methodischen Qualität einer Publikation ist deshalb ein wichtiger Prozessschritt, um die Aussagekraft dieser Publikation zur Beantwortung der Fragestellung beurteilen bzw. einschätzen zu können.

### **Kritische Bewertung von Primärliteratur**

Im Wesentlichen lassen sich mögliche Verzerrungen folgendermaßen einteilen (nach Fletcher et al. 2007):

- Informationsfehler, die bei der Erhebung der Merkmale (z. B. Erkrankungsstatus, Exposition) entstehen können
- Selektionsfehler bei der Auswahl und beim Follow-up der Studienteilnehmerinnen und -teilnehmer
- Confounding, welches bei mangelnder Berücksichtigung von Störgrößen entstehen kann

Mithilfe standardisierter Bewertungsinstrumente bewerten mindestens zwei Personen unabhängig voneinander das Verzerrungspotenzial von Einzelstudien. Uneinheitliche Bewertungen werden diskutiert und für den Fall einer fehlenden Einigung wird eine dritte Person hinzugezogen.

### **Kritische Bewertung von systematischen Übersichtsarbeiten**

Methodische Mängel können auch in systematischen Übersichtsarbeiten zu Verzerrungen der Ergebnisse führen. Dies betrifft insbesondere die Studiensuche und -auswahl, die Qualitätsbewertung bzw. das Verzerrungsrisiko der eingeschlossenen Einzelstudien sowie die Zusammenfassung der Daten/Ergebnisse.

Mithilfe standardisierter Bewertungsinstrumente wie z. B. dem AMSTAR-Instrument (Assessment of Multiple Systematic Reviews) (Shea et al. 2007) oder dem ROBIS-Instrument (Risk Of Bias In Systematic reviews) (Whiting et al. 2016) bewerten mindestens zwei Personen unabhängig voneinander die methodische Qualität systematischer Übersichtsarbeiten. Uneinheitliche Bewertungen werden diskutiert und für den Fall einer fehlenden Einigung wird eine dritte Person hinzugezogen.

### **Kritische Bewertung von Leitlinien**

Um die methodische Genauigkeit und Transparenz der Leitlinienentwicklung zu beurteilen, verwendet das IQTIG standardmäßig die deutsche Übersetzung (Brouwers et al. 2014) des AGREE-II-Instruments (Appraisal of Guidelines for Research & Evaluation). Das Instrument besteht aus 23 Items, die sechs Domänen zugeordnet sind. Jede einzelne Domäne deckt dabei eine Qualitätsdeterminante der Leitlinien ab.

- Domäne 1: Geltungsbereich und Zweck
- Domäne 2: Beteiligung von Interessengruppen
- Domäne 3: Genauigkeit der Leitlinienentwicklung
- Domäne 4: Klarheit der Gestaltung
- Domäne 5: Anwendbarkeit
- Domäne 6: Redaktionelle Unabhängigkeit

Zusätzlich erfolgen noch zwei Gesamtbewertungen der Leitlinie.

Die Leitlinien werden anhand dieses Instruments von mindestens zwei Personen unabhängig voneinander bewertet. Aus der Bewertung jedes Items auf einer Punkteskala von eins bis sieben wird für jede Domäne ein separater Wert errechnet. Diese sechs Domänenwerte werden zwischen den Leitlinien verglichen; sie sind unabhängig voneinander und dürfen nicht zu einem Gesamtwert zusammengefasst werden.

## **6.3 Zusammenfassende Qualitätsbewertung der Gesamtevidenz**

Um die Qualität der Evidenz, die für eine Fragestellung identifiziert werden konnte, in ihrer Gesamtheit einschätzen zu können, ist es notwendig, sie zusammenzufassen. Derzeit existiert hierfür kein allgemein angewendeter internationaler Standard. Häufig wird in diesem Zusammenhang auf international verschiedene Klassifikationsschemata, sogenannte Evidenzhierarchien, zurückgegriffen (z. B. Oxford CEBM Levels of Evidence, SIGN, G-BA-Klassifizierung). Die Darstellung der Evidenzgrade entspricht hierbei einer Rangfolge, basierend auf dem Verzerrungspotenzial, welches den verschiedenen Studiendesigns zugesprochen wird. All diesen Klassifikationsschemata gemeinsam ist das Problem, dass sie auf der Ebene von Einzelstudien verbleiben und wichtige Aspekte wie klinische Relevanz und Konsistenz der Studienergebnisse sowie in der Regel auch die Studienqualität unberücksichtigt lassen. Evidenzhierarchien können daher auch nicht die kritische Bewertung von Studien ersetzen und sind lediglich als Orientierungshilfen bei der Darstellung der Studienlage anzusehen (Droste und Lühmann 2014). Soweit das IQTIG Evidenzgrade zu Zwecken der ergänzenden Information benennt, wird einheitlich das Schema des G-BA (G-BA 2016 2. Kapitel 3. Abschnitt § 11) verwendet.

Neuere Ansätze zur Darstellung der Qualität gehen von einer endpunktorientierten Betrachtung der Evidenz aus, wobei für jeden relevanten Endpunkt die Qualität der Gesamtheit der Evidenz („body of evidence“) bewertet wird (siehe hierzu „Grading of Recommendations, Assessment, Development and Evaluation“ [GRADE], Schönemann 2009). Bei dieser Form der Evidenzklassifizierung ist die Qualität der Evidenz der Gradmesser für das Vertrauen, dass ein ermittelter Effekt korrekt ist (Schönemann 2009).

### 6.3.1 Primärstudien

Das Institut orientiert sich bei der zusammenfassenden Evidenzbewertung von Primärstudien an dem methodischen Vorgehen der internationalen GRADE-Gruppe und des National Quality Forum (NQF). Für die Bewertung der Evidenz von Qualitätsindikatoren zur Messung von Prozessen, Strukturen und Zwischenendpunkten wendet das NQF ein Schema an, welches die Quantität, Qualität und Konsistenz der Gesamtevidenz bewertet und darstellt (NQF 2015).

Angelehnt an die Schemata von GRADE und vom NQF wird die Gesamtheit der Evidenz („body of evidence“) unter Einschätzung der Kriterien Quantität, Qualität und Konsistenz bewertet. Die Einstufungen werden in den folgenden Tabellen dargestellt.

#### Kriterium Quantität

Definition: Gesamtzahl der Studien (nicht Artikel oder Veröffentlichungen)

Tabelle 3: Schema zur Evidenzbewertung in Bezug auf das Kriterium Quantität

Bewertung	Quantität
hoch	≥ 5 Studien
mäßig	2–4 Studien
niedrig	1 Studie
unzureichend	keine empirische Studien

#### Kriterium Qualität

Definition: Sicherheit oder Vertrauen in die Effektschätzungen von Nutzen und Schaden für die Patientinnen und Patienten über alle Studien hinweg, bezogen auf Studienfaktoren wie Studiendesign und methodische Qualität der Studien.

Tabelle 4: Schema zur Evidenzbewertung in Bezug auf das Kriterium Qualität

Bewertung	Qualität
hoch	randomisierte kontrollierte Studien (RCT), die einen direkten Nachweis für die spezifische Maßnahme liefern, mit ausreichender Größe, um präzise Schätzungen der Wirkung zu erzielen, und ohne gravierende Mängel, die zu einer Verzerrung führen könnten

Bewertung	Qualität
mäßig	<ul style="list-style-type: none"> <li>▪ Nicht-RCT mit Kontrolle für Störfaktoren und großen, präzisen Schätzungen des Effekts</li> </ul> <p>oder</p> <ul style="list-style-type: none"> <li>▪ RCT ohne gravierende Mängel, die zu Verzerrungen führen könnten, aber mit entweder indirekten Nachweisen oder ungenauen Schätzungen des Effekts</li> </ul>
niedrig	<ul style="list-style-type: none"> <li>▪ RCT mit Mängeln, die zu Verzerrungen führen</li> </ul> <p>oder</p> <ul style="list-style-type: none"> <li>▪ Nicht-RCT mit kleinen oder unpräzisen Schätzungen des Effekts oder ohne Kontrolle von Störfaktoren</li> </ul>
unzureichend	keine empirische Studien

### Kriterium Konsistenz

Definition: Stabilität sowohl in Richtung und Größe des klinisch/praktisch bedeutsamen Nutzens und Schadens für die Patientinnen und Patienten über alle Studien hinweg in der Gesamtheit der Evidenz.

Tabelle 5: Schema zur Evidenzbewertung in Bezug auf das Kriterium Konsistenz

Bewertung	Konsistenz
hoch	die Schätzungen des klinisch/praktisch bedeutsamen Nutzens und Schadens für die Patientinnen und Patienten sind konsistent in der Richtung und ähnlich in der Größenordnung bei der Mehrzahl der Studien in der Gesamtheit der Evidenz
mäßig	<ul style="list-style-type: none"> <li>▪ die Schätzungen des klinisch/praktisch bedeutsamen Nutzens und Schadens für die Patientinnen und Patienten sind konsistent in der Richtung bei der Mehrzahl der Studien in der Gesamtheit der Evidenz, unterscheiden sich jedoch in der Größenordnung</li> <li>▪ wenn nur eine Studie vorliegt, dann sollte die Schätzung des Nutzens den potenziellen Schaden für den Patientinnen und Patienten überwiegen (nur eine vorliegende Studie kann nicht die Bewertung „hoch“ für Konsistenz erhalten)</li> </ul>
niedrig	<ul style="list-style-type: none"> <li>▪ die Schätzungen des klinisch/praktisch bedeutsamen Nutzens und Schadens für die Patientinnen und Patienten unterscheiden sich sowohl in der Richtung und in der Größenordnung bei der Mehrzahl der Studien in der Gesamtheit der Evidenz</li> </ul> <p>oder</p> <ul style="list-style-type: none"> <li>▪ breite Konfidenzintervalle erschweren die Schätzung des Nutzens</li> <li>▪ wenn nur eine Studie vorliegt, dann überwiegt die Schätzung des Nutzens nicht den potenziellen Schaden für die Patientinnen und Patienten</li> </ul>

Bei ausreichender Studienanzahl wird themenspezifisch entschieden, ob die Studien, wenn sinnvoll und möglich (z. B. nach Prüfung der Heterogenität), für einzelne Endpunkte metaanalytisch zusammengefasst werden.

### **6.3.2 Systematische Übersichtsarbeiten**

Bei einer systematischen Übersichtsarbeit werden zu einer strukturierten, präzisen Frage alle verfügbaren Primärstudien nach einer vorab festgelegten Methodik systematisch identifiziert, ausgewählt, kritisch bewertet und zusammengefasst. Abhängig von der Fragestellung stellen sie somit eine wichtige Evidenzbasis im Rahmen der Verfahrensentwicklung bzw. -pflege dar. Die methodische Qualität (z. B. anhand des AMSTAR- oder ROBIS-Instruments) und die Konsistenz der Ergebnisse aller zu einer Fragestellung eingeschlossenen systematischen Übersichtsarbeiten werden eingeschätzt und tabellarisch dargestellt. Über die Mindestzahl relevanter systematischer Übersichtsarbeiten wird jeweils themenspezifisch entschieden.

### **6.3.3 Leitlinien**

Mithilfe des AGREE-II-Instruments lassen sich alle zu einer Fragestellung eingeschlossenen Leitlinien hinsichtlich ihrer methodischen Qualität innerhalb ihrer sechs Domänenwerte miteinander vergleichen und tabellarisch darstellen. Die Konsistenz der Ergebnisse wird anhand der Dokumentation der Leitlinieninhalte und Empfehlungen überprüft und eingeschätzt.

### **6.3.4 Qualitative Studien**

Je nach Thema können Forschungsergebnisse aus qualitativen Studien wesentlicher Teil der Evidenzbewertung bzw. -aufbereitung sein. Insbesondere zur Ermittlung und Abbildung der Patientenperspektive hinsichtlich Erfahrungen, Einstellungen, Haltungen und Akzeptanz werden im Rahmen der Neu- und Weiterentwicklung sowie Pflege von QS-Verfahren qualitative Studien einbezogen. Ausgehend von der Form der Evidenzaufbereitung werden eingeschlossene qualitative Studien von mindestens zwei Personen bzgl. ihrer methodischen Qualität mithilfe eines Bewertungsinstruments (z. B. der Checkliste des Critical Appraisal Skills Programme / CASP) bewertet. Uneinheitliche Bewertungen werden diskutiert und für den Fall einer fehlenden Einigung wird eine dritte Person hinzugezogen.

## 7 Einbeziehung externen Wissens und externer Erfahrung

### 7.1 Ziele der Einbindung externen Wissens

Die Breite des Spektrums an Themen sowie die fachliche und organisatorische Komplexität der Gesundheitsversorgung erfordern die Einbindung externen Wissens und externer Erfahrung bei der Entwicklung und Durchführung von QS-Verfahren. Dies ist umso mehr von Bedeutung, als die Qualitätssicherung weitreichende Einflüsse auf die Versorgung von Patientinnen und Patienten haben soll.

Ein einzelnes Institut kann trotz fachlicher und methodischer Expertise seiner Mitarbeiterinnen und Mitarbeiter nicht über die Gesamtheit medizinischen Wissens, klinischer Erfahrung, Erfahrung aus Patientenperspektive und weitere Spezialkenntnisse verfügen. Zudem kann dieses Wissen aus folgenden Gründen nur teilweise durch Recherchen generiert werden:

- Nicht zu allen für die Qualitätssicherung relevanten Themenbereichen und Fragestellungen liegt ausreichend Evidenz oder liegen überhaupt Untersuchungen vor.
- Qualität und Konsistenz der Evidenz aus der wissenschaftlichen Literatur sind für viele Fragestellungen begrenzt.
- Qualitätssicherung zielt auf die Versorgungspraxis, sodass die Übertragbarkeit von Studienergebnissen eingeschränkt sein kann.

Als Ergänzung zur institutseigenen Expertise sind daher bei der Entwicklung von QS-Verfahren Perspektiven und Erfahrungen z. B. von klinischen Expertinnen und Experten ein wichtiges Element (Donabedian 2003). Das Gleiche gilt, analog zum Vorgehen der evidenzbasierten Medizin (Sackett et al. 1997), für die Berücksichtigung von Patientenpräferenzen.

Durch die Einbindung von Personen, die nicht dem IQTIG angehören und über ergänzende Erfahrung und Kompetenz verfügen, werden demnach folgende Ziele verfolgt:

- Richtigkeit, Aktualität und Vollständigkeit der Inhalte
- Einbezug von nicht publizierten Inhalten und von Erfahrungswissen, insbesondere Praxiswissen und Patientenpräferenzen
- Vermeidung einseitiger Betrachtungsweisen durch breitgefächerte Berücksichtigung von Wissen und Perspektiven verschiedener Personen und/oder Institutionen
- Akzeptanz der Arbeitsergebnisse durch Einbeziehung der relevanten Beteiligten in die Verfahrensentwicklung

### 7.2 Beratungszwecke bei der Einbindung externen Wissens

Bei der Einbindung von Personen mit ergänzendem Wissen und Erfahrung in Anlehnung an Niederberger und Wassermann (2015: 37 ff.) werden folgende Kategorien von Beratungszwecken unterschieden:

**Ermittlung oder Prüfung von Fakten**

In der Entwicklung und Umsetzung von QS-Verfahren entstehen häufig Fragen fachlicher oder methodischer Art. Sofern in diesen Fällen lediglich eine bestimmte, unstrittige Sachinformation ermittelt oder auf Korrektheit geprüft werden soll, werden einzelne Personen mit entsprechender Kompetenz zur Klärung des Sachverhalts hinzugezogen, z. B. ein Experte für ein bestimmtes Operationsverfahren. Die Aufgabe der hinzugezogenen Personen besteht in der Bereitstellung von Wissen (Fakten).

**Ermittlung unterschiedlicher Perspektiven**

Ein weiteres Ziel der Einbindung externen Wissens kann die möglichst umfassende Ermittlung und Abbildung unterschiedlicher Ideen und Perspektiven zu einem Thema sein. In diesem Fall werden das Wissen und die Erfahrungen mehrerer externer Personen eingebunden. Dabei kann es sich entweder um mehrere Einzelpersonen (z. B. in Form von Einzelinterviews) handeln oder, wenn ein Dialog und Austausch zwischen den Personen stattfinden soll, um eine formale Gruppe (z. B. ein Expertengremium). In beiden Fällen werden zusätzlich zu personenbezogenen Kriterien für die Einbindung weitere Auswahlkriterien angewendet, die eine für die Fragestellung sachgerechte Zusammensetzung der Gruppe sicherstellen sollen.

**Bewertung diskussionsbedürftiger Sachverhalte**

Auch die Zusammenführung diskussionsbedürftiger Aspekte und unterschiedlicher Ansichten zu einem gemeinsamen Ergebnis sowie eine, ggf. quantitative, Bewertung bestimmter Aspekte können Zweck der Einbindung externen Wissens und externer Erfahrung sein. In diesem Fall wird grundsätzlich nicht auf einzelne Personen zurückgegriffen, sondern eine Gruppe externer Personen mit entsprechendem Wissen oder entsprechender Erfahrung eingebunden. Für solche Gruppen werden besondere Anforderungen hinsichtlich der Ausgewogenheit der Gruppe gestellt.

In Tabelle 6 sind Beispiele für die verschiedenen Arten von Beratungszwecken aufgeführt. Aus diesen drei Kategorien werden die Form der Einbindung externen Wissens und die Auswahlkriterien für die Personen mit entsprechender Kompetenz abgeleitet.

*Tabelle 6: Beratungszwecke der Einbindung externen Wissens mit Beispielen für die Einsatzbereiche und für die verwendeten Methoden*

	<b>Ermittlung und Prüfung von Fakten</b>	<b>Ermittlung unterschiedlicher Perspektiven</b>	<b>Bewertung diskussionsbedürftiger Sachverhalte</b>
<b>Form der Einbindung</b>	einzelne Personen	mehrere Personen	mehrere Personen als Gruppe
<b>Beispiele Einsatzbereich</b>	<ul style="list-style-type: none"> <li>▪ Erläuterung eines Therapieverfahrens</li> <li>▪ Klärung eines Datenfelds</li> </ul>	<ul style="list-style-type: none"> <li>▪ Themenerschließung</li> <li>▪ Ermittlung von Fragebogendimensionen</li> </ul>	<ul style="list-style-type: none"> <li>▪ Beurteilung von Qualitätsindikatoren</li> <li>▪ Ermittlung des Handlungsbedarfs</li> </ul>

	<b>Ermittlung und Prüfung von Fakten</b>	<b>Ermittlung unterschiedlicher Perspektiven</b>	<b>Bewertung diskussionsbedürftiger Sachverhalte</b>
		<ul style="list-style-type: none"> <li>▪ Interpretation von Indikatorergebnissen</li> </ul>	
<b>Beispiele Methode der Einbindung</b>	<ul style="list-style-type: none"> <li>▪ Einzelbefragung (Hausärztin)</li> <li>▪ Einzelbefragung (Medizin-Controller)</li> </ul>	<ul style="list-style-type: none"> <li>▪ Scoping Workshop</li> <li>▪ Fokusgruppe</li> <li>▪ Bundesfachgruppe</li> <li>▪ qualitatives Einzelinterview</li> </ul>	<ul style="list-style-type: none"> <li>▪ Entwicklungspanel</li> <li>▪ Bundesfachgruppe</li> </ul>

### 7.3 Methoden der Einbindung externen Wissens und externer Kompetenzen

Die Art und Weise, wie die externen Personen für die spezifischen Aufgaben eingebunden werden, leitet sich aus den Beratungszwecken ab. Sofern aus Gründen der praktischen Umsetzung dieselbe Gruppe von Personen für Beratungen in verschiedenen der o. g. Kategorien tätig wird (z. B. sowohl Beratung zu medizinischen Inhalten als auch Beurteilung von Indikatoreigenschaften), müssen für die Auswahl der Personen und die Zusammensetzung der Personengruppe alle maßgeblichen Auswahlkriterien erfüllt sein.

In Abhängigkeit von der Fragestellung haben die beratenden Personen unterschiedliche Rollen. Dementsprechend erfolgt die Hinzuziehung dieser Personen mittels verschiedener Methoden, die in den folgenden Abschnitten erläutert sind:

- Beziehen sich die zu ermittelnden Sachverhalte in erster Linie auf die persönliche Erfahrung und das individuelle Wissen der Personen, so werden leitfadengestützte Interviews und Fokusgruppen genutzt (siehe Abschnitte 7.3.1 und 7.3.2).
- Sollen die beratenden Personen dagegen mit ihrem besonderen Wissen stellvertretend für Personengruppen, Institutionen oder Organisationen Auskunft geben, so nehmen sie in erster Linie eine Expertenrolle ein (vgl. Meuser und Nagel 1991: 444). Das Vorgehen für die Einbindung von Expertengruppen ist in Abschnitt 7.3.3 beschrieben.

#### 7.3.1 Leitfadengestützte Einzelinterviews

Das leitfadengestützte Interview ist eine qualitative Forschungsmethode, die regelhaft im Rahmen der Entwicklungsarbeiten eingesetzt wird. Grundsätzliches Ziel von leitfadengestützten Einzelinterviews ist der Erkenntnisgewinn durch die Befragung bestimmter Vertreterinnen und Vertreter einer Zielgruppe, die den Untersuchungsgegenstand repräsentieren (Flick 2016). So dienen die Interviews zur Überprüfung und Ergänzung vorab getroffener Annahmen genauso wie zur Ableitung neuer Erkenntnisse. Auch wird das leitfadengestützte Interview als ein qualitativer Methodenbaustein einzeln oder in Kombination mit anderen qualitativen oder quantitativen Methoden verwendet.

Kern des Interviews ist der Interviewleitfaden, der die zentralen Themen anhand konkreter Fragen behandelt und das Interview strukturiert. Die Offenheit der Frageformulierung richtet sich nach dem Erkenntnisinteresse, d. h. danach, ob Wissen eher überprüft (geschlossene Fragen) oder Themen und Sichtweisen eruiert werden sollen (offene Fragen).

Leitfadengestützte Interviews werden hauptsächlich zu folgenden Zwecken eingesetzt:

- Zur Entwicklung von Qualitätsmodellen, d. h. bei der Identifikation der Qualitätsaspekte (siehe Abschnitt 3.1): Aus den in der Interviewsituation direkt berichteten Sichtweisen, Erfahrungen und Erkenntnissen werden qualitätsrelevante Themen abgeleitet, die in Ergänzung zu den anderen Wissensbeständen als Qualitätsaspekte in das Qualitätsmodell eingehen.
- Zur Ableitung von Qualitätsmerkmalen im Rahmen von Fragebogenentwicklungen für Patientenbefragungen: Anhand der Qualitätsmerkmale werden Fragebogenitems entwickelt, die die Grundlage der Qualitätsindikatoren darstellen (siehe Abschnitt 4.4).

Die leitfadengestützten Interviews werden für diese Zwecke meist in Kombination mit Fokusgruppen eingesetzt (siehe Abschnitt 7.3.2). Im Sinne des explorativen Charakters der Anwendungsfelder geben die Fragen der Interviewleitfäden in diesen Fällen Themenbereiche vor, die von den interviewten Personen offen beantwortet werden.

Leitfadengestützte Interviews werden protokolliert bzw. aufgezeichnet und anschließend transkribiert und anonymisiert.<sup>14</sup> Zur Unterstützung der qualitativen Analyse wird eine Software eingesetzt, anhand derer die Kernaussagen abgeleitet und bei mehreren Interviews zusammengeführt werden.

### 7.3.2 Fokusgruppen

Fokusgruppen gelten als eine Methode der qualitativen Sozialforschung, bei der eine Gruppe moderiert und fokussiert zu bestimmten, vorgegebenen Themen diskutiert (Kühn und Koschel 2011, Tausch und Menold 2015). Allen Fokusgruppen ist üblicherweise gemein, dass eine begrenzte Anzahl von Personen, die ein bestimmtes Thema oder Charakteristikum verbindet, zielgerichtet zu bestimmten, vorgegebenen Themen diskutiert (Krueger und Casey 2015). Durch die Gruppendiskussion und die damit verbundene diskursive Erschließung eines Themas werden im Unterschied zu Einzelinterviews Gedanken, Erlebnisse, Einstellungen, Wahrnehmungen, Bedürfnisse und Orientierungen der Gruppe in einer größeren Breite transparent. Die offene und wenig steuernde Moderation der Gruppe ermöglicht einen Einblick in die Realität der Fokusgruppenmitglieder, durch die Berichte der Betroffenen werden die ansonsten vornehmlich theoriebasierten Grundannahmen über das Verhalten oder über die Einstellungen bestimmter Gruppen oder einer bestimmten Population selbst auf ein breiteres Fundament gestellt. Folglich helfen Fokusgruppen, die Forschungsfrage bzw. den Untersuchungsgegenstand besser zu verstehen. Somit können sie als einzelne qualitative Methode (*stand alone*) genauso wie in Kombination mit anderen qualitativen oder quantitativen Methoden (Kühn und Koschel 2011) angewandt werden.

---

<sup>14</sup> Die Teilnehmerinnen und Teilnehmer werden vorab darüber informiert und geben ihr Einverständnis. Die Aufzeichnungen werden nach der Transkription gelöscht.

Fokusgruppen werden zum einen bei der Erstellung eines Qualitätsmodells (siehe Abschnitt 3.2) zur Erschließung der Qualitätsaspekte allein oder in Kombination mit leitfadengestützten Interviews durchgeführt (siehe Abschnitt 4.1). Zum anderen bieten Fokusgruppen bei der Entwicklung von quantitativen Patientenbefragungen eine Grundlage für die Itemgenerierung. Hierbei werden (patienten-)relevante Begriffe und Themen identifiziert, aus denen Qualitätsmerkmale abgeleitet werden, die später als Items in die Befragung eingehen und anhand derer die Qualitätsindikatoren definiert werden (siehe Abschnitt 4.4).

Um ein ausgewogenes Diskussionsbild zu erhalten, bestehen die Fokusgruppen jeweils aus 5 bis 10 für das betreffende QS-Verfahren charakteristischen Patientinnen oder Patienten bzw. Vertreterinnen und Vertretern einer medizinischen Berufsgruppe (Dreher und Dreher 1982, Krueger und Casey 2015). Zudem müssen die Personen physisch und psychisch in der Lage sein, an einer Gruppendiskussion teilzunehmen.

Um dem explorativen Charakter des Einsatzbereichs gerecht zu werden, erfolgt die Moderation der Fokusgruppen anhand eines Topic Guides, d. h., es werden lediglich bestimmte Themen vorgegeben, anhand derer die Gruppe diskutiert, der Topic Guide gibt somit den Rahmen der Diskussion vor. Auch wenn Fokusgruppen z. B. im Rahmen der Fragebogenentwicklung für Patientenbefragungen eingesetzt werden, werden von der Moderatorin oder dem Moderator lediglich bestimmte Themen vorgegeben, die sich an den zuvor entwickelten Qualitätsaspekten orientieren und die in der Gruppe vertiefend frei diskutiert werden.

Alle Gruppendiskussionen werden von einem Co-Moderator bzw. einer Co-Moderatorin protokolliert. Am Ende einer jeden Gruppendiskussion füllen alle Teilnehmerinnen und Teilnehmer auf freiwilliger Basis einen Kurzfragebogen aus, in dem soziodemografische Merkmale, wie Alter, Geschlecht und Angaben zur Gesundheit bzw. bei den Fokusgruppen mit medizinischem Fachpersonal Angaben zur Funktion u. Ä., erfragt werden. Anhand dieser Angaben können im Rahmen der Auswertungen Aussagen zur Gruppenzusammensetzung getroffen werden.

Die Fokusgruppendiskussionen werden aufgezeichnet und im Anschluss transkribiert und anonymisiert.<sup>15</sup> Für die qualitative Analyse wird eine Software eingesetzt. Mithilfe der schriftlichen Aufzeichnungen werden die Transkripte mit Blick auf die Hauptfragestellung der Fokusgruppen gruppenübergreifend ausgewertet. Dabei werden zentrale Diskussionsthemen erfasst, anhand derer sich die zentralen Aussagen mit Ankerbeispielen ableiten lassen.

### **7.3.3 Einbindung externer Expertengruppen**

#### **Begriffsdefinition „Expertin/Experte“**

Für den Begriff „Expertin/Experte“ gibt es keine allgemein konsentiertere, von der Fragestellung unabhängige Definition und in der Konsequenz keine allgemeingültige Operationalisierung (siehe z. B. Baker et al. 2006, Mullen 2003, Niederberger und Wassermann 2015). Der Status als

---

<sup>15</sup> Es werden in jedem Fall Audioaufzeichnungen angefertigt. Bei den meisten Fokusgruppen erfolgt eine zusätzliche Videoaufzeichnung. Auf diese wird im Falle vulnerabler Patientengruppen ggf. verzichtet. Die Teilnehmerinnen und Teilnehmer werden vorab darüber informiert und geben ihr Einverständnis. Die Aufzeichnungen werden nach der Transkription gelöscht.

Expertin oder Experte hängt unter anderem davon ab, ob die betreffende Person bezogen auf die jeweilige Fragestellung über besonderes Wissen oder Kompetenzen verfügt (Meuser und Nagel 1991: 443 f.). Hier werden als „externe Expertin/externer Experte“ Personen bezeichnet, die nicht dem IQTIG angehören, die für die Beratung oder Bewertung eines oder mehrerer Aspekte hinzugezogen werden und die für mindestens einen dieser Aspekte ein ausgewiesenes Wissen oder Erfahrung aufweisen.

Bei Expertinnen und Experten kann es sich im Kontext der Qualitätssicherung demnach nicht nur um Angehörige der Gesundheitsberufe handeln, sondern auch um andere Berufsgruppen (z. B. Medizincontrollerinnen und -controller, Kodierexpertinnen und -experten, Datenschutzbeauftragte, Juristinnen und Juristen, Sozialarbeiterinnen und -arbeiter, Informatikerinnen und Informatiker, Qualitätsmanagerinnen und -manager) und insbesondere auch um Patientinnen und Patienten sowie Patientenvertreterinnen und -vertreter.

### **Aufgaben von externen Expertinnen und Experten**

Für die Arbeit des Instituts werden externe Expertinnen und Experten für verschiedene Aufgaben benötigt. Bei der Durchführung, Pflege und Weiterentwicklung bestehender QS-Verfahren sind dies insbesondere:

- Beratung zu fachlichen Fragen, die im Rahmen der Verfahrensdurchführung und -pflege entstehen
- Bewertung von Einrichtungsergebnissen in QS-Verfahren, die direkt vom IQTIG durchgeführt werden
- Interpretation der Indikatorergebnisse auf Bundesebene
- Ermittlung von Handlungsbedarf auf Bundesebene
- Beratung zu möglichem Anpassungs- und Entwicklungsbedarf der Qualitätsindikatoren
- Präsentation der Verfahrensergebnisse in Fachkreisen und Öffentlichkeit

Bei der Entwicklung neuer QS-Verfahren werden Expertinnen und Experten für folgende Aufgaben konsultiert:

- Beratung bezüglich der Versorgungspraxis
- Beratung bei der Entwicklung der Qualitätsmodelle
- Beratung bei der Entwicklung der Instrumente zur Messung bzw. Datenerfassung
- Einschätzung der Eigenschaften von Qualitätsindikatoren
- Gesamtbeurteilung neuer QS-Verfahren

Bei Aufgaben des Instituts, die unabhängig von konkreten Verfahren der einrichtungsübergreifenden Qualitätssicherung sind, werden Expertinnen und Experten aus folgenden Gründen hinzugezogen:

- Beratung bei der Publikation von sekundären Auswertungen
- Beratung zu methodischen Fragestellungen

Details zur Auswahl und zu den Aufgaben bei den verschiedenen Einsatzgebieten finden sich in den jeweiligen Abschnitten der „Methodischen Grundlagen“ und des Verfahrenshandbuchs. Im vorliegenden Kapitel wird auf die übergreifenden Aspekte und deren methodische Begründung eingegangen.

Allgemein gilt, dass alle Beratungsergebnisse der Expertinnen und Experten empfehlenden Charakter haben. Das IQTIG ist als fachlich unabhängiges, wissenschaftliches Institut für die ordnungsgemäße Wahrnehmung seiner Aufgaben letztverantwortlich und kann in seinen Berichten und Empfehlungen in begründeten Fällen von den Empfehlungen der beratenden Expertinnen und Experten abweichen.

### **Kriterien für die Tätigkeit als Expertin oder Experte**

Es wird bei der Hinzuziehung von externen Expertinnen und Experten und der Bildung beratender Expertengruppen den folgenden Festlegungen gefolgt. Sofern die Richtlinien des G-BA hinsichtlich der Benennung und Zusammensetzung von Expertengruppen in bestimmten Verfahren Vorgaben machen, werden diese entsprechend berücksichtigt. Die Kriterien für die Auswahl von Expertinnen und Experten hängen von dem jeweiligen Einsatzbereich, d. h. von Zweck und Fragestellung der Einbindung ab (siehe Abschnitt 7.2). Bei der Einbindung von Expertinnen und Experten und bei der Auswahl von Mitgliedern für Expertengruppen werden zum einen Kriterien berücksichtigt, die von den Expertinnen und Experten als Person zu erfüllen sind, unabhängig von der Form der Einbindung (als Einzelexpertin/Einzelexperte oder als Mitglied einer Expertengruppe). Zum anderen werden Kriterien für die Zusammensetzung von Expertengruppen beachtet. Das Ziel dabei ist, dass die beratenden Expertinnen und Experten eine möglichst hohe fachliche Kompetenz in die jeweiligen Aufgaben einbringen und die Ergebnisse nicht durch Einzelmeinungen verzerrt werden. Da die Expertise der Personen und deren Gewichtung untereinander quantitativ nicht angemessen zu beschreiben ist, erfolgt die Auswahl von Expertinnen und Experten in Form einer Abwägung der Kriterien durch das Institut. Die Erfüllung der Kriterien durch die ausgewählten Expertinnen und Experten wird durch das Institut dokumentiert.

### **Einschlusskriterien**

Alle Expertinnen und Experten sollen über Fachwissen oder Erfahrung im vorgesehenen Einsatzbereich verfügen. Welche Erfahrung für die Tätigkeit als Expertin/Experte Voraussetzung ist, wird themenspezifisch zu Beginn des Bewerbungs- und Berufungsverfahrens für die Expertinnen und Experten festgelegt. Für medizinische Fachexpertinnen und Fachexperten kann dies z. B. bedeuten, dass formale Qualifikationen, Erfahrung mit einem bestimmten Diagnose- oder Therapieverfahren oder wissenschaftliche Expertise gefordert sind. Für Patientenvertreterinnen und Patientenvertreter oder Expertinnen und Experten aus den Institutionen des Gesundheitswesens können beispielsweise Grundkenntnisse der Qualitätssicherung erforderlich sein.

Angesichts des sich schnell entwickelnden medizinischen Wissens ist das Institut darauf angewiesen, sich von Expertinnen und Experten beraten zu lassen, deren Wissen und Erfahrungen möglichst aktuell sind. Für Expertinnen und Experten, die aufgrund ihrer beruflichen Kenntnisse in die Aufgaben des Instituts eingebunden werden, kann daher eine Anforderung sein, dass sie in einem bestimmten Zeitraum vor Beginn des Berufungsverfahrens noch in ihrem Beruf tätig waren.

### **Ausschlusskriterium**

Als allgemeines Ausschlusskriterium gilt das Vorliegen von relevanten Interessenkonflikten. Denn bereits gesetzlich ist das Institut angehalten, Interessenkonflikte von Beschäftigten und

von allen anderen an der Aufgabenerfüllung des Instituts beteiligten Personen und Institutionen zu vermeiden (§ 137a Abs. 9 SGB V). Die fachliche Unabhängigkeit ist darüber hinaus auch in der Satzung der Stiftung verankert. Die Beratungsleistung der Expertinnen und Experten kann durch sekundäre Interessen unangemessen beeinflusst werden. Solche sekundären Interessen können materieller und nicht materieller Natur sein, wobei die Kategorien nicht trennscharf sind und sich häufig gegenseitig bedingen (Klemperer 2011: 14 ff.). Materielle Interessenkonflikte entstehen z. B. durch Vergütungsanreize in Dienstverträgen, nicht materielle Interessenkonflikte können beispielsweise durch Streben nach öffentlichem Ansehen entstehen oder wenn eine Person eine berufliche Position oder ein Amt innehat, das diese zum Vertreten spezifischer Interessen verpflichtet. Interessenkonflikte sind alltäglich und unvermeidlich. Sie erhöhen das Risiko für ein verzerrtes Urteil und können die Expertinnen oder Experten beeinflussen, ohne dass sie es bemerken (Thompson 1993). Damit eine unabhängige Beratung gewährleistet werden kann, müssen Interessenkonflikte, die zu einer unsachgemäßen Verzerrung der Beratungsleistung führen könnten, von den bewerbenden Personen im Vorfeld einer Berufung als Expertin oder Experte offengelegt werden. Dies bedeutet im Umkehrschluss, dass bestehende Interessenkonflikte nicht notwendigerweise zu einem Ausschluss führen, was angesichts des meist breiten Betätigungsfeldes ausgewiesener Experten auch nur zulasten der Fachlichkeit gelingen würde (Roters in: Kasseler Kommentar, § 137a, Rdnr. 36).

Es wird geprüft, ob die angegebenen Interessenkonflikte für die vorgesehenen Beratungsthemen relevant sind und als wie ausgeprägt diese zu bewerten sind. Sofern sich aufgrund der vorgelegten Informationen erhebliche Bedenken hinsichtlich einer Verzerrung des Beratungsergebnisses ergeben, wird diese Person nicht als Mitglied von Expertengruppen berufen.

Die Bewertung von potenziellen Interessenkonflikten erfolgt durch die Selbstauskunft der als Expertinnen und Experten einzubeziehenden Personen. Hierzu verwendet das Institut ein Formblatt zur „Offenlegung potenzieller Interessenkonflikte“, wie es bereits seit vielen Jahren bei der Entwicklung und Durchführung von Qualitätssicherungsverfahren verwendet wird. Dieses ist auf der Website des Instituts einsehbar und kann heruntergeladen werden. Die Selbstauskunft bezieht sich auf folgende Arten von Beziehungen:

- Arbeitsverhältnisse oder ehrenamtliche Tätigkeiten
- Beratungsverhältnisse
- Honorare, z. B. für Vorträge und Stellungnahmen
- finanzielle Unterstützung von Forschungsaktivitäten
- sonstige Unterstützung, z. B. durch Übernahme von Reisekosten oder Teilnahmegebühren ohne wissenschaftliche Gegenleistung, bei der Erstellung von Broschüren, bei der Einrichtung und Unterhaltung von Beratungsstellen oder bei Ausrichtung und Organisation einer Veranstaltung
- Aktienbesitz

Die Bewertung der Angaben erfolgt in zwei Schritten. Zunächst fasst der Fachbereich im Institut die Angaben zusammen und unterzieht sie einer ersten Bewertung im Kontext des Beratungsthemas. In einem zweiten Schritt wird diese Bewertung durch eine neutrale Stelle im Institut (z. B. Justizariat) überprüft und ggf. ergänzt. Bestehen Zweifel oder wird ein Interessenkonflikt

als relevant eingestuft, wird die Institutsleitung eingebunden und gemeinsam eine Entscheidung herbeigeführt.

Die Namen der Experten sowie eine Zusammenfassung der Angaben zu den potenziellen Interessenkonflikten werden auf der Website des Instituts veröffentlicht. Dabei werden die Informationen aus den Formblättern nur allgemein wiedergegeben, d. h. es wird dargelegt, ob Interessenkonflikte bestehen, jedoch keine Einzelheiten z. B. über die Höhe von Zuwendungen oder über Partner.

### **Anforderungen während der Tätigkeit als Expertin oder Experte**

Die beratenden Expertinnen und Experten müssen in der Lage sein, die Beratung in dem vorgesehenen Umfang und während des gesamten vorgesehenen Zeitraums der Einbindung leisten zu können (Kontinuität). Für Mitglieder von Expertengruppen schließt dies insbesondere die Teilnahme an Sitzungen sowie deren Vor- und Nachbereitung ein.

Im Rahmen der Expertentätigkeit werden teilweise vertrauliche Unterlagen und Diskussionsinhalte behandelt. Personen, die sich als beratende Expertinnen und Experten bewerben, müssen sich daher zur strikten Wahrung der Vertraulichkeit dieser Inhalte verpflichten. Dies ist auch notwendig, um die Beratungsprozesse nach sachlichen Gesichtspunkten und frei von externer Beeinflussung durchführen zu können.

### **Kriterien für die Zusammensetzung von Expertengruppen**

Für die Zusammenstellung von Expertengruppen werden zusätzlich zu den personenbezogenen Anforderungen an Expertinnen und Experten weitere Kriterien benötigt, die die Auswahl der Gruppenmitglieder beschreiben.

### **Heterogene Gruppenzusammensetzung**

In der Regel wird bei der Benennung von Expertengruppen eine möglichst heterogene Gruppenzusammensetzung angestrebt (so wie es z. B. auch die AHRQ im Rahmen der Indikatorenentwicklung handhabt: CHP/PCOR/Battelle Memorial Institute 2011). Heterogenität soll in Abhängigkeit von dem Einsatzbereich der jeweiligen Expertengruppe insbesondere erzielt werden hinsichtlich folgender Aspekte:

- Versorgungssektor
- Einrichtungsart und -größe
- Rolle hinsichtlich des Verfahrens (z. B. Leistungserbringer, Patientin/Patient, Qualitätsmanagerin/Qualitätsmanager, Institution)
- Beruf und Fachrichtung
- Geschlecht

Eine heterogene Gruppenzusammensetzung ermöglicht, dass viele verschiedene Ansichten und Ideen in die Gruppe eingebracht werden. In bestimmten Einsatzbereichen kann für eine Expertengruppe auch Homogenität in einem einzelnen Merkmal notwendig sein, z. B. wenn über die Expertinnen und Experten die Ansichten einer bestimmten Zielgruppe wie Patientinnen und Patienten oder Leistungserbringer ermittelt werden sollen (Häder 2000). Welche Anforderungen

für eine spezifische Expertengruppe relevant sind und für welche Merkmale ein besonderes Gewicht auf Heterogenität gelegt wird, wird während des konkreten Berufungsverfahrens entsprechend dokumentiert.

### **Gruppengröße**

Mit zunehmender Zahl an Expertinnen und Experten in einer Gruppe kann einerseits leichter Heterogenität in verschiedenen Eigenschaften erreicht werden, andererseits muss die Gruppengröße zusätzliche, eher praktische Aspekte berücksichtigen, wie z. B. den Zeitbedarf, der für Beratungen und Bewertungen zur Verfügung steht, den personellen und finanziellen Aufwand bei den externen Expertinnen und Experten und dem Institut und ggf. regulatorische Vorgaben z. B. in G-BA-Richtlinien. Die angestrebten Mindest- und Höchstzahlen an Mitgliedern für die Expertengruppen können in den verschiedenen Aufgabenbereichen unterschiedlich sein. Dies wird in den spezifischen Auswahlverfahren konkret angegeben.

### **Besonderheiten bei Expertengruppen mit Bewertungsaufgaben**

Einige Aufgaben von Expertengruppen umfassen nicht nur die Sammlung von Ideen und von Informationen, sondern sollen als Ergebnis auch eine Bewertung bestimmter Aspekte liefern (siehe Abschnitt 7.2). Diese Bewertungen sollen soweit wie möglich anhand fachlich-inhaltlicher Kriterien vorgenommen werden, d. h., einer Einseitigkeit der Bewertungsergebnisse bedingt durch die Zusammensetzung der Gruppe oder einer Verzerrung aufgrund des Bewertungsverhaltens einzelner Mitglieder soll entgegengewirkt werden. Ansätze hierfür sind:

- der Ausschluss relevanter Interessenkonflikte
- die Sicherstellung der Ausgewogenheit der Gruppenzusammensetzung
- die Nutzung formaler Gruppenprozesse zur Bewertungsfindung (siehe Abschnitt 7.3.3 ab S. 107).

Auch hier werden heterogene Gruppen angestrebt. Insgesamt ist jedoch kaum systematisch evaluiert, wie Expertengremien mit Bewertungsaufgaben in der Gesundheitsversorgung optimal zusammengesetzt werden. Reviews, die sich z. B. mit der Expertenauswahl für Delphi-Panels beschäftigten, konnten keine allgemein akzeptierten Kriterien beschreiben (Keeney et al. 2006, Mullen 2003). Es besteht zwar in der Literatur Einigkeit, dass die Gruppenzusammensetzung einen Einfluss auf das Ergebnis formaler Gruppenprozesse (oft auch als Konsensusverfahren bezeichnet) hat (Campbell et al. 2003, Campbell et al. 2004, Stelfox und Straus 2013), hochwertige und konsistente Evidenz findet sich aber weder für einen Vorteil von heterogenen gegenüber homogenen Gruppen noch für konkrete Gruppengrößen (Mullen 2003, Murphy et al. 1998). Zweck der Nutzung von Expertengruppen mit Bewertungsaufgaben ist daher auch nicht, Repräsentativität des Bewertungsergebnisses für eine Population zu erreichen, sondern möglichst viele relevante Ansichten einzubeziehen. Für Gruppen mit Bewertungsaufgaben wird angestrebt, die Zusammensetzung der Gruppen wie oben erläutert gezielt zu strukturieren (vgl. Häder 2000).

### **Vorgehen bei der Berufung von Expertengruppen**

Um die Ziele der Experteneinbindung – breiter Einbezug von Wissen und Erfahrungen sowie Akzeptanz der Beratungsergebnisse – zu erreichen, werden Expertengruppen – soweit es in den

übergeordneten rechtlichen Regelungen nicht anders bestimmt ist – in einem kriteriengebundenen Bewerbungsverfahren besetzt. Die Institutionen des Gesundheitswesens werden auf die Bewerbungsmöglichkeit jeweils hingewiesen, um zu ermöglichen, dass alle relevanten Stakeholder über von ihnen zur Bewerbung motivierte Expertinnen und Experten ihre Informationen und Perspektiven in den Beratungsprozess einbringen können.

Die Auswahl und Berufung der beratenden Expertinnen und Experten erfolgt nach Maßgabe der oben angegebenen Kriterien. Diese werden in geeigneter Form transparent gemacht. Sind die Kriterien für die Tätigkeit als Expertin/Experte nicht mehr erfüllt, endet die Beratungstätigkeit der betreffenden Person. Sofern für ein dauerhaft eingerichtetes Expertengremium eine Nachberufung erforderlich ist, folgt diese dem Vorgehen und den Kriterien für die erstmalige Berufung.

### **Personenbindung der Kriterien**

Die externen Expertinnen und Experten werden auf Grundlage ihrer persönlichen und beruflichen Kenntnisse hinzugezogen bzw. für Expertengruppen berufen. Die Kontinuität der Beratung und die Vermeidung von Verzerrungen in der Beratung durch Expertengruppen hängen daher unmittelbar mit der Person der Expertinnen und Experten zusammen. Es sind daher keine Vertretungsregelungen für Expertinnen und Experten vorgesehen, die vorübergehend an Beratungs- oder Sitzungsterminen nicht teilnehmen können. Ausnahmen hiervon können sich ergeben, wenn dies für den Fortgang der Beratungen erforderlich ist (z. B. Präsentation von Materialien durch eine Vertretung, wenn die Person mit Expertise kurzfristig verhindert ist). Sofern Expertinnen und Experten dauerhaft an Beratungen nicht teilnehmen können, wird eine Neuberufung nach den genannten Kriterien vorgenommen.

### **Formale Gruppenprozesse**

Eine Aufgabe von Expertengruppen kann sein, quantitative Bewertungen eines Sachverhalts abzugeben. In diesem Fall wird für die Bewertung durch die Gruppe ein geeigneter formaler Gruppenprozess genutzt. Für quantitative Bewertungsprozesse durch Gruppen gibt es keinen Goldstandard (Boulkedid et al. 2011, Mullen 2003, Murphy et al. 1998, Waggoner et al. 2016). Es wird von der Annahme ausgegangen, dass persönliche Treffen einer Gruppe durch die gemeinsame Diskussion eines Sachverhalts zu erwünschten, inhaltlich bedingten Meinungsänderungen der Gruppenmitglieder führen. Außerdem wird angenommen, dass die anonyme Abgabe der Bewertungen durch die Gruppenmitglieder gruppendynamisch oder psychologisch bedingte Einflüsse auf das Bewertungsergebnis verringern kann. Der Bewertungsprozess kann daher beispielsweise in Form eines Delphi-Verfahrens (Mullen 2003) durchgeführt werden, mit der Modifikation, dass postalische und persönliche Bewertungsrunden mit anonymer Abgabe der Bewertung kombiniert werden. Da es keine allgemein akzeptierte Methode gibt, der Expertise einzelner Gruppenmitglieder individuelle Gewichte zuzuweisen, werden die Einzelbewertungen gleich gewichtet. Der Algorithmus, nach welchem die Bewertungen der Gruppenmitglieder aggregiert werden und nach welchem Übereinstimmung oder Dissens der Einzelmeinungen dargestellt wird, wird dabei a priori festgelegt.

**Transparenz des formalen Gruppenprozesses**

Bei formalen Bewertungen durch Gruppen ist davon auszugehen, dass die Informationsgrundlage für die Gruppe, die Größe und Zusammensetzung der Gruppe, die Kriterien, nach denen eine Bewertung abgegeben werden soll, die Anzahl der Bewertungsrunden und der Umgang mit verschiedenen Meinungen im Bewertungsprozess Einfluss auf das Ergebnis der Bewertung haben (Campbell et al. 2003, Murphy et al. 1998, Sackman 1974) und deswegen explizit beschrieben werden sollten (Diamond et al. 2014). Sofern formale Gruppenprozesse durchgeführt werden, werden diese Aspekte sowie das Gesamtergebnis mit Richtung und Grad der Übereinstimmung detailliert dokumentiert.

## 8 Eignung von Qualitätsindikatoren

Kern der Aufgaben des Instituts ist die Qualitätsmessung in der Gesundheitsversorgung mittels Qualitätsindikatoren. Die Messung der Qualität und Darstellung der Ergebnisse soll Qualitätsverbesserungen bewirken und hat folglich einen potenziell großen Einfluss auf die Versorgung von Patientinnen und Patienten. Aus diesem Grund ist es wichtig sicherzustellen, dass die verwendeten Qualitätsindikatoren auch tatsächlich für diese Zwecke geeignet sind. Dieses Kapitel gibt einen Überblick über die Kriterien, anhand derer die Eignung der Qualitätsindikatoren für die Qualitätsmessung in der Gesundheitsversorgung beurteilt wird.

### 8.1 Konzepte zur Eignungsbeurteilung in der Literatur

In Veröffentlichungen zur medizinischen Qualitätssicherung sind eine Reihe von Kriterienkatalogen zu finden, oft als „Gütekriterien“ für Qualitätsindikatoren bezeichnet, die die Eignung von Indikatoren für Qualitätsmessungen beschreiben sollen (siehe z. B. AQUA 2015, de Koning 2007b, Geraedts et al. 2003, Jones et al. 2014, Mainz 2003, NQF 2015, Reiter et al. 2007, Stelfox und Straus 2013, Wollersheim et al. 2007). Diese Kriterienkataloge variieren bezüglich der Anzahl der Kriterien, der Definition der Kriterien und der Funktion der Kriterien (Rode et al. 2016), was u. a. durch unterschiedliche Schwerpunktsetzungen bedingt ist. Beispielsweise fokussieren manche Kriterienkataloge mehr auf die Güte des Entwicklungsprozesses der Indikatoren als auf die Anwendung der Indikatoren (z. B. de Koning 2007a). Wieder andere Kriterienkataloge differenzieren bei der Beurteilung der Eignung nicht zwischen verschiedenen Anwendungsbereichen von Indikatoren (z. B. NQF 2015).

Die in der Literatur aufgeführten Eignungskriterien weisen zudem sehr unterschiedliche Detaillierungsgrade auf. Beispiele hierfür sind weite Definitionen des Kriteriums „Relevanz“ (AQUA 2015, Jones et al. 2014), bei denen das Kriterium eher eine Sammelkategorie sehr verschiedener Aspekte darstellt, sowie Zusammenfassungen von „Nutzung“ und „Nutzen“ von Indikatorwerten in einem einzelnen Kriterium (NQF 2015, Reiter et al. 2007). Dementsprechend hängt auch die Anzahl der Kriterien, die verschiedene Autoren auf Grundlage systematischer Literatursuchen abgeleitet haben, vom Vorgehen bei der Zusammenfassung inhaltlich verwandter Kriterien ab (Jones et al. 2014, Reiter et al. 2007).

Es wird deutlich, dass kein allgemein anerkannter Standard für die Beurteilung der Eignung von Qualitätsindikatoren existiert (Rode et al. 2016). Es wurde daher unter Berücksichtigung der Literatur eine eigene Systematik für die Eignungsbeurteilung von Qualitätsindikatoren entwickelt, die möglichst klar und trennscharf definierte Eignungskriterien verwendet.

### 8.2 Grundsätze zur Eignungsbeurteilung

Ob ein Qualitätsindikator für ein bestimmtes QS-Verfahren geeignet ist, bestimmt sich danach, inwieweit er dem spezifischen Zweck des QS-Verfahrens (in den Kategorien Qualitätsförderung, Information, Regulierung; siehe Abschnitt 2.4) dient. Es wird dabei davon ausgegangen, dass die

Eignung von Qualitätsindikatoren nur im Hinblick auf den Kontext ihres Einsatzes beurteilt werden kann. Ein Qualitätsindikator kann folglich für unterschiedliche Anwendungszwecke auch unterschiedlich gut geeignet sein (Meyer 2004: 24). Beispielsweise ist ein Qualitätsindikator zur Schlaganfallbehandlung in der Notaufnahme für die Qualitätsförderung geeignet, aber nicht für die Unterstützung der Auswahlentscheidungen von Patientinnen und Patienten, da diese im Notfall meist keine Wahlmöglichkeit haben. Weiterhin erfordern Indikatoren, die für Zwecke der Qualitätsförderung (z. B. durch den Strukturierten Dialog) eingesetzt werden, andere Bewertungskonzepte als Indikatoren, die beispielsweise zwischen guter Qualität und exzellenter Qualität unterscheiden sollen. Durch die Eignungsbeurteilung der Qualitätsindikatoren soll für alle Komponenten eines Qualitätsindikators (Qualitätsziel, Messverfahren, Bewertungskonzept; siehe Abschnitt 3.4) eine Aussage ermöglicht werden, inwieweit diese für den Zweck eines bestimmten QS-Verfahrens geeignet sind.

Die Beurteilung der Eignung erfolgt anhand von Kriterien, die je nach Anwendungszeitpunkt auf verschiedene Weisen genutzt werden:

- prospektiv als Leitfaden für die Neuentwicklung geeigneter Qualitätsmessungen
- retrospektiv für die Beurteilung der Eignung bestehender und ggf. schon in Gebrauch befindlicher Indikatoren und Indikatorensets, z. B. bei der Überprüfung von Indikatoren am Ende der Erprobungsphase eines neuen QS-Verfahrens oder im Rahmen der Verfahrenspflege

Da Qualitätsindikatoren letztendlich die Operationalisierung von Qualitätsaspekten und Qualitätsmerkmalen darstellen, muss eine Beurteilung der Indikatoreignung auch diese theoretischen, zugrunde liegenden Konzepte umfassen. Zudem ist die Eignungsbeurteilung nur dann adäquat möglich, wenn nicht allein einzelne Qualitätsindikatoren, sondern auch deren Auswahl, Zusammenspiel und Verwendung betrachtet werden. Das bedeutet, dass ein Teil der in den folgenden Abschnitten dargestellten Eignungskriterien sich nicht nur auf den Qualitätsindikator im engeren Sinne bezieht, sondern auch auf indikatorübergreifende Aspekte der Qualitätsmessung. Beispielsweise muss die Patientenrelevanz der Inhalte einer Qualitätsmessung bereits zu einem frühen Zeitpunkt in der Entwicklung eines QS-Verfahrens beurteilt werden und dient dabei als Kriterium für die Auswahl von Qualitätsaspekten, sie ist aber auch für den konkret formulierten Qualitätsindikator nochmals sicherzustellen. Der besseren Lesbarkeit halber wird in der folgenden Darstellung vereinfachend von Eignung bzw. Eignungskriterien der Qualitätsindikatoren gesprochen; wo es für das Verständnis wichtig ist, wird aber auf die verschiedenen Bezugsebenen der Eignungskriterien hingewiesen.

Die Eignung von Qualitätsindikatoren ist nicht nur von der Konstruktion des Indikators (Konzept, Rechenregeln etc.), sondern auch von der konkreten Datengrundlage, der betrachteten Population, den Rahmenbedingungen der Qualitätsmessung und -bewertung sowie der Zielsetzung des QS-Verfahrens abhängig. So hängt etwa die Reliabilität eines Indikators maßgeblich von der Varianz der gemessenen Indikatorwerte zwischen Leistungserbringern ab und potenzielle Nebenwirkungen können bei Verfahren zur Qualitätsförderung und Verfahren der Regulierung unterschiedlich sein. Bei der Eignung eines Indikators handelt es sich also nicht um eine inhärente Eigenschaft des Indikators, sondern um die Fähigkeit, bestimmte Anforderungen für den vorgesehenen Anwendungsbereich unter den gegebenen Umständen zu erfüllen. Aus diesem Grunde

wird nicht der Begriff „Gütekriterien“, sondern „Eignungskriterien“ verwendet. Zudem kann eine vollständige Beurteilung der Eignung eines Qualitätsindikators erst erfolgen, wenn praktische Erfahrungen mit ihm vorliegen. Dies betrifft z. B. Fragen der Praktikabilität und der Brauchbarkeit für den Verfahrenszweck. Diese Kriterien können also erst in der Anwendung, d. h. im Regelbetrieb der QS-Verfahren beurteilt werden.

Bei der Beurteilung der Eignung von Qualitätsindikatoren für ein QS-Verfahren werden vier Kategorien unterschieden, wobei drei davon den Komponenten der Qualitätsindikatoren (Qualitätsziel, Messverfahren, Bewertungskonzept; siehe Abschnitt 3.4) entsprechen und die vierte Kategorie Eignungskriterien enthält, die sich komponentenübergreifend auf den Qualitätsindikator als Ganzes beziehen (siehe Abbildung 9).

Eignung			
des Qualitätsziels	des Messverfahrens	des Bewertungskonzepts	Bezug zum Verfahrenszweck
Bedeutung für die Patientinnen und Patienten	Objektivität der Messung	Zweckeignung des Bewertungskonzepts	Brauchbarkeit
Potenzial zur Verbesserung	Reliabilität der Messung	Erreichbarkeit des Referenzbereichs	potenzielle Nebenwirkungen
Zusammenhang mit patientenrelevantem Ziel	Datenvalidität	Klassifikationsgüte	Zuschreibbarkeit der Verantwortung
Übereinstimmung mit Qualitätsaspekt	Validität der Messung	Angemessenheit d. Risikoadjustierung	
Beeinflussbarkeit	Praktikabilität der Messung		

Abbildung 9: Überblick über die Eignungskriterien für Qualitätsindikatoren

Für einen Teil dieser Eignungskriterien existiert keine Kennzahl, die den Erfüllungsgrad des Kriteriums sinnvoll quantifizieren könnte, weil sie entweder qualitativer Natur sind oder weil keine entsprechenden Daten zur Verfügung stehen. So ist beispielsweise die Beeinflussbarkeit der Indikatorwerte durch die Leistungserbringer ein wichtiges Kriterium für die Eignung von Qualitätsindikatoren. Das Ausmaß, zu dem Indikatorwerte durch die Leistungserbringer beeinflussbar sind, lässt sich jedoch nicht quantifizieren. Die Beurteilung, ob das jeweilige Kriterium angemessen erfüllt ist, kann in diesen Fällen nur durch eine qualitative Einschätzung unter Zuhilfenahme ergänzender Informationen wie Literatur und externer Expertise erfolgen. Ein formales Verfahren, das Expertenbewertungen in einem Gruppenprozess quantifiziert, kann hier eine nützliche Zusatzinformation darstellen (siehe Abschnitt 7.3.3 ab S. 107), insbesondere beim Vergleich ver-

schiedener Qualitätsindikatoren. Die Expertinnen und Experten übernehmen dabei die Funktion, die eingebrachten Informationen zusammenfassend zu bewerten und ihre persönliche Einschätzung als Zahl auszudrücken. Am Beispiel der Beeinflussbarkeit der Indikatorergebnisse durch die Leistungserbringer wird allerdings auch deutlich, dass durch eine Quantifizierung im Rahmen eines Gruppenprozesses nicht die tatsächliche Beeinflussbarkeit der Indikatorergebnisse ermittelt wird, sondern lediglich die Meinung der Expertengruppe bezüglich der Beeinflussbarkeit quantifiziert wird. Das Ergebnis eines solchen Verfahrens ist somit immer auch von den impliziten Überzeugungen der Gruppenmitglieder und der Zusammensetzung der Gruppe abhängig. Die abschließende Einschätzung hinsichtlich der Erfüllung eines Eignungskriteriums nimmt daher grundsätzlich das Institut durch qualitative Beurteilung, Einordnung und Zusammenschau der verfügbaren Informationen vor. Aus der fehlenden Quantifizierbarkeit eines Teils der Eignungskriterien folgt außerdem, dass ein objektiver Nachweis der Erfüllung der Kriterien, wie es bei Kennzahlen etwa über das Erreichen eines Schwellenwerts definiert werden kann, nicht möglich ist. Ein einzelnes Eignungskriterium wird daher als angemessen erfüllt angesehen, wenn unter Anwendung der beschriebenen Methoden in der Gesamtbetrachtung die positiven Belege und Hinweise gegenüber möglichen gegenteiligen Hinweisen überwiegen.

Die Gesamteignung eines Qualitätsindikators für ein QS-Verfahren bemisst sich danach, ob alle auf diesen Indikator anwendbaren Eignungskriterien angemessen erfüllt sind. Auf der Basis möglichst reichhaltiger Informationen (z. B. Literatur, Expertengespräche) muss die Auswahl von Indikatoren durch Abwägung von Vor- und Nachteilen für einen Indikator erfolgen und lässt sich nicht sinnvoll durch einen festen Algorithmus treffen (Meyer 2004: 40). Aufgrund der qualitativen Natur vieler Eignungskriterien wird von einer numerischen Gewichtung der Kriterien untereinander abgesehen. Die Gesamtbeurteilung der Indikatoreignung erfolgt stattdessen in Form einer qualitativen Abwägung der Einzelkriterien, die in den Berichten zur Verfahrens- bzw. Indikatorenentwicklung z. B. in Form eines Indikatorprofils dargestellt werden. Das gleiche gilt, wenn zwischen zwei oder mehr möglichen Indikatoren zur Abbildung eines Sachverhalts entschieden werden muss. Transparenz über das entwickelte QS-Verfahren und die Indikatoren entsteht somit nicht durch die Quantifizierung kaum quantifizierbarer Kriterien, sondern durch die nachvollziehbare Darlegung der Gründe, weshalb das IQTIG einen Indikator für geeignet hält.

### **8.3 Eignungskriterien des Qualitätsziels des Indikators**

Voraussetzung für eine Qualitätsmessung mit Indikatoren ist, dass das Qualitätsziel des Indikators (kurz: Indikatorziel) tatsächlich auch Qualität widerspiegelt und dass Verbesserungen dieser Qualität erreichbar sind. Vereinfacht ausgedrückt muss beurteilt werden, ob mit den Qualitätsindikatoren das Richtige gemessen wird. Diese Beurteilung lässt sich dabei in fünf einzelne Eignungskriterien untergliedern, die in den folgenden Abschnitten beschrieben werden.

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

Anhand dieses Kriteriums wird beurteilt, wie bedeutend ein Qualitätsaspekt oder ein Qualitätsmerkmal für die Patientinnen und Patienten ist. Diese Beurteilung erfolgt zum einen in der Konzeptstudie bezogen auf die ermittelten Qualitätsaspekte, zum anderen im Prozess der Indikatorenentwicklung bezogen auf die Qualitätsmerkmale. Die Notwendigkeit einer nochmaligen Be-

urteilung nach diesem Kriterium bei der Entwicklung der Indikatoren ergibt sich aus der höheren Konkretheit in diesem Bearbeitungsschritt. Es soll sichergestellt werden, dass das konkrete Qualitätsmerkmal und der darauf basierende Indikator (z. B. „Anteil an Wundinfektionen“) eine für die Patientinnen und Patienten bedeutsame Konkretisierung des Qualitätsaspekts (z. B. „Komplikationen der Behandlung“) ist und nicht nur eine weniger bedeutsame (z. B. „blaue Flecken“) darstellt. Die Bedeutung für die Patientinnen und Patienten muss also auch nach Konkretisierung des zu messenden Qualitätsmerkmals und Operationalisierung in Form eines Qualitätsindikators noch gegeben sein. Eine hohe Bedeutung kann ein Qualitätsmerkmal beispielsweise besitzen bei:

- großer individueller Schwere (z. B. hohe Beeinträchtigung der einzelnen Patientin oder des einzelnen Patienten durch eine Erkrankung oder durch den Eintritt einer Komplikation; Verkürzung oder Verlängerung von Lebensdauer oder Krankheitsdauer)
- Aspekten der Patientensicherheit
- hoher Fallzahl (d. h. große Krankheitslast bezogen auf die Gesamtpopulation)
- wichtigen ethischen Fragen (z. B. Verteilung von Spenderorganen zur Transplantation)
- Hinweisen auf große Qualitätsdefizite der Versorgung

Die Bedeutung für Patientinnen und Patienten muss für jedes Thema gezielt ermittelt werden (ähnlich: Mühlhauser und Müller 2009: 36 f.). In den Entwicklungsprojekten des IQTIG sind die Informationsgrundlage hierfür insbesondere die Ergebnisse der Fokusgruppen mit Patientinnen und Patienten sowie ergänzend Hinweise aus Literatur und von Expertinnen und Experten, wobei wiederum die Patientenvertretung in die beratenden Expertengruppen eingebunden ist. Für bereits in Verwendung befindliche Qualitätsindikatoren, die nicht explizit aus einem Qualitätsmodell abgeleitet wurden, kann ggf. anhand der Rationale und des formulierten Ziels des Indikators rückgeschlossen werden, welches Qualitätsmerkmal durch diesen Indikator gemessen werden soll, und dieses einer entsprechenden Beurteilung unterzogen werden.

### **Potenzial zur Verbesserung**

Neben dem Kriterium „Bedeutung für die Patientinnen und Patienten“, das Patientenrelevanz auf persönlicher Ebene adressiert, ist die Wichtigkeit einer Qualitätsmessung auch auf einer übergeordneten Ebene zu beurteilen.<sup>16</sup> Dabei ist für die Qualitätsmerkmale und die zugehörigen Qualitätsindikatoren die Frage zu klären, ob es sich um ein Merkmal handelt, durch dessen Messung und Nutzung in einem QS-Verfahren überhaupt positive Effekte auf die Gesundheitsversorgung auf Systemebene denkbar sind. Wenngleich eine angemessene Antwort auf diese Fragestellung allenfalls im Rahmen einer Evaluation des QS-Verfahrens vorgenommen werden könnte, so sind doch bereits bei der Indikatorenentwicklung Hinweise zu suchen, ob zumindest ein nennenswertes Potenzial für Verbesserungen besteht. So kann es Qualitätsmerkmale geben, die zwar eine hohe Bedeutung für die betroffenen Patienten haben, aber deren Messung wegen bei allen Leistungserbringern hoher Versorgungsqualität keine nützlichen Zusatzinformationen

---

<sup>16</sup> Es wird also – anders als z. B. im QUALIFY-Instrument (Reiter et al. 2007) oder im AQUA-Verfahren (AQUA 2015) – bei der Beurteilung zwischen Patientenrelevanz und Bedeutung auf Systemebene getrennt.

bietet. Umgekehrt kann Verbesserungspotenzial bei Merkmalen bestehen, die aber aus Patientensicht wenig bedeutsam sind. Aus Aufwand-Nutzen-Gründen ist die Einführung oder Beibehaltung von Indikatoren zu vermeiden, deren Nützlichkeit begrenzt ist oder deren Ergebnisse eine Sättigungsgrenze erreicht haben (Meyer 2004: 39). Die Beurteilung des Potenzials zur Verbesserung erfolgt dementsprechend nicht nur in der Verfahrensentwicklung bei der Erstellung des Qualitätsmodells und bei der Indikatorenentwicklung, sondern auch wiederkehrend für Qualitätsindikatoren in laufenden QS-Verfahren.

Als Informationsgrundlage für die Beurteilung des Verbesserungspotenzials werden in der Verfahrensentwicklung die Ergebnisse aus Fokusgruppen, Literaturrecherche und aus der Beratung durch Expertinnen und Experten genutzt, bei laufenden QS-Verfahren können zudem Auswertungen der bestehenden Daten erstellt werden. Anhaltspunkte, dass positive Effekte durch die Messung und Nutzung in einem QS-Verfahren in einem relevanten Ausmaß möglich sind, bestehen z. B. bei

- Belegen für eine hohe Variabilität in der Versorgungsqualität (z. B. aus Studien oder Auswertungen von QS-Daten)
- gehäuften Berichten über Qualitätsdefizite oder Fehlentwicklungen (z. B. von Expertinnen und Experten oder in Fokusgruppen)
- hohen Kosten (d. h. hohem Verbrauch von Ressourcen, die dann für andere Maßnahmen nicht zur Verfügung stehen), z. B. für nicht indizierte Behandlungen
- hoher Fallzahl (d. h. großem Gewicht bezogen auf die Gesamtpopulation)
- großem öffentlichen Interesse aus einem sonstigen Grund

#### **Zusammenhang mit einem unmittelbar patientenrelevanten Ziel**

Bei den Indikatorzielen werden mittelbar und unmittelbar für Patientinnen und Patienten relevante unterschieden (siehe Tabelle 7). Unmittelbar patientenrelevante Indikatorziele liegen bei Qualitätsindikatoren vor, die ein Qualitätsmerkmal messen, das *per se* Bedeutung für die Patientinnen und Patienten hat und das in der Regel unmittelbar von diesen erfahren wird oder berichtet werden kann, wie z. B. Mortalität, Komplikationen einer Operation, oder Aufklärung über Behandlungsalternativen. Unmittelbar patientenrelevante Qualitätsmerkmale basieren auf ethischem oder sozialem Konsens. Demgegenüber bezeichnen mittelbar patientenrelevante Indikatorziele solche Ziele und Qualitätsmerkmale, die erst durch den Nachweis eines Zusammenhangs mit einem unmittelbar patientenrelevanten Ziel als relevant angesehen werden können. Hierbei handelt es sich z. B. um Laborparameter, die als Surrogate für ein patientenrelevantes Ergebnis verwendet werden, oder um Strukturen und Prozesse, die die Wahrscheinlichkeit des Auftretens eines bestimmten patientenrelevanten Ergebnisses erhöhen.

Tabelle 7: Beispiele für unmittelbar und mittelbar patientenrelevante Indikatorziele

	<b>unmittelbar patientenrelevantes Indikatorziel</b>	<b>mittelbar patientenrelevantes Indikatorziel</b>
<b>Strukturen</b>	barrierefreier Zugang	Desinfektionsmittelspender in Patientenzimmer
<b>Prozesse</b>	partizipative Entscheidungsfindung	Antibiotikaprophylaxe bei Operation
<b>Ergebnisse</b>	niedrige Mortalität	CRP-Abnahme bei Rheumabehandlung

Mittelbar patientenrelevante Qualitätsmerkmale und darauf basierende Qualitätsindikatoren werden nur dann als geeignet für die Qualitätsmessung angesehen, wenn der Zusammenhang mit einem unmittelbar patientenrelevanten Ziel ausreichend sicher ist.<sup>17</sup> So kann beispielsweise das Qualitätsziel, bestimmte Laborwerte zu erreichen (mittelbar patientenrelevantes Ziel), mit einer verminderten Sterblichkeit einhergehen (unmittelbar patientenrelevantes Ziel). Um diesen Zusammenhang zwischen mittelbar und unmittelbar patientenrelevanten Zielen zu belegen, erfolgt eine systematische Literaturrecherche und -bewertung im Rahmen der Neuentwicklung von Indikatoren. Dieser Zusammenhang wird zudem regelmäßig im Rahmen der Verfahrenspflege überprüft. Falls keine ausreichende externe Evidenz für diesen Zusammenhang vorliegt, kann auf eine konsenterte Einschätzung von Expertinnen und Experten zurückgegriffen werden. Ein fester Grenzwert, ab welcher Qualität und Konsistenz der recherchierten Evidenz der Zusammenhang mit einem unmittelbar patientenrelevanten Ziel als ausreichend belegt anzusehen ist, wird nicht definiert (vgl. Grundsätze zur Eignungsbeurteilung in Abschnitt 8.2). Die Bewertung von Qualität und Konsistenz der Evidenz wird in der Beschreibung der Indikatoren dokumentiert.

### Übereinstimmung mit dem zugehörigen Qualitätsaspekt

Aus den Qualitätsaspekten, die im Qualitätsmodell als bedeutsam festgestellt wurden, werden in der Verfahrensentwicklung Qualitätsmerkmale abgeleitet und in Form von Qualitätsindikatoren operationalisiert (siehe Kapitel 4). Entsprechend wird vorgegangen, wenn im Rahmen der Verfahrenspflege oder -weiterentwicklung einzelne Qualitätsindikatoren neu entwickelt werden müssen (siehe Abschnitt 5.2.3). In diesem Schritt der Konkretisierung muss beurteilt werden, ob ein mögliches Qualitätsmerkmal (z. B. „transfusionspflichtige postoperative Blutung“) in Bezug zu einem als relevant ermittelten Qualitätsaspekt (z. B. „Komplikationen der Behandlung“) steht. Dies ist erforderlich, um sicherzustellen, dass jeder Qualitätsaspekt des Qualitätsmodells auch durch mindestens einen Qualitätsindikator abgebildet wird.<sup>18</sup> Gleichzeitig wird durch die Beurteilung mittels dieses Eignungskriteriums die Entwicklung von Indikatoren vermieden, die

<sup>17</sup> Für Struktur- und Prozessindikatoren entspricht dieses Eignungskriterium dem in QUALIFY als Indikator-evidenz bezeichneten Kriterium (Reiter et al. 2007).

<sup>18</sup> Das Ausmaß, in dem alle Qualitätsaspekte des Qualitätsmodells durch das Qualitätsindikatorenset eines QS-Verfahrens abgebildet werden, könnte auch als Inhaltsvalidität des Indikatorensets für die Abbildung des Qualitätsmodells bezeichnet werden.

Aspekte adressieren, die nicht im Qualitätsmodell aufgenommen wurden. Die Beurteilung der Übereinstimmung erfolgt durch sachlogische Überlegungen zum Versorgungsgeschehen.

### **Beeinflussbarkeit durch den Leistungserbringer**

Die Qualität der Versorgung besteht in der Fähigkeit, bestimmte legitime Ziele zu erreichen, wobei diese Ziele unter anderem im Rahmen des Möglichen sein müssen (Steffen 1988). In der externen Qualitätssicherung werden diese Ziele als Anforderungen an die Leistungserbringer formuliert und konkret anhand der Qualitätsaspekte und Qualitätsmerkmale dargestellt (siehe Kapitel 3). Ein Indikator beschreibt folglich nur dann Qualität, wenn die Beeinflussung des von ihm gemessenen Qualitätsmerkmals auch innerhalb der Handlungsmöglichkeiten des Leistungserbringers liegt.

Bei der Abbildung von Qualitätsmerkmalen mittels Prozess- und Ergebnisindikatoren werden die leistungserbringerbezogenen Indikatorwerte nicht nur von Faktoren des Leistungserbringers selbst beeinflusst, sondern potenziell auch von patientenbedingten und sonstigen Faktoren sowie von Wechselwirkungen zwischen diesen Faktoren (siehe auch Abschnitt 13.1). Der Anteil des Einflusses, den die Leistungserbringer auf die Indikatorwerte haben, wird als Beeinflussbarkeit der Indikатораusrprägung bezeichnet und ist bei verschiedenen Indikatoren unterschiedlich groß. Wie Abbildung 10 zeigt, haben sowohl bei Ergebnisindikatoren (z. B. die Krankenhaussterblichkeit von Organempfängern im Rahmen der Transplantation) als auch bei einigen Prozessindikatoren (z. B. die Eingriffsdauer beim Austausch eines Herzschrittmachers) patientenbedingte Faktoren einen starken Einfluss auf den Indikatorwert. Für andere Prozess- und Ergebnisindikatoren (z. B. Eingehen auf Informationsbedürfnisse der Patientinnen und Patienten, Tod bei elektiver Operation ohne Vorerkrankungen) kann dagegen bei ausreichend spezifischer Formulierung des Indikators unterstellt werden, dass patientenbedingte und sonstige Einflüsse weitestgehend vernachlässigbar sind und dass der Indikatorwert im Wesentlichen auf das Handeln des Leistungserbringers zurückgeführt werden kann. Dazwischen bewegen sich Ergebnisindikatoren (z. B. Wundinfektionen bei Operationen) und Prozessindikatoren (z. B. Antibiotikaphylaxe), bei denen der Indikatorwert in stärkerem Maße sowohl von patientenbedingten oder sonstigen Faktoren als auch vom Handeln des Leistungserbringers beeinflusst ist.

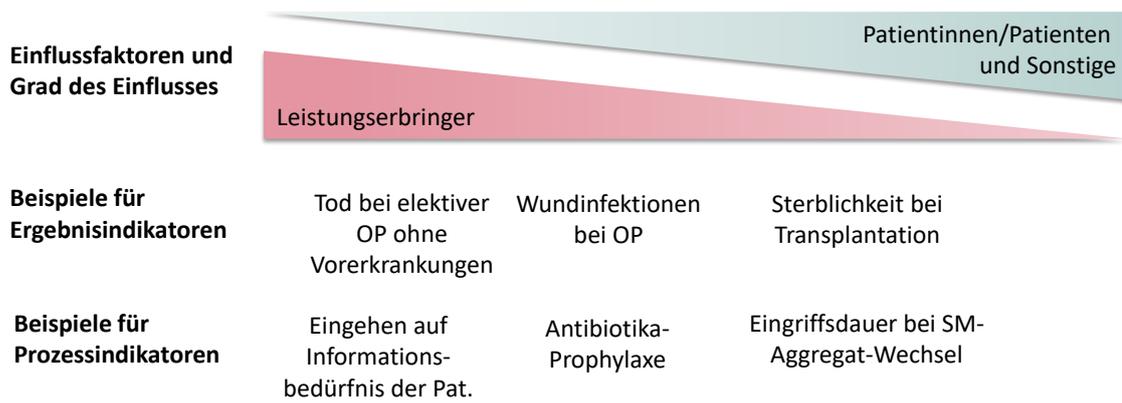


Abbildung 10: Schematische Darstellung des Ausmaßes der leistungserbringerbedingten und nicht leistungserbringerbedingten Einflüsse auf den Indikatorwert anhand von Beispielen fallbezogener Qualitätsindikatoren

Die Leitfrage dieses Eignungskriteriums ist also, inwieweit die Leistungserbringer einen Qualitätsaspekt bzw. ein Qualitätsmerkmal und damit den Indikatorwert der zugehörigen Qualitätsindikatoren in Richtung des Qualitätsziels beeinflussen können. Diese Beeinflussbarkeit lässt sich nicht quantifizieren, sondern wird auf Grundlage der Informationen aus Literatur, Fokusgruppen und aus der Beratung durch Expertinnen und Experten eingeschätzt. Dies erfolgt in der Verfahrensentwicklung zunächst bei der Entwicklung der Qualitätsaspekte und wird auf der konkreteren Ebene der Qualitätsmerkmale/Qualitätsindikatoren nochmals überprüft. Bei der Einschätzung des Ausmaßes der Beeinflussbarkeit wird auch berücksichtigt, ob eine angemessene Risikoadjustierung der Indikatorwerte vorgenommen wird. Das Ausmaß der Beeinflussbarkeit hat u. a. Auswirkung auf die Festlegung des Bewertungskonzepts eines Qualitätsindikators (siehe Kapitel 10). Konkrete Einflüsse von Rahmenbedingungen auf die Erfüllbarkeit von Qualitätsanforderungen werden ebenfalls bei der Festlegung des Bewertungskonzepts berücksichtigt und durch das Kriterium „Erreichbarkeit des Referenzbereichs“ adressiert (siehe dort).

#### 8.4 Eignungskriterien des Messverfahrens

Unter den Eignungskriterien des Messverfahrens werden Anforderungen an Qualitätsindikatoren zusammengefasst, die die messtheoretischen Eigenschaften und die Praktikabilität des Indikators beschreiben. Nur wenn diese Messung das Konzept des Indikators adäquat umsetzt, können anhand des Indikatorwerts sinnvoll Entscheidungen getroffen werden.

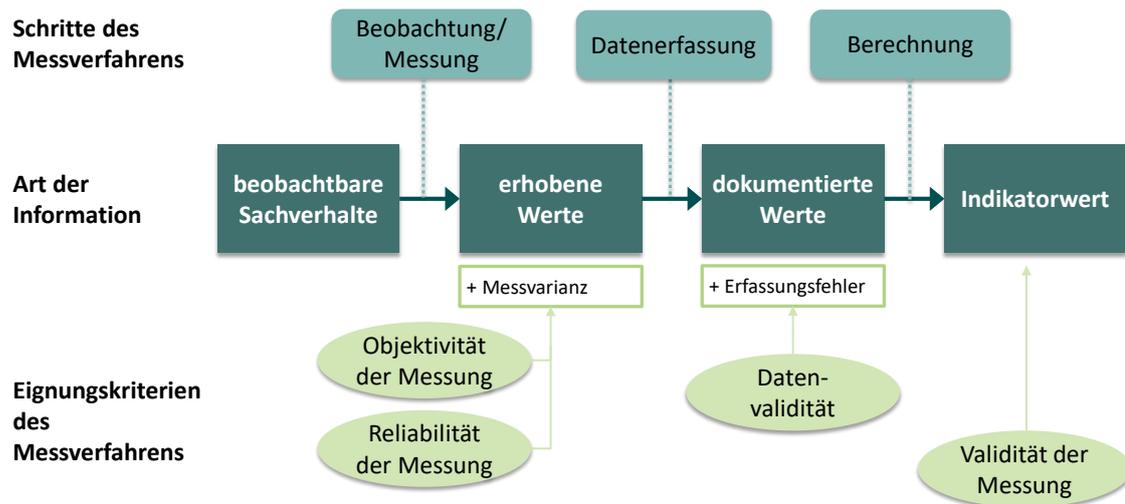


Abbildung 11: Schritte und zugehörige Eignungskriterien bei der Ermittlung und Aufarbeitung von Informationen für die Berechnung von Indikatorwerten

Bei der Qualitätsmessung über Prozess- oder Ergebnisindikatoren ist zu berücksichtigen, dass sich diese über zwei Ebenen erstreckt: Indikatorwerte zu Qualitätsmerkmalen der Versorgung auf Ebene der Leistungserbringer werden auf Grundlage von Daten auf Patienten- bzw. Fallebene berechnet. Die Gewinnung von Informationen für die Qualitätssicherung und Berechnung von Indikatorwerten kann dabei als Kette von drei aufeinanderfolgenden Schritten beschrieben werden (Abbildung 11). Der erste Schritt besteht in der Beobachtung eines Sachverhalts und Erhebung eines Wertes für diesen Sachverhalt auf Patientenebene, wie z. B. die Bestimmung des Blutdrucks oder die Befragung nach Schmerzen. Dieser Vorgang der Zuordnung von Zahlen zu einem Sachverhalt erfüllt bereits die Definition einer Messung (Schnell et al. 2013: 8 f.). Bei Daten, die primär für die Patientenversorgung erhoben werden, wird erst in einem zweiten Schritt der gemessene Wert für den Zweck der Qualitätssicherung dokumentiert und übermittelt. Dieser Vorgang wird als Datenerfassung bezeichnet und kann direkt, z. B. über die Dokumentation des Blutdruckwerts in der QS-Dokumentationssoftware, oder indirekt erfolgen, wie z. B. bei der ICD-Kodierung eines Bluthochdrucks in der Einrichtung zu Abrechnungszwecken und nachfolgender Übermittlung von den Krankenkassen an die Auswertungsstelle für die Qualitätssicherung.<sup>19</sup> Bei Patientenbefragungen für die Qualitätssicherung fallen Messung auf Patientenebene und Datenerfassung zusammen, d. h., beides erfolgt gleichzeitig beim Ausfüllen des Fragebogens. Der dritte Schritt schließlich besteht in der Berechnung des Indikatorwerts gemäß den Festlegungen zu den verwendeten Datenfeldern, Dateninhalten, Rechenregeln und Risiko-adjustierungsmethoden, kurz als Berechnungsspezifikation des Indikators bezeichnet.

Bei den Schritten des Messverfahrens einschließlich der Berechnung des Indikatorwerts bestehen mögliche Fehlerquellen, die im Fokus der hier beschriebenen Eignungskriterien stehen. Diese Eignungskriterien des Messverfahrens bauen aufeinander auf, d. h., nur wenn alle Kriterien angemessen erfüllt sind, ist der Indikatorwert zur Abbildung des Qualitätsmerkmals geeignet. Im Folgenden werden diese Kriterien näher erläutert.

<sup>19</sup> Genau genommen besteht hier also die Datenerfassung ihrerseits aus zwei Schritten.

### **Objektivität der Messung**

Die Objektivität einer Messung beschreibt das Ausmaß, in dem der erhobene (beobachtete) Wert für einen Sachverhalt unabhängig von der beobachtenden Person und von den Begleitumständen der Messung ist (Bühner 2011: 58). Bereits der Messvorgang auf Patienten- bzw. Fallebene kann beim gleichen Sachverhalt zu unterschiedlichen Messwerten führen. Beispielsweise kann die Messung des Blutdrucks einer Patientin oder eines Patienten unterschiedliche Werte ergeben, wenn verschiedene Größen der Messmanschette verwendet werden oder wenn der systolische Wert nicht durch Auskultation, sondern palpatorisch bestimmt wird. Auf ähnliche Weise kann ein Messwert dadurch variieren, dass die Messung durch unterschiedliche beobachtende Personen durchgeführt wird, z. B. bei der Größenbestimmung eines Organs mittels Ultraschalluntersuchung. Ähnliche Überlegungen sind bei der Datenerfassung auf Ebene der Leistungserbringer zu berücksichtigen: Diese entspricht für Prozess- und Ergebnisindikatoren einer Einrichtungsbefragung auf Fallebene, und die Durchführungsobjektivität kann z. B. durch Interpretationsspielraum bei Datenfeldern beeinträchtigt sein.

Sowohl für Qualitätsindikatoren, die auf der Dokumentation der Leistungserbringer basieren, als auch für Qualitätsindikatoren, die mittels Patientenbefragung erhoben werden, steigern bestimmte Maßnahmen der Erfassungs- bzw. Fragebogengestaltung deren Objektivität (siehe auch Abschnitt 4.4). Um die Objektivität der Daten für die Qualitätsmessung weitestgehend sicherzustellen, werden daher solche Sachverhalte zur Erfassung ausgewählt, für die eine weitgehend objektive Messung angenommen werden kann (siehe auch Abschnitt 4.4.2), und eindeutige Spezifikationen zum Erhebungsprozess einschließlich leicht verständlicher und eindeutiger Ausfüllhinweise erstellt.

### **Datenvalidität**

Dieses Kriterium bezieht sich auf Qualitätsmessungen, bei denen die Beobachtung der Sachverhalte für die Indikatorberechnung und die Datenerfassung für die Qualitätssicherung in getrennten Schritten erfolgen, wie z. B. bei Indikatoren, die mittels QS-Dokumentation durch die Leistungserbringer oder über Sozialdaten bei den Krankenkassen gemessen werden. Die Datenvalidität beschreibt dabei, ob die zur Berechnung des Indikatorwerts verwendeten Daten den Vorgaben des Indikators hinsichtlich Vollzähligkeit, Vollständigkeit und Dateninhalten entsprechen oder ob es z. B. durch Erfassungs- oder Übertragungsfehler zu Unterschieden zwischen erhobenem Wert und dokumentiertem Wert gekommen ist. Voraussetzung für hohe Datenvalidität sind damit eindeutige Spezifikationen der Instrumente zur Datenerfassung (QS-Filter, QS-Dokumentationssoftware, Sozialdatenspezifikation, Patientenfragebogen) und der zugehörigen Datenflüsse.

Formal kann die Datenvalidität als Teilaspekt der Durchführungsobjektivität auf Ebene der dokumentierenden Einrichtungen aufgefasst werden: Einschränkungen der Datenvalidität führen dazu, dass die Ergebnisse des Messverfahrens nicht unabhängig von den Durchführenden sind. Sie wird dennoch als separates Eignungskriterium aufgeführt, da sie in der Verantwortung der

Dokumentierenden liegt und erst in laufenden QS-Verfahren adäquat empirisch beurteilt werden kann. Dies erfolgt mittels Auffälligkeitskriterien und anhand eines Vergleichs der Dokumentation mit Daten aus der Patientenakte (siehe Kapitel 12).

### **Reliabilität der Messung**

Die Reliabilität beschreibt die Genauigkeit eines Messinstruments, verstanden als Anteil der Varianz in den Messwerten, der durch tatsächliche Unterschiede und nicht durch zufällige Messfehler bedingt ist (Adams et al. 2010, Bühner 2011: 60, Meyer 2004: 28, Schermelleh-Engel und Werner 2012: 120). Mathematisch wird Reliabilität als ein Varianzverhältnis definiert, das ausdrückt, wie viel der Varianz in den beobachteten Messwerten auf Varianz in den wahren Werten zurückgeht, wobei der wahre Wert als der Erwartungswert unendlich vieler Messungen des gleichen Objekts unter den gleichen Umständen definiert ist (Raykov und Marcoulides 2011: 138 bzw. 118).

Soll Reliabilität für die Messung eines Sachverhalts auf Patienten- bzw. Fallebene adressiert werden, muss geprüft werden, ob diese Trennung von wahren und beobachtetem Wert für eine Messung plausibel ist. Dies betrifft vor allem die Messung direkt beobachtbarer Merkmale einzelner Patientinnen und Patienten, wie z. B., ob ein Patient verstorben ist. Es erscheint nicht sinnvoll anzunehmen, dass ein Patient über einen wahren Wert (tot/lebendig) verfügt und parallel dazu ein (z. B. durch die Ärztin) beobachteter Wert vorliegt, dass es also zu Messfehlern kommt.<sup>20</sup> Bei der sonografischen Bestimmung der Organgröße dagegen ist die Reliabilität z. B. dann eingeschränkt, wenn dieselbe untersuchende Person bei wiederholten Messungen an derselben Patientin oder demselben Patienten unterschiedliche Messwerte bestimmt (Test-Retest-Reliabilität). Die Reliabilität ist also ein Maß für die Präzision der Messung bei sonst gleichen Untersuchungsumständen. Das Ausmaß, in dem Qualitätsindikatoren anfällig für reliabilitätsmindernde Einflüsse sind, ist dabei unter anderem abhängig von den zu messenden Sachverhalten und der Datenquelle.

Aufgrund der hierarchischen Struktur der Qualitätsmessung kann die Reliabilität der Messung auch auf Ebene der Leistungserbringer adressiert werden. In Analogie zur obigen Definition kann der wahre Wert eines Leistungserbringers definiert werden als der Erwartungswert unendlich vieler Messungen von Patientinnen und Patienten dieses Leistungserbringers. Da diese wahren oder zugrunde liegenden Indikatorwerte nicht direkt beobachtbar sind (siehe auch Abschnitt 10.4), kann die Reliabilität, auch auf Leistungserbringerebene, nur geschätzt werden. Entsprechende Verfahren sind Gegenstand zukünftiger Entwicklungen.

### **Validität der Messung**

Indikatoren machen theoretische Konzepte durch beobachtbare Sachverhalte messbar (Döring und Bortz 2016: 223 und 228, Meyer 2004: 28 und 51, Schnell et al. 2013: 117 und 221). Damit

---

<sup>20</sup> Nicht zu verwechseln ist der wahre Wert mit Werten aus anderen Datenquellen, wie beispielsweise der Patientenakte. Stimmt etwa der für die Qualitätssicherung dokumentierte Wert nicht mit dem Wert aus einer Patientenakte überein, ist das auf der Messebene einzelner Patientinnen und Patienten kein Reliabilitätsproblem im Sinne eines zufälligen Messfehlers, sondern ein Dokumentationsproblem, das durch das Eignungskriterium Datenvalidität adressiert wird.

die Ergebnisse dieser Messung verwertbar sind, müssen Rückschlüsse von den Messergebnissen auf den theoretischen Sachverhalt gerechtfertigt sein. Validität bezeichnet dabei das Ausmaß, in dem die Angemessenheit von Interpretationen und Maßnahmen auf Basis von Messergebnissen empirisch und theoretisch gestützt ist (Messick 1994). Validität kann unter verschiedenen Gesichtspunkten beurteilt werden und sich z. B. auf die Bewertung eines Ergebnisses, seine Verallgemeinerung oder das Fällen von Entscheidungen beziehen (Kane 2001). Das Konzept der Eignungsprüfung von Qualitätsindikatoren trennt die Beurteilung dieser unterschiedlichen Facetten des Validitätsbegriffs mittels der verschiedenen Eignungskriterien, um sie transparenter zu machen.

Das Eignungskriterium „Validität der Messung“ wird daher in Anlehnung an Hartig et al. (2012: 144), Reiter et al. (2007), Schnell et al. (2013: 144) in einer eng gefassten Bedeutung als das Ausmaß definiert, in dem ein Qualitätsindikator das Qualitätsmerkmal, das er abbilden soll, tatsächlich erfasst. Validität meint also insbesondere kein Gesamturteil über die Eignung eines Qualitätsindikators. Das Kriterium entspricht dem ursprünglich aus der Psychometrie stammenden Konzept der Inhaltsvalidität und bezieht sich dort auf latente, nicht direkt beobachtbare Sachverhalte wie etwa Intelligenz oder Persönlichkeitsmerkmale. Auf Qualitätsindikatoren, die direkt messbare Sachverhalte erfassen, wie etwa Mortalität, Komplikationsraten oder Ähnliches, lässt sich dieses Konzept insoweit übertragen, als auch diese direkt messbaren Größen letztlich der Abbildung eines theoretischen Konzepts dienen: Beispielsweise ist der Indikator „Krankenhaussterblichkeit bei Behandlung ambulant erworbener Pneumonien“ nur zu dem Ausmaß valide, zu dem korrekt festgelegt wurde, welche Daten und Rechenregeln die Konzepte „Sterblichkeit“, „Krankenhausaufenthalt“, „Pneumonie“ und „ambulant erworben“ im Sinne des interessierenden Qualitätsmerkmals beschreiben. Die Validität der Messung solcher Indikatoren wird im Wesentlichen dadurch bestimmt, dass die Berechnungsvorschrift für den Indikatorwert angemessen und korrekt spezifiziert ist.

Die Korrektheit der Berechnungsspezifikation wird bei der Neuentwicklung von Qualitätsindikatoren und bei der Überprüfung von Qualitätsindikatoren im Regelbetrieb in erster Linie mittels sachlogischer Überlegungen eingeschätzt. Dabei werden Publikationen, externe Expertinnen und Experten sowie eigene Analysen und Erfahrung zu den verfügbaren Datenquellen einbezogen, um die am besten geeigneten Daten für die Berechnung zu ermitteln und Besonderheiten bestimmter Daten (z. B. wegen Kodierung zu Abrechnungszwecken) zu berücksichtigen. Für die Validität von Messungen mittels Patientenbefragungen wird ergänzend auf den Abschnitt 4.4.4 verwiesen.

### **Praktikabilität der Messung**

Die Praktikabilität der Messung beschreibt den Aufwand, der für die Erfassung der Informationen zur Indikatorberechnung benötigt wird. Dazu zählen nicht nur laufende Aufwände bei der Durchführung eines QS-Verfahrens, sondern auch einmalige Aufwände, die bei der erstmaligen Einrichtung der Messung für einen Indikator anfallen, z. B. für die Spezifikation neuer Datenfelder, für die Einrichtung neuer Datenflüsse oder für die Information der an der Datenerfassung Beteiligten. In die Beurteilung der Praktikabilität fließt auch die Verfügbarkeit der Informationen für einen Indikator ein. Die Praktikabilität der Messung ist damit keine Messeigenschaft eines

Qualitätsindikator im eigentlichen Sinn, ist aber ein wichtiger Faktor bei der Beurteilung, ob das Messverfahren dieses Indikators für die Anwendung in einem QS-Verfahren geeignet ist. Hierzu muss ein angemessenes Verhältnis von Aufwand und Nutzen hinsichtlich der zu erwartenden Qualitätsverbesserungen bestehen.

Bei der Einschätzung der Praktikabilität der Messung werden beide Schritte der Informationsbeschaffung berücksichtigt, d. h. sowohl die Messung der Merkmale im eigentlichen Sinne als auch die Datenerfassung und -übermittlung (siehe oben und Abbildung 11). Je nach Datenquelle bzw. Erfassungsinstrument ist der Aufwand für die Informationsbeschaffung unterschiedlich auf diese Schritte verteilt. Beispielsweise wird in der Qualitätssicherung mit fallbezogenen QS-Dokumentationsdaten oder Sozialdaten bei den Krankenkassen auf Befunde zurückgegriffen, die bereits im Rahmen der Patientenbehandlung erhoben und in der Patientenakte dokumentiert wurden. Der Aufwand entsteht bei den Datensätzen der QS-Dokumentation also durch die Datenerfassung und -übermittlung seitens der Leistungserbringer, nicht durch die Messung der Merkmale. Die Nutzung von Abrechnungsdaten erfordert demgegenüber keine zusätzliche Datenerfassung. Hier erfolgt die Datenübermittlung aus den bereits vorliegenden Datenbeständen mit sehr viel geringerem Aufwand durch die Krankenkassen. Bei Patientenbefragungen wiederum werden Informationen erfragt, die sonst nicht erhoben werden. Hier sind Messung des Merkmals und Datenerfassung unmittelbar miteinander verbunden. Der Aufwand für die Informationsbeschaffung liegt also vor allem bei der die Befragung durchführenden Stelle und bei den Patientinnen und Patienten.

## **8.5 Eignungskriterien des Bewertungskonzepts**

Damit Qualitätsindikatoren Handlungen zur Qualitätsverbesserung anstoßen können, muss die Bewertung der Indikatorwerte in geeigneter Weise zu Handlungsempfehlungen führen, d. h. die Aufgabe von Qualitätsindikatoren besteht in einer entsprechenden Klassifikation des Indikatorwerts (siehe Kapitel 10). Die Eignung des Bewertungskonzepts eines Indikators im Rahmen eines QS-Verfahrens lässt sich zusammenfassend anhand der nachfolgend beschriebenen vier Kriterien beurteilen.

### **Zweckeignung des Bewertungskonzepts**

Leitfrage für dieses Kriterium ist, ob das Bewertungskonzept des Indikators so gewählt wurde, dass die daraus abzuleitenden Handlungsempfehlungen dem Zweck des QS-Verfahrens (in den Kategorien Qualitätsförderung, Information, Regulierung) dienen. Dabei stellen unterschiedliche Kontexte unterschiedliche Anforderungen an die Bewertungskonzepte. Beispielsweise kann es sinnvoll sein, den Indikatorwert eines Leistungserbringers in Verfahren mit dem Zweck Qualitätsförderung anders zu bewerten als hinsichtlich der Frage, ob auf Grundlage des Indikatorwerts ein qualitätsbezogener Vergütungszuschlag gezahlt werden soll (regulierender Verfahrenszweck). Diesen sehr unterschiedlichen Anforderungen der verschiedenen Verfahrenszwecke wird dadurch Rechnung getragen, dass jeweils speziell für den Zweck angepasste Bewertungskonzepte entwickelt werden (siehe Kapitel 10).

### **Erreichbarkeit des Referenzbereichs**

Neben der grundsätzlichen Beeinflussbarkeit des Indikatorwerts durch die Leistungserbringer (siehe Abschnitt 8.3) ist es wichtig, dass auch die konkrete Zielvorgabe des Indikators, die durch den Referenzbereich zum Ausdruck kommt, durch die Leistungserbringer prinzipiell erreichbar ist. Anderenfalls würde sich die Klassifikation mittels des Indikators erübrigen, da schon vor der Messung feststünde, dass die Ergebnisse nicht im Referenzbereich liegen. Beispielsweise ist ein Bewertungskonzept für einen Indikator „In-Hospital-Sterblichkeit bei Lebertransplantation“ mit einer (unrealistisch niedrigen) oberen Referenzbereichsgrenze von 1 % für den Zweck der Qualitätsförderung nicht geeignet, da dieser Wert in der Regel nicht erreicht werden kann. Somit könnte der Indikator nicht zwischen Leistungserbringern mit erwartbarer Qualität und solchen mit mangelnder Qualität differenzieren, und wäre also für diese Klassifikation wegen fehlenden Informationsgehalts ungeeignet.

Je nach Zielsetzung und Kontext des QS-Verfahrens werden Indikatorwerte auch für andere Zwecke als den Vergleich mit dem Referenzbereich genutzt, z. B. für den Vergleich der Leistungserbringer untereinander. In diesen Fällen ist die Erreichbarkeit des Referenzbereichs nicht von Belang.

### **Klassifikationsgüte**

Für Prozess- und Ergebnisindikatoren, bei denen die Fallzahl des Leistungserbringers bei der Bewertung des gemessenen Indikatorwerts zu berücksichtigen ist (siehe Abschnitt 13.1), wird die Eignung des Bewertungskonzepts auch von der Klassifikationsgüte bestimmt. Die Klassifikationsgüte gibt an, wie gut der Indikator unter den gegebenen Rahmenbedingungen zwischen dem zugrunde liegenden Indikatorwert und einem Vergleichswert unterscheidet, wenn der gemessene Indikatorwert nur eine Schätzung für den zugrunde liegenden Indikatorwert darstellt (siehe Abschnitt 10.4). Dies entspricht dem Konzept der Teststärkeanalyse (vgl. Döring und Bortz 2016: 807 ff.). Je nach Zweck des QS-Verfahrens kann sich die Klassifikationsgüte auf unterschiedliche Vergleichskategorien beziehen: z. B. auf den Referenzbereich des Indikators (für die Qualitätsförderung), auf den Indikatorwert eines anderen Leistungserbringers (für den Vergleich von Leistungserbringern) oder auf den Mittelwert aller Leistungserbringer. Die Klassifikationsgüte kann anhand der Parameter Sensitivität und Spezifität beschrieben werden, sofern empirische Daten aus der Anwendung des Indikators vorliegen.<sup>21</sup>

### **Angemessenheit der Risikoadjustierung**

Damit die Klassifikation der Indikatorwerte verschiedener Leistungserbringer (z. B. der Vergleich mit einem Referenzbereich) sowie der Vergleich der Indikatorwerte zwischen verschiedenen

---

<sup>21</sup> Dies ist von anderen Verständnissen von Klassifikationsgüte, Sensitivität und Spezifität abzugrenzen, die diese Begriffe für den Vergleich von Indikatorergebnissen mit einem Außenkriterium nutzen (z. B. bei QUALIFY, Reiter et al. 2007) und damit nicht Teststärke, sondern Kriteriumsvalidität beschreiben. Ein solches Außenkriterium ist aber nicht verfügbar, da ein Goldstandard zur Beschreibung von Qualität nicht vorliegt (Kerr et al. 2007, ebenfalls Reiter et al. 2007). Das Konzept des IQTIG zur Eignungsbeurteilung sieht daher keine Globalbeurteilung der Angemessenheit der indikatorgestützten Klassifikation vor, sondern unterscheidet stattdessen zwischen den verschiedenen Gesichtspunkten, die Einfluss auf die Klassifikation nehmen, anhand der Eignungskriterien.

Leistungserbringern fair ist, müssen die Indikatorwerte um die Einflüsse von patientenseitigen, nicht durch die Leistungserbringer beeinflussbaren Faktoren bereinigt werden. Dies geschieht mittels Methoden der Risikoadjustierung (siehe Abschnitt 13.2). Das Kriterium beschreibt somit das Ausmaß, in dem patientenseitige Einflussfaktoren auf die Indikatorwerte durch Risikoadjustierung kontrolliert sind.

## 8.6 Bezug zum Verfahrenszweck

In die vierte Kategorie von Eignungskriterien für Qualitätsindikatoren fallen Anforderungen, die die Eignung des Indikators als Ganzes für den Zweck des QS-Verfahrens beschreiben und nicht nur auf eine einzelne Komponente des Indikators (Qualitätsziel, Messverfahren, Bewertungskonzept) bezogen sind.

### Brauchbarkeit

Dieses Kriterium fasst eine Reihe von Aspekten zusammen, die die Eignung der Indikatoren für die praktische Anwendung im QS-Verfahren beschreiben. Leitfrage ist, ob die Messung mit dem Qualitätsindikator dem Zweck des Verfahrens dienlich ist, vorausgesetzt, dass Qualitätsziel, Messverfahren und Bewertungskonzept geeignet sind. Beispielsweise könnten zwei Indikatoren A und B die Krankenhaussterblichkeit bei einer Erkrankung gleichermaßen zuverlässig und valide messen, aber die Daten aus unterschiedlichen Datenflüssen nutzen. Falls die Werte für Indikator A erst ein Jahr später als für Indikator B zur Verfügung stehen, wäre seine Brauchbarkeit für die Qualitätsförderung geringer als die von Indikator B.

Die Brauchbarkeit ist teilweise auf Ebene eines einzelnen Indikators und teilweise auch in Zusammenschau der Indikatoren eines QS-Verfahrens (d. h. für ein Indikatorenset) oder mit Blick auf den Prozess der Qualitätsmessung zu beurteilen. Folgende Aspekte sind für die Brauchbarkeit der Indikatoren von Bedeutung:

- **Zeitnahe Nutzbarkeit:** Der Prozess der Qualitätsmessung soll daraufhin beurteilt werden, ob die Indikatorwerte als Grundlage für Handlungsempfehlungen auch zeitnah zur Verfügung stehen.
- **Interpretierbarkeit:** Es soll beurteilt werden, ob die Aussage des Indikators für die Leistungserbringer verständlich ist, damit diese daraus Maßnahmen für ihr internes Qualitätsmanagement ableiten können (Verfahrenszweck Qualitätsförderung).
- **Nützlichkeit:** Es soll beurteilt werden, ob die berichteten Indikatorwerte eine Information zur Verfügung stellen, die für die Patientinnen und Patienten bzw. die Öffentlichkeit nützlich ist (Verfahrenszweck Patienteninformation).
- **Ausgewogenheit:** Soll ein Set von Qualitätsindikatoren dazu genutzt werden, die Versorgungsqualität in einem Versorgungsbereich (z. B. „Geburtshilfe“) repräsentativ abzubilden, so schließt die Prüfung der Brauchbarkeit auch eine Beurteilung ein, ob das Set an Qualitätsindikatoren in der Zusammenschau ausgewogen ist.

### **Potenzielle Nebenwirkungen**

Bei der Neuentwicklung von Qualitätsindikatoren und im Rahmen der Verfahrenspflege wird geprüft, ob neben den erwünschten Wirkungen durch die Verwendung des Indikators, d. h. Qualitätsverbesserungen, auch unerwünschte Wirkungen auftreten können. Eine solche unerwünschte Wirkung kann z. B. der Fehlanreiz sein, Patientinnen und Patienten, deren Behandlung mit einem erhöhten Risiko von Komplikationen und ggf. von auffälligen Indikatorwerten einhergeht, nicht zu behandeln (Patientenselektion). Die Prüfung dieses Kriteriums dient dazu, geeignete Maßnahmen zum Ausgleich solcher Nebenwirkungen zu ergreifen. Dies kann z. B. durch die Verwendung von weiteren Indikatoren geschehen, deren Anreiz entgegengesetzt wirkt, oder durch eine Optimierung der Risikoadjustierung. Als Grundlage für die Einschätzung nach diesem Kriterium dienen in der Verfahrensentwicklung die Informationen aus der Literatur und aus der Beratung durch Experten. In laufenden QS-Verfahren kommen Rückmeldungen zum Verfahren sowie ggf. eigene Datenauswertungen als weitere Informationsquelle hinzu.

### **Zuschreibbarkeit der Verantwortung**

Am Zustandekommen eines Indikatorwerts können mehrere Leistungserbringer beteiligt sein, insbesondere wenn an der Patientenversorgung mehrere Fachdisziplinen oder Berufsgruppen beteiligt sind oder wenn lange Beobachtungszeiträume vorliegen (Langzeit-Follow-up). Beispielsweise kann ein Ergebnisindikator „Häufigkeit von Abstoßungsreaktionen einer Transplantatnieren“ gemessen werden, dessen Indikatorwert sowohl von der Qualität des Krankenhauses (Nierentransplantationszentrum) als auch der Qualität der vertragsärztlichen Betreuung (z. B. niedergelassene Nephrologin oder niedergelassener Nephrologe) abhängt. Für jeden Indikator wird untersucht, welchen Anteil am Zustandekommen eines Indikatorwerts jeder der beteiligten Leistungserbringer hat. Dabei kann der jeweilige Anteil nicht während der Indikatorenentwicklung quantifiziert werden, sondern erfordert eine qualitative Abschätzung auf Grundlage der Informationen zum Versorgungsgeschehen, insbesondere aus Literatur und Beratung durch Expertinnen und Experten. Damit die Verantwortung für einen Indikatorwert einem Leistungserbringer zugewiesen werden kann, ist also nicht nur die Beeinflussbarkeit des Indikatorwerts Voraussetzung (siehe Abschnitt 8.3), sondern muss die Beeinflussung des Indikatorwerts im Wesentlichen diesem Leistungserbringer zugeschrieben werden können. Dieses Kriterium wird als Zuschreibbarkeit der Verantwortung bezeichnet. Es ist dann wichtig, wenn in einem QS-Verfahren die auf die Qualitätsmessung folgenden Maßnahmen auf einzelne konkret benennbare Leistungserbringer ausgerichtet sind. Eine mangelnde Zuschreibbarkeit der Verantwortung kann dazu führen, dass ein Indikator für solche QS-Verfahren nicht verwendet werden kann. Qualitätsindikatoren mit auf mehrere Leistungserbringer verteilter Verantwortung für die Versorgung können aber perspektivisch ggf. für andere Zwecke verwendet werden, etwa wenn eine Gruppe von beteiligten Leistungserbringern gemeinschaftlicher Adressat von Qualitätssicherungsmaßnahmen sein soll, so wie es z. B. in den USA bei den Accountable Care Organizations erfolgt, oder wenn Versorgungsqualität auf regionaler Ebene dargestellt und gefördert werden soll.

## 9 Veröffentlichung von Indikatorergebnissen

Eine der Kernaufgaben des Instituts ist die Schaffung von Transparenz über die Qualität der Gesundheitsversorgung. Dahinter steht der Wille des Gesetzgebers, mehr als bisher Qualitätsinformationen für die Öffentlichkeit zugänglich zu machen, d. h., mehr Transparenz für die Patientinnen und Patienten herzustellen (siehe Begründung zum Gesetz zur Weiterentwicklung der Finanzstruktur und der Qualität in der gesetzlichen Krankenversicherung, GKV-FQW, BR-Drucksache 151/14 vom 11.04.2014: : 25). Die Ausführungen in diesem Kapitel beziehen sich dabei auf die Schaffung von Transparenz durch Veröffentlichung der Indikatorergebnisse der einzelnen Leistungserbringer. Im Folgenden wird hierfür auch der international gebräuchliche Begriff „Public Reporting“ verwendet.

In den Verfahren der Qualitätssicherung nach §§ 136 ff. SGB V erfolgt eine Veröffentlichung von Indikatorergebnissen in den Qualitätsberichten der Krankenhäuser nach den Regelungen des G-BA (Qb-R)<sup>22</sup>. Zu diesem Zweck spricht das Institut jährlich eine Empfehlung an den G-BA aus, für welche Qualitätsindikatoren die Indikatorwerte, die Referenzbereiche und die Ergebnisse aus dem Strukturierten Dialog veröffentlicht werden sollen. Die Grundsätze für die Empfehlungen zur Veröffentlichung sind im Folgenden beschrieben.

### Ziele der Transparenz von Qualitätsinformationen

Die Transparenz von Qualitätsinformationen gilt als ein gesellschaftlicher Wert an sich, da die Leistungserbringer im Gesundheitswesen damit einer Rechenschaftspflicht gegenüber den Patientinnen und Patienten sowie Beitragszahlerinnen und -zahlern nachkommen (Murphy und Scheu 2010, Schwenk und Schmidt-Kaehler 2016) und deutlich machen, dass sie eine qualitativ hochwertige und effiziente Versorgung anstreben. In diesem Sinne sieht das Institut eine Veröffentlichung von leistungserbringerbezogenen Indikatorergebnissen bereits deswegen als notwendig an, weil diese der Verpflichtung gegenüber der Allgemeinheit nachkommt, Transparenz über die Qualität der Leistungserbringer und über den Prozess der Qualitätssicherung zu schaffen, und weil sie eine Voraussetzung für die emanzipierte Teilhabe durch Patientinnen und Patienten an Informationen über die Gesundheitsversorgung bildet.

Die durch Public Reporting geschaffene Transparenz bildet zudem den Ausgangspunkt für weitere intendierte Wirkungen auf die Gesundheitsversorgung. Dies sind die Ermöglichung von Auswahlentscheidungen zwischen Leistungserbringern durch Patientinnen und Patienten und die sie beratenden Personen sowie die Motivation der Leistungserbringer zur Qualitätsverbesserung (Geraedts et al. 2009, Schwenk und Schmidt-Kaehler 2016, Totten et al. 2012). Das Institut strebt an, für einen solchen eigenständigen Verfahrenszweck „Patienteninformation“ zukünftig explizit bedarfs- und nutzerorientierte Qualitätsinformationen mit einer spezifischen Methodik zu entwickeln.

---

<sup>22</sup> Regelungen des Gemeinsamen Bundesausschusses gemäß § 137 Abs. 3 Satz 1 Nr. 4 SGB V über Inhalt, Umfang und Datenformat eines strukturierten Qualitätsberichts für nach § 108 SGB V zugelassene Krankenhäuser (Regelungen zum Qualitätsbericht der Krankenhäuser, Qb-R).

### **Kriterien zur Veröffentlichung**

Systematische Literaturübersichten zeigen, dass es nur wenige Studien zu Wirkungen von Public Reporting gibt und diese eine unterschiedliche Qualität aufweisen (Fung et al. 2008, Totten et al. 2012). Unerwünschte Wirkungen von Public Reporting wurden in den Studien bisher aber nicht schlüssig nachgewiesen (Totten et al. 2012). In der Literatur werden unerwünschte Wirkungen besonders dann diskutiert, wenn die Indikatoren methodische Schwächen aufweisen (z. B. Ennker et al. 2009, Lilford und Pronovost 2010). Für die Herstellung von Transparenz im oben genannten Sinne ist aus Sicht des IQTIG lediglich Voraussetzung, dass die veröffentlichten Informationen tatsächlich Qualität abbilden, dass also eine hinreichende Sicherheit besteht, dass Qualitätsziel, Messung und Bewertungskonzept der Qualitätsindikatoren für die Qualitätsmessung geeignet sind (siehe Kapitel 8). Die Verständlichkeit der Qualitätsinformationen ist dagegen für die Auswahl zur Veröffentlichung zunächst nicht maßgeblich. Sie ist nur durch die potenziellen Informationsempfänger zu beurteilen und kann zudem für verschiedene Gruppen unterschiedlich sein (Schaeffer 2006). Allgemeine Bewertungen der Verständlichkeit der Indikatorwerte für Patientinnen und Patienten bzw. für die allgemeine Öffentlichkeit mit dem Zweck, als Kriterium für die Auswahl der zu veröffentlichenden Indikatoregebnisse zu dienen, werden daher nicht vorgenommen. Alle veröffentlichten Informationen sollen stattdessen möglichst laienverständlich dargestellt werden. Eine ergänzende Abschätzung der Verständlichkeit und Nutzerorientierung wird aber zukünftig angestrebt, wenn Qualitätsaussagen eigens für den Zweck gezielter Patienteninformation entwickelt oder zusammengestellt werden sollen.

Das Institut empfiehlt daher, grundsätzlich alle Qualitätsindikatoregebnisse leistungserbringerbezogen zu veröffentlichen, die sich für die routinemäßige Anwendung in einem QS-Verfahren qualifiziert haben, sofern nicht bei sehr geringen Fallzahlen der Datenschutz einzelner Patientinnen oder Patienten dem entgegensteht. Werden Qualitätsindikatoren in einem gesetzlich verpflichtenden QS-Verfahren genutzt, so ist damit potenziell ein erheblicher und unmittelbarer Einfluss auf die Versorgung verbunden. Für diese Indikatoren muss davon ausgegangen werden, dass sie Qualität angemessen darstellen. Indikatoren, bezüglich deren Eignung für die externe Qualitätssicherung nennenswerte Bedenken bestehen, werden nicht zur externen Qualitätssicherung und Veröffentlichung empfohlen.

### **Veröffentlichung der Ergebnisse neu entwickelter Indikatoren**

Für Qualitätsindikatoren, die neu entwickelt werden, wird die Eignung für das jeweilige QS-Verfahren durch die Berücksichtigung der Eignungskriterien (siehe Kapitel 8) konzeptuell sichergestellt. Teilweise können die Eignungskriterien jedoch erst anhand der Erfahrung aus den ersten Jahren des Routinebetriebs abschließend beurteilt werden. Das Gleiche trifft für Indikatoren zu, die im Rahmen der Verfahrenspflege oder -weiterentwicklung derart weitgehenden Änderungen unterzogen wurden, dass Auswirkungen auf die Validität des Indikators nicht ausgeschlossen werden können. In beiden Fällen gelten diese Indikatoren als „neu entwickelt“ und werden vorerst nicht zur Veröffentlichung empfohlen. Bei der kontinuierlichen Beurteilung der Indikatoren im Rahmen der Verfahrenspflege (siehe Abschnitt 5.2) wird jährlich individuell geprüft, ob die Eignungskriterien hinreichend sicher als erfüllt gelten und die Indikatoren damit auch zur Veröffentlichung empfohlen werden können.

**Veröffentlichung begleitender Informationen**

Es gibt Qualitätsindikatoren, die für die externe Qualitätssicherung zwar insgesamt geeignet sind, aber Schwächen in der Erfüllung einzelner Eignungskriterien aufweisen. Für diese Indikatoren werden nach Beratung mit den jeweiligen fachlichen Expertinnen und Experten ergänzende Texte veröffentlicht. Diese erläutern methodische Limitationen und fachliche Besonderheiten der Indikatoren. Dadurch soll eine sachgerechte und kritische Interpretation der veröffentlichten Indikatorergebnisse durch die Öffentlichkeit ermöglicht werden. Um dieses Ziel unabhängig vom Publikationsmedium für die Indikatorergebnisse zu erreichen, ist es aus Sicht des Instituts geboten, dass Veröffentlichungen von Indikatorergebnissen eine begleitende Darstellung dieser Erläuterungen verpflichtend enthalten.

## 10 Bewertungskonzepte und Referenzbereiche

Die Messung der Qualität der Gesundheitsversorgung mittels Qualitätsindikatoren hat das Ziel, die Versorgungsqualität zu verbessern. Um auf Grundlage der Ergebnisse der Qualitätsmessungen zu Handlungsempfehlungen für die Leistungserbringer oder für die an der Qualitätssicherung beteiligten Institutionen des Gesundheitswesens zu kommen, müssen die Qualitätsindikatoren das richtige Qualitätsziel verfolgen und angemessen operationalisiert sein (siehe Kapitel 8), insbesondere eine geeignete Messskala aufweisen. Nach erfolgter Messung müssen die Messergebnisse zunächst nach einer objektiven Regel kategorisiert werden (z. B. als „auffällig“ vs. „unauffällig“). Anhand der Kategorisierung werden dann die Empfehlungen für weitere Schritte, wie z. B. den Strukturierten Dialog oder eine Vergabe qualitätsbezogener Zu- oder Abschläge, abgeleitet. Teil dieser folgenden Schritte kann auch eine zusätzliche einzelfallbezogene Analyse der Behandlungsfälle sein, die zum Messergebnis geführt haben (siehe Kapitel 13). Diese Analyse stellt zum einen eine ergänzende Sichtweise auf die Behandlungsqualität dar, zum anderen gibt sie Hinweise auf wichtige Messeigenschaften der Qualitätsindikatoren, wie z. B. Datenvalidität oder Angemessenheit der Risikoadjustierung.

Um zu entscheiden, wann welche Handlungsempfehlungen abzuleiten sind, werden standardisierte Bewertungskonzepte und Referenzbereiche benötigt. In diesem Kapitel wird das Vorgehen bei der Entwicklung von Bewertungskonzepten für Messergebnisse von Qualitätsindikatoren dargestellt.

### 10.1 Begriffsbestimmung und Funktion von Bewertungskonzepten

Qualitätsindikatoren beschreiben die Erfüllung von Qualitätsanforderungen auf Ebene der Leistungserbringer anhand eines Zahlenwerts, des Indikatorwerts. Unter dem Bewertungskonzept eines Qualitätsindikators wird die Methode verstanden, mit der dieser Indikatorwert eines Leistungserbringers im Hinblick auf zweckbezogene Handlungsoptionen eingestuft wird. Die Funktion des Bewertungskonzepts eines Qualitätsindikators besteht somit in der algorithmischen Zuordnung des Indikatorwerts zu einer Kategorie, es handelt sich also um ein Klassifikationsschema (vgl. Meyer 2007: 199). Beispielsweise können gemessene Indikatorwerte als „rechnerisch auffällig“ oder „rechnerisch unauffällig“ klassifiziert werden, sodass daraus abgeleitet wird, ob ein Strukturiertes Dialog mit dem jeweiligen Leistungserbringer aufzunehmen ist oder nicht. Das Bewertungskonzept muss drei Komponenten, nämlich mögliche Kategorien, Vergleichswerte und eine Zuordnungsregel, beinhalten, damit die Klassifikation eines Indikatorwerts vorgenommen werden kann (siehe Abbildung 12).

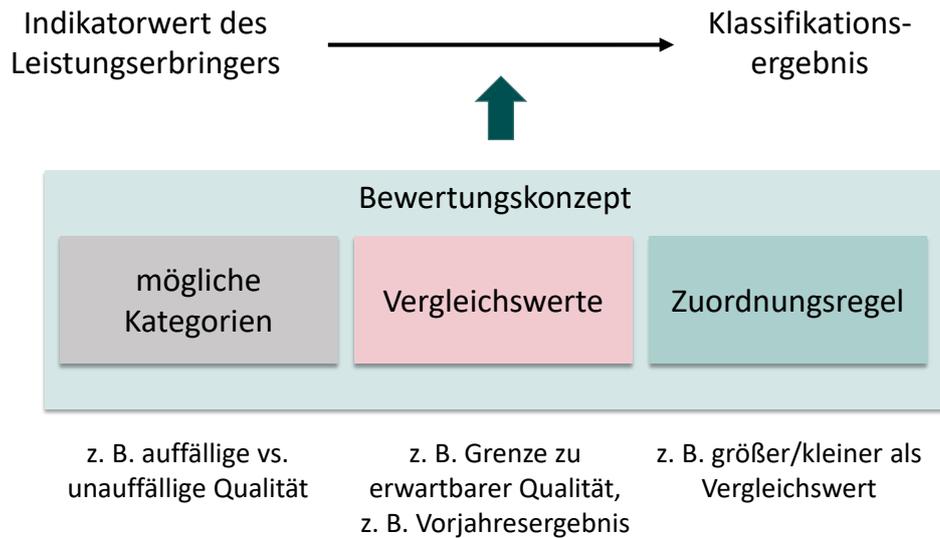


Abbildung 12: Funktion und Bestandteile des Bewertungskonzepts von Qualitätsindikatoren

Diese drei Komponenten werden im Folgenden einzeln erläutert.

### Mögliche Kategorien

Aus einer umschriebenen Menge möglicher Kategorien wird dem Indikatorwert eines Leistungserbringers genau eine Kategorie als Ergebnis der Klassifikation zugeordnet. Bisher werden im Rahmen der Qualitätssicherung zwei Kategorien unterschieden: „(rechnerisch) auffällige“ oder „(rechnerisch) unauffällige“<sup>23</sup> Qualität als Grundlage für die Entscheidung, ob ein Strukturierter Dialog bzw. ein Stellungnahmeverfahren als Qualitätssicherungsmaßnahme durchgeführt werden soll oder nicht. Die Konzeptentwicklung für andere Kategorien (z. B. „außerordentlich gute“ und „unzureichende“ Qualität), wie sie etwa im Rahmen von qualitätsbezogenen Zu- oder Abschlägen oder bei planungsrelevanten Qualitätsindikatoren denkbar sind, ist derzeit noch nicht abgeschlossen (siehe Abschnitt 2.4). Obwohl das grundlegende Prinzip auf solche alternativen Kategorien anwendbar ist, wird daher im Folgenden nicht näher auf diese eingegangen.

Das Entscheidungskriterium für die Zuordnung eines Indikatorwerts in eine Kategorie besteht aus einem oder mehreren Vergleichswerten und einer Zuordnungsregel, die im Folgenden erläutert werden.

### Vergleichswerte

Für die Einordnung eines Indikatorwerts werden Vergleichswerte benötigt, welche die Kategorien als Bereiche auf der Indikatorkala abbilden und voneinander abgrenzen. Die Vergleichswerte stellen also Grenzwerte für die Entscheidung dar, in welche Kategorie ein Indikatorwert klassifiziert wird (vgl. auch die Ausführungen der Joint Commission zu Entscheidungsschwellen, JCAHO 1990: 85 ff.).

<sup>23</sup> Während in der QSKH-RL der Begriff „rechnerische Auffälligkeit“ Verwendung findet, wird in der Qesü-RL allgemeiner von „Auffälligkeiten“ gesprochen. Zum besseren Verständnis wird hier der allgemeinere Begriff „Auffälligkeit“ verwendet.

Der Bereich der Indikatorkala, der eine Kategorie repräsentiert (z. B. „erwartbare Qualität“), wird als *Referenzbereich* des Indikators für diese Kategorie bezeichnet, oder auch als *Referenzwert*, falls der Referenzbereich nur aus einem einzigen Wert besteht (z. B. 100 %). Die Referenzbereichsgrenzen entsprechen folglich den Vergleichswerten für die Klassifikation des Indikatorwerts.

In den bisherigen Verfahren der Qualitätssicherung nach §§ 136 ff. SGB V mit dem Zweck der Qualitätsförderung wird in der Regel nur ein einziger Vergleichswert definiert, der die Schwelle zu erwartbarer Qualität im Sinne eines einzuhaltenden Standards beschreibt. In diesem Fall gibt es genau einen Referenzbereich je Indikator, der zwischen dem Vergleichswert (einschließlich) und dem Ende der Indikatorkala mit guter Qualität liegt.<sup>24</sup> Zukünftige QS-Verfahren können möglicherweise verschiedene Qualitätsabstufungen beschreiben und folglich mehrere Referenzbereiche für einen Indikator definieren, z. B. für unzureichende, durchschnittliche und besonders gute Qualität.

### **Zuordnungsregel**

Die Zuordnungsregel beschreibt, auf welche Weise Vergleichswert und Indikatorwert miteinander verglichen werden: Im einfachsten Fall handelt es sich um den mathematischen Vergleich, ob der Indikatorwert z. B. größer als der Vergleichswert ist. Es kann sich aber auch um komplexere Regeln handeln, etwa wenn der Grad des Vertrauens in die Lage des Indikatorwerts berücksichtigt werden soll, z. B. mittels Konfidenzintervallen oder statistischer Tests, oder wenn mehrere Indikatorvariablen über einen Index zusammengefasst werden.

## **10.2 Entwicklung von Bewertungskonzepten**

Wie die Bewertungskonzepte der Qualitätsindikatoren sinnvoll zu spezifizieren sind, leitet sich aus den Entscheidungen ab, die damit getroffen werden sollen, also aus dem Zweck des QS-Verfahrens (in den Kategorien Qualitätsförderung, Information, Regulierung). Beispielsweise kann es für die Qualitätsförderung sinnvoll sein, Qualitätsdefizite möglichst niedrigschwellig zu erfassen, um Verbesserungen frühzeitig anzustoßen, während im Rahmen regulierender QS-Verfahren für die Klassifikation eines Indikatorwerts als Qualitätsdefizit andere Kriterien angelegt werden können. Ein Qualitätsindikator kann also, je nachdem in welchem QS-Verfahren er zum Einsatz kommt, unterschiedliche Bewertungskonzepte haben. Bei der Entwicklung von Bewertungskonzepten für neue QS-Verfahren legt das IQTIG daher zunächst fest, in welche Kategorien klassifiziert werden soll bzw. welches Qualitätsniveau die Referenzbereiche der Indikatoren repräsentieren sollen (z. B. erwartbare Qualität) und wie die Parameter der Klassifikationsgüte (z. B. Sensitivität oder Spezifität) gewichtet werden sollen. Diese Festlegungen gelten einheitlich für alle Indikatoren, die für denselben Verfahrenszweck eingesetzt werden.

Die Spezifikation des konkreten Referenzbereichs und der Zuordnungsregel erfolgt dagegen für jeden Qualitätsindikator spezifisch unter Berücksichtigung der Besonderheiten des Indikators (z. B. Qualitätsdimension nach Donabedian (1966) und Grad der Beeinflussbarkeit durch die

---

<sup>24</sup> Auch andere Vergleichswerte und Referenzbereiche können sinnvoll sein, wie etwa Vorjahreswerte, wenn auf Veränderungen im Zeitverlauf reagiert werden soll.

Leistungserbringer) und der Informationen zur Versorgungssituation aus Literatur und Expertenberatung. Dabei berücksichtigt das IQTIG außerdem formale Anforderungen an das Bewertungskonzept sowie praktische Auswirkungen des Bewertungskonzepts.

### **Formale Anforderungen**

Der gesetzliche bzw. durch Richtlinien des G-BA gesetzte Rahmen für QS-Verfahren macht an einigen Stellen Vorgaben für die Bewertungskonzepte. Einige Beispiele hierfür sind:

- die Verwendung einer Klassifikation in zwei Kategorien („rechnerisch auffällige“ und „unauffällige“ Ergebnisse) in § 10 Abs. 1 QSKH-RL
- die Nichtberücksichtigung von Fallzahleinflüssen bei der Bewertung der Indikatorwerte nach QSKH-RL
- die Verwendung prospektiv vom G-BA beschlossener, veröffentlichter Rechenregeln in den beiden bisher vom G-BA beschlossenen Verfahren nach der Qesü-RL

Sofern aus inhaltlichen oder methodischen Gesichtspunkten Anpassungen solcher formalen Vorgaben erforderlich erscheinen, werden in den Berichten vom IQTIG entsprechende Empfehlungen ausgesprochen.

### **Praktische Auswirkungen**

Da die Bewertungskonzepte der Qualitätsindikatoren zur Entscheidungsfindung eingesetzt werden, haben sie immer auch unmittelbar praktische Auswirkungen. Dies betrifft insbesondere den Aufwand, der dadurch entsteht, dass es einen hohen Anteil von Klassifikationsergebnissen gibt, die zu Handlungsempfehlungen führen (z. B. durch eine hohe Anzahl zu führender strukturierter Dialoge wegen auffälliger Indikatorwerte). Ergänzend zu den inhaltlichen Überlegungen bei der Konstruktion eines Bewertungskonzepts wird daher sofern möglich eine Einschätzung der praktischen Konsequenzen für die QS-Verfahren vorgenommen und diese wird in den Berichten dargestellt.

Die Bewertungskonzepte und Referenzbereiche der Qualitätsindikatoren werden nicht nur bei der Neuentwicklung von Qualitätsindikatoren definiert. Aufgrund ihrer Handlungsrelevanz werden diese darüber hinaus wiederkehrend im Rahmen der Verfahrenspflege – wie andere Bestandteile der Indikatoren auch – auf ihre Eignung für die Entscheidungsfindung überprüft (siehe hierzu Abschnitt 5.2).

## **10.3 Festlegung von Referenzbereichen**

Die Referenzbereiche der Qualitätsindikatoren in den bestehenden Verfahren der Qualitätssicherung nach §§ 136 ff. SGB V (d. h. mit dem Zweck der externen Qualitätsförderung) sollen erwartbare Qualität, also einen zu fordernden Standard in der Patientenversorgung, abbilden, unterhalb dessen das interne Qualitätsmanagement extern unterstützt werden soll, z. B. durch einen Dialog über Ursachen für die auffälligen Ergebnisse. Die Erwartung, dass ein bestimmtes Qualitätsniveau erreicht werden kann, z. B. dass die Sterblichkeit bei der Behandlung von Pneumonien eine bestimmte Rate nicht überschreiten sollte, beruht auf dem Wissen über den Status

quo der Versorgungssituation. Diese Referenzbereiche werden also empirisch ermittelt. Bei manchen Qualitätsindikatoren leitet sich der erwartete Standard aber nicht aus der aktuellen Versorgungssituation ab, sondern aus wissenschaftlich begründeten Forderungen z. B. entsprechender Leitlinien oder aus einem ethischen bzw. gesellschaftlichen Konsens, dass bestimmte Anforderungen zu 100 % erfüllt werden sollten, z. B. solche zur Transparenz oder Patiententeilhabe. Solche Referenzbereiche setzen eine vollständige Beeinflussbarkeit der Indikatorwerte durch den Leistungserbringer voraus (siehe Abschnitt 8.3).

### 10.3.1 Informationsbewertung für die Festlegung von Referenzbereichen

Für die Beurteilung des Status quo der Qualitätsmerkmale in der Versorgung stehen als Informationsquellen wissenschaftliche Literatur und Hinweise aus der Beratung durch Expertinnen und Experten zur Verfügung. Sofern es sich um Neuentwicklungen von Qualitätsindikatoren auf Basis von Sozialdaten bei den Krankenkassen, um Patientenbefragungen oder um die Überprüfung bestehender Indikatoren im Rahmen der Verfahrenspflege (siehe Abschnitt 5.2) handelt, kann die Verteilung der Indikatorwerte zusätzlich auch anhand von Auswertungen der entsprechenden Daten untersucht werden.

Auf Grundlage der genannten Informationsquellen wird insbesondere folgenden Fragestellungen nachgegangen, um zu einer Einschätzung zu gelangen, welcher Referenzbereich angemessen ist:

- Wie sind die Indikatorwerte (aggregiert je Leistungserbringer; z. B. Pneumoniesterblichkeit je Einrichtung) bei den Leistungserbringern verteilt und mit welcher statistischen Sicherheit entsprechen die beobachteten Indikatorwerte den zugrunde liegenden Werten der Leistungserbringer? Dies schließt Maße der zentralen Tendenz (z. B. Mittelwert, Median), der Streuung (z. B. Standardabweichung), Maße für die mehrheitlich erreichbaren Werte (z. B. 95. Perzentil) sowie die besten und die schlechtesten Werte, die erreicht wurden, ein.
- Wie sind die erhobenen Daten auf Ebene der Behandlungsfälle (z. B. Durchleuchtungszeit bei Schrittmacherimplantation) verteilt? Gibt es einen patientenbezogenen Wert, der nicht überschritten werden sollte (z. B. 99. Perzentil der Durchleuchtungszeiten)?
- Unter welchen Bedingungen wurde die Verteilung der Indikatorwerte ermittelt, d. h. für welche Patientenpopulation und in welchem Untersuchungszeitraum? Inwieweit entsprechen diese Bedingungen den Konstellationen des QS-Verfahrens?

Für Qualitätsindikatoren, deren Patientenrelevanz erst durch den Nachweis eines Zusammenhangs zwischen Indikatorziel und einem unmittelbar patientenrelevantem Ziel begründet wird (z. B. zwischen Antibiotikaphylaxe und verminderten OP-Wundinfektionsraten, siehe Abschnitt 8.3), werden außerdem folgende Aspekte untersucht:

- Unter welchen Bedingungen wurde der Zusammenhang zwischen Qualitätsmerkmal und unmittelbar patientenrelevantem Ziel gezeigt, d. h. für welche Patientenpopulation und in welchem Untersuchungszeitraum? Dies schließt die Beurteilung ein, welche Einflussfaktoren in der Definition der Population, welche in der Risikoadjustierung und welche nicht berücksichtigt wurden.

- Welche Art des Zusammenhangs besteht zwischen Indikatorziel und unmittelbar patientenrelevantem Ziel? Handelt es sich um einen monotonen Zusammenhang (z. B. je stärker das gemessene Qualitätsmerkmal ausgeprägt, desto eher wird ein unmittelbar relevantes Patientenziel erreicht)?
- Wurde dieser Zusammenhang in den zugrunde liegenden Studien für alle denkbaren Werte der gemessenen Größe untersucht (z. B. mittels eines Regressionsmodells) oder nur für einen bestimmten Ausschnitt des Messbereichs des Indikators?

### 10.3.2 Spezifikation des Referenzbereichs

Der geeignete Referenzbereich wird für jeden Qualitätsindikator individuell auf Grundlage der oben genannten Informationen und Fragestellungen festgelegt. Bei der formalen Spezifikation der Referenzbereichsgrenze eines Indikators können dabei zwei prinzipielle Möglichkeiten unterschieden werden: die Angabe eines konkreten Zahlenwerts auf der Messskala des Indikators, der unabhängig von der Verteilung der Indikatorwerte der Leistungserbringer festgelegt wird, oder die Angabe eines Zahlenwerts auf der Messskala des Indikators, der sich auf einen speziellen Punkt der Verteilung (z. B. ein Perzentil) der Indikatorwerte aller Leistungserbringer im Betrachtungszeitraum bezieht (verteilungsabhängig).

#### Verteilungsunabhängige Referenzbereiche

Sollen Entscheidungen anhand der Klassifikation getroffen werden, ob ein Leistungserbringer ein festgelegtes Maß an Qualität erreicht (z. B. den erwartbaren Standard), so muss dieses Maß zu Beginn der Qualitätsmessung bekannt sein. In solchen Fällen werden daher verteilungsunabhängige (feste) Referenzbereiche festgelegt. Beispielsweise könnte für den Indikator „Komplikationsrate nach einem operativen Eingriff“ als Referenzbereichsgrenze eine Komplikationsrate von 5 % festgelegt werden. Damit wird im Intervall 0–5 % von begründet zu erwartender Qualität gesprochen, die zu erreichen von jedem Leistungserbringer (nach Risikoadjustierung) gefordert wird. Überschreitet die Komplikationsrate eines Leistungserbringers (nach Risikoadjustierung) in einem gegebenen Betrachtungszeitraum 5 %, erfolgt die Klassifikation des Indikatorwerts in die Kategorie „(rechnerisch) auffällig“. Im Rahmen der Verfahrenspflege (siehe Abschnitt 5.2) werden diese festen Referenzbereiche regelmäßig überprüft und dann eine Anpassung empfohlen, wenn sich Hinweise ergeben, dass sich seit der letzten Überprüfung des Indikators und seines Referenzbereichs das Maß der erwartbaren Qualität geändert hat.

#### Verteilungsabhängige Referenzbereiche

Perzentilbasierte Referenzbereichsgrenzen definieren Qualität prospektiv über einen bestimmten Bereich der Ergebnisverteilung, indem z. B. die 90 % besten Leistungserbringerergebnisse als Standardqualität oder die 5 % besten Ergebnisse als Exzellenzqualität festgelegt werden. Im Gegensatz zu verteilungsunabhängigen (festen) Referenzbereichen geht die perzentilbasierte Festlegung mit dynamischen, im Zeitverlauf wechselnden Bereichsgrenzen einher und führt zwangsläufig dazu, dass ein bestimmter Prozentsatz an Leistungserbringern als vom Referenzbereich abweichend klassifiziert wird. Im Beispielfall des 90. Perzentils als Referenzbereichsgrenze erreichen per Definition 10 % der Leistungserbringer den Referenzbereich nicht. Verteilungsabhängige Referenzbereiche werden verwandt, wenn es zunächst nicht möglich erscheint, ein direkt benennbares Qualitätsniveau zu definieren (beispielsweise „begründet erwartbare

Qualität“). Sie führen dazu, dass ein Leistungserbringer z. B. als „auffällig“ klassifiziert wird, weil die anderen Leistungserbringer „bessere“ Indikatorwerte erzielen. Die Wahl eines Referenzbereichs, der nur relativ zu den Indikatorwerten anderer Leistungserbringer definiert ist, kommt folglich nur dann infrage, wenn

- das Bewertungskonzept ausdrücklich den relativen Vergleich der Leistungserbringer untereinander bezweckt und nicht das Erreichen eines absoluten Maßes an Qualität in den Vordergrund stellt, oder wenn
- aus praktischen Gründen ein fester Vergleichswert nicht angegeben werden kann. Dies ist häufig bei neu entwickelten Qualitätsindikatoren der Fall, falls zu Beginn des Regelbetriebs noch keine ausreichenden externen, empirischen Informationen aus den oben (Abschnitt 10.3.1) genannten Quellen vorliegen. Solche Referenzbereiche werden im Rahmen der Verfahrenspflege überprüft und eine Überführung in feste Referenzbereiche empfohlen, sobald eine ausreichende Datenbasis aus dem Regelbetrieb zur Verfügung steht.

#### **Berücksichtigung von Unsicherheit bezüglich der Lage des Referenzbereichs**

Die Einschätzung, welcher Wert auf der Indikatorkala adäquat das im QS-Verfahren adressierte Maß an Qualität abbildet, kann mit Unsicherheit behaftet sein. Dafür kann es mehrere Gründe geben:

- Es liegt nur eine unzureichende Menge an Informationen vor: z. B. wenn nur wenige Studien zum Status quo der Qualitätsmerkmale verfügbar sind oder wenn der Zusammenhang zwischen einem Prozess (z. B. Menge des gegebenen Kontrastmittels) und einem patientenrelevanten Ergebnis (z. B. Lungenödem) nur kategorial untersucht wurde (Vergleich nur zweier Gruppen mit Kontrastmittelmengen größer oder kleiner als ein bestimmter Grenzwert).
- Die Qualität und Konsistenz der empirischen Daten bzw. der Studienpublikationen ist eingeschränkt (siehe Kapitel 6): z. B. wenn nur Studien mit niedrigem Evidenzgrad verfügbar sind.
- Die Übertragbarkeit der Informationen auf die Zielpopulation des Qualitätsindikators ist eingeschränkt: z. B. wenn Daten nur aus anderen Ländern oder für Patientinnen und Patienten mit einem anderen Spektrum an Begleiterkrankungen vorliegen.

Dies wird berücksichtigt, indem gegebenenfalls ein von der Literatur abweichender, großzügiger Wert als Referenzbereichsgrenze festgelegt wird.

Ebenso passt das IQTIG den Referenzbereich an, wenn bestimmte systematische Einflussfaktoren auf die Indikatorwerte zwar bekannt sind, aber aus praktischen Gründen nicht berücksichtigt werden können, z. B. wenn sie in der Datengrundlage des Indikators nicht zur Verfügung stehen. Erfasst beispielsweise ein Indikator die prophylaktische Antibiotikagabe bei Operationen (Qualitätsziel: Antibiotikagabe soll immer erfolgen), müssten eigentlich Behandlungsfälle von Patientinnen und Patienten mit einer Kontraindikation zur Antibiotikagabe ausgeschlossen werden. Nur dann wäre ein Referenzwert von 100 % bei der Antibiotikagabe gerechtfertigt. Werden Kontraindikationen jedoch nicht erfasst, um beispielsweise den Dokumentationsaufwand für die Leistungserbringer gering zu halten, wirkt sich das auf die Festlegung des Referenzbereichs aus. In diesem Fall wären die Richtung der möglichen Verzerrung (der Anteil der Patientinnen und Patienten mit Antibiotikaphylaxe ist niedriger) und ggf. auch das mittlere Ausmaß der Verzerrung über alle Leistungserbringer (z. B. eine mittlere Häufigkeit von Kontraindikationen in

Höhe von 5 % der Patientinnen und Patienten) bekannt. Das IQTIG passt in diesem Fall die Referenzbereichsgrenze für den Qualitätsindikator unter Abschätzung des Ausmaßes der Verzerrung anhand externer Daten und Literatur an (z. B. Antibiotikagabe nicht bei 100 % der Patientinnen und Patienten, sondern nur bei 95 %).

#### 10.4 Festlegung der Zuordnungsregel

Im Bewertungskonzept der Qualitätsindikatoren erfüllt die Zuordnungsregel die Funktion, den Indikatorwert eines Leistungserbringers und den Vergleichswert miteinander in Beziehung zu setzen, um eine Klassifikation des Indikatorwerts des Leistungserbringers zu erreichen und daraus eine Handlungsempfehlung abzuleiten. Für Prozess- und Ergebnisindikatoren ist dabei in der Regel ein einfacher mathematischer Größer/kleiner-Vergleich zwischen gemessenem Indikatorwert und Referenzbereichsgrenze (z. B. „Ist die Pneumoniesterblichkeit bei einem Leistungserbringer in Höhe von 7 % größer als die Referenzbereichsgrenze 5 %?“) nicht ausreichend, da im Gegensatz zu Strukturindikatoren bei diesen Indikatoren fallbezogene, nicht dem Leistungserbringer zuzuschreibende Faktoren den gemessenen Indikatorwert des Leistungserbringers beeinflussen (siehe Abschnitt 13.1). Dabei handelt es sich zum einen um Faktoren, für die Richtung und Stärke des Einflusses auf den Indikatorwert bekannt sind, beispielsweise medizinische Risikofaktoren der Patientinnen und Patienten. Der Einfluss dieser Faktoren wirkt systematisch und wird als Verzerrung (Bias) bezeichnet und mit Methoden der Risikoadjustierung adressiert. Zum anderen handelt es sich um unbekannte oder nicht gemessene Einflussfaktoren auf den Indikatorwert. Beide letztgenannten Faktoren verringern die Präzision der Messung auch bei risikoadjustierter Berechnung des Indikatorwerts und erfordern eine Berücksichtigung mittels Methoden der Inferenzstatistik.

Bei der Klassifikation der Indikatorwerte der Leistungserbringer wird daher zwischen dem beobachteten Indikatorwert und dem diesem Wert „zugrunde liegenden Indikatorwert“ unterschieden. Der zugrunde liegende Indikatorwert kann als der Wert verstanden werden, den ein Leistungserbringer aufweisen würde, wenn es keine unbekannt oder nicht gemessenen Einflüsse gäbe. Da Qualitätsindikatoren nur die vom Leistungserbringer zu verantwortenden Einflüsse auf die Versorgung abbilden sollen, ist es nicht der beobachtete Indikatorwert, sondern der zugrunde liegende (ggf. risikoadjustierte) Indikatorwert eines Leistungserbringers, über den bezüglich der Lage innerhalb oder außerhalb des Referenzbereichs eine Aussage getroffen werden soll. Der beobachtete Indikatorwert ist eine Schätzung für den zugrunde liegenden Wert und nähert sich mit steigender Zahl beobachteter Fälle diesem an.

Eine Ausnahme bilden hier lediglich diejenigen Qualitätsindikatoren, für die alle fallbezogenen Faktoren dem Leistungserbringer zugeschrieben werden können (siehe Abschnitt 8.3) und deren Referenzbereich daher normativ gesetzt wird. Für diese Indikatoren werden beobachtete und zugrunde liegende Indikatorwerte als identisch angesehen.

##### **Klassifikationsgüte bei fallbezogenen Einflüssen**

Da der zugrunde liegende Indikatorwert unbekannt ist, handelt es sich bei der Feststellung, ob der zugrunde liegende Indikatorwert eines Leistungserbringers vom Referenzbereich abweicht, um eine Entscheidung unter Unsicherheit. Dies bedeutet, dass Fehlklassifikationen dadurch

möglich sind, dass unbekannte oder nicht gemessene fallbezogene Einflüsse eine Auffälligkeit herbeiführen, die nur durch die eingeschränkte Präzision der Messung bedingt ist. Tabelle 8 stellt die Entscheidungssituation bei der Feststellung, ob der zugrunde liegende Indikatorwert eines Leistungserbringers innerhalb eines Referenzbereichs liegt, dar.

Tabelle 8: Darstellung von Sensitivität und Spezifität bei der Feststellung von Abweichungen vom Referenzbereich

		zugrunde liegender Indikatorwert	
		liegt <i>nicht</i> innerhalb des Referenzbereichs	liegt innerhalb des Referenzbereichs
Entscheidung auf Basis des beobachteten Indikatorwerts	liegt <i>nicht</i> innerhalb des Referenzbereichs	Sensitivität	Fehler 1. Art
	liegt innerhalb des Referenzbereichs	Fehler 2. Art	Spezifität

Bei der Feststellung, ob der zugrunde liegende Indikatorwert eines Leistungserbringers innerhalb des Referenzbereichs des Indikators liegt, spielen die Sensitivität und die Spezifität dieser Klassifikation eine wichtige Rolle. Als Sensitivität wird die Wahrscheinlichkeit bezeichnet, mit der zugrunde liegende Indikatorwerte, die tatsächlich außerhalb des Referenzbereichs liegen, auch auf Basis der beobachteten Indikatorwerte als außerhalb des Referenzbereichs liegend klassifiziert werden (siehe Tabelle 8). Als Spezifität wird die Wahrscheinlichkeit bezeichnet, mit der zugrunde liegende Indikatorwerte, die innerhalb des Referenzbereichs liegen, auf Basis der beobachteten Indikatorwerte tatsächlich auch als innerhalb des Referenzbereichs liegend klassifiziert werden. Dabei ist zu beachten, dass Sensitivität und Spezifität in einer Austauschbeziehung stehen – es ist meist nicht möglich, beide unabhängig voneinander zu maximieren.

Die Berücksichtigung statistischer Unsicherheit geschieht mittels Verfahren der Inferenzstatistik, beispielsweise indem für die Indikatorwerte der Leistungserbringer Vertrauensbereiche (frequentistische oder bayesianische Konfidenzintervalle) berichtet werden oder mittels inferenzstatistischer Tests.<sup>25</sup>

Das Instrument zur Optimierung von Sensitivität und Spezifität in diesem Kontext ist die Teststärkeanalyse (im Rahmen frequentistischer Statistik). Abbildung 13 stellt beispielhaft eine Teststärkeanalyse für einen exakten einseitigen Binomialtest mit einer gewählten Spezifität von 85 % ( $\alpha = 1 - \text{Spezifität} = 15\%$ ) dar (Faul et al. 2009). In Abhängigkeit von der Fallzahl (x-Achse) werden für zwei verschiedene gegebene Sensitivitäten (Quadrate und Dreiecke) diejenigen zu-

<sup>25</sup> In der QSKH-RL ist festgelegt, dass bei der Klassifikation in „rechnerisch auffällige“ und „rechnerisch unauffällige“ Indikatorwerte die Anzahl der dem Indikatormesswert zugrunde liegenden Behandlungsfälle keine Berücksichtigung findet. Die Klassifikation und damit der Einstieg in den Strukturierten Dialog wird demnach auf Basis eines einfachen Größer/kleiner-Vergleichs des beobachteten Werts des Leistungserbringers mit der Referenzbereichsgrenze vorgenommen. Aus den hier dargelegten methodischen Gründen werden zusätzlich die Klassifikationsergebnisse auf Basis inferenzstatistischer Verfahren berichtet. Dies geschieht auch derzeit bereits in den Qualitätsberichten der Krankenhäuser gemäß G-BA-Regelung (Qb-R).

grunde liegenden Indikatorwerte dargestellt, die als signifikant kleiner als die Referenzbereichsgrenze von 90 % (durchgezogene Linie) klassifiziert würden. Aus dieser Darstellung wird deutlich, dass substantielle Abweichungen vom Referenzbereich auch für kleine Fallzahlen mit ausreichender Sensitivität und Spezifität erkannt werden können.

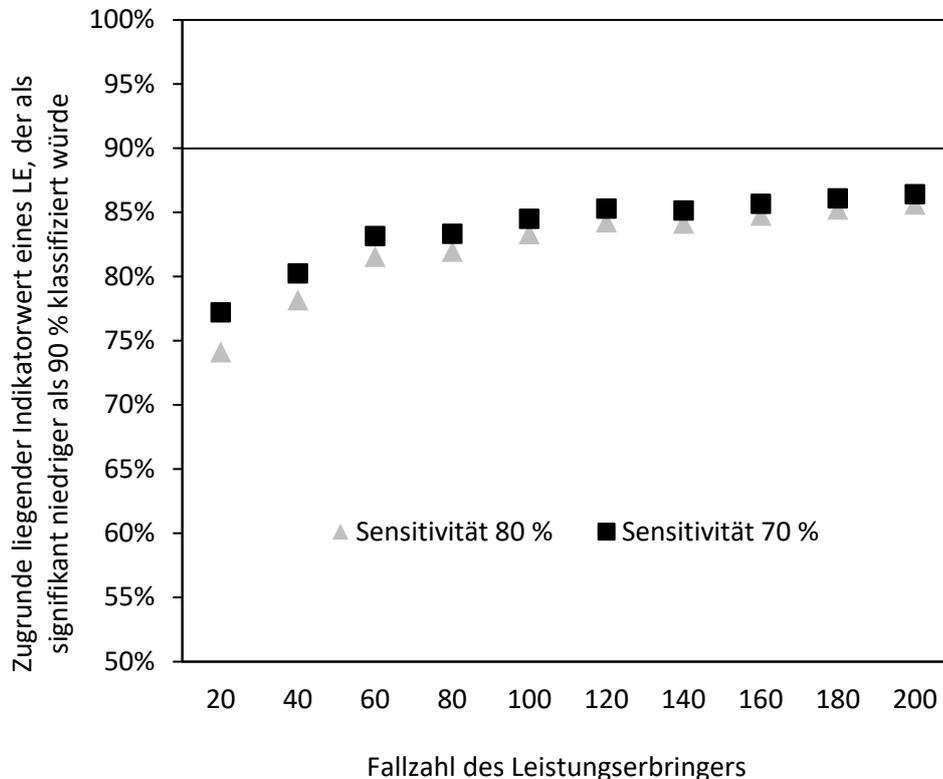


Abbildung 13: Beispielhafte Darstellung der zugrunde liegenden Indikatorwerte (y-Achse), die mit einer Spezifität von 85 % in Abhängigkeit von Fallzahl (x-Achse) und Sensitivität (80 % vs. 70 %) mit einem einseitigen exakten Binomialtest als signifikant verschieden von einem Wert von 90 % klassifiziert werden

Lesebeispiel: Bei einem Leistungserbringer (LE) mit einer Fallzahl von  $n = 40$  würde ein zugrunde liegende Indikatorwert von ca. 80,3 % (Quadrat) mit einer Wahrscheinlichkeit von 70 % (Sensitivität) als signifikant niedriger als die Referenzbereichsgrenze von 90 % (durchgezogene Linie) klassifiziert werden. Wird an die Sensitivität eine höhere Anforderung gestellt (z. B. 80 % Sensitivität; Dreiecke), während die Spezifität konstant gehalten wird (im Beispiel 85 %), würde erst ein Indikatorwert von ca. 78,2 % oder niedriger als signifikant abweichend klassifiziert. Dies verdeutlicht, dass Sensitivität und Spezifität nicht unabhängig voneinander optimiert werden können. Weiterhin bietet diese Betrachtungsweise die Möglichkeit, Sensitivität und Spezifität je nach Zweck des QS-Verfahrens (in den Kategorien Qualitätsförderung, Information, Regulierung) unterschiedlich zu gewichten. So könnte beispielsweise für die Qualitätsförderung im Strukturierten Dialog für Einrichtungen mit kleinen Fallzahlen die Spezifität zugunsten einer höheren Sensitivität herabgesetzt werden, um bei möglichst vielen Leistungserbringern, bei denen der Indikator auf Nicht-Erfüllung der erwartbaren Qualität hinweist, Qualitätsverbesserungen anzustoßen.

Um die unvermeidbare statistische Unsicherheit, mit der die Klassifikation fallbezogener Indikatorwerte bei einer endlichen Anzahl von erfassten Behandlungsfällen behaftet ist, transparent darzustellen, berechnet das IQTIG Konfidenzintervalle um die Lagemaße der Leistungserbringer auf den Qualitätsindikatoren.

## 10.5 Sentinel Events

Sentinel Events sind eine besondere Form von Bewertungskonzept für Prozess- oder Ergebnisindikatoren. Dabei wird jedes aufgetretene Ereignis, das der Merkmalsdefinition des Indikators entspricht, in die Kategorie „auffällig“ klassifiziert.<sup>26</sup>

Sentinel Events werden dann eingesetzt, wenn es sich um schwerwiegende patientenrelevante, seltene Ereignisse handelt (z. B. Sterbefälle bei elektiven Operationen mit niedrigem Risiko) und außerdem für einen nennenswerten Anteil der Ereignisse angenommen wird, dass sie der Qualität der Leistungserbringer zuzuschreiben sind (JCAHO 1989: 332). Die Nutzung von Sentinel Events erfolgt aus praktischen Gründen, z. B. wenn gemessen an der geringen Ereignisrate ein unverhältnismäßig großer Aufwand für die Erhebung von Risikofaktoren bei allen potenziell gefährdeten Patientinnen und Patienten entstände. Hier sind kasuistische Analysen der wenigen auftretenden Fälle die effizientere Vorgehensweise.

Beim Auftreten bestimmter Sentinel Events kann unmittelbar ein Qualitätsdefizit postuliert werden, z. B. bei Operationen auf der falschen Körperseite. In den anderen Fällen ist die bereits erwähnte Einzelfallanalyse durchzuführen, da Sentinel Events zwar auf potenziell gravierende Qualitätsmängel der Versorgung hinweisen können (Mainz 2003), diese aber auch schicksalhaft auftreten können (z. B. manche Fälle von Müttersterblichkeit nach Geburt).

---

<sup>26</sup> Formal entspricht ein Sentinel Event einem Referenzwert von 0 % und einem einfachen mathematischen Größer-Vergleich als Zuordnungsregel.

# 11 Grundgesamtheit, Stichprobenziehung und Vollerhebung

Im Rahmen der Entwicklung von QS-Verfahren und deren Durchführung kommen an verschiedenen Stellen Datenerhebungsverfahren zum Einsatz, die sowohl von Stichprobentechniken als auch von Vollerhebungen Gebrauch machen. In diesem Kapitel werden die Prinzipien dargelegt, nach denen Stichprobenziehungen oder Vollerhebungen durchgeführt werden, etwa im Rahmen von Entwicklungsleistungen oder beispielsweise im Rahmen des Regelbetriebs von QS-Verfahren. Der Fokus dieses Kapitels liegt auf der Datenerhebung bei Patientinnen und Patienten, beispielsweise mittels Patientenbefragungen oder Falldokumentationen durch die Leistungserbringer. Andere Formen der Datenerhebung, wie etwa bei der Datenvalidierung (siehe Kapitel 12), bei denen auch Stichprobenziehungen vorgenommen werden, sind nicht Gegenstand dieses Kapitels.

Nach § 299 SGB V sind prinzipiell auf Patientenebene Stichprobenverfahren in der Qualitätssicherung durchzuführen, es sei denn, dass der Verfahrenszweck nur mit einer Vollerhebung erfüllt werden kann. In jedem Fall muss die Datenerhebung in der medizinischen Qualitätssicherung nach §§ 136 ff. SGB V eine verzerrungsfreie Messung und Darstellung der Qualität der Versorgung ermöglichen. Eine zentrale Frage von Qualitätsmessungen anhand von Patientendaten ist daher, ob eine Stichprobenerhebung ausreicht, oder ob nur eine Vollerhebung der Grundgesamtheit zielführend ist und ob diese auch praktisch umsetzbar ist.

## 11.1 Ebenen der Stichprobenziehung und Vollerhebung

Grundsätzlich sind zwei Ebenen zu unterscheiden, auf denen im Kontext von QS-Verfahren in der medizinischen Qualitätssicherung eine Stichprobenziehung oder Vollerhebung vorgenommen werden kann: das sind zum einen die Leistungserbringer und zum anderen deren Behandlungsfälle. Theoretisch sind noch weitere, übergeordnete und zwischengeschaltete Ebenen denkbar, wie z. B. Bundesländer auf übergeordneter Ebene, Fachabteilungen innerhalb von Leistungserbringern, Ärzte und Ärztinnen innerhalb von Fachabteilungen/Praxen sowie Patientinnen und Patienten, die mehrmals behandelt werden. Hier wird deutlich, dass sich Qualitätsmessungen prinzipiell hierarchisch gliedern lassen, sodass sich die Frage nach Stichprobenziehung oder Vollerhebung auf verschiedenen Ebenen stellt.

Dabei ist zu beachten, dass die Datenerhebung auf Ebene der Patientinnen und Patienten hier vor allem der Beschreibung der Aggregateinheiten, also der Leistungserbringer dient (O'Brien 1990). Die Berechnung von Indikatorwerten für die Leistungserbringer erfolgt meist durch Aggregation der Daten der Behandlungsfälle eines bestimmten Zeitraums (siehe Kapitel 13), um quantitative Aussagen über die Qualität der medizinischen Versorgung durch den Leistungserbringer zu ermöglichen. Aussagen auf Ebene der Patientinnen und Patienten werden im Rahmen der medizinischen Qualitätssicherung meist nicht angestrebt. Eine Ausnahme hiervon bildet die qualitative Einzelfallbetrachtung im Strukturierten Dialog, bei der einzelne Behandlungsfälle (anonymisiert) analysiert werden.

Bei Qualitätsmessungen im Rahmen der QS-Verfahren nach §§ 136 ff. SGB V ist eine Vollerhebung der Leistungserbringer durch die Gesetzgebung vorgegeben, die alle Leistungserbringer zur Teilnahme an den Verfahren verpflichtet. Ist aber beispielsweise die Darstellung der Versorgungsqualität auf einer übergeordneten Ebene (etwa Regionen, Länder oder dem Bund) das einzige Ziel, eignen sich auch auf Ebene der Leistungserbringer Stichprobenverfahren (siehe etwa Lohr 2006). Auf Patientenebene regelt dagegen § 299 Abs. 1 SGB V, dass grundsätzlich Stichprobenziehungen anstelle von Vollerhebungen anzustreben sind, sofern dem keine wichtigen methodischen Gründe entgegenstehen.

Als Grundgesamtheit (= Zielpopulation) der Datenerhebung auf dieser Ebene werden diejenigen Patientinnen und Patienten bezeichnet, deren Daten für die Qualitätsmessung der Leistungserbringer relevant sind. Bevor aus dieser Grundgesamtheit eine Stichprobe gezogen oder eine Vollerhebung durchgeführt werden kann, muss die sogenannte Auswahlgesamtheit (engl. „Sampling Frame“) definiert werden. Diese stellt die tatsächlich verfügbare Grundlage zur Stichprobenziehung dar (Lohr 2006, Schnell et al. 2013). Im Kontext der medizinischen Qualitätssicherung nach §§ 136 ff. SGB V entsprechen dieser Auswahlgesamtheit diejenigen Behandlungsfälle, die im Regelbetrieb durch die Filter zur Selektion der für das jeweilige QS-Verfahren relevanten Behandlungsfälle (QS-Filter) in der Software der Leistungserbringer oder aus den Sozialdaten bei den Krankenkassen ausgewählt werden können. Als Stichprobe wird eine (zufällige oder systematisch) gewonnene Auswahl von Patientinnen und Patienten aus dieser Auswahlgesamtheit bezeichnet. Werden von allen Patientinnen und Patienten der Auswahlgesamtheit Daten erhoben, spricht man von einer Vollerhebung. Dies ist beispielsweise dann der Fall, wenn zu allen Behandlungsfällen, die den Spezifikationen des QS-Filters entsprechen (Nenner-Definition), Daten erhoben werden. Werden nur zu einer Teilmenge der Behandlungsfälle, die dieser sogenannten Nenner-Definition entsprechen, Daten erhoben, handelt es sich um eine Stichprobenziehung. Im Fall einer Stichprobenziehung erlauben Zufallsstichproben den inferenzstatistischen Verallgemeinerungsschluss auf die Grundgesamtheit (sogenannte Repräsentativität). Eine Zufallsstichprobe liegt dann vor, wenn jeder Merkmalsträger der Grundgesamtheit die gleiche Chance hatte, in die Stichprobe zu gelangen. In diesem Fall können Stichprobenparameter (z. B. Mittelwert) als unverzerrte Schätzungen für die Parameter der Grundgesamtheit interpretiert werden (Lohr 2006, Schnell et al. 2013). Bei gelenkten Stichproben werden besonders interessierende Merkmalsträger mit höherer Wahrscheinlichkeit gezogen als andere, wenn bei ihnen z. B. zusätzliche Auffälligkeiten bestehen, die eine besondere Problemkonstellation erkennen lassen (z. B. anhand einer auffallend langen Verweildauer).

Während die Auswahlgesamtheit als eine Operationalisierung der Grundgesamtheit verstanden werden kann, ist die Grundgesamtheit selbst in Abhängigkeit der jeweiligen Fragestellung der Datenerhebung eindeutig zu definieren. Denn nur in dem Ausmaß, in dem die Grundgesamtheit eindeutig beschrieben ist, ist die Definition der Auswahlgesamtheit möglich und nur dann können Stichprobenziehungen oder Vollerhebungen durchgeführt werden, die einen Rückschluss auf die interessierende Grundgesamtheit erlauben. Bei der Übertragung des Konzepts der Auswahlgesamtheit auf den Kontext der medizinischen Qualitätssicherung wird auch deutlich, dass die mittels der softwarebasierten QS-Filter erfassbaren Behandlungsfälle nicht deckungsgleich mit der interessierenden Grundgesamtheit sein müssen. Beispielsweise kann es sein, dass eine

interessierende Grundgesamtheit nur unscharf mittels einer Kombination aus verschiedenen Informationen abgebildet werden kann, die für den QS-Filter zugänglich sind (z. B. ICD- und OPS-Kodes, Alter, etc.). Diese Abbildung muss dann nicht unbedingt identisch mit der Grundgesamtheit sein, sondern kann interessierende Fälle ausschließen („Undercoverage“) sowie nicht interessierende Fälle einschließen („Overcoverage“; Schnell et al. 2013). Diese Situation besteht grundsätzlich bei allen Datenerhebungen, bei denen keine vollständige Liste der Grundgesamtheit der Merkmalsträger vorliegt und findet im Rahmen der Entwicklung neuer QS-Verfahren bei der Definition des QS-Filters sowie von Zähler und Nenner der Qualitätsindikatoren Berücksichtigung (siehe Kapitel 4).

Sofern Stichprobenziehungen durchgeführt werden, sind Zufallsstichproben oder Varianten von Zufallsstichproben (z. B. stratifizierte Zufallsstichproben) zu präferieren, um den angestrebten inferenzstatistischen Verallgemeinerungsschluss auf die Grundgesamtheit zu ermöglichen.

Davon abweichend sind Stichproben mit spezifischer Fragestellung zu sehen, bei denen es nicht um eine repräsentative Abbildung der Grundgesamtheit geht, sondern um das Auffinden besonderer Einzelfälle für kasuistische Analysen. So kann es z. B. im Rahmen der Datenvalidierung Ziel sein, die Häufung auffällig langer Aufenthaltsdauern ohne dokumentierten Grund anhand der Patientenakte zu klären. In solchen Fällen sind gezielte Stichproben das Verfahren der Wahl (siehe Kapitel 12)

## 11.2 Kriterien für Stichprobenziehung und Vollerhebungen

Durch § 299 Abs. 1 SGB V wird vorgegeben, dass Stichprobenziehungen die Regel darstellen sollen und Vollerhebungen nur bei Vorliegen gewichtiger medizinisch-fachlicher oder gewichtiger methodischer Gründe zum Einsatz kommen sollen. Innerhalb dieser gesetzlichen Rahmenbedingungen wird daher geprüft, ob Gründe gegen eine Stichprobenziehung vorliegen. Die Kriterien für diese Prüfung sind im Einzelnen:

### **Ressourcenbedarf**

Der Ressourcenbedarf ist für Vollerhebungen fast immer deutlich höher als für Stichprobenziehungen. Daher kann eine Stichprobenziehung die Möglichkeit eröffnen, umfangreichere Datenerhebungen pro Behandlungsfall zu realisieren als bei einer Vollerhebung, da die umfangreiche Datenerhebung nicht für alle Behandlungsfälle der Grundgesamtheit durchgeführt werden muss.

### **Praktische Durchführbarkeit**

Obwohl Stichprobenziehungen in der Regel weniger Ressourcen benötigen als Vollerhebungen, kann die praktische Durchführbarkeit für Stichprobenziehungen unter Umständen behindert sein, da sich beispielsweise die Patientenauswahlgesamtheit, aus der eine Zufallsstichprobe gezogen werden soll, erst über den Lauf des Betrachtungszeitraums akkumuliert und nicht schon prospektiv absehbar ist. Dies macht die Ziehung einer Zufallsstichprobe zwar nicht unmöglich, erschwert sie jedoch.

**Dokumentationsgüte**

Die Dokumentationsgüte und die daraus resultierende Datenqualität können in Abhängigkeit von der Erhebungsmethode variieren. Beispielsweise kann durch die niedrigere Dokumentationsbelastung bei Stichproben die Dokumentationsgüte höher ausfallen als bei Vollerhebungen. Andererseits kann aber auch bei Vollerhebungen die Dokumentationsgüte größer sein, wenn mehr und kontinuierlichere „Übung“ im Umgang mit dem Dokumentationsbogen vorliegt.

**Stichprobenverzerrungen**

Bei einer Stichprobenziehung stellt sich grundsätzlich die Frage nach der Verallgemeinerbarkeit der Ergebnisse auf die Grundgesamtheit („Repräsentativität“, siehe oben). Dieser Verallgemeinerungsschluss ist prinzipiell nur bei Zufallsstichproben und deren Varianten (z. B. stratifizierte Zufallsstichproben) zulässig (Schnell et al. 2013). Bestehen daher berechtigte Zweifel an der Realisierbarkeit einer Zufallsstichprobe, wird auf eine Vollerhebung zurückgegriffen.

**Seltene Ereignisse**

Sollen mit der Datenerhebung seltene Ereignisse erfasst werden, können Stichproben kontraindiziert sein. Beinhaltet ein QS-Verfahren beispielsweise die Erfassung von Sentinel Events, bei denen schon das einmalige Auftreten eines Ereignisses (z. B. Versterben nach Hüft-Endoprothesen-Erstimplantation) eine Auffälligkeit impliziert, der in jedem Fall nachzugehen ist, sind Stichprobenverfahren nicht geeignet, da diese nicht gewährleisten können, jedes Auftreten eines sogenannten Sentinel Event zu erfassen.

**11.3 Vorgehen bei Stichprobenziehungen**

Sofern eine Stichprobenziehung für ein QS-Verfahren als zielführend eingestuft wird, wird eine Fallzahlberechnung durchgeführt, um die Mindestzahl an Behandlungsfällen pro Leistungserbringer zu schätzen, die nötig ist, um die gewünschte Präzision der Parameterschätzung zu erreichen. Als Parameterschätzung wird die Berechnung eines Stichprobenparameters (z. B. Anteilswert in der Stichprobe) als Schätzung für den Populationsparameter (z. B. Anteilswert in der Grundgesamtheit) bezeichnet. Die Präzision der Parameterschätzung drückt sich u. a. in der Größe des Standardfehlers (synonym: Stichprobenfehler) oder des Vertrauensintervalls aus (Lohr 2006). Da die Größe des Standardfehlers durch die Stichprobengröße und die Varianz innerhalb der Grundgesamtheit determiniert ist, können bei geringer Varianz innerhalb der Grundgesamtheit verhältnismäßig kleine Stichproben ausreichen, um ähnlich kleine Standardfehler zu erreichen wie in Fällen großer Varianz innerhalb der Grundgesamtheit. Die im Regelbetrieb benötigte Stichprobengröße pro Leistungserbringer ist demzufolge abhängig von folgenden Größen:

- Varianz des Merkmals in der Grundgesamtheit
- gewünschte Breite des Konfidenzintervalls
- erwartete Rücklaufquote (bei Befragungen)

Die mindestens benötigte Fallzahl pro Leistungserbringer ist dabei unabhängig von der Anzahl der Fälle, die der Leistungserbringer im Betrachtungszeitraum behandelt (siehe Kapitel 13). Dies

kann dazu führen, dass bei Leistungserbringern mit kleinen Fallzahlen alle Behandlungsfälle erhoben werden müssen, um die benötigte Präzision der Parameterschätzung zu erreichen, während bei Leistungserbringern mit größeren Fallzahlen Stichproben ausreichend sind.

## 12 Datenvalidierung von QS-Dokumentationsdaten und Strukturierter Dialog

Für quantitative Messungen von Versorgungsqualität mithilfe von Qualitätsindikatoren stellt die Validität der zugrunde liegenden Daten eine wichtige Voraussetzung dar (siehe Kapitel 88). Das vorliegende Kapitel beschreibt, mit welchem Verfahren die Validität von Daten geprüft wird, die von den Leistungserbringern zum Zweck der externen Qualitätssicherung erfasst werden. Diese Prüfung wird im folgenden Text kurz als „Datenvalidierung“ bezeichnet. Das Verfahren dient dazu, die Datenvalidität einzuschätzen, die Leistungserbringer für die korrekte Dokumentation zu sensibilisieren und zu schulen sowie Erkenntnisse für Verbesserungen der Spezifikation zu sammeln (z. B. Präzisierung der Ausfüllhinweise, Gestaltung von Datenfeldern, Präzisierung des QS-Filters). In bestimmten Fällen kann ein Strukturierter Dialog eingeleitet werden. Im Folgenden werden die methodischen Grundlagen des Datenvalidierungsverfahrens für die fallbezogene QS-Dokumentation nach der QSKH-RL und QFR-RL beschrieben und begründet. Für die sektorenübergreifenden Verfahren nach der Qesü-RL und die damit verbundenen neuen Datenquellen (einrichtungsbezogene QS-Dokumentation, Patientenbefragung, Sozialdaten bei den Krankenkassen) wird derzeit eine Methodik zur Datenvalidierung durch das IQTIG entwickelt, die in den zukünftigen Versionen der „Methodischen Grundlagen“ beschrieben wird.

Kein Bestandteil des Datenvalidierungsverfahrens sind:

- Psychometrische Testung von Messinstrumenten (z. B. Fragebögen oder Qualitätsindikatoren): dies ist Gegenstand der Instrumenten- bzw. Indikatorenentwicklung.
- Validierung des Prozesses der Dokumentation: Gegenstand des Validierungsverfahrens ist das Ergebnis der Dokumentation (= die Datengrundlage zur Berechnung der Qualitätsindikatoren).

In der nationalen und internationalen Literatur werden verschiedene Aspekte der Datenvalidität benannt und überprüft. Diese variieren in Abhängigkeit davon, um welche Datengrundlage es sich handelt, wie viele Ressourcen und rechtliche Möglichkeiten für die Validierung zur Verfügung stehen und welches Ziel in den jeweiligen Untersuchungen verfolgt wird (Arts et al. 2002, Chen et al. 2014, Ehling et al. 2007, Nonnemacher et al. 2014).

Vor dem Hintergrund der international etablierten Datenqualitätsaspekte sowie den Rahmenbedingungen und Anforderungen der externen Qualitätssicherung prüft das IQTIG die folgenden Aspekte der Datenvalidität:

- Vollständigkeit: alle erforderlichen Angaben zu einem einzelnen Datensatz (Behandlungsfall) sind vorhanden
- Vollzähligkeit: Übereinstimmung der Anzahl der gelieferten QS-Datensätze mit der Anzahl der (erwarteten) dokumentationspflichtigen Datensätze
- Plausibilität: erlaubte Werte, plausible Werteverteilungen (datenfeldbezogen) und Wertekombinationen (datenfeldübergreifend)
- Konkordanz/Übereinstimmungsvalidität: Übereinstimmung mit einer Referenzquelle

Zur Datenvalidierung stehen grundsätzlich zwei Methoden zur Verfügung: die interne und die externe Validierung. Bei der internen Validierung finden Validitätsprüfungen innerhalb der zu validierenden Datenquelle statt. Die externe Validierung hingegen sieht den Abgleich der zu validierenden Daten mit einer externen Referenzquelle (Primär- oder Sekundärdaten) vor (Horenkamp-Sonntag et al. 2014). Die externe Validierung gilt im Allgemeinen als Goldstandard für die Validierung (Nonnemacher et al. 2014, Schubert et al. 2010). Sie ist jedoch häufig aufgrund von zeitlichen, finanziellen, organisatorischen und datenschutzrechtlichen Gründen nicht möglich (Hartmann et al. 2016). Die interne Validierung der QS-Daten ist aufwandsärmer als die externe Validierung und ermöglicht so die statistische Prüfung einer größeren Anzahl an Leistungserbringern und Fällen. Aus diesem Grund wird in der externen Qualitätssicherung ein Mix aus interner und externer Validierung angewendet.

Der Prozess der Datenvalidierung in der externen Qualitätssicherung besteht aus drei Elementen, die in den folgenden Abschnitten beschrieben werden:

- einer technischen Prüfung bei Dateneingabe und -annahme (interne Validierung)
- einer statistischen Prüfung der erhobenen Daten nach Datenannahme (interne Validierung)
- einem Datenabgleich mit Referenzquelle nach Datenannahme (externe Validierung).

### **12.1 Technische Prüfung bei Dateneingabe und -annahme**

Die technische Prüfung bei Dateneingabe und -annahme prüft die QS-Daten prospektiv und findet für jeden Fall jedes Leistungserbringers Anwendung. Somit kann bestimmten Fehldokumentationen und Missings flächendeckend vorgebeugt werden. In der retrospektiven Prüfung nach Datenannahme finden anschließend Prüfungen statt, die (noch) nicht während der Dateneingabe und -annahme möglich sind.

Die Spezifikationen zur Datenerfassung und -übermittlung für die externe Qualitätssicherung beinhalten Prüfungen zur Vollständigkeit der Datensätze und zur Plausibilität der Daten. Die Erfassungssoftware ist entsprechend der Spezifikationsvorgaben so konfiguriert, dass fehlende Werte bei verpflichtend auszufüllenden Datenfeldern bei der Dateneingabe vor Ort beim Leistungserbringer nicht akzeptiert werden. Ein Datensatz kann erst bei Vollständigkeit abgeschlossen und an die Datenannahmestelle übermittelt werden. Außerdem beinhalten die Spezifikationen Plausibilitätsprüfungen zu nicht zulässigen Werteeingaben (datenfeldbezogen) und nicht plausiblen Kombinationen von Werten (datenfeldübergreifend).

Bei der Datenannahme durch die zuständigen Datenannahmestellen wird geprüft, ob die Daten in der spezifizierten Form übermittelt wurden. Zudem werden auch an dieser Stelle Vollständigkeits- und Plausibilitätsprüfungen durchgeführt. Dabei finden sowohl „weiche“ als auch „harte“ Plausibilitätsregeln Anwendung. Bei der Verletzung einer weichen Plausibilitätsregel erscheint eine Warnung, der Datenfluss wird jedoch nicht gestoppt. Wird eine harte Plausibilitätsregel verletzt, wird der Datenfluss angehalten und eine Datennachlieferung bzw. -korrektur ist erforderlich.

## 12.2 Statistische Prüfung nach Datenannahme

Nach der Datenannahme werden die gelieferten Daten durch die zuständigen Datenannahmestellen einer formalen und inhaltlichen Prüfung anhand verfahrensspezifischer und verfahrensübergreifender Kriterien (Auffälligkeitskriterien) unterzogen. Mittels der Auffälligkeitskriterien werden die Leistungserbringer auf Basis von definierten Referenzbereichen als rechnerisch auffällig oder unauffällig eingestuft. Die rechnerischen Auffälligkeiten werden im Rahmen eines Strukturierten Dialogs mit den jeweiligen Leistungserbringern evaluiert und abschließend als korrekte oder fehlerhafte Dokumentation eingestuft.

### Auffälligkeitskriterien der Vollzähligkeit

Die Prüfung auf Vollzähligkeit findet in allen QS-Verfahren statt. Sie erfolgt mithilfe des Abgleichs der gelieferten Daten mit einer Sollstatistik, die auf den Abrechnungsdaten der Leistungserbringer beruht. In diese Sollstatistik werden alle Fälle aufgenommen, die durch den QS-Filter in den Abrechnungsdaten des Erfassungsjahres als dokumentationspflichtig eingeordnet wurden. Ein Leistungserbringer wird als „rechnerisch auffällig“ klassifiziert, wenn die gelieferten Daten in einem QS-Verfahren weniger als 95 % der Datensätze der Sollstatistik enthalten. Andererseits werden auch Leistungserbringer als „rechnerisch auffällig“ klassifiziert, wenn die gelieferten Daten mehr als 110 % der Anzahl der Datensätze der Sollstatistik enthalten. Die Vollzähligkeit wird nur bei Leistungserbringern überprüft, die eine bestimmte Mindestanzahl an Fällen aufweisen. Das Krankenhausstrukturgesetz (KHSG) sieht vor, dass die geforderte Dokumentationsrate auf 100 % gesetzt wird – wie das bei den QS-Verfahren zu Transplantationen bereits der Fall ist –, sodass zukünftig alle dokumentationspflichtigen Fälle – und nicht wie bislang mindestens 95 % – durch die Leistungserbringer geliefert werden müssen.

Ein weiteres Kriterium der Vollzähligkeitsprüfung ist die Anzahl der Minimaldatensätze. Ein Minimaldatensatz kann im Ausnahmefall übermittelt werden, wenn ein Behandlungsfall als dokumentationspflichtig klassifiziert wurde, jedoch nicht für die Dokumentation in dem betroffenen QS-Verfahren geeignet ist. Minimaldatensätze können nicht zur Berechnung von Qualitätsindikatoren verwendet werden. Leistungserbringer werden dann als „rechnerisch auffällig“ klassifiziert, wenn die gelieferten Minimaldatensätze in einem QS-Verfahren mehr als 5 % der gelieferten Datensätze ausmachen.

### Auffälligkeitskriterien zur Plausibilität

Die Auffälligkeitskriterien zur Plausibilität sind verfahrensspezifisch und beziehen sich auf die inhaltliche Plausibilität der Angaben. Sie prüfen im Regelfall unwahrscheinliche oder widersprüchliche Werteverteilungen und Kombinationen von Werten einzelner Datenfelder (sehr seltene oder sehr häufige Angaben, z. B. nie Komplikationen bei hoher Verweildauer). Über Auffälligkeitskriterien zur Plausibilität werden gezielte Stichproben von Leistungserbringern und Fällen mit bestimmten Problemkonstellationen gezogen, die im Strukturierten Dialog geklärt werden sollen.

## Neu- und Weiterentwicklung von Auffälligkeitskriterien

Neben der Neuentwicklung findet auch die Überprüfung bereits eingeführter Auffälligkeitskriterien auf ihre Eignung zur Weiterführung im Rahmen der Verfahrenspflege statt. In beiden Fällen werden externe Expertinnen und Experten beratend hinzugezogen.

Sowohl für die Neuentwicklung von Auffälligkeitskriterien als auch für die Evaluation bestehender Auffälligkeitskriterien werden folgende Aspekte berücksichtigt:

- **Keine Abbildung über harte Plausibilitätsregel(n) in der Spezifikation möglich:** Wenn die Möglichkeit besteht, eine Plausibilitätsregel in der Spezifikation als harte Regel zu verankern, ist dies einem Auffälligkeitskriterium vorzuziehen, denn die Prüfung über eine Plausibilitätsregel findet bereits bei der Dateneingabe statt und verhindert die Eingabe unplausibler Angaben. Mit einem Auffälligkeitskriterium können Angaben immer nur retrospektiv nach der Datenannahme geprüft werden.
- **Relevanz für Qualitätsindikatoren:** Insbesondere die Angaben, die unmittelbar in die Indikatorenberechnung einfließen, sollten bei der Entwicklung berücksichtigt werden.
- **Abbildung eines Dokumentationsproblems:** Bei der Operationalisierung sollte geprüft werden, ob das vermutete Dokumentationsproblem durch das Auffälligkeitskriterium hinreichend erfasst werden kann.
- **Bestätigte Fehldokumentationen:** Es sollte auch berücksichtigt werden, ob sich tatsächlich Dokumentationsfehler identifizieren lassen. Dabei wird der Anteil der im Strukturierten Dialog bestätigten Fehldokumentationen an der Anzahl der angeforderten Stellungnahmen pro Auffälligkeitskriterium betrachtet.
- **Brauchbarkeit:** Auffälligkeitskriterien sollten für die Leistungserbringer nachvollziehbar und das Dokumentationsproblem überprüfbar sein. Die Auffälligkeitskriterien sollten in der Lage sein, Erkenntnisse zu generieren, die zu einer Verbesserung der Dokumentationsprobleme beitragen.
- **Berücksichtigung von Fallzahlen:** Bei einigen Auffälligkeitskriterien ist die rechnerische Auffälligkeit stark durch die Anzahl der behandelten Patientinnen und Patienten beeinflusst. So werden Krankenhäuser mit geringen Fallzahlen beispielsweise bei einem Auffälligkeitskriterium „Nie Komplikationen bei hoher Verweildauer“ häufiger rechnerisch auffällig als Krankenhäuser mit höheren Fallzahlen. Deshalb sollten hier die Fallzahlen Berücksichtigung bei der Berechnung finden.

## 12.3 Datenabgleich mit Referenzquelle

Als Referenzquelle zur Validierung von Gesundheitsdaten werden verschiedene Datenquellen verwendet (ärztliche Originaldokumentationen, Befragungsdaten, Abrechnungsdaten, amtliche Statistiken oder Registerdaten) (Benchimol et al. 2011, Horenkamp-Sonntag et al. 2014). Der Goldstandard als Referenzquelle für medizinische Daten ist die ärztliche Originaldokumentation

(= Patientenakte) (Arts et al. 2002, Stange et al. 1998). Je nach Referenzquelle kann ein individueller Abgleich (Ebene Einzelfall) oder aggregierter Abgleich (Ebene Gesamtdatensatz) vorgenommen werden.

Aufgrund des hohen Aufwands einer externen Validierung wird diese häufig anhand von Stichproben durchgeführt. Zufallsstichproben erlauben einen inferenzstatistischen Verallgemeinerungsschluss auf die Grundgesamtheit (sogenannte Repräsentativität). Das heißt Stichprobenparameter (z. B. Mittelwert) können als unverzerrte Schätzungen für die Parameter der Grundgesamtheit interpretiert werden – ohne dass eine Vollerhebung durchgeführt werden muss (siehe Kapitel 11). Aus diesem Grund wird in der externen stationären Qualitätssicherung jährlich für ausgewählte QS-Verfahren eine Zufallsstichprobe aus Leistungserbringern und Fällen gezogen, für die ein Abgleich auf Einzelfallebene zwischen den Angaben aus den gelieferten QS-Daten und den Angaben aus der Patientenakte stattfindet. So können quantitative Aussagen über die Datenvalidität in einem QS-Verfahren getroffen werden. Bei der Auswahl der QS-Verfahren werden Kriterien herangezogen, die beispielsweise berücksichtigen, ob sich ein hoher Anteil an Dokumentationsproblemen im Strukturierten Dialog gezeigt hat, ein QS-Verfahren noch nie umfassend validiert wurde oder ein QS-Verfahren Qualitätsindikatoren mit besonderem Handlungsbedarf aufweist. Stehen größere Umstrukturierungen in einem QS-Verfahren an, so ist eine umfassende Datenvalidierung weniger sinnvoll. Ergänzend werden zur Auswahl der QS-Verfahren, in denen ein Datenabgleich erfolgen soll, externe Expertinnen und Experten beratend hinzugezogen.

Der Datenabgleich findet als Stichprobenverfahren statt. Hierbei wird in einem zweistufigen Verfahren zunächst eine Zufallsstichprobe von Krankenhausstandorten und anschließend eine Zufallsstichprobe von Fällen je Standort gezogen. Die anhand der Patientenakte zu überprüfenden Datenfelder sind insbesondere solche, die unmittelbar in die Indikatorenberechnung einfließen. Zudem werden unter Einbindung von Fachexpertinnen und -experten Datenfelder ausgewählt, in denen sich Dokumentationsprobleme gezeigt haben bzw. vermutet werden.

Nach dem Datenabgleich wird die Übereinstimmungsrate zwischen den Daten aus der QS-Dokumentation und der Patientenakte pro Datenfeld bestimmt. Bei der Bewertung der Ergebnisse des Stichprobenverfahrens mit Datenabgleich ist zu unterscheiden zwischen der Ebene des geprüften QS-Verfahrens (verfahrensbezogene Bewertung der Datenvalidität und des Handlungsbedarfs) und der Ebene der geprüften Leistungserbringer (leistungserbringerbezogene Bewertung der Datenvalidität und des Handlungsbedarfs). Zur Bewertung der Datenvalidität reicht die Übereinstimmungsrate alleine nicht aus, es müssen weitere Informationen hinzugezogen werden:

#### **Verfahrensbezogene Bewertung der Datenvalidität und des Handlungsbedarfs**

- Anzahl/Anteil der Leistungserbringer mit Abweichungen
- Relevanz der Abweichungen für die Berechnung der Qualitätsindikatoren und Risikoadjustierung (Unter-, Über-, Fehldokumentation)
- Beeinflussbarkeit der Dokumentationsmängel durch Maßnahmen der Verfahrenspflege (z. B. Einführung/Änderung Ausfüllhinweise, Einführung Plausibilitätsregeln, Präzisierung Datenfelder, Vereinfachung der Struktur des Erfassungsbogens)

Nach der verfahrensbezogenen Bewertung können bei Bedarf Maßnahmen durch die Fachgruppen auf Bundesebene im Rahmen der Verfahrenspflege ergriffen werden (siehe Abschnitt 5.2).

#### **Leistungserbringerbezogene Bewertung der Datenvalidität und des Handlungsbedarfs**

- Anzahl/Anteil der Datenfelder mit Abweichungen
- Art und Auswirkung der Abweichungen auf die rechnerischen Ergebnisse der Qualitätsindikatoren und Risikoadjustierung (Unter-, Über-, Fehldokumentation)
- Beeinflussbarkeit der Dokumentationsmängel durch Maßnahmen des internen Struktur- und Prozessmanagements der Leistungserbringer
- Datenmanipulation oder versehentliche Dokumentationsfehler

Nach der leistungserbringerbezogenen Bewertung können bei Bedarf Maßnahmen des Strukturierten Dialogs eingeleitet werden.

Im Rahmen der Qualitätssicherungs-Richtlinie Früh- und Reifgeborene (QFR-RL) findet zur Vollzähligkeitsprüfung neonataler Sterbefälle ein Abgleich zwischen den für die QS-Dokumentation Neonatologie übermittelten Daten und Leistungsdaten des Instituts für das Entgeltsystem im Krankenhaus (InEK) nach § 21 Krankenhausentgeltgesetz statt. Durch den Einsatz eines deterministischen Record-Linkage-Verfahrens (= Matching über exakt übereinstimmende Angaben) werden die Fälle aus den beiden Datensätzen über die Angaben „Institutskennzeichen“, „Aufnahme- und Entlassungsdatum“, „Geschlecht“ und „Aufnahmegewicht“ zusammengeführt. Einige Sterbefälle lassen sich aufgrund von Dokumentationsfehlern in einem der Datensätze nicht über ein exaktes Matching zusammenführen (z. B. Abweichung des Aufnahmegewichts um bspw. 30 g bei gleichem Institutionskennzeichen, gleichem Geschlecht, gleichem Aufnahme- und Entlassungsdatum). Aus diesem Grund werden in einem zweiten Schritt mittels eines probabilistischen Linkage-Verfahrens (= übereinstimmende Angaben mit Toleranz von geringfügigen Abweichungen) hinsichtlich der Möglichkeit der Zusammenführung überprüft. Abschließend lassen sich so drei Gruppen neonataler Sterbefälle bilden, die den betroffenen Leistungserbringern zurückgespiegelt werden:

- Sterbefälle, die nur in den §21-Daten dokumentiert sind
- Sterbefälle, die nur in den QS-Daten dokumentiert sind
- Sterbefälle, die mithilfe des probabilistischen Linkage-Verfahrens zusammengeführt werden konnten

Sterbefälle, die nur in einem Datensatz eindeutig identifiziert werden konnten, werden im Rahmen eines Validierungsverfahrens bei den Leistungserbringern vor Ort durch Einbeziehung der Originaldokumentation (Patientenakten und KIS) geprüft und ggf. nachdokumentiert. Der Abgleich wurde erstmalig für die Daten der Erfassungsjahre 2010 bis 2014 durchgeführt.

## **12.4 Strukturierter Dialog / Stellungnahmeverfahren**

Die primäre Bewertung der Versorgungsqualität anhand von Versorgungsdaten und den darauf angewandten Algorithmen von Qualitätsindikatoren basiert auf standardisierten Beschreibungsweisen der Versorgung: welche Versorgung geleistet wurde und wie ihre Durchführung und Ergebnisse zu bewerten sind. Liegen die Ergebnisse einer Versorgungseinrichtung außerhalb

des geforderten Referenzbereichs, dann wird sie zunächst als rechnerisch auffällig eingestuft. Dabei wird zwar eine gewisse Vielfalt der medizinischen Praxis berücksichtigt, z. B. durch Methoden der Risikoadjustierung. Dennoch kann nie ausgeschlossen werden, dass die Versorgungsrealität eines einzelnen Leistungserbringers so stark vom Durchschnitt abweicht, dass diese Methoden nicht ausreichen, um eine faire, vergleichbare Qualitätsbewertung gewährleisten zu können. Daher kann der Leistungserbringer im Strukturierten Dialog (QSKH-RL) bzw. Stellungnahmeverfahren (Qesü-RL) Gründe vortragen, wegen derer die rechnerische Auffälligkeit nicht auf ein tatsächliches Qualitätsdefizit hinweist.

Der Strukturierte Dialog bzw. das Stellungnahmeverfahren hat drei wesentliche Ziele:

- die Klärung, ob ein qualitatives Problem vorliegt, das behoben werden muss
- die Förderung der Verbesserungsmaßnahmen beim Leistungserbringer
- die Kontrolle der erfolgreichen Problembehebung

Hierzu stehen richtlinienbasiert u. a. die Instrumente des schriftlichen Hinweises und Dialogs, der Einladung zum persönlichen Dialog mit der Arbeitsgruppe oder der Fachkommission, der Begehung mit Dialog vor Ort sowie dem Abschluss einer Zielvereinbarung zur Verfügung. Strukturierter Dialog bzw. Stellungnahmeverfahren werden in den indirekten Verfahren von der zuständigen Einrichtung auf Landesebene oder in den direkten Verfahren durch das IQTIG initiiert und geführt. Die Bewertungsergebnisse werden dem Leistungserbringer kommuniziert sowie in einem Bericht über alle Leistungserbringer jährlich zusammengefasst. Die Rückmeldungen aus diesen Verfahren dienen gleichzeitig der Weiterentwicklung der QS-Verfahren, in dem die Instrumente zur spezifischen Erfassung von Fällen, von Prozeduren und Ereignissen sowie deren Bewertung kontinuierlich verbessert werden, um die Aussagekraft von Indikatoren zu erhöhen und falsch positive rechnerische Auffälligkeiten zu reduzieren. Auch die Verfahren für Stellungnahme und Strukturierten Dialog selbst werden systematisch weiterentwickelt.

## 13 Methoden der Datenanalyse

Die Analyse der aus den Verfahren der Qualitätssicherung nach §§ 136 ff. SGB V mit dem Zweck der Qualitätsförderung anfallenden Daten erfolgt in zwei Schritten. Im ersten Schritt findet eine quantitative Analyse der Daten in Form der Berechnung von Qualitätsindikatorwerten statt. Wird in diesem ersten Schritt festgestellt, dass ein Indikatorwert eines Leistungserbringers auffällig<sup>27</sup> ist (siehe Kapitel 10), erfolgt im zweiten Schritt eine fachlich-qualitative Bewertung der auffälligen Ergebnisse im Hinblick auf ihre Vermeidbarkeit. Es geht darum, ob eine rechnerische angezeigte Auffälligkeit auf ein Qualitätsproblem hinweist oder ob andere Faktoren, die nicht vom Leistungserbringer zu verantworten sind, zur Auffälligkeit geführt haben. Diese Analyse im Strukturierten Dialog ergänzt die rein quantitative Analyse mittels Qualitätsindikatoren durch eine fachliche Bewertung der Behandlungsqualität eines Leistungserbringers. Gegenstand dieses Kapitels sind die quantitativen Methoden der Datenanalyse.

Die grundlegende Voraussetzung für den Einsatz datenanalytischer Methoden ist eine modellartige Vorstellung der Struktur der zu analysierenden Daten und der relevanten Einflussfaktoren. Annahmen über die Einflussgrößen auf die zu analysierenden Messwerte bestimmen dabei die Fragen, die eine Auswertungsmethode beantworten soll und kann. So führt beispielsweise die Annahme, dass patientenseitige Faktoren wie Alter oder Morbidität ein Behandlungsergebnis systematisch beeinflussen, zur Wahl bestimmter Methoden der Risikoadjustierung für Ergebnisindikatoren. In diesem Abschnitt wird daher zunächst die Modellvorstellung des Zustandekommens von Daten der externen Qualitätssicherung nach §§ 136 ff. SGB V dargelegt und im Anschluss das prinzipielle Vorgehen bei zentralen datenanalytischen Methoden wie Risikoadjustierung (Abschnitt 13.2) und Indexbildung (Abschnitt 13.3) erläutert.

Die vorliegende Beschreibung der datenanalytischen Methoden muss allerdings zunächst vorläufig und auf Grundlagen beschränkt bleiben, da viele Fragestellungen erst in zukünftigen Entwicklungsarbeiten des Instituts entsprechend fundiert beantwortet werden können. Beispiele hierfür sind fortgeschrittene Verfahren der Risikoadjustierung von Qualitätsindikatoren oder die Entwicklung eines Auswertungskonzepts für Follow-up-Indikatoren (IQTIG 2017).

### 13.1 Grundlegende Annahmen

Ein zentrales Ziel von Qualitätsmessungen nach §§ 136 ff. SGB V ist die Messung und vergleichende Darstellung der Versorgungsqualität der Leistungserbringer. Dies beinhaltet meist auch den Vergleich der Leistungserbringer mit einem Referenzbereich (siehe Kapitel 10). Daher fokussiert die folgende Darstellung auf die Einflussgrößen, die bei einer Darstellung der Indikatorergebnisse von Leistungserbringern relevant sind.

---

<sup>27</sup> Während in der QSKH-RL der Begriff „rechnerische Auffälligkeit“ Verwendung findet, wird in der Qesü-RL allgemeiner von „Auffälligkeiten“ gesprochen. Zum besseren Verständnis wird hier der allgemeinere Begriff „Auffälligkeit“ verwendet.

Die Ergebnisse von Qualitätsmessungen können prinzipiell als komplexe Funktion verschiedener Einflussgrößen aufgefasst werden. Manche dieser Einflussgrößen lassen sich klar den Leistungserbringern zuordnen. Andere Einflussgrößen sind den Patientinnen und Patienten zuzuschreiben. Des Weiteren gibt es auch Einflussgrößen die weder klar den Leistungserbringern noch den Patientinnen und Patienten zuzuschreiben sind, welche hier als „weitere Einflüsse“ bezeichnet werden. Diese drei Gruppen von Einflussgrößen werden im Folgenden kurz dargestellt.

- **Leistungserbringer**

Die Strukturen und Handlungen seitens der Leistungserbringer, wie z. B. Personalausstattung, Equipment, Prozesse der Indikationsstellung sowie der medizinischen und pflegerischen Versorgung ergeben zusammen den Beitrag der Leistungserbringer zum Messergebnis. Dieser Einfluss ist es, der durch Qualitätsmessungen abgebildet werden soll – er stellt die Qualität der medizinischen Versorgung durch die Leistungserbringer dar.

- **Patientinnen und Patienten**

Darüber hinaus wirken sich auch patientenseitige Charakteristika auf das Ergebnis von Qualitätsmessungen aus. Unterschiede in der Zusammensetzung der Patientengruppen (Patienten-Mix) der Leistungserbringer können Unterschiede der Qualitätsergebnisse der Leistungserbringer bewirken und sind somit zu berücksichtigen. Dies ist beispielsweise der Fall, wenn Patientinnen oder Patienten mit mehreren Risikofaktoren bei einem untersuchten Ergebnisindikator geringere Heilungschancen haben als risikoarme Patientinnen und Patienten und wenn es Leistungserbringer gibt, die anteilig mehr risikobehaftete Patientinnen und Patienten behandeln als andere Leistungserbringer. Methoden der Risikoadjustierung (Details siehe Abschnitt 13.2) dienen dazu, Indikatormessergebnisse um solche systematischen, patientenseitigen Einflüsse – soweit möglich – zu bereinigen und damit einen faireren Vergleich der Ergebnisse der Leistungserbringer zu ermöglichen.

- **Weitere Einflüsse**

Weitere Einflüsse, welche auf Qualitätsmessungen wirken und nicht auf Patienteneinflüsse oder Leistungserbringer zurückgehen, können i. d. R. nicht gemessen und somit auch nicht berücksichtigt werden. Beispiele für solche Einflüsse sind externe Einflüsse, die die Leistungserbringer nicht zu vertreten haben, wie etwa überlastende, unvorhersehbare Situationen, etwa große Krankheitsausbrüche oder Naturkatastrophen. Auch können sporadische Produktionsfehler bei Medizinprodukten oder Eigenschaften von Spenderorganen bei Transplantationen teilweise als solche weiteren Einflüsse gesehen werden. In der Einzelsituation ist zu diskutieren, wie weit die Verantwortung des Leistungserbringers reicht, und inwieweit solche Faktoren gemessen und somit ggf. berücksichtigt werden sollen.

Nicht gemessene Patientenmerkmale und weitere Einflüsse zu denen keine Daten vorliegen und für welche entschieden wird, dass der Leistungserbringer nicht dafür in Verantwortung gezogen werden soll, werden oft vereinfachend unter der Kategorie „stochastische Einflüsse“ zusammengefasst (vgl. Iezzoni 2013b: 5).

Bei der Erfassung jedes einzelnen Behandlungsfalls wirken potenziell alle drei oben genannten Gruppen von Einflussgrößen und es ist zu bedenken, inwiefern beispielsweise das Versterben

bzw. Überleben einer Patientin oder eines Patienten der Qualität der Versorgung durch den Leistungserbringer, patientenseitigen Faktoren oder anderen Einflüssen (z. B. fehlerhaftes Medizinprodukt) zuzuschreiben ist.

### **Struktur-, Prozess- und Ergebnisindikatoren**

Das Ausmaß, mit dem die Einflussgrößen die Ergebnisse von Qualitätsmessungen bedingen, ist u. a. abhängig von der gemessenen Qualitätsdimension. So sind Messungen der Strukturqualität unabhängig von einzelnen Behandlungsfällen, d. h. es gibt keine fallbezogenen Einflüsse auf die Messung. Dagegen ist es meist plausibel, dass für Ergebnisindikatoren von Fall zu Fall variierende, vielfältige und teils unbekannte Einflüsse und Charakteristika der Patientinnen und Patienten eine Rolle spielen und auf den Indikatorwert eines Leistungserbringers einwirken. Für Prozessindikatoren kann dagegen die Frage, ob fallbezogene Einflüsse beim Zustandekommen des Messergebnisses plausibel sind und inwieweit diese berücksichtigt werden sollen, oft nur indikatorspezifisch beantwortet werden.

Sind systematische Einflüsse auf Indikatorergebnisse durch patientenseitige Charakteristika plausibel und nicht vernachlässigbar, werden Methoden der Risikoadjustierung eingesetzt.

## **13.2 Risikoadjustierung**

Unter der Rubrik Risikoadjustierung werden datenanalytische Verfahren subsumiert, die bei der Auswertung von Qualitätsindikatoren Unterschiede in der Zusammensetzung der Patientengruppen zwischen den verschiedenen Leistungserbringern berücksichtigen. Patientenseitige Einflüsse werden oft als Risikofaktoren bezeichnet (vgl. Iezzoni 2013a). So sind beispielsweise für multimorbide Patientinnen und Patienten häufig weniger positive Behandlungsergebnisse erreichbar als für weniger stark belastete Patientinnen und Patienten (Smith et al. 2012). Aber auch Faktoren wie Alter und Geschlecht der Patientinnen und Patienten sind Variablen, die häufig in relevantem und signifikantem Zusammenhang mit dem Behandlungsergebnis stehen (Smith et al. 2012). Dabei sind nicht das Alter oder das Geschlecht selbst der Risikofaktor, vielmehr sind beide Faktoren indirekt mit den eigentlichen Risiken verbunden. Da sich die Zusammensetzung der Patientengruppen zwischen den verschiedenen Leistungserbringern erheblich unterscheiden kann, kann dies zu entsprechenden Unterschieden in den beobachteten Indikatorwerten der Leistungserbringer führen, die nicht auf Qualitätsunterschiede zwischen den Leistungserbringern zurückzuführen sind. Um daher die Darstellung der Indikatorwerte der Leistungserbringer von Verzerrungspotenzial durch solche Unterschiede in der Zusammensetzung der Patientengruppen zu bereinigen, werden verschiedene Verfahren der Risikoadjustierung eingesetzt und aktuell auch weiterentwickelt.

Die Auswahl der verwendeten Methode ist dabei abhängig von der Anzahl und dem Zusammenspiel der einzelnen Risikofaktoren (bzw. deren Datenstruktur), sowie dem Skalenniveau der Risikofaktoren und dem des Indikators. Eine erste Auswahl potenzieller Risikofaktoren erfolgt dabei getrennt für jeden Qualitätsindikator schon bei der Entwicklung des Auswertungskonzeptes anhand einer ausführlichen Literaturrecherche und unter Einbeziehung der fachlich-inhaltlichen Expertise der entsprechenden Fachgremien. Die Auswahl und Operationalisierung der einzelnen

Risikofaktoren bei der Erhebung erfolgt dabei fachlich begründet und danach, ob eine zuverlässige Messung des einzelnen Faktors möglich ist. In diesem Rahmen wird auch die mögliche Beeinflussbarkeit der Risikofaktoren durch die Leistungserbringer berücksichtigt. Des Weiteren erfolgt in Zusammenarbeit mit Fachexpertinnen und -experten eine erste Einschätzung hinsichtlich des Zusammenwirkens verschiedener potenzieller Risikofaktoren auf das Indikatorergebnis. Außerdem kann die Relevanz der Prävalenz einzelner Faktoren und das erwartete Verteilungsmuster über die Leistungserbringer diskutiert werden.

Im Folgenden werden die aktuell verwendeten Verfahren zur Risikoadjustierung dargestellt. Diese lassen sich in Stratifizierung, additive Scores und Regressionsmodelle untergliedern. Insgesamt handelt es sich bei den verwendeten Methoden um klassische Verfahren der Biostatistik zur Standardisierung und Adjustierung von Störfaktoren (siehe z. B. Keiding und Clayton 2014).

### 13.2.1 Stratifizierung

Bei einer Risikoadjustierung mittels Stratifizierung wird die dem Qualitätsindikator zu Grunde liegende Grundgesamtheit zunächst eindeutig nach den zu berücksichtigenden Ausprägungen der Risikofaktoren aufgeteilt. Hierbei ist das Ziel, bezüglich der vorhandenen Patientenrisiken, in sich homogene und vergleichbare Untergruppen (Strata) zu bilden. Die Ergebnisse zu den Qualitätsindikatoren können dann in den Untergruppen getrennt berechnet und ausgewertet werden. Ein Vergleich zwischen Leistungserbringern innerhalb jedes Stratums ist damit unabhängig vom Einfluss der Stratifizierungsvariablen. Dieses Vorgehen wird exemplarisch anhand des Risikofaktors „Dringlichkeit des Eingriffs“ für den Qualitätsindikator „Anteil im Krankenhaus verstorbener Patientinnen und Patienten“ für isolierte koronarchirurgische Eingriffe in Tabelle 9 dargestellt.

Tabelle 9: Im Krankenhaus verstorbene Patientinnen und Patienten mit isoliertem koronarchirurgischem Eingriff geschichtet nach Dringlichkeit des Eingriffs im Jahr 2015

Dringlichkeit des Eingriffs	Anzahl im Krankenhaus verstorbene Patientinnen und Patienten	Anteil
Elektiv/dringlicher Eingriff	598 / 32.331	1,85 %
Notfalleingriff	506 / 5.776	8,76 %
Gesamt	1.104 / 38.107	2,90 %

Im Jahr 2015 verstarben von allen isoliert koronarchirurgisch behandelten Patientinnen und Patienten bundesweit 2,90 % während des Aufenthalts im Krankenhaus. In der Gruppe der elektiv/dringlichen Eingriffe ist der Anteil verstorbener Patientinnen und Patienten mit 1,85 % deutlich geringer als bei Patientinnen und Patienten mit Notfalleingriffen mit 8,76 %. Behandelt ein Leistungserbringer nun überdurchschnittlich viele Notfallpatientinnen und -patienten, ist hiermit auch eine insgesamt erhöhte Mortalitätsrate bei dieser Einrichtung zu erwarten, da bei Notfalleingriffen das Risiko entsprechend höher ist. Die erhöhte Mortalitätsrate kann nicht als Hinweis auf weniger gute Versorgungsqualität interpretiert werden. Ein getrennter Vergleich

innerhalb der Strata „elektiv/dringlicher Eingriff“ und „Notfalleingriff“ ermöglicht dagegen einen bezüglich der Schichtungsvariablen unverzerrten Vergleich zwischen den Einrichtungen.

Ein weiterer Vorteil dieser Methode ist, dass die Ergebnisse für unterschiedliche Patientengruppen unmittelbar sichtbar werden und sich somit ggf. ein unterschiedlicher Handlungsbedarf für die verschiedenen Risikogruppen erkennen lässt. Auch die Bewertungskonzepte einschließlich Referenzbereiche (siehe Kapitel 10) lassen sich spezifisch für bestimmte Risikogruppen festlegen oder auf besonders relevante Patientengruppen beschränken.

Diese Methode ist jedoch kritisch zu bewerten, wenn mehrere Risikofaktoren gleichzeitig zu berücksichtigen sind. Die vollständige Anzahl möglicher Untergruppen ergibt sich dann als Produkt der Anzahl der Ausprägungen über alle Risikofaktoren (Kreienbrock et al. 2012). Sind z. B. neben den zwei oben dargestellten Ausprägungen zur Dringlichkeit des Eingriffs zusätzlich drei Kategorien für den Risikofaktor „linksventrikuläre Ejektionsfraktion“ ( $\leq 30\%$ ; 31-50%;  $> 50\%$ ) und das Geschlecht der Patientinnen und Patienten als Risikofaktor zu berücksichtigen, steigt die Anzahl der möglichen Strata auf zwölf ( $2 \times 3 \times 2$ ). Die Fallzahl der Patientinnen und Patienten in den einzelnen Strata kann mit steigender Anzahl an Risikofaktoren soweit ausgedünnt werden, dass ein Vergleich nur noch für Patientinnen und Patienten in stark vertretenen Risikogruppen sinnvoll erscheint oder gar unmöglich wird. Als Alternativen bieten sich dann additive Scores oder Regressionsverfahren zur Risikoadjustierung an. Letztere erlauben auch eine flexiblere Handhabung von stetig gemessenen Risikofaktoren gegenüber der Stratifizierung, welche immer kategoriale Risikofaktoren voraussetzt.

### **Additive Scores**

Additive Scores zur Risikoadjustierung werden gebildet, indem den Ausprägungen mehrerer Risikofaktoren jeweils Punktwerte zugeordnet werden, die dann für jede Patientin und jeden Patienten über die Risikofaktoren addiert werden, sodass am Ende für jede Patientin und jeden Patienten statt mehrerer Risikovariablen nur ein additiver Risikopunktwert steht. Höhere Punktwerte sind dabei in der Regel mit einem erhöhten Patientenrisiko assoziiert, und zwar derart, dass eine höhere Summe über alle Risikofaktoren ebenfalls ein schlechteres Ergebnis für einen Qualitätsindikator erwarten lässt. Die zu berücksichtigenden Risikofaktoren sowie das Bewertungssystem nach Punkten werden in der Regel aus der Literatur entnommen.

Beispielsweise wird im Rahmen des QS-Verfahrens *Ambulant erworbene Pneumonie* der Schweregrad der Pneumonie einer Patientin oder eines Patienten mittels CRB-65-Score (Lim et al. 2003) bestimmt. Als risikoerhöhende Faktoren gelten eine pneumoniebedingte Desorientiertheit, eine spontane Atemfrequenz  $\geq 30/\text{min}$ , ein Blutdruck von systolisch  $< 90\text{mmHg}$  oder von diastolisch  $\leq 60\text{mmHg}$  sowie ein Alter  $\geq 65$  Jahre. Für jeden bei einer Patientin oder einem Patienten auftretenden Risikofaktor wird ein Punkt vergeben, sodass in der Summe für jede Patientin und jeden Patienten zwischen 0 Punkte (geringster Schweregrad) und 4 Punkte (höchster Schweregrad) vergeben werden können.

Im nächsten Schritt kann dann ein risikoadjustierter Vergleich der Leistungserbringer anhand der ermittelten Punktwerte bzw. nach einer Einteilung der Punktwerte in Risikoklassen erfolgen,

wobei hierzu Stratifizierungsverfahren verwendet werden. So werden für Patientinnen und Patienten mit ambulant erworbener Pneumonie die Ergebnisse zur Frühmobilisation nach der Aufnahme im Krankenhaus getrennt nach Risikoklassen des CRB-65-Score betrachtet.

Wesentliche Vorteile dieser Methode sind eine im Vergleich zur Anzahl der Risikofaktoren weiterhin überschaubare Anzahl an möglichen Risikoklassen, sowie die Möglichkeit auf in der Literatur bereits etablierte additive Risiko-Scores (z. B. CRB-65) zurückzugreifen. Eigene datengestützte Analysen der Risikofaktoren sind daher bei diesem Verfahren nicht unbedingt notwendig, sollten aber dennoch im Nachgang zur Überprüfung einer Übertragbarkeit für die Messung des jeweiligen Qualitätsindikators genutzt werden.

### 13.2.2 Regressionsmodelle

Sind viele potenzielle Einflussfaktoren bei der Messung eines Qualitätsindikators gemeinsam zu berücksichtigen, stellen multiple Regressionsmodelle eine geeignete Methode dar, um Risikofaktoren zu identifizieren und den gemeinsamen Einfluss auf eine Zielvariable zu quantifizieren. Liegt die Zielvariable des Qualitätsindikators auf binärem Messniveau vor (z. B. das Ereignis tritt ein oder das Ereignis tritt nicht ein), kommen hier vor allem multiple logistische Regressionsmodelle zum Einsatz (Hosmer et al. 2013). Dabei wird das binäre Ereignis für jede Patientin und jeden Patienten nicht direkt modelliert, sondern es wird eine Funktion der Eintrittswahrscheinlichkeit  $P$  für das Ereignis in Abhängigkeit der Risikofaktoren modelliert.

Um mit dieser Methode die Indikatorwerte der Leistungserbringer zu adjustieren, werden im ersten Schritt auf Basis der Daten aller Patientinnen und Patienten im gewählten Auswertungszeitraum – ohne Berücksichtigung der Leistungserbringer – die Regressionskoeffizienten bestimmt. Im nächsten Schritt wird durch Umrechnung zur logistischen Verteilungsfunktion für jede Patientin und jeden Patienten die individuelle Wahrscheinlichkeit  $P$  für das Eintreten des Ereignisses bei der Patientin oder beim Patienten unter Berücksichtigung seiner gemessenen Risikofaktoren bestimmt. Diese Ergebnisse werden dann im weiteren Verlauf zur Berechnung der risikoadjustierten Indikatorwerte zusammengefasst. So ergibt sich die für einen Leistungserbringer *erwartete* Anzahl an Ereignissen (z. B. verstorbene Patientinnen und Patienten) aus der Summe aller modellbasierten Wahrscheinlichkeiten seiner Patientinnen und Patienten für das Ereignis. Dadurch wird das Risikoprofil über alle Patientinnen und Patienten der Einrichtung berücksichtigt. Die erwartete Anzahl an Ereignissen  $E$  lässt sich dann mit der beobachteten Anzahl an Ereignissen ( $O$ ) vergleichen. Als risikoadjustierte Vergleichskennzahlen zwischen den Einrichtungen stehen z. B. die Differenz zwischen beobachteter und erwarteter Rate ( $O/n - E/n$ ) mit  $n$  = Anzahl der Fälle der Einrichtung, das Verhältnis ( $O/E$ ) oder die risikoadjustierte Rate ( $O/E$ ) $\times(O_G/n_G)$  [mit  $n_G$  = Anzahl der Fälle Gesamt] zur Verfügung (Ash et al. 2013).

Die Differenz  $O/n - E/n$  kann als Abweichung von der erwarteten Ereignisrate betrachtet werden, wobei positive Werte als Prozentpunkte oberhalb der erwarteten Ereignisrate und negative Werte als Prozentpunkte unterhalb der erwarteten Ereignisrate zu interpretieren sind. Über das Verhältnis  $O/E$  lässt sich erkennen, um welchen Faktor die beobachtete von der erwarteten Ereignishäufigkeit abweicht. Da hierbei die beobachtete Rate in Bezug zur erwarteten Rate einer Standardpopulation gesetzt wird, entspricht die Interpretation hierbei dem Vorgehen einer

standardisierten Mortalitäts- oder Morbiditätsrate (SMR) bzw. einer indirekten Standardisierung. Das Verhältnis O/E ist als Faktor einer erhöhten ( $O/E > 1$ ) bzw. verminderten ( $O/E < 1$ ) beobachteten Ereignisrate im Vergleich zur erwarteten Rate in der Vergleichspopulation aufzufassen.

Die Risikoadjustierung mittels Regressionsmodellierung hat gegenüber der Methode der Stratifizierung den Vorteil, dass für den risikoadjustierten Vergleich von O/E die Grundgesamtheit des Qualitätsindikators nicht in weitere Subgruppen unterteilt werden muss und somit direkt anhand einer Kennzahl beurteilt werden kann. Dabei kann eine große Anzahl an Risikofaktoren gleichzeitig und mit ihrem gemeinsamen Einfluss auf das Ergebnis berücksichtigt werden. Die Risikofaktoren müssen nicht notwendigerweise kategorisiert werden, da nicht nur diskrete, sondern auch Einflussfaktoren mit stetigem Messniveau modelliert werden können. Die Unterschiede im Risiko auf das Ereignis zwischen zwei unterschiedlichen Kombinationen von Einflussfaktoren sind als Odds-Ratio gut interpretierbar und dabei relativ einfach zu ermitteln. Demgegenüber steht ein erhöhter Aufwand im Rahmen der Verfahrenspflege (siehe Abschnitt 5.2), wenn Regressionskoeffizienten jährlich neu berechnet werden, und dementsprechend Anpassungen der Rechenregeln notwendig sind, sowie ein erhöhter Aufwand bei der Dokumentation durch die Leistungserbringer, die zusätzliche Informationen erfassen und dokumentieren müssen.

### 13.2.3 Zusammenfassung und Ausblick

In der vorliegenden Fassung der „Methodischen Grundlagen“ werden lediglich Grundbegriffe der zum Zeitpunkt der Erstellung eingesetzten Methoden der Risikoadjustierung eingesetzt. Wichtige Aspekte der gegenwärtig bereits verwendeten Verfahren werden derzeit durch das Institut wissenschaftlich weiterentwickelt. Dazu zählen folgende Problemstellungen: Wie werden die notwendigen Variablen einer Risikoadjustierung mit noch größerer Verlässlichkeit gefunden und wie kann die Güte eines Modells zur Risikoadjustierung hinsichtlich der Ziele der Risikoadjustierung beurteilt werden? Bei der Wahl der Risikofaktoren ist wichtig, dass alle wichtigen aus der Literatur bekannten Faktoren grundsätzlich die Möglichkeit haben, in eine Risikoadjustierung mit einzugehen. Andererseits müssen die Risikofaktoren auch tatsächlich im Rahmen der Verfahren nach §§ 136 ff. SGB V messbar sein. Die Variablenselektion erfolgt aus einer Kombination von inhaltlichen Überlegungen und datengestützten Modelselektionsverfahren.

Zu einem besseren Verständnis der Abhängigkeitsstrukturen zwischen Risikofaktoren und Indikator wird die Häufigkeit und der Einfluss der einzelnen potenziellen Risikofaktoren als auch der gemeinsame Einfluss analysiert. Die Variablenselektion bei logistischen Regressionsmodellen betrachtet dabei nicht nur den p-Wert eines Signifikanztests auf zusätzlichen Einfluss eines Risikofaktors über die berücksichtigten Faktoren hinaus. Auch Werte auf Grundlage der Likelihood-Funktion, wie z. B. Akaikes Informationskriterium sowie Shrinkage-Verfahren, werden für eine Abwägung zwischen Modellanpassung und Modellkomplexität herangezogen. Für die Beurteilung der Modellanpassung und Prognosegüte werden Methoden wie z. B. der Brier-Score, der Hosmer-Lemeshow-Test oder die c-Statistik berücksichtigt.

Insgesamt kann die Wahl eines adäquaten Risikoadjustierungsmodells zwar zu einem gewissen Grad standardisiert werden, aber eine auf jede Fragestellung oder Situation passende, allgemeingültige Beurteilung der Modellgüte erscheint nicht sinnvoll. Damit wird letztendlich immer auch ein Grad an inhaltlich begründetem subjektivem Spielraum notwendig. Umso wichtiger ist es, z. B. die Definition der Risikofaktoren oder die Berechnungsmodelle zur Risikoadjustierung öffentlich zur Verfügung zu stellen und die entsprechende Transparenz beim gewählten Vorgehen herzustellen. Um die vergleichenden Darstellungen der Leistungserbringer durch Risikoadjustierung fair zu gestalten, besteht insgesamt eine Austauschbeziehung zwischen der Erfassung von wichtigen Variablen vs. dem damit verbundenen erhöhten Aufwand bei der Dokumentation durch die Leistungserbringer.

Abschließend ist anzumerken, dass die Risikoadjustierung auch dazu beiträgt, Fehlanreize im Sinne der Patientenselektion zu minimieren. Jedoch können solche Effekte trotz detaillierter Adjustierungen nicht ausgeschlossen werden und sind somit entsprechend sorgfältig bei der Entwicklung mitzudiskutieren.

### **13.3 Aggregation mehrerer Qualitätsmessungen zum Zweck der Qualitätsförderung**

Einzelne Qualitätsindikatoren und Kennzahlen geben meist nur einen fragmentierten Einblick in die Qualität des Versorgungsgeschehens. Um einen umfassenden Eindruck von der Versorgungsqualität zu erhalten, müssen in der Regel mehrere Indikatoren und Kennzahlen zusammen betrachtet und bewertet werden. Mit steigender Anzahl an Qualitätsindikatoren und Kennzahlen wird dies jedoch zunehmend schwieriger. Dieses Problem wird zusätzlich verschärft, wenn Patientinnen und Patienten über Qualität informiert werden sollen oder wenn beispielsweise eine regulative Entscheidung auf Basis einer Vielzahl von Indikatoren getroffen werden soll. Vor allem für Patientinnen und Patienten kann es mitunter eine Herausforderung sein, aus einer großen Menge von Indikatoren die für sie relevanten zu finden und die Informationen entsprechend zu gewichten, um zu einer Entscheidung zu kommen, bei welchem Leistungserbringer sie die beste Qualität erwarten können. In derartigen Situationen, in denen eine umfassende, über einen einzelnen Qualitätsindikator hinausgehende Betrachtung der Versorgungsqualität im Vordergrund steht, eignen sich aggregierte Maße, die Qualitätsinformationen über mehrere Dimensionen hinweg aggregiert darstellen (Goddard und Jacobs 2010). Zu diesem Zweck stehen zwei Methoden zur Verfügung, die Bildung hierarchischer Ergebnisgrößen (siehe Abschnitt 13.3.1) und die Indexbildung (siehe Abschnitt 13.3.2), die in diesem Abschnitt erläutert werden.

Qualitätsmessungen lassen sich prinzipiell immer als aggregierte Messungen auffassen. So werden etwa bei der Ermittlung der mittleren Operationsdauer bei einem Leistungserbringer die Einzelwerte der Operationen aller Patientinnen und Patienten dieses Leistungserbringers in einem bestimmten Zeitraum (meist ein Jahr) aggregiert. Dementsprechend geht bei der Aggregation über den Betrachtungszeitraum eines Jahres die Information über zeitliche Schwankungen eines Indikators innerhalb des Jahres verloren. Dies geschieht jedoch zugunsten der übersichtlichen Darstellung für das Betrachtungsjahr. Daher sind Qualitätsindikatoren und Kennzahlen

immer als Aggregatmaße zu verstehen, bei denen über eine oder mehrere Dimensionen zu Gunsten einer übersichtlicheren Betrachtung aggregiert wird.

Grundsätzlich können in der medizinischen Qualitätsmessung folgende Aggregationsdimensionen unterschieden werden, die auch in Kombination vorkommen können:

- Über Behandlungsfälle (z. B. Mittelwertbildung innerhalb von Leistungserbringern)
- Über Zeitspannen (z. B. über den Erfassungszeitraum eines Jahres)
- Über Leistungserbringer (z. B. Mittelwertbildung für Bundesauswertungen)
- Über Indikatoren und/oder Kennzahlen (z. B. Indexbildung)

Viele dieser Aggregationen werden implizit vorgenommen, wie beispielsweise bei der Berechnung eines mittleren Indikatorwerts pro Leistungserbringer für den Betrachtungszeitraum von einem Jahr. Die explizite Darstellung dieser Aggregationsdimensionen macht deutlich, dass durch jede Aggregation eine übersichtlichere Darstellung zugunsten eines Informationsverlusts erkauft wird. Dies kann für eine umfassendere Betrachtungsweise erwünscht sein oder aber für die Ableitung konkreter Handlungen unerwünscht sein. So kann beispielsweise ein Index „Partizipative Entscheidungsfindung“ Auskunft geben, ob das Ausmaß an Patientenbeteiligung bei einem Leistungserbringer ähnlich ist wie im Bundesdurchschnitt. Um jedoch im Rahmen der Qualitätsförderung konkrete Verbesserungsmöglichkeiten für diesen Leistungserbringer in diesem Qualitätsbereich abzuleiten, ist ein solch globales Maß weniger geeignet als Einzelindikatoren zu konkreten Qualitätsmerkmalen. Die zweckmäßige Aggregationsebene ist folglich notwendigerweise vom geplanten Einsatzzweck des Indikators/Index abhängig und kann nur vor diesem Hintergrund beurteilt werden. Es wird daher geprüft, ob die Bildung eines oder mehrerer Indizes oder einer oder mehrerer hierarchischer Ergebnisgrößen für die gegebene Fragestellung der Qualitätssicherung zweckmäßig und umsetzbar ist. Im Folgenden werden diese beiden Methoden näher erläutert.

### 13.3.1 Hierarchisierung von Ergebnisgrößen

Häufig stehen mehrere Ergebnisindikatoren eines QS-Verfahrens in einer systematischen Beziehung, beispielsweise wenn mehrere Indikatoren Komplikationen aufsteigenden Schweregrads nach einem Eingriff darstellen. Dies wäre etwa der Fall bei den Indikatoren „Entzündung“ (ja/nein), „Sepsis“ (ja/nein) und „Tod“ (ja/nein) als eine Reihe von Ergebnisindikatoren, die Komplikationen nach einem Eingriff erfassen. Dabei können die weniger schweren Komplikationen zum Teil als Vorstufen der schwereren Komplikationen verstanden werden. Um eine derart hierarchisch strukturierte Menge von Qualitätsindikatoren zusammenzufassen ist es möglich, die Indikatorwerte auf Patientenebene zu einer ordinalen Ergebnisgröße zu aggregieren. Im obigen Beispiel könnte die Ergebnisgröße „Schweregrad der postoperativen Komplikationen“ mit den vier Stufen (A) „keine Komplikation“, (B) „Entzündung“, (C) „Sepsis“, (D) „Versterben“ gebildet werden. Im Gegensatz zur unabhängigen Betrachtung der einzelnen Qualitätsindikatoren bietet eine derartige Zusammenfassung von Indikatoren insbesondere im Hinblick auf die Durchführung eines Strukturierten Dialoges den Vorteil, dass auffällige Indikatorwerte in einem gemeinsamen Kontext dargestellt werden und für jeden Behandlungsfall nur einmal betrachtet werden

müssen. Weiterhin wird eine Mehrfachauslösung des Strukturierten Dialoges vermieden, da die (rechnerische) Auffälligkeit anhand nur eines Wertes beurteilt wird.

Ähnlich wie bei anderen Formen der Aggregation von Variablen erfordert ein Zusammenfassen der Ergebnisse auf Leistungserbringerebene immer eine Gewichtung der konstituierenden Variablen (siehe auch Abschnitt 13.3.2). Beispielsweise ließe sich der mittlere Schweregrad der Komplikationen eines Leistungserbringers nur berechnen, wenn den Stufen (A) bis (D) explizite Gewichte zugeschrieben werden. Die Bestimmung solcher Gewichte kann einerseits datengestützt erfolgen, etwa durch Regression auf einen bestimmten Endpunkt, oder andererseits durch Expertinnen und Experten nach fachlich-heuristischen Aspekten.

Hängen die betrachteten Qualitätsindikatoren auch von den individuellen Voraussetzungen der Patientinnen und Patienten ab, erfordert deren Aggregation zu einem ordinalen Indikatorwert eine Risikoadjustierung. Ähnlich wie bei einfachen Ergebnisindikatoren besteht hier die Möglichkeit der Regressionsmodellierung oder Stratifizierung (siehe Abschnitt 13.2). Zur regressionsbasierten Risikoadjustierung ordinaler Ergebnisgrößen stehen multikategoriale Regressionsmodelle zur Verfügung (Fahrmeir und Tutz 2010). Ein Spezialfall dieser übergeordneten Modellklasse bilden ordinale Regressionsmodelle, die genutzt werden können, um eine patientenindividuelle Wahrscheinlichkeitsverteilung für eine ordinale Ergebnisgröße zu modellieren. Analog zur Rolle der logistischen Regression bei dichotomen Ergebnissen, können derartige Regressionsmodelle genutzt werden, um beobachtete und erwartete Ergebnisse zu vergleichen. Auch dies erfordert allerdings eine Gewichtung der ordinalen Stufen.

Derartige Ansätze sind abzugrenzen von der im folgenden Abschnitt beschriebenen Methode zur Bildung von Indizes, da ordinale Ergebnisgrößen eine Hierarchie der konstituierenden Indikatoren voraussetzen und damit der Beschreibung einer *eindimensionalen* latenten Variable dienen. Indizes hingegen sind inhärent multidimensional und setzen keine kausale oder hierarchische Struktur der einzelnen Indikatoren voraus.

### 13.3.2 Indizes in der Qualitätsmessung

Das grundsätzliche Ziel bei der Aggregation mehrerer Indikatoren und/oder Kennzahlen zu einem Index ist die gebündelte und verständliche Darstellung einer ansonsten schwer überschaubaren Zahl an Informationen. Im Folgenden wird von einem Index (im Englischen auch als „Composite Indicator“, „Compound Measure“ etc. bezeichnet) gesprochen, wenn mehrere Variablen (hier: Qualitätsindikatoren oder Kennzahlen) zu einer neuen Variable, dem Index, zusammengefasst werden (Schnell et al. 2013: 156). Dabei stellt diese neue Variable zunächst eine Kennzahl dar. Soll ein solcher Index auch eine Bewertung der Leistungserbringer anhand eines Referenzbereichs ermöglichen, wird dieser Index selbst zu einem Qualitätsindikator. Indikatoren und/oder Kennzahlen können also selbst wieder zu einem Kennzahl-Index oder einem Qualitätsindikator-Index zusammengefasst werden. Entsprechende methodische Anforderungen sind an den Index zu stellen. In diesem Abschnitt wird die Vorgehensweise bei der Bildung von Indizes dargestellt, die über mehrere Indikatoren und/oder Kennzahlen aggregieren. Zunächst werden jedoch grundlegende Konzepte zur Dimensionalität von Indizes erläutert.

### Dimensionalität von Indizes

Bei der Aggregation mehrerer Indikatoren und/oder Kennzahlen zu einer übergeordneten Größe wird zwischen eindimensionalen und mehrdimensionalen Indizes unterschieden (Bühner 2011, Diekmann 2007, Shwartz et al. 2015). Eindimensionale Indizes werden z. B. in der Psychometrie verwandt und sind dadurch definiert, dass alle verrechneten Variablen Manifestationen des gleichen eindimensionalen latenten Konstrukts darstellen (z. B. Konzentrationsfähigkeit). Unter Eindimensionalität versteht man die Annahme, dass für die Ausprägung jeder Variable, die zur Indexbildung herangezogen wird, nur das zugrunde liegende latente Konstrukt verantwortlich ist, das vom Index abgebildet werden soll. Wenn diese Voraussetzung erfüllt ist, können Unterschiede zwischen Indexwerten eindeutig unterschieden in der Ausprägung des latenten Konstrukts zugeschrieben werden. Dies ist beispielsweise der Fall bei Konzentrationsfähigkeitstests: Unterschiede in den Testwerten können eindeutig auf Unterschiede in der Konzentrationsfähigkeit zurückgeführt werden; Unterschiede in anderen Fähigkeiten, wie etwa Merkfähigkeit oder Sprachgewandtheit sind als Alternativerklärungen durch die Eindimensionalität des Tests ausgeschlossen (Bühner 2011, Shwartz et al. 2015).

Mehrdimensionalität bei Indizes bedeutet, dass Merkmalsausprägungen aus verschiedenen Quellen, Bereichen oder Domänen zusammengefasst werden, die nicht durch ein einzelnes zugrunde liegendes, latentes Konstrukt verursacht sind. Ein Beispiel hierfür wäre ein Index, der Indikatoren für die Qualität der Versorgungsprozesse, Indikatoren für Patientenerfahrungen und Ergebnisindikatoren zusammenfasst. Die in einen solchen Index eingehenden Variablen stellen somit keine Manifestationen *eines* zugrunde liegenden Konstrukts dar, sondern decken sehr heterogene Dimensionen der Qualitätsmessung ab, die nicht unbedingt statistisch miteinander zusammenhängen müssen. Dieses Beispiel macht auch auf einen wichtigen Aspekt mehrdimensionaler Indizes aufmerksam: Da unterschiedliche Dimensionen in ein Maß aggregiert werden, muss eine Gewichtung dieser Dimensionen vorgenommen werden und die spätere inhaltliche Aussagekraft des Index hängt maßgeblich von der Gewichtung der eingehenden Variablen ab.

Um zu inhaltsvaliden, verständlichen und nützlichen Indizes zu gelangen, wird ein Konzept verfolgt, das aus vier Schritten bei der Entwicklung von Indizes (vgl. OECD 2008) besteht, welches im Folgenden dargestellt wird.

#### ■ Theoretisches Indexmodell

Im ersten Schritt der Indexbildung wird das zu messende Konstrukt (z. B. „Indikationsstellung“) und seine konstituierenden Dimensionen (z. B. „Partizipative Entscheidungsfindung“, „Nicht-invasive Diagnostik“, etc.) spezifiziert. Grundsätzlich schließt dieses Konzept auch die Konstruktion aspektübergreifender Indizes ein. Da bei mehrdimensionalen Indizes die Inhaltsvalidität des Index maßgeblich von den in den Index eingehenden Indikatoren und/oder Kennzahlen abhängt, ist die präzise Definition des zu operationalisierenden Konstrukts und seiner Dimensionen im ersten Schritt unerlässlich. Die Entwicklung des theoretischen Indexmodells ist dabei von der inhaltlichen Fragestellung, die der Index beantworten soll (z. B. „Wie gut ist die Indikationsstellung?“), geleitet und zunächst unabhängig von den zur Verfügung stehenden Daten. Damit wird der Grundstein dafür gelegt, dass der zu entwickelnde Index später inhaltlich sinnvoll interpretierbar ist.

## ■ Variablenselektion und Normalisierung

Im nächsten Schritt wird geprüft, welche Indikatoren und Kennzahlen für die Indexkonstruktion zur Verfügung stehen bzw. entwickelt werden können. Da die Güte eines Index maßgeblich von der Güte der eingehenden Indikatoren und/oder Kennzahlen abhängt, werden nur solche Indikatoren berücksichtigt, die die Anforderungen an Indikatoren bzw. Kennzahlen erfüllen (siehe Kapitel 8). Der Abgleich mit den nach dem theoretischen Indexmodell erforderlichen Indikatoren und den zur Verfügung stehenden und/oder entwickelbaren Indikatoren erlaubt an dieser Stelle eine erste Richtungsentscheidung. Zeigt sich, dass nicht für alle dem Index theoretisch zugerechneten Dimensionen Indikatoren oder Kennzahlen zur Verfügung stehen, wird abgewogen, ob die Vorteile eines Index überwiegen und ob trotz suboptimaler Datenlage mit der Indexkonstruktion fortgeschritten werden soll.

Handelt es sich bei den ausgewählten Indikatoren um Variablen, die auf unterschiedlichen Skalen vorliegen (z. B. Komplikationsrate [%] und Kontrastmittelmenge [ml]) oder die unterschiedliche Varianzen aufweisen, müssen diese zuerst in eine einheitliche Metrik überführt werden. Ohne die Anpassung der Metrik der Indikatoren würden diese implizit anhand ihrer Varianz gewichtet werden – Indikatoren mit größerer Varianz erhielten automatisch ein größeres Gewicht im Index (OECD 2008, Shwartz et al. 2015). Übliche Methoden zur Normalisierung sind beispielsweise z-Standardisierung und Spannweiten-Standardisierung.

## ■ Gewichtung der Dimensionen und Aggregation

Im nächsten Schritt müssen die konstituierenden Variablen für die Aggregation gewichtet werden. Da der Index mehrdimensional ist, hängt die inhaltliche Interpretation des Index entscheidend von der Gewichtung der Dimensionen, d. h. der einzelnen Variablen, ab. Da die Gewichtung selten eindeutig anhand des theoretischen Indexmodells möglich sein wird, werden zur Ermittlung der Gewichte Skalierungsverfahren mit Expertinnen und Experten eingesetzt, wie etwa die Budget-Allocation-Methode (OECD 2008). Der Einsatz solcher expertenbasierter Gewichtungen führt zu validen Indizes, sofern der Prozess der Gewichtung und die resultierenden Gewichte transparent dargestellt werden – dies stellt das Institut in seinen Berichtsformaten sicher.

Während zur Überprüfung der Güte von eindimensionalen Indizes Methoden der psychometrischen Messtheorie wie etwa konfirmatorische Faktorenanalyse eingesetzt werden können, lässt sich die Güte mehrdimensionaler Indizes nicht anhand messtheoretischer Eigenschaften (z. B. Einfachstruktur und unkorrelierte Fehler in einer Faktorenanalyse) bestimmen, sondern allein anhand ihrer Inhaltsvalidität. Beispielsweise hängt die inhaltliche Güte eines Index, der die Qualität der Patientenaufklärung abbilden soll, allein davon ab, ob alle relevanten Aspekte von Patientenaufklärung darin eingehen und dass keine irrelevanten Dimensionen berücksichtigt wurden. Dabei müssen die einzelnen Dimensionen jedoch nicht unbedingt hoch miteinander korreliert sein – dies wäre ein Merkmal von Eindimensionalität. Die inhaltliche Güte eines mehrdimensionalen Index wird daran gemessen, inwieweit sich die bei der Entwicklung des theoretischen Indexmodells als relevant erachteten Dimensionen später auch im Index als konstituierende Variablen wiederfinden. Dabei gibt das theoretische Indexmodell das gewünschte Optimum vor. Da allerdings häufig mit Einschränkungen in der Datenverfügbarkeit

und ggf. Datenqualität zu rechnen ist, wird ein Index immer einen Kompromiss zwischen dem gewünschten Indexmodell und dem Realisierbaren darstellen.

#### ▪ **Indexevaluation**

Mit dem Festlegen der Gewichte für die konstituierenden Variablen ist der Index fertig konstruiert. Mit dem Vorliegen erster Daten werden dann im letzten Schritt Analysen zur kritischen Prüfung des Index durchgeführt. Diese Analysen dienen als Qualitätssicherungsschritte und beinhalten zum Beispiel Prüfungen der Korrelationsstruktur der konstituierenden Variablen untereinander, Analysen der Variable-Index-Korrelationen (analog zu Trennschärfeanalysen) sowie Sensitivitätsanalysen der Gewichte. Gerade letzteres ermöglicht eine Abschätzung der Stabilität der Werte der Leistungserbringer auf dem Index in Abhängigkeit von den gewählten Gewichten für die konstituierenden Variablen. Ein weiterer wichtiger Schritt ist die Konstruktion und Darstellung von Vertrauensintervallen für fallzahlabhängige Indizes.

Zusammenfassend stellt das hier dargestellte Konzept zur Indexbildung sicher, dass zweckmäßige und vor allem inhaltsvalide Indizes konstruiert werden. Darüber hinaus wird durch intuitiv verständliche Indexbenennungen und Polung der Indizes sichergestellt, dass diese für die Zielgruppe interpretierbar und verständlich sind. So sollten etwa höhere Werte auf einem fiktiven Index „Indikationsstellung“ beispielsweise eine bessere Indikationsstellung bedeuten und niedrigere Werte eine schlechtere Indikationsstellung bedeuten. Nach diesen Maßstäben konstruierte Indizes ermöglichen eine integrative Darstellung der Qualität in der Gesundheitsversorgung und verhindern eine fragmentierte Betrachtung. Wie oben beschrieben, wird für jeden entwickelten Index geprüft, ob er dem Zweck des QS-Verfahrens methodisch gerecht wird. So kann ein übergreifender Index über eine Vielzahl von Indikatoren für die Zuordnung von Zu- oder Abschlägen im Rahmen von QS-Verfahren mit einem regulierenden Zweck sinnvoll sein, während er für die gezielte Qualitätsförderung von nur geringem Wert ist, da einzelne Probleme und Verbesserungspotenziale in der Gesamtschau nicht mehr erkennbar sind.

Somit ist bei der Indexbildung insbesondere die Art der Darstellung der Ergebnisse wichtig, um trotz der Informationskondensierung dem Empfänger – sei es eine Patientin / ein Patient, ein Leistungserbringer oder eine den Strukturierten Dialog durchführende Stelle – nachvollziehbare und verständliche Inhalte zu bieten. Wie bei allen datenanalytischen Verfahren spielt hier die zweckgebundene Praxistauglichkeit bei der Auswahl geeigneter Methoden und Darstellungsformen eine große Rolle.

Der Einsatz von adäquaten datenanalytischen Verfahren erfordert generell eine Abwägung zwischen Komplexität und fachlicher Genauigkeit auf der einen und Umsetzbarkeit, Reproduzierbarkeit und Kommunizierbarkeit auf der anderen Seite. Dies sollte jedoch nicht dazu führen, dass zugunsten vermeintlicher Transparenz auf wissenschaftlich unzureichend fundierte Verfahren zurückgegriffen wird. Es kann durchaus angemessen sein, die Zielgruppe mit komplexeren Methoden zu fordern und diese dann mit entsprechendem Informationsmaterial und Dialog zu begleiten.

# Glossar

## **Auffälligkeit, rechnerische (auffälliger Indikatorwert)**

Bezeichnet in der externen Qualitätssicherung nach §§ 136 ff. SGB V den Fall, dass der Indikatorwert eines Leistungserbringers außerhalb des Referenzbereichs für diesen Indikator liegt. Rechnerische Auffälligkeiten, denen eine statistisch signifikante Abweichung vom Referenzbereich zugrunde liegt, werden auch kurz als „statistische Auffälligkeit“ bezeichnet.

## **Datenvalidität**

Die Datenvalidität beschreibt, ob die zur Berechnung des Indikatorwerts verwendeten Daten den Vorgaben des Indikators hinsichtlich Vollzähligkeit, Vollständigkeit und Dateninhalten entsprechen.

## **Gemeinsamer Bundesausschuss (G-BA)**

Oberstes Beschlussgremium der gemeinsamen Selbstverwaltung der Ärzte und Ärztinnen, Zahnärzte und Zahnärztinnen, Psychotherapeuten und Psychotherapeutinnen, Krankenhäuser und Krankenkassen in Deutschland. Er bestimmt unter anderem den Leistungskatalog der Gesetzlichen Krankenversicherung und beschließt die Qualitätssicherungsmaßnahmen nach §§ 136 ff. SGB V.

## **Indikatorwert**

Das numerische Ergebnis, das durch Anwendung der Rechenregeln eines Qualitätsindikators auf die qualitätsbezogenen Daten resultiert.

## **Interessenkonflikt**

Gegebenheiten, die ein Risiko dafür schaffen, dass professionelles Urteilsvermögen oder Handeln, welche sich auf ein primäres Interesse beziehen, durch ein sekundäres Interesse unangemessen beeinflusst werden.

## **Kennzahl**

Im Rahmen der externen Qualitätssicherung nach §§ 136 ff. SGB V eine numerische Größe, die auf Grundlage von qualitätsbezogenen Daten errechnet wird.

## **Konfidenzintervall (Vertrauensbereich)**

Bereich um einen empirisch bestimmten Wert (z. B. Indikatorwert), innerhalb dessen unter Berücksichtigung aller zufälligen Einflüsse mit einer bestimmten Wahrscheinlichkeit (meist festgelegt als 95 %) der zugrunde liegende Wert liegt.

**Konstrukt, latentes**

Ein nicht durch direkte Beobachtung zugänglicher Sachverhalt innerhalb einer Theorie (z. B. Intelligenz, Lebensqualität).

**Leistungserbringer**

Einrichtungen (z. B. Krankenhäuser, Praxen, Medizinische Versorgungszentren) oder Personen (z. B. Ärztinnen und Ärzte, Pflegekräfte), die Patientinnen und Patienten medizinisch-pflegerisch versorgen.

**Operationalisierung**

Festlegung, wie ein theoretisches Konzept messbar gemacht werden soll. In der externen Qualitätssicherung werden beispielsweise Qualitätsaspekte durch Qualitätsindikatoren operationalisiert.

**Patient-Reported Experience Measure (PREM)**

Messinstrument basierend auf durch Patientinnen und Patienten berichteten Erfahrungen (z. B. Wartezeit, Stattfinden eines Aufklärungsgesprächs), meist mittels Fragebogen.

**Patient-Reported Outcome Measure (PROM)**

Messinstrument basierend auf durch Patientinnen und Patienten berichteten Ergebnissen der Versorgung (z. B. Komplikationen, Lebensqualität), meist mittels Fragebogen.

**Qualität**

Grad, in dem Merkmale eines Objekts Anforderungen erfüllen. Mit Objekt sind dabei beispielsweise auch Dienstleistungen wie eine medizinische Behandlung gemeint. Siehe auch: Versorgungsqualität.

**Qualitätsaspekt**

Thema der Versorgungspraxis (z. B. Indikationsstellung zur Herzkatheteruntersuchung, Komplikationen bei Hüftoperation), anhand dessen sich für einen bestimmten Versorgungsbereich die Qualität der Gesundheitsversorgung beschreiben lässt. Im Gegensatz zu Qualitätsdimensionen (z. B. nach Donabedian) handelt es sich bei Qualitätsaspekten nicht um generelle Kategorien, sondern um spezifisch für jedes Verfahren abgeleitete Themen.

**Qualitätsdimension (nach Donabedian)**

Kategorie von Qualität anhand der gemessenen Informationen<sup>28</sup>: Strukturqualität, Prozessqualität und Ergebnisqualität.

---

<sup>28</sup> Der Begriff Qualitätsdimension wird in der Literatur unterschiedlich verwendet, z. B. auch für allgemeine Teilaspekte von Qualität (z. B. bei Arah et al. (2006), Wiig et al. (2014)) wie Wirksamkeit (effectiveness), Patientensicherheit, Patientenorientierung, Zugänglichkeit. Diese Aspekte werden von anderen Autoren wiederum als Komponenten der Qualität (Donabedian 2003), Qualitätsmerkmale (DIN EN 15224), Kernziele

**Qualitätsindikator**

Beschreibt den Grad der Erfüllung von Qualitätszielen über eine empirisch bestimmte Größe (den Indikatorwert), und bewertet diesen anhand von Vergleichsmaßstäben. Qualitätsindikatoren umfassen somit die drei Komponenten Qualitätsziel, Messverfahren und Bewertungskonzept.

**Qualitätsmerkmal**

Eigenschaft der Versorgung (z. B. Erlangen von Gehfähigkeit), die sich auf Anforderungen bezieht.<sup>29</sup>

**Qualitätsmessung**

Empirische Beschreibung der Erfüllung von Anforderungen anhand von Qualitätsindikatoren.

**Qualitätsmodell**

Die zur Beschreibung der Qualität eines Versorgungsbereichs ausgewählten Qualitätsaspekte. Für jedes neue QS-Verfahren wird ein individuelles Qualitätsmodell entwickelt, welches als Basis für die Entwicklung von Qualitätsindikatoren dient.

**Qualitätssicherung**

Unter Qualitätssicherung im Gesundheitswesen werden in diesem Dokument alle Prozesse und Maßnahmen verstanden, durch die die Qualität der Versorgung gewährleistet oder verbessert werden soll. Auf das Gesundheitswesen bezogen soll dadurch Vertrauen in die Qualität der medizinischen Versorgung erzeugt werden. Auf dieser Ebene entspricht das Verständnis dem der DIN EN ISO 9000:2015. Diese versteht unter Qualitätssicherung den „Teil des Qualitätsmanagements, der auf das Erzeugen von Vertrauen darauf gerichtet ist, dass Qualitätsanforderungen erfüllt werden“.

**Qualitätssicherung, externe**

Prozesse und Maßnahmen, durch die für ausgewählte Versorgungsbereiche die Qualität der teilnehmenden Einrichtungen bewertet wird und ggf. Handlungskonsequenzen aus den Ergebnissen initiiert werden. Externe Qualitätssicherung wird in der Regel einrichtungsübergreifend durchgeführt und mit einem Feedback an die teilnehmenden Einrichtungen verbunden.

**Qualitätssicherungsverfahren (QS-Verfahren)**

Vom G-BA in Richtlinien festgelegtes Bündel an Maßnahmen der externen Qualitätssicherung in Bezug auf bestimmte medizinisch-pflegerische Leistungen und Leistungsbereiche. Dazu gehören

---

des Gesundheitssystems (Commonwealth Fund 2006) oder Domänen (NQMC-Website der AHRQ [kein Datum]) bezeichnet.

<sup>29</sup> Der Merkmalsbegriff entspricht dem der DIN EN ISO 9000 ff. und der sozialwissenschaftlichen Literatur und ist nicht gleichbedeutend mit den elf grundlegenden Qualitätsanforderungen („Qualitätsmerkmalen“), die in der DIN EN 15224 genannt werden.

im Wesentlichen die Festlegung der einbezogenen Leistungen oder Leistungsbereiche, Qualitätsindikatoren (mit Qualitätszielen, Mess- und Auswertungsmethoden sowie Regeln zur Bewertung), Spezifikationen (bspw. der QS-Dokumentation, des QS-Filters oder der Nutzung von Sozialdaten, sowie der zugehörigen Datenflüsse), die Auswertung, die Vorgehensweise zur Bewertung sowie die Regelung qualitätsverbessernder Maßnahmen.

**Qualitätsziel**

Konkrete Anforderung an Strukturen, Prozesse oder Ergebnisse einer medizinischen Versorgungsleistung.

**Referenzbereich**

Der Bereich auf der Messskala eines Qualitätsindikators, der ein bestimmtes Qualitätsniveau repräsentiert.

**Reliabilität**

Die Zuverlässigkeit einer Messung. Sie ist definiert als der Anteil der Varianz, der durch tatsächliche Unterschiede verursacht ist. Geschätzt werden kann die Reliabilität mit Methoden wie der Test-Retest-Methode, der Paralleltest-Methode oder Methoden der internen Konsistenz.

**Sozialdaten bei den Krankenkassen**

Daten, die die Krankenkassen nach § 284 SGB V erheben und speichern. Dazu gehören u. a. Abrechnungsdaten für medizinische Behandlungen und Versichertenstammdaten. Nach § 299 Abs. 1a SGB V dürfen diese in pseudonymisierter Form für die Qualitätssicherung genutzt werden.

**Strukturierter Dialog**

Prozess, in dem Ursachen für auffällige Indikatorwerte zwischen der für die Qualitätssicherung zuständigen externen Einrichtung (z. B. LQS, LAG, IQTIG) und den Leistungserbringern nachgegangen wird und eine abschließende Bewertung vorgenommen wird.

**Validität**

Integriertes bewertendes Urteil über das Ausmaß, in dem die Angemessenheit und die Güte von Interpretationen und Maßnahmen auf Basis von Testwerten oder anderen diagnostischen Verfahren durch empirische Belege und theoretische Argumente gestützt sind (Messick 1989 in der Übersetzung durch Hartig et al. 2012). Validität beschreibt z. B., ob ein Indikator das misst, was er messen soll.

**Verfahrenszweck**

Handlungsrahmen, für den die Ergebnisse der Qualitätsmessungen eines QS-Verfahrens verwendet werden sollen. Die Regelungen zur externen verpflichtenden Qualitätssicherung sehen

eine Verwendung von Qualitätsmessungen z. B. für Zwecke der Qualitätsförderung (wie Verfahren nach der QSKH-RL und Qesü-RL), der Information von Patientinnen und Patienten sowie der Öffentlichkeit, der qualitätsorientierten Vergütung und der Krankenhausplanung vor.

**Versorgungsbereich**

Bereich thematisch eng miteinander verbundener Anteile der Gesundheitsversorgung, z. B. Versorgung von Patientinnen und Patienten mit Diabetes mellitus.

**Versorgungsqualität**

Grad, in dem die Gesundheitsversorgung patientenorientierte, begründete Anforderungen erfüllt.

## Literatur

- Adams, JL; Mehrotra, A; McGlynn, EA (2010): Estimating Reliability and Misclassification in Physician Profiling [*Technical Report*]. Santa Monica, US-CA: RAND. URL: [https://www.rand.org/content/dam/rand/pubs/technical\\_reports/2010/RAND\\_TR863.pdf](https://www.rand.org/content/dam/rand/pubs/technical_reports/2010/RAND_TR863.pdf) (abgerufen am: 18.08.2017).
- AHRQ [Agency for Healthcare Research and Quality] (2011): Timeliness. Chapter 4. In: AHRQ [Agency for Healthcare Research and Quality]: *National Healthcare Quality Report 2011*. [Stand:] March 2012. (AHRQ Publication, No. 12-0005). Rockville, US-MD: AHRQ, 153-159. URL: <https://archive.ahrq.gov/research/findings/nhqrdr/nhqr11/nhqr11.pdf> (abgerufen am: 22.08.2017).
- AHRQ [Agency for Healthcare Research and Quality] ([kein Datum]): Browse by Measure Domain. Rockville, US-MD. URL: <https://www.qualitymeasures.ahrq.gov/browse/domain> (abgerufen am: 29.08.2017).
- Altenhofen, L; Blumenstock, G; Diel, F; Döbler, K; Geraedts, M; Jäckel, WH; et al. (2009): Programm für Nationale VersorgungsLeitlinien von BÄK, KBV und AWMF. Qualitätsindikatoren – Manual für Autoren. (ÄZQ Schriftenreihe, Band 36). Neukirchen: Make a Book. ISBN: 978-3940218-551. URL: <http://www.leitlinien.de/mdb/edocs/pdf/schriftenreihe/schriftenreihe36.pdf> (abgerufen am: 30.01.2017).
- AQUA [Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen]; Hrsg. (2015): Allgemeine Methoden im Rahmen der sektorenübergreifenden Qualitätssicherung im Gesundheitswesen nach §137a SGB V. Version 4.0. Stand: 17.02.2015. Göttingen: AQUA. Signatur: 14-SQG-029. URL: <http://www.sqg.de/sqg/upload/CONTENT/Hintergrund/Methodenpapier/AQUA-Methodenpapier-4.0.pdf> (abgerufen am: 30.01.2017).
- Arah, OA; Westert, GP; Hurst, J; Klazinga, NS (2006): A conceptual framework for the OECD Health Care Quality Indicators Project. *International Journal for Quality in Health Care* 18(Suppl. 1): 5-13. DOI: 10.1093/intqhc/mzl024.
- Arts, DGT; de Keizer, NF; Scheffer, G-J (2002): Defining and Improving Data Quality in Medical Registries: A Literature Review, Case Study, and Generic Framework. *Journal of the American Medical Informatics Association* 9(6): 600-611. DOI: 10.1197/jamia.M1087.
- Ash, AS; Schwartz, M; Peköz, EA (2013): Comparing Outcomes Across Providers. Chapter 12. In: Iezzoni, LI; Hrsg.: *Risk Adjustment for Measuring Healthcare Outcomes*. 3rd edition. Chicago, US-IL: Health Admission Press, 297-333. ISBN: 9781567934373.
- Baker, J; Lovell, K; Harris, N (2006): How expert are the experts? An exploration of the concept of 'expert' within Delphi panel techniques. *Nurse Researcher* 14(1): 59-70. DOI: 10.7748/nr2006.10.14.1.59.c6010.
- Benchimol, EI; Manuel, DG; To, T; Griffiths, AM; Rabeneck, L; Guttman, A (2011): Development and use of reporting guidelines for assessing the quality of validation studies of health

- administrative data. *Journal of Clinical Epidemiology* 64(8): 821-829. DOI: 10.1016/j.jclinepi.2010.10.006.
- Boulkedid, R; Abdoul, H; Loustau, M; Sibony, O; Alberti, C (2011): Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. *PLoS One* 6(6): e20476. DOI: 10.1371/journal.pone.0020476.
- BR-Drucksache 151/14 vom 11.04.2014. Gesetzentwurf der Bundesregierung. Entwurf eines Gesetzes zur Weiterentwicklung der Finanzstruktur und der Qualität in der gesetzlichen Krankenversicherung (GKV-Finanzstruktur- und Qualitäts-Weiterentwicklungsgesetz – GKV-FQWG). Köln: BundesanzeigerVerlagsgesellschaft. URL: [http://www.bundesrat.de/SharedDocs/drucksachen/2014/0101-0200/151-14.pdf?\\_\\_blob=publicationFile&v=1](http://www.bundesrat.de/SharedDocs/drucksachen/2014/0101-0200/151-14.pdf?__blob=publicationFile&v=1) (abgerufen am: 30.01.2017).
- Bradburn, NM; Sudman, S; Wansink, B (2004): Asking Questions. The Definitive Guide to Questionnaire Design – For Market Research, Political Polls, and Social and Health Questionnaires. San Francisco: Jossey-Bass. ISBN: 978-0-7879-7088-8.
- Brouwers, M; Kho, ME; Browman, GP; Cluzeau, F; Fervers, B; Hanna, S; et al. (2014): AGREE II. The AGREE Next Steps Consortium. Update. Deutsche Version: 2014. AGREE Research Trust. URL: [http://www.agreetrust.org/wp-content/uploads/2014/03/AGREE\\_II\\_German-Version.pdf](http://www.agreetrust.org/wp-content/uploads/2014/03/AGREE_II_German-Version.pdf) (abgerufen am: 30.01.2017).
- Bühner, M (2011): Einführung in die Test- und Fragebogenkonstruktion. 3., aktualisierte und erweiterte Auflage. München [u. a.]: Pearson. ISBN: 978-3-86894-033-6.
- Campbell, SM; Braspenning, J; Hutchinson, A; Marshall, MN (2003): Research methods used in developing and applying quality indicators in primary care. *BMJ – British Medical Journal* 326(7393): 816-819. DOI: 10.1136/bmj.326.7393.816.
- Campbell, SM; Shield, T; Rogers, A; Gask, L (2004): How do stakeholder groups vary in a Delphi technique about primary mental health care and what factors influence their ratings? *Quality & Safety in Health Care* 13(6): 428-434. DOI: 10.1136/qshc.2003.007815.
- Chen, H; Hailey, D; Wang, N; Yu, P (2014): A review of data quality assessment methods for public health information systems. *International Journal of Environmental Research and Public Health* 11(5): 5170-5207. DOI: 10.3390/ijerph110505170.
- CHP/PCOR [Center for Health Policy/Center for Primary Care and Outcomes Research]; Battelle Memorial Institute (2011): Quality Indicator Measure Development, Implementation, Maintenance, and Retirement (Prepared by Battelle, under Contract No. 290-04-0020). [Stand:] May 2011. Rockville, US-MD: AHRQ [Agency for Healthcare Research and Quality]. URL: [http://www.qualityindicators.ahrq.gov/Downloads/Resources/Publications/2011/QI\\_Measure\\_Development\\_Implementation\\_Maintenance\\_Retirement\\_Full\\_5-3-11.pdf](http://www.qualityindicators.ahrq.gov/Downloads/Resources/Publications/2011/QI_Measure_Development_Implementation_Maintenance_Retirement_Full_5-3-11.pdf) (abgerufen am: 30.01.2017).
- Cleary, PD (1999): The increasing importance of patient surveys. Now that sound methods exist, patient surveys can facilitate improvement. *BMJ – British Medical Journal* 319(7212): 720-721. DOI: 10.1136/bmj.319.7212.720.

- Colquhoun, HL; Levac, D; O'Brien, KK; Straus, S; Tricco, AC; Perrier, L; et al. (2014): Scoping reviews: time for clarity in definition, methods, and reporting. *Journal of Clinical Epidemiology* 67(12): 1291-1294. DOI: 10.1016/j.jclinepi.2014.03.013.
- Committee on Quality of Health Care in America; Institute of Medicine (2001): Crossing the Quality Chasm. A New Health System for the 21st Century. Washington, D.C.: National Academy Press. ISBN: 0-309-07280-8
- Commonwealth Fund, Commission on a High Performance Health System (2006): Framework for a high performance health system for the United States. [Stand:] August 2006. (Commonwealth Fund pub., no. 943). New York [u. a.]: Commonwealth Fund. URL: [http://www.commonwealthfund.org/~media/files/publications/fund-report/2006/avg/framework-for-a-high-performance-health-system-for-the-united-states/commission\\_framework\\_high\\_performance\\_943-pdf.pdf](http://www.commonwealthfund.org/~media/files/publications/fund-report/2006/avg/framework-for-a-high-performance-health-system-for-the-united-states/commission_framework_high_performance_943-pdf.pdf) (abgerufen am: 30.01.2017).
- Conrad, F (1977): Die Münchner Perinatalstudie. *Deutsches Ärzteblatt International* 74(51): 3015-3020. URL: <https://www.aerzteblatt.de/pdf/74/51/a3015.pdf?ts=07%2E01%2E2014+15%3A59%3A39> (abgerufen am: 22.08.2017).
- Davis, K (2005): Toward A High Performance Health System: The Commonwealth Fund's New Commission. The commission will monitor insurance coverage, quality of care, and health system efficiency. *Health Affairs* 24(5): 1356-1360. DOI: 10.1377/hlthaff.24.5.1356.
- de Koning, J; Smulders, A; Klazinga, N (2007a): Appraisal of Indicators through Research and Evaluation (AIRE). Versie 2.0. [Stand:] Januari 2007. Amsterdam: Academisch Medisch Centrum. URL: <http://zorginzicht.garansys.nl/kennisbank/PublishingImages/Paginas/AIRE-instrument/AIRE%20Instrument%202.0.pdf> (abgerufen am: 30.01.2017).
- de Koning, J (2007b): Development and validation of a measurement instrument for appraising indicator quality: appraisal of indicators through research and evaluation (AIRE) instrument [Abstract]. Kongress Medizin und Gesellschaft 2007. 17.-21.09.2007. Augsburg. URL: <http://www.egms.de/static/de/meetings/gmds2007/07gmds798.shtml> (abgerufen am: 30.01.2017).
- Diamond, IR; Grant, RC; Feldman, BM; Pencharz, PB; Ling, SC; Moore, AM; et al. (2014): Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology* 67(4): 401-409. DOI: 10.1016/j.jclinepi.2013.12.002.
- Diekmann, A (2007): Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen. 18. Aufl., vollst. überarb. und erw. Neuausg. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag. ISBN: 978-3-499-55678-4.
- DIN [Deutsches Institut für Normung] (2015): Qualitätsmanagementsysteme – Grundlagen und Begriffe (ISO 9000:2015); Deutsche und Englische Fassung EN ISO 9000:2015. [Stand:] November 2015. Berlin: Beuth.
- DIN [Deutsches Institut für Normung] (2017): Qualitätsmanagementsysteme – EN ISO 9001:2015 für die Gesundheitsversorgung; Deutsche Fassung EN 15224:2016. [Stand:] Mai 2017. Berlin: Beuth.

- Donabedian, A (1966): Evaluating the Quality of Medical Care. *Milbank Memorial Fund Quarterly* 44(3 Suppl.): 166-206.
- Donabedian, A (2003): An introduction to quality assurance in health care. Oxford [u. a.]: Oxford University Press. ISBN: 0-19-515809-1.
- 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.
- Dreher, M; Dreher, E (1982): Gruppendiskussion. Kapitel 7. In: Huber, GL; Mandl, H: *Verbale Daten*. Weinheim [u. a.]: Beltz Verlag, 141-164. ISBN: 9783407546326.
- Droste, S; Lühmann, D (2014): Informationsquellen und systematische Informationsgewinnung. Kapitel 5. In: Perleth, M; Busse, R; Gerhardus, A; Gibis, B; Lühmann, D; Zentner, A; Hrsg.: *Health Technology Assessment. Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung. 2.*, aktualisierte und erweiterte Auflage. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft, 162-208. ISBN: 978-3-941468-71-9.
- Ehling, M; Körner, T; Bergdahl, M; Elvers, E; Földesi, E; Kron, A; et al. (2007): Handbook on Data Quality Assessment: Methods and Tools Wiesbaden: European Commission, eurostat. URL: [https://unstats.un.org/unsd/dnss/docs-nqaf/Eurostat-HANDBOOK ON DATA QUALITY ASSESSMENT METHODS AND TOOLS I.pdf](https://unstats.un.org/unsd/dnss/docs-nqaf/Eurostat-HANDBOOK_ON_DATA_QUALITY_ASSESSMENT_METHODS_AND_TOOLS_I.pdf) (abgerufen am: 05.05.2017).
- Ennker, J; Albert, A; Thanner, M; Nagel, E; Rosendahl, U; Ennker, I (2009): Transparenz medizinischer Qualitätsdaten. Ethische Bringschuld oder Instrument zur Schaffung unlauterer Wettbewerbsvorteile? *Deutsche Medizinische Wochenschrift* 134(Suppl. 6): S188-S189. DOI: 10.1055/s-0029-1241904.
- Fahrmeir, L; Tutz, G (2010): Multivariate Statistical Modelling Based on Generalized Linear Models. New York: Springer. ISBN: 978-1-4419-2900-6
- Faul, F; Erdfelder, E; Buchner, A; Lang, A-G (2009): Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41(4): 1149-1160. DOI: 10.3758/BRM.41.4.1149.
- Fitch, K; Bernstein, SJ; Auguilar, MD; Burnand, B; LaCalle, JR; Lázaro, P; et al. (2001): The RAND/UCLA Appropriateness Method User's Manual. Santa Monica, US-CA: RAND. URL: [https://www.rand.org/content/dam/rand/pubs/monograph\\_reports/2011/MR1269.pdf](https://www.rand.org/content/dam/rand/pubs/monograph_reports/2011/MR1269.pdf) (abgerufen am: 30.01.2017).
- Fletcher, RH; Fletcher, SW (2007): Klinische Epidemiologie: Grundlagen und Anwendungen. 2., vollständig überarbeitete deutschsprachige Ausgabe. Bern: Verlag Hans Huber. ISBN: 978-3-456-84374-2.
- Flick, U (2016): Qualitative Sozialforschung. Eine Einführung. 7. Auflage. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag. ISBN: 978-3-499-55694-4.

- Frost, MH; Reeve, BB; Liepa, AM; Stauffer, JW; Hays, RD (2007): What Is Sufficient Evidence for the Reliability and Validity of Patient-Reported Outcome Measures? *Value in Health* 10(Suppl. 2): S94-S105. DOI: 10.1111/j.1524-4733.2007.00272.x.
- Fung, CH; Lim, Y-W; Mattke, S; Damberg, C; Shekelle, PG (2008): Systematic Review: The Evidence That Publishing Patient Care Performance Data Improves Quality of Care. *Annals of Internal Medicine* 148(2): 111-123, W17-W23. DOI: 10.7326/0003-4819-148-2-200801150-00006.
- G-BA [Gemeinsamer Bundesausschuss] (2015): Beschluss des Gemeinsamen Bundesausschusses über die Beauftragung des IQTIG zur Beteiligung an der Durchführung der einrichtungsübergreifenden Qualitätssicherung und zur Veröffentlichung der Ergebnisse der Qualitätssicherungsmaßnahmen. [Stand:] 27.11.2015. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/39-261-2408/2015-11-27\\_IQTIG-Beauftragung\\_Beteiligung-Qesue.pdf](https://www.g-ba.de/downloads/39-261-2408/2015-11-27_IQTIG-Beauftragung_Beteiligung-Qesue.pdf) (abgerufen am: 30.01.2017).
- G-BA [Gemeinsamer Bundesausschuss] (2016): Verfahrensordnung des Gemeinsamen Bundesausschusses. In der Fassung vom 18. Dezember 2008, veröffentlicht im Bundesanzeiger Nr. 84a (Beilage) vom 10. Juni 2009, in Kraft getreten am 1. April 2009. Zuletzt geändert am 21. Juli 2016, veröffentlicht im Bundesanzeiger BAnz AT 28.10.2016 B5, in Kraft getreten am 29. Oktober 2016. [Stand:] 01.11.2016. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/62-492-1282/VerfO\\_2016-07-21\\_iK-2016-10-29.pdf](https://www.g-ba.de/downloads/62-492-1282/VerfO_2016-07-21_iK-2016-10-29.pdf) (abgerufen am: 30.01.2017).
- Gechter, D; Nothacker, M; Khan, C; Rütters, D; Mavergames, C; Motschall, E; et al. (2013): Manual Systematische Literaturrecherche für die Erstellung von Leitlinien. [Stand:] 10.05.2013. Freiburg [u. a.]: DCZ [Deutsches Cochrane-Zentrum], Institut für Medizinische Biometrie und Medizinische Informatik; AWMF [Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften], Institut für Medizinisches Wissensmanagement; ÄZQ [Ärztliche Zentrum für Qualität in der Medizin]. DOI: 10.6094/UNIFR/2013/2.
- Geraedts, M; Selbmann, H-K; Ollenschlaeger, G (2003): Critical appraisal of clinical performance measures in Germany. *International Journal for Quality in Health Care* 15(1): 79-85. DOI: 10.1093/intqhc/15.1.79.
- Geraedts, M; Auras, S; Hermeling, P; de Cruppé, W (2009): Public Reporting – Formen und Effekte öffentlicher Berichterstattung. *DMW – Deutsche Medizinische Wochenschrift* 134(Suppl. 6): S232-S233. DOI: 10.1055/s-0029-1241920.
- Gerteis, M; Edgman-Levitan, S; Daley, J; Delbanco, T (2017): Through the patient's eyes. Understanding and promoting patient-centered care. San Francisco: Jossey-Bass. ISBN: 978-0-7879-5699-8.
- Goddard, M; Jacobs, R (2010): Using Composite Indicators to Measure Performance. Chapter 3.4. In: Smith, PC; Mossialos, E; Papanicolas, I; Leatherman, S: *Performance Measurement for Health System Improvement: Experiences, Challenges, and Prospects*. 3rd edition. Cambridge [u. a.]: Cambridge University Press, 339-368. ISBN: 978-0-521-11676-3.

- Häder, M (2000): Die Expertenauswahl bei Delphi-Befragungen. (ZUMA How-to-Reihe, Nr. 5). Mannheim: ZUMA [Zentrum für Umfragen, Methoden und Analysen]. URL: [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/howto/how-to5mh.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/how-to5mh.pdf) (abgerufen am: 30.01.2017).
- Härter, M; Loh, A; Spies, C (2005): Gemeinsam entscheiden – erfolgreich behandeln. Neue Wege für Ärzte und Patienten im Gesundheitswesen. Köln: Deutscher Ärzte-Verlag. ISBN: 978-3-7691-3250-2.
- Hartig, J; Frey, A; Jude, N (2012): Validität. Kapitel 7. In: Moosbrugger, H; Kelava, A; Hrsg.: *Testtheorie und Fragebogenkonstruktion*. 2., aktualisierte und überarbeitete Auflage. Berlin [u. a.]: Springer, 143-171. ISBN: 3-642-20071-0.
- Hartmann, J; Weidmann, C; Biehle, R (2016): Validierung von GKV-Routinedaten am Beispiel von geschlechtsspezifischen Diagnosen. *Gesundheitswesen* 78(10): e162-e167. DOI: 10.1055/s-0035-1565072.
- Haynes, RB; McKibbin, KA; Kanani, R (1996): Systematic review of randomised trials of interventions to assist patients to follow prescriptions for medications. *Lancet* 348(9024): 383-386. DOI: 10.1016/S0140-6736(96)01073-2.
- Heidegger, T; Nuebling, M; Saal, D; Kreienbühl, G (2008): Patient-centred outcomes in clinical research: does it really matter? *British Journal of Anaesthesia* 100(1): 1-3. DOI: 10.1093/bja/aem351.
- HHS.gov [U.S. Department of Health and Human Services]; FDA [Food and Drug Administration]; CDER [Center for Drug Evaluation and Research]; CBER [Center for Biologics Evaluation and Research]; CDRH [Center for Devices and Radiological Health] (2009): Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. [Stand:] December 2009. Washington, D. C.: HHS.gov. URL: <http://www.fda.gov/downloads/Drugs/.../Guidances/UCM193282.pdf> (abgerufen am: 30.01.2017).
- Hippler, H-J; Schwarz, N; Noelle-Neumann, E; Knäuper, B; Clark, L (1991): Der Einfluss numerischer Werte auf die Bedeutung verbaler Skalenendpunkte. *ZUMA Nachrichten* 28(15): 54-64. URL: [http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma\\_nachrichten/zn\\_28.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_28.pdf) (abgerufen am: 30.01.2017).
- Hodson, M; Andrew, S; Michael Roberts, C (2013): Towards an understanding of PREMS and PROMS in COPD. *Breathe* 9(5): 358-364. DOI: 10.1183/20734735.006813.
- Horenkamp-Sonntag, D; Schneider, U; Engel, S; Linder, R; Gerste, B; Ihle, P (2014): Validität von GKV-Routinedaten: In welchem Umfang muss bei der wissenschaftlichen Nutzung von Sekundärdaten die Daten-Qualität geprüft werden? *Zeitschrift für Palliativmedizin* 15(3): PD293. DOI: 10.1055/s-0034-1374464.
- Hosmer, DW; Lemeshow, S; Sturdivant, RX (2013): Applied Logistic Regression. Third Edition. Hoboken, US-NJ [u. a.]: Wiley. ISBN: 978-0-470-58247-3.

- Iezzoni, LI (2013a): Range of Risk Factors. Chapter 3. In: Iezzoni, LI; Hrsg.: *Risk Adjustment for Measuring Healthcare Outcomes*. 4th edition. Chicago, US-IL: Health Administration Press, 29-76. ISBN: 978-1-56793-437-3.
- Iezzoni, LI (2013b): Reasons for Risk Adjustment. Chapter 1. In: Iezzoni, LI; Hrsg.: *Risk Adjustment for Measuring Healthcare Outcome*. 4th edition. Chicago, US-IL: Health Administration Press, 1-14. ISBN: 978-1-56793-437-3.
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2016): Planungsrelevante Qualitätsindikatoren [Abschlussbericht zur Auswahl und Umsetzung]. Stand: 31.08.2016. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2016/IQTIG\\_Planungsrelevante-Qualitätsindikatoren\\_Abschlussbericht.pdf](https://iqtig.org/downloads/berichte/2016/IQTIG_Planungsrelevante-Qualitätsindikatoren_Abschlussbericht.pdf) (abgerufen am: 25.08.2017).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2017): Ereigniszeitanalyse-Methodik für die Follow-up-Indikatoren nach QSKH-RL. Stand: 06.04.2017. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2017/IQTIG\\_Ereigniszeitanalyse-Methodik-f%C3%BCr-Follow-up-Indikatoren-nach-QSKH-RL\\_2017-04-06.pdf](https://iqtig.org/downloads/berichte/2017/IQTIG_Ereigniszeitanalyse-Methodik-f%C3%BCr-Follow-up-Indikatoren-nach-QSKH-RL_2017-04-06.pdf) (abgerufen am: 18.08.2017).
- Jacob, G; Bengel, J (2000): Das Konstrukt Patientenzufriedenheit: Eine kritische Bestandsaufnahme. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie* 48(3): 280-301.
- JCAHO [Joint Commission on Accreditation of Healthcare Organisations] (1989): Characteristics of Clinical Indicators. *QRB – Quality Review Bulletin* 15(11): 330-339.
- JCAHO [Joint Commission on Accreditation of Healthcare Organizations] (1990): Primer on Indicator Development and Application: Measuring Quality in Health Care. Oakbrook Terrace, US-IL: JCAHO. ISBN: 0-86688-213-8.
- Joint Commission (2016): Standards for Behavioral Health Care 2016. Oak Brook, US-IL: Joint Commission Resources. ISBN: 978-1-59940-912-2.
- Jones, P; Shepherd, M; Wells, S; Le Fevre, J; Ameratunga, S (2014): Review Article: What makes a good healthcare quality indicator? A systematic review and validation study. *Emergency Medicine Australasia* 26(2): 113-124. DOI: 10.1111/1742-6723.12195.
- Kane, MT (2001): Current Concerns in Validity Theory. *JEM – Journal of Educational Measurement* 38(4): 319-342. DOI: 10.1111/j.1745-3984.2001.tb01130.x.
- Keeney, S; Hasson, F; McKenna, H (2006): Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *Journal of Advanced Nursing* 53(2): 205-212. DOI: 10.1111/j.1365-2648.2006.03716.x.
- Keiding, N; Clayton, D (2014): Standardization and Control for Confounding in Observational Studies: A Historical Perspective. *Statistical Science* 29(4): 529-558. DOI: 10.1214/13-STS453.
- Kerr, EA; Hofer, TP; Hayward, RA; Adams, JL; Hogan, MM; McGlynn, EA; et al. (2007): Quality by Any Other Name?: A Comparison of Three Profiling Systems for Assessing Health Care Quality. *Health Services Research* 42(5): 2070-2087. DOI: 10.1111/j.1475-6773.2007.00730.x.

- Khan, KS; Kunz, R; Kleijnen, J; Antes, G (2004): Systematische Übersichten und Meta-Analysen. Ein Handbuch für Ärzte in Klinik und Praxis sowie Experten im Gesundheitswesen. Berlin [u. a.]: Springer. ISBN: 978-3-540-43936-3.
- Klein, K (2004): Ereignisorientierte Patientenbefragung – Entwicklung und Validierung eines ereignisorientierten Fragebogens zur Bewertung der stationären medizinischen Rehabilitation [Dissertation]. Freiburg: Universität Freiburg, Institut für Soziologie, Philosophische Fakultät. URL: <https://www.deutsche-digitale-bibliothek.de/binary/BBEF3UNBRUGFILOIKZJ2AMCF4ILXZQNA/full/1.pdf> (abgerufen am: 30.01.2017).
- Klemperer, D (2000): Patientenorientierung im Gesundheitssystem. *Qualität in der Gesundheitsversorgung – Newsletter der GQMG* 1(7): 15-16.
- Klemperer, D (2011): Was ist ein Interessenkonflikt und wie stellt man ihn fest? Kapitel 2. In: Lieb, K; Klemperer, D; Ludwig, W-D; Hrsg.: *Interessenkonflikte in der Medizin: Hintergründe und Lösungsmöglichkeiten*. Berlin [u. a.]: Springer Verlag, 11-25. ISBN: 978-3-642-19841-0.
- Kloda, LA; Bartlett, JC (2013): Formulating Answerable Questions: Question Negotiation in Evidence-based Practice. *Journal of the Canadian Health Libraries Association* 34(2): 55-60. DOI: 10.5596/c13-019.
- Kreienbrock, L; Pigeot, I; Ahrens, W (2012): *Epidemiologische Methoden*. 5. Auflage. Berlin [u. a.]: Springer Spektrum. ISBN: 978-3-8274-2333-7.
- Krick, E (2013): Verhandlungen im Konsensverfahren: Varianten kollektiver Entscheidung in Expertengremien. Wiesbaden: Springer VS. ISBN: 978-3-658-00355-5.
- Krippendorff, K (1970): Estimating the Reliability, Systematic Error and Random Error of Interval Data. *Educational and Psychological Measurement* 30(1): 61-70. DOI: 10.1177/001316447003000105.
- Krueger, RA; Casey, MA (2015): *Focus Groups. A Practical Guide for Applied Research*. 5th ed. Thousand Oaks, US-CA: SAGE. ISBN: 978-1-4833-6524-4.
- Kühn, T; Koschel, K-V (2011): *Gruppendiskussionen. Ein Praxis-Handbuch*. Wiesbaden: VS Verlag. ISBN: 978-3-531-16921-7.
- Lamnek, S (2010): *Qualitative Sozialforschung. Lehrbuch*. 5. überarbeitete Auflage. Weinheim [u. a.]: Beltz. ISBN: 978-3-621-27770-9.
- Lenzner, T; Neuert, C; Otto, W (2015): *Kognitives Pretesting. Version 1.1*. [Stand:] Januar 2015. Mannheim: GESIS [Leibniz-Institut für Sozialwissenschaften]. DOI: 10.15465/gesis-sg\_010.
- Liberati, A; Altman, DG; Tetzlaff, J; Mulrow, C; Gøtzsche, PC; Ioannidis, JPA; et al. (2009): The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *Journal of Clinical Epidemiology* 151(4): e1-e34. DOI: 10.7326/0003-4819-151-4-200908180-00136.
- Lilford, R; Pronovost, P (2010): Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ – British Medical Journal* 340: c2016. DOI: 10.1136/bmj.c2016.

- Lim, WS; van der Eerden, MM; Laing, R; Boersma, WG; Karalus, N; Town, GI; et al. (2003): Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* 58(5): 377-382. DOI: 10.1136/thorax.58.5.377.
- Linder-Pelz, S (1982): Social psychological determinants of patient satisfaction: A test of five hypotheses. *Social Science & Medicine* 16(5): 583-589.
- Lingenfelder, M; Schneider, W (1990): Die Kundenzufriedenheit: Bedeutung, Meßkonzept und empirische Befunde. (Arbeitspapier, Nr. 80). Mannheim: Universität Mannheim, Institut für Marketing. ISBN: 3-89333-110-7.
- Lohr, SL (2006): Sampling: Design and Analysis. 2. International Edition. Boston, US-MA: Brooks/Cole. ISBN: 978-0-495-11084-2.
- Mainz, J (2003): Defining and classifying clinical indicators for quality improvement. *International Journal for Quality in Health Care* 15(6): 523-530. DOI: 10.1093/intqhc/mzg081.
- McGowan, J; Sampson, M; Lefebvre, C (2010): An Evidence Based Checklist for the Peer Review of Electronic Search Strategies (PRESS EBC). *Evidence Based Library and Information Practice* 5(1): 149-154. DOI: 10.18438/B8SG8R.
- Messick, S (1994): Validity of Psychological Assessment: Validation of Inferences From Persons' Responses and Performances as Scientific Inquiry into Score Meaning. *ETS – Research Report Series* 1994(2): i-28. RR-94-45. DOI: 10.1002/j.2333-8504.1994.tb01618.x.
- Meuser, M; Nagel, U (1991): Experteninterviews – vielfach erprobt, wenig bedacht: ein Beitrag zur qualitativen Methodendiskussion. In: Garz, D; Kraimer, K; Hrsg.: *Qualitativ-empirische Sozialforschung: Konzepte, Methoden, Analysen*. Opladen: Westdeutscher Verlag, 441-471. ISBN: 3-531-12289-4. URL: [http://www.ssoar.info/ssoar/bitstream/handle/document/2402/ssoar-1991-meuser\\_et\\_al-expertinneninterviews\\_-\\_vielfach\\_erprobt.pdf?sequence=1](http://www.ssoar.info/ssoar/bitstream/handle/document/2402/ssoar-1991-meuser_et_al-expertinneninterviews_-_vielfach_erprobt.pdf?sequence=1) (abgerufen am: 30.01.2017).
- Meyer, W (2004): Indikatorenentwicklung: eine praxisorientierte Einführung. 2. Auflage. (CEval-Arbeitspapier, 10). Saarbrücken: Universität des Saarlandes, Fak. 05 Empirische Humanwissenschaften, CEval [Centrum für Evaluation]. URL: <http://www.ssoar.info/ssoar/bitstream/handle/document/11124/ssoar-2004-meyer-indikatorenentwicklung.pdf?sequence=1> (abgerufen am: 28.06.2017).
- Meyer, W (2007): Messen: Indikatoren – Skalen – Indizes – Interpretationen. Kapitel 7. In: Stockmann, R; Hrsg.: *Handbuch zur Evaluation. Eine praktische Handlungsanleitung*. Münster [u. a.]: Waxmann, 195-222. ISBN: 978-3-8309-1766-3.
- Moosbrugger, H; Kelava, A (2012): Testtheorie und Fragebogenkonstruktion. 2. Auflage. Berlin [u. a.]: Springer-Verlag. ISBN: 978-3-642-20071-7.
- Muche-Borowski, C; Selbmann, HK; Nothacker, M; Müller, W; Kopp, I; Hrsg. (2012): AWMF-Regelwerk Leitlinien. [Stand:] 06.11.2012. Marburg: AWMF [Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften], Ständige Kommission Leitlinien.

- URL: [http://www.awmf.org/fileadmin/user\\_upload/Leitlinien/AWMF-Regelwerk/AWMF-Regelwerk.pdf](http://www.awmf.org/fileadmin/user_upload/Leitlinien/AWMF-Regelwerk/AWMF-Regelwerk.pdf) (abgerufen am: 30.01.2017).
- Mühlhauser, I; Müller, H (2009): Patientenrelevante Endpunkte und patient-reported outcomes in klinischer Forschung und medizinischer Praxis. In: Klusen, N; Fließgarten, A; Nebling, T: *Informiert und selbstbestimmt. Der mündige Bürger als mündiger Patient.* (Beiträge zum Gesundheitsmanagement, 24). Baden-Baden: Nomos, 34-65. ISBN: 978-3-8329-4474-2.
- Mullen, PM (2003): Delphi: myths and reality. *Journal of Health Organization and Management* 17(1): 37-52. DOI: 10.1108/14777260310469319.
- Murphy, FG; Scheu, C (2010): Warum ist Transparenz so wichtig? Kapitel 5. In: Kuhlen, R; Rink, O; Zacher, J; Hrsg.: *Jahrbuch Qualitätsmedizin 2010.* Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft, 33-38. ISBN: 978-3-941468-30-6.
- Murphy, MK; Black, NA; Lamping, DL; McKee, CM; Sanderson, CFB; Askham, J; et al. (1998): Consensus development methods, and their use in clinical guideline development: a review. *Health Technology Assessment* 2(3). DOI: 10.3310/hta2030.
- Niederberger, M; Wassermann, S (2015): Methoden der Experten- und Stakeholdereinbindung in der sozialwissenschaftlichen Forschung. Wiesbaden: Springer VS. ISBN: 978-3-658-01686-9.
- Nonnemacher, M; Nasseh, D; Stausberg, J (2014): Datenqualität in der medizinischen Forschung. Leitlinien zum adaptiven Management von Datenqualität in Kohortenstudien und Registern. 2., aktualisierte und erweiterte Auflage. (Schriftenreihe der TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung, 4). Berlin: MWV [Medizinisch Wissenschaftliche Verlagsgesellschaft]. ISBN: 978-3-95466-121-3.
- NQF [National Quality Forum] (2015): Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement. Effective April 2015. Washington, D. C.: NQF. URL: <http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=79434> (abgerufen am: 30.01.2017).
- O'Brien, RM (1990): Estimating the Reliability of Aggregate-Level Variables Based on Individual-Level Characteristics. *Sociological Methods and Research* 18(4): 473-504. DOI: 10.1177/0049124190018004004
- OECD (2008): Handbook on Constructing Composite Indicators: Methodology and User Guide. Paris: OECD Publications. ISBN: 978-92-64-04345-9.
- Pascoe, GC (1983): Patient satisfaction in primary health care: a literature review and analysis. *Evaluation and Program Planning* 6(3-4): 185-210. DOI: 10.1016/0149-7189(83)90002-2.
- Perleth, M; Busse, R; Gerhardus, A; Gibis, B; Lühmann, D; Zenter, A; Hrsg. (2014): Health Technology Assessment. Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung. 2., aktualisierte und erweiterte Auflage. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft. ISBN: 978-3-941468-71-9.

- Porst, R (1996): Fragebogenerstellung. Kapitel 90. In: Goebel, H; Ungeheuer, G; Burkhardt, A; Hrsg.: *Kontaktlinguistik*. (Handbücher zur Sprach- und Kommunikationswissenschaft, 12/1). Berlin: de Gruyter Mouton, 737-744. ISBN: 978-3-11-013264-9.
- Porst, R (2014): Fragebogen. Ein Arbeitsbuch. 4. erweiterte Auflage. Wiesbaden: Springer VS. ISBN: 978-3-658-02117-7.
- Prüfer, P; Rexroth, M (1996): Verfahren zur Evaluation von Survey – Fragen: Ein Überblick. [Stand:] Mai 1996. (ZUMA-Arbeitsbericht, Nr. 96/05). Mannheim: ZUMA [Zentrum für Umfragen, Methoden und Analysen]. URL: [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/zuma\\_arbeitsberichte/96\\_05.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/zuma_arbeitsberichte/96_05.pdf) (abgerufen am: 30.01.2017).
- Prüfer, P; Rexroth, M; Hrsg. (2005): Kognitive Interviews. (ZUMA How-to-Reihe, 15). Mannheim: ZUMA [Zentrum für Umfragen, Methoden und Analysen]. URL: [http://www.ssoar.info/ssoar/bitstream/handle/document/20147/ssoar-2005-prufer\\_et\\_al-kognitive\\_interviews.pdf?sequence=1](http://www.ssoar.info/ssoar/bitstream/handle/document/20147/ssoar-2005-prufer_et_al-kognitive_interviews.pdf?sequence=1) (abgerufen am: 30.01.2017).
- Raykov, T; Marcoulides, GA (2011): Introduction to Psychometric Theory. London: Routledge. ISBN: 978-0-415-87822-7.
- Reiter, A; Fischer, B; Kötting, J; Geraedts, M; Jäckel, WH (2007): QUALIFY: Ein Instrument zur Bewertung von Qualitätsindikatoren. Düsseldorf: BQS [Bundesgeschäftsstelle Qualitätssicherung].
- Rode, S; Ries, V; Petzold, T; Buch, U; Untersweg, F; Fischer, B (2016): Anforderungen und Anwendungshinweise für den Einsatz von Qualitätsindikatoren in der medizinischen Versorgung: Ergebnisse eines systematischen Reviews. *Gesundheitswesen* 78(Suppl. 1): e110-e119. DOI: 10.1055/s-0042-102347.
- Sackett, D; Rosenberg, WMC; Gray, JAM; Hayses, RB; Richardson, WS (1997): Was ist Evidenzbasierte Medizin und was nicht? *MMW – Münchener Medizinische Wochenschrift* 139(44): 644-645.
- Sackman, H (1974): Delphi Assessment: Expert Opinion, Forecasting, and Group Process. [Stand:] April 1974. Santa Monica, US-CA: RAND. R-1283-PR. URL: <https://www.rand.org/content/dam/rand/pubs/reports/2006/R1283.pdf> (abgerufen am: 25.08.2017).
- Schaeffer, D (2006): Bedarf an Patienteninformationen über das Krankenhaus. Eine Literaturanalyse. Gütersloh: Bertelsmann Stiftung. URL: [https://www.weisse-liste.de/export/sites/weisseliste/de/.content/pdf/service/xcms\\_bst\\_dms\\_20028\\_20029\\_2.pdf](https://www.weisse-liste.de/export/sites/weisseliste/de/.content/pdf/service/xcms_bst_dms_20028_20029_2.pdf) (abgerufen am: 30.01.2017).
- Scheibler, F; Janssen, C; Pfaff, H (2003): Shared decision making: ein Überblicksartikel über die internationale Forschungsliteratur. *Sozial- und Präventivmedizin* 48(1): 11-24. DOI: 10.1007/s000380300002.

- 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.
- Schmucker, C; Motschall, E; Antes, G; Meerpohl, JJ (2013): Methoden des Evidence Mappings. Eine systematische Übersichtsarbeit. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 56(10): 1390-1397. DOI: 10.1007/s00103-013-1818-y.
- Schnell, R; Hill, PB; Esser, E (2013): Methoden der empirischen Sozialforschung. 10., überarbeitete Auflage. München: Oldenbourg. ISBN: 978-3-486-72899-6.
- Schubert, I; Ihle, P; Köster, I (2010): Interne Validierung von Diagnosen in GKV-Routinedaten: Konzeption mit Beispielen und Falldefinition. *Gesundheitswesen* 72(6): 316-322. DOI: 10.1055/s-0030-1249688.
- Schünemann, HJ (2009): GRADE: Von der Evidenz zur Empfehlung. Beschreibung des Systems und Lösungsbeitrag zur Übertragbarkeit von Studienergebnissen. *ZEFQ – Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 103(6): 391-400. DOI: 10.1016/j.zefq.2009.05.023.
- Schwenk, U; Schmidt-Kaehler, S (2016): Public Reporting: Transparenz über Gesundheitsanbieter erhöht Qualität der Versorgung. [Stand:] Januar 2016. (Spotlight Gesundheit. Daten, Analysen, Perspektiven, Nr. 1/2016). Gütersloh: Bertelsmann Stiftung. URL: [https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/SpotGes\\_PubRep\\_dt\\_final\\_web.pdf](https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/SpotGes_PubRep_dt_final_web.pdf) (abgerufen am: 30.01.2017).
- 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: 30.01.2017).
- Shea, BJ; Grimshaw, JM; Wells, GA; Boers, M; Andersson, N; Hamel, C; et al. (2007): Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC – Medical Research Methodology* 7(10): 1-7. DOI: 10.1186/1471-2288-7-10.
- Shrout, PE; Fleiss, JL (1979): Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 86(2): 420-428. DOI: 10.1037/0033-2909.86.2.420.
- Shwartz, M; Restuccia, JD; Rosen, AK (2015): Composite Measures of Health Care Provider Performance: A Description of Approaches. *Milbank Quarterly* 93(4): 788-825. DOI: 10.1111/1468-0009.12165.
- Smith, PC; Mossialos, E; Papanicolas, I; Leatherman, S; Hrsg. (2012): Performance Measurement for Health System Improvement: Experiences, Challenges, and Prospects. 3rd edition. New York: Cambridge University Press. ISBN: 978-0-521-11676-3.

- Stange, KC; Zyzanski, SJ; Smith, TF; Kelly, R; Langa, DM; Flocke, SA; et al. (1998): How Valid Are Medical Records and Patient Questionnaires for Physician Profiling and Health Services Research? A Comparison With Direct Observation of Patient Visits. *Medical Care* 36(6): 851-867.
- Steffen, GE (1988): Quality Medical Care. A Definition. *JAMA – Journal of the American Medical Association* 260(1): 56-61. DOI: 10.1001/jama.1988.03410010064036.
- Stelfox, HT; Straus, SE (2013): Measuring quality of care: considering conceptual approaches to quality indicator development and evaluation. *Journal of Clinical Epidemiology* 66(12): 1328-1337. DOI: 10.1016/j.jclinepi.2013.05.017.
- Tausch, A; Menold, N (2015): Methodische Aspekte der Durchführung der Fokusgruppen in der Gesundheitsforschung. Welche Anforderungen ergeben sich aufgrund der besonderen Zielgruppe und Fragestellungen? (GESIS Papers, 2015/12). Mannheim [u. a.]: GESIS [Leibniz-Institut für Sozialwissenschaften]. URL: [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_papers/GESIS-Papers\\_2015-12.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_papers/GESIS-Papers_2015-12.pdf) (abgerufen am: 30.01.2017).
- Thompson, DF (1993): Understanding financial conflicts of interest. *NEJM – New England Journal of Medicine* 329(8): 573-576. DOI: 10.1056/nejm199308193290812.
- Totten, AM; Wagner, J; Tiwari, A; O’Haire, C; Griffin, J; Walker, M (2012): Public Reporting as a Quality Improvement Strategy. Closing the Quality Gap: Revisiting the State of the Science. Evidence Report No. 208. (Prepared by the Oregon Evidence-based Practice Center under Contract No. 290-2007-10057-I.). [Stand:] July 2012. (AHRQ Publication, No. 12-E011-EF). Rockville, US-MD: AHRQ [Agency for Healthcare Research and Quality]. URL: [https://www.ncbi.nlm.nih.gov/books/NBK99879/pdf/Bookshelf\\_NBK99879.pdf](https://www.ncbi.nlm.nih.gov/books/NBK99879/pdf/Bookshelf_NBK99879.pdf) (abgerufen am: 30.01.2017).
- Waggoner, J; Carline, JD; Durning, SJ (2016): Is There a Consensus on Consensus Methodology? Descriptions and Recommendations for Future Consensus Research. *Academic Medicine* 91(5). DOI: 10.1097/ACM.0000000000001092.
- Whiting, P; Savović, J; Higgins, JP; Caldwell, DM; Reeves, BC; Shea, B; et al. (2016): ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology* 69: 225-234. DOI: 10.1016/j.jclinepi.2015.06.005.
- Wiig, S; Aase, K; von Plessen, C; Burnett, S; Nunes, F; Weggelaar, AM; et al. (2014): Talking about quality: exploring how ‘quality’ is conceptualized in European hospitals and healthcare systems. *BMC: Health Service Research* 14: 478. DOI: 10.1186/1472-6963-14-478.
- Williams, B (1994): Patient satisfaction: A valid concept? *Social Science & Medicine* 38(4): 509-516. DOI: 10.1016/0277-9536(94)90247-X.
- Willis, GB (2004): Cognitive Interviewing Revisited. A Useful Technique in Theory? Chapter 2. In: Presser, S; Hrsg.: *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, US-NJ: Wiley, 23-43. ISBN: 978-0-471-45841-8.

Willis, GB (2005): Cognitive Interviewing. A Tool to improve Questionnaire Design. Thousand Oaks, US-CA: SAGE Publications. ISBN: 0-7619-2803-0.

Wollersheim, H; Hermens, R; Hulscher, M; Braspenning, J; Ouwens, M; Schouten, J; et al. (2007): Clinical indicators: development and applications. *Netherlands Journal of Medicine* 65(1): 15-22. URL: <http://www.njmonline.nl/getpdf.php?id=488> (abgerufen am: 30.01.2017).