



Institut für Qualitätssicherung und
Transparenz im Gesundheitswesen

Methodische Grundlagen V1.1

Stand: 15. April 2019

Impressum

Thema:

Methodische Grundlagen V1.1

Ansprechpartner:

Dr. Dennis Boywitt

Stand:

15. April 2019

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

info@iqtig.org

<https://www.iqtig.org>

Übersicht

Tabellenverzeichnis.....	9
Abbildungsverzeichnis.....	10
Abkürzungsverzeichnis.....	11
Über die „Methodischen Grundlagen“	13
Teil A: Grundlagen der Qualitätssicherung und -förderung	15
1 Qualität der Gesundheitsversorgung	16
2 Qualitätssicherung, Qualitätsmanagement und Qualitätsverbesserung	25
3 Patientenzentrierung in der Qualitätssicherung.....	30
4 Verwendungszwecke von Qualitätsmessungen.....	38
5 Grundlegende Methodik der Qualitätsmessung.....	42
Teil B: Entwicklung und Durchführung von Qualitätssicherungsverfahren	49
6 Neuentwicklung von Qualitätssicherungsverfahren	50
7 Durchführung, Pflege und Weiterentwicklung von Qualitätssicherungsverfahren	94
Teil C: Methodische Elemente	112
8 Grundlagen der systematischen Literatur- und Informationsbeschaffung	113
9 Einbeziehung externen Wissens und externer Erfahrung	123
10 Eignung von Qualitätsindikatoren.....	135
11 Veröffentlichung von Indikator- und Kennzahlergebnissen	152
12 Bewertungskonzepte und Referenzbereiche.....	156
13 Grundgesamtheit, Stichprobenziehung und Vollerhebung	168
14 Datenvalidierung von QS-Dokumentationsdaten und Strukturierter Dialog.....	173
15 Methoden der Datenanalyse	180
Glossar.....	197
Literatur.....	202

Inhaltsverzeichnis

Tabellenverzeichnis.....	9
Abbildungsverzeichnis.....	10
Abkürzungsverzeichnis.....	11
Über die „Methodischen Grundlagen“	13
Teil A: Grundlagen der Qualitätssicherung und -förderung	15
1 Qualität der Gesundheitsversorgung	16
1.1 Rahmenkonzept für Qualität.....	16
1.2 Qualität auf Systemebene.....	21
1.3 Begründung der Qualitätsanforderungen.....	23
2 Qualitätssicherung, Qualitätsmanagement und Qualitätsverbesserung	25
2.1 Der Begriff der Qualitätssicherung	25
2.2 Rechtliche Grundlagen der gesetzlichen Qualitätssicherung.....	26
2.3 Die Rolle der externen Qualitätssicherung für die Verbesserung der Versorgungsqualität	27
3 Patientenzentrierung in der Qualitätssicherung.....	30
3.1 Verständnis von Patientenzentrierung	30
3.2 Umsetzung in den Arbeitsprozessen und -produkten des IQTIG	33
3.2.1 Methoden zur Förderung der Patientenzentrierung	33
3.2.2 Inhalte zur Förderung der Patientenzentrierung	37
3.2.3 Informationen für Patientinnen und Patienten über die Arbeitsprozesse des IQTIG und Ergebnisse der Qualitätssicherung	37
4 Verwendungszwecke von Qualitätsmessungen.....	38
5 Grundlegende Methodik der Qualitätsmessung.....	42
5.1 Methoden zur Messung von Qualität.....	42
5.2 Qualitätsmodelle als Ausgangspunkt für die Verfahrensentwicklung.....	44
5.2.1 Qualitätsaspekte der Gesundheitsversorgung.....	44
5.2.2 Auswahl der Qualitätsaspekte für ein Qualitätsmodell.....	45
5.2.3 Qualitätsmerkmale als Zwischenschritt der Konkretisierung	46
5.2.4 Zusammenfassung.....	47

Teil B: Entwicklung und Durchführung von Qualitätssicherungsverfahren	49
6 Neuentwicklung von Qualitätssicherungsverfahren	50
6.1 Konzeptstudie für Qualitätssicherungsverfahren.....	53
6.1.1 Themenerschließung	54
6.1.2 Identifikation von Qualitätsaspekten	55
6.1.3 Auswahl von Qualitätsaspekten für das Qualitätsmodell.....	56
6.2 Dokumentations- und sozialdatenbasierte Neuentwicklung von Qualitätsindikatoren	58
6.2.1 Kriteriengeleitete Konkretisierung der Qualitätsaspekte zu Qualitätsmerkmalen.....	59
6.2.2 Operationalisierung der Qualitätsmerkmale	62
6.2.3 Zusammenführung zum Qualitätsindikatorensatz	65
6.2.4 Schriftliches Stellungnahmeverfahren und Ergebnisbericht	65
6.3 Machbarkeitsprüfung für Dokumentation und Auswertung	66
6.3.1 Gegenstand der Machbarkeitsprüfung	67
6.3.2 Ablauf der Machbarkeitsprüfung	68
6.3.3 Ergebnisse der Machbarkeitsprüfung und Rückmeldeberichte.....	74
6.4 Entwicklung von Patientenbefragungen im Rahmen von Qualitätssicherungsverfahren	75
6.4.1 Faktororientierte Befragungsansätze	75
6.4.2 Methodische Schritte der Fragebogen- und Qualitätsindikatorenentwicklung.....	77
6.4.3 Eignungskriterien für Qualitätsindikatoren auf Basis von Patientenbefragungen.....	88
6.4.4 Risikoadjustierung von Qualitätsindikatoren der Patientenbefragung.....	93
7 Durchführung, Pflege und Weiterentwicklung von Qualitätssicherungsverfahren	94
7.1 Verfahrensdurchführung	94
7.2 Verfahrenspflege	95
7.2.1 Informationsquellen.....	97
7.2.2 Einbindung von Expertengruppen.....	97
7.2.3 Handlungsoptionen	98
7.3 Verfahrensweiterentwicklung.....	100
7.3.1 Anlässe für Verfahrensweiterentwicklungen.....	101
7.3.2 Durchführung von Verfahrensweiterentwicklungen.....	102

7.4	Feststellung von besonderem Handlungsbedarf.....	102
7.4.1	Zweck der Einstufung des besonderen Handlungsbedarfs.....	102
7.4.2	Definition von besonderem Handlungsbedarf.....	102
7.4.3	Vorgehen zur Einstufung des besonderen Handlungsbedarfs.....	104
7.4.4	Leitkriterien.....	105
7.4.5	Qualitätsindikatoren, für die keine Aussage zu besonderem Handlungsbedarf getroffen wird.....	109
7.4.6	Prozess der Einstufung des besonderen Handlungsbedarfs.....	110
7.4.7	Konsequenzen aus der Einstufung des besonderen Handlungsbedarfs	111
Teil C: Methodische Elemente		112
8	Grundlagen der systematischen Literatur- und Informationsbeschaffung	113
8.1	Systematische Recherche	114
8.1.1	Präzisierung der Fragestellung.....	114
8.1.2	Informationsquellen der systematischen Recherche nach wissenschaftlicher Literatur und Leitlinien.....	115
8.1.3	Weitere Recherchequellen	116
8.1.4	Ableitung der Suchwörter, Suchstrategie und Dokumentation.....	116
8.1.5	Auswahlprozess.....	117
8.2	Kritische Bewertung relevanter Publikationen	117
8.3	Zusammenfassende Qualitätsbewertung der Gesamtevidenz	119
8.3.1	Einzelstudien.....	119
8.3.2	Systematische Übersichtsarbeiten.....	121
8.3.3	Leitlinien	122
8.4	Qualitative Studien	122
9	Einbeziehung externen Wissens und externer Erfahrung	123
9.1	Ziele der Einbindung externen Wissens.....	123
9.2	Beratungszwecke bei der Einbindung externen Wissens	124
9.3	Methoden der Einbindung externen Wissens und externer Kompetenzen	125
9.3.1	Leitfadengestützte Interviews	126
9.3.2	Fokusgruppen	126
9.3.3	Einbindung externer Expertengruppen	128

10	Eignung von Qualitätsindikatoren.....	135
10.1	Grundsätze zur Eignungsbeurteilung.....	135
10.2	Eignungskriterien des Qualitätsziels des Indikators	138
10.2.1	Bedeutung für die Patientinnen und Patienten	138
10.2.2	Zusammenhang mit einem unmittelbar patientenrelevanten Ziel.....	139
10.2.3	Potenzial zur Verbesserung	140
10.2.4	Übereinstimmung mit einem Qualitätsaspekt.....	141
10.2.5	Beeinflussbarkeit durch den Leistungserbringer	141
10.3	Eignungskriterien des Messverfahrens.....	143
10.3.1	Objektivität der Messung	144
10.3.2	Datenqualität	145
10.3.3	Reliabilität der Messung	145
10.3.4	Validität der Messung.....	146
10.3.5	Praktikabilität der Messung.....	147
10.4	Eignungskriterien des Bewertungskonzepts	147
10.4.1	Angemessenheit des Referenzbereichs.....	148
10.4.2	Klassifikationsgüte.....	148
10.4.3	Angemessenheit der Risikoadjustierung	149
10.5	Übergreifende Eignungskriterien	149
10.5.1	Brauchbarkeit.....	149
10.5.2	Berücksichtigung unerwünschter Wirkungen.....	150
10.5.3	Zuschreibbarkeit der Verantwortung	150
11	Veröffentlichung von Indikator- und Kennzahlergebnissen	152
12	Bewertungskonzepte und Referenzbereiche.....	156
12.1	Begriffsbestimmung und Funktion von Bewertungskonzepten.....	156
12.2	Entwicklung von Bewertungskonzepten.....	158
12.3	Festlegung von Referenzbereichen	159
12.3.1	Informationsbewertung für die Festlegung von Referenzbereichen	160
12.3.2	Spezifikation des Referenzbereichs.....	161
12.4	Festlegung der Zuordnungsregel.....	163
12.5	Sentinel Events	166

13	Grundgesamtheit, Stichprobenziehung und Vollerhebung	168
13.1	Ebenen der Stichprobenziehung und Vollerhebung.....	168
13.2	Kriterien für Stichprobenziehung und Vollerhebungen.....	170
13.3	Vorgehen bei Stichprobenziehungen	171
14	Datenvalidierung von QS-Dokumentationsdaten und Strukturierter Dialog.....	173
14.1	Technische Prüfung bei Dateneingabe und -annahme.....	174
14.2	Statistische Prüfung nach Datenannahme.....	175
14.3	Datenabgleich mit Referenzquelle nach Datenannahme	177
14.4	Strukturierter Dialog / Stellungnahmeverfahren	179
15	Methoden der Datenanalyse	180
15.1	Grundlegende Annahmen.....	180
15.2	Risikoadjustierung	182
15.2.1	Stratifizierung.....	183
15.2.2	Regressionsmodelle.....	185
15.2.3	Zusammenfassung und Ausblick	186
15.3	Auswertung von Follow-up-Indikatoren	187
15.3.1	Aktives und passives Follow-up	187
15.3.2	Lexis-Diagramme.....	188
15.3.3	Auswertung nach Kohortensichtweise und nach Periodensichtweise.....	189
15.4	Aggregation mehrerer Qualitätsmessungen zum Zweck der Qualitätsförderung ..	191
15.4.1	Hierarchisierung von Ergebnisgrößen	192
15.4.2	Indizes in der Qualitätsmessung	193
	Glossar.....	197
	Literatur.....	202

Tabellenverzeichnis

Tabelle 1: Fiktive Beispiele für die methodischen Elemente der Qualitätsmessung.....	47
Tabelle 2: Beispiel für die Zuordnung der Qualitätsaspekte zu den Qualitätsdimensionen des IQTIG-Rahmenkonzepts	58
Tabelle 3: Eignungskriterien für Qualitätsindikatoren und dazugehörige Maßnahmen und Informationsquellen der Patientenbefragung.....	89
Tabelle 4: Qualitätsdarstellung auf Einrichtungs- und auf Systemebene im Vergleich	104
Tabelle 5: Anwendung der Leitkriterien 1a, 1b und 2 und der Zeitverläufe bei Indikatoren mit verschiedenen Bewertungskonzepten.....	109
Tabelle 6: Schema zur Evidenzbewertung in Bezug auf das Kriterium Quantität	120
Tabelle 7: Schema zur Evidenzbewertung in Bezug auf das Kriterium Qualität.....	120
Tabelle 8: Schema zur Evidenzbewertung in Bezug auf das Kriterium Konsistenz.....	121
Tabelle 9: Beratungszwecke der Einbindung externen Wissens mit Beispielen für die Einsatzbereiche und für die verwendeten Methoden	125
Tabelle 10: Beispiele für unmittelbar und mittelbar patientenrelevante Indikatorziele.....	140
Tabelle 11: Darstellung von Sensitivität und Spezifität bei der Feststellung von Abweichungen vom Referenzbereich	164
Tabelle 12: Im Krankenhaus verstorbene Patientinnen und Patienten mit isoliertem koronarchirurgischem Eingriff geschichtet nach Dringlichkeit des Eingriffs im Jahr 2015 ..	183

Abbildungsverzeichnis

Abbildung 1: Verbesserungspotenzial ausgehend von der erreichten Qualität bei einem Leistungserbringer und der derzeit bestmöglichen Qualität als Ansatzpunkt von Qualitätssicherung und internem Qualitätsmanagement	28
Abbildung 2: Qualitätsaspekte, Qualitätsmodell, Qualitätsmerkmale und Qualitätsindikatoren	46
Abbildung 3: Schritte und Produkte der Entwicklung eines QS-Verfahrens	52
Abbildung 4: Entwicklungsschritte vom Qualitätsmodell bis zum Qualitätsindikatorenset	59
Abbildung 5: Ablauf einer Machbarkeitsprüfung	69
Abbildung 6: Schematische Darstellung der Entwicklungsschritte einer Patientenbefragung ...	77
Abbildung 7: Methodische Schritte und Zwischenprodukte der Fragebogenentwicklung.....	78
Abbildung 8: Formative und reflektive Konstrukte	82
Abbildung 9: Hierarchische Struktur von Qualitätsaspekten, Qualitätsmerkmalen und Items und hierarchische Aggregation zu Qualitätsindikatoren.....	88
Abbildung 10: Überblick über die Eignungskriterien für Qualitätsindikatoren	137
Abbildung 11: Schematische Darstellung des Ausmaßes der leistungserbringerbedingten und nicht leistungserbringerbedingten Einflüsse auf den Indikatorwert anhand von Beispielen fallbezogener Qualitätsindikatoren	142
Abbildung 12: Schritte und zugehörige Eignungskriterien bei der Ermittlung und Aufarbeitung von Informationen für die Berechnung von Indikatorwerten.....	143
Abbildung 13: Funktion und Bestandteile des Bewertungskonzepts von Qualitätsindikatoren	157
Abbildung 14: 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	165
Abbildung 15: Beispielhafte Darstellung der Daten eines 1-Jahres-Follow-up-Indikators mit 100 Fällen in Form von Lebenslinien in einem Lexis-Diagramm	189
Abbildung 16: Lexis-Diagramme für einen 1-Jahres-Follow-up-Indikator in Kohortensichtweise und in Periodensichtweise.....	189

Abkürzungsverzeichnis

Abkürzung	Erläuterung
AHRQ	Agency for Healthcare Research and Quality
ATC	Anatomisch-Therapeutisch-Chemisches Klassifikationssystem (Anatomical Therapeutic Chemical Classification System)
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
DeQS-RL	Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung
DIMDI	Deutsches Institut für Medizinische Information und Dokumentation
G-BA	Gemeinsamer Bundesausschuss
GOP	Gebührenordnungsposition
HTA	Health Technology Assessment
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
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KHG	Gesetz zur wirtschaftlichen Sicherung der Krankenhäuser und zur Regelung der Krankenhauspflegesätze (Krankenhausfinanzierungsgesetz)
KHSG	Gesetz zur Reform der Strukturen der Krankenhausversorgung (Krankenhausstrukturgesetz)
OECD	Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (Organisation for Economic Co-operation and Development)
OPS	Operationen- und Prozedurenschlüssel
PCI	perkutane Koronarintervention (<i>percutaneous coronary intervention</i>)
plan. QI-RL	Richtlinie zu planungsrelevanten Qualitätsindikatoren
PREM	Patient-Reported Experience Measure
PROM	Patient-Reported Outcome Measure
Qb-R	Regelungen zum Qualitätsbericht der Krankenhäuser
Qesü-RL	Richtlinie zur einrichtungs- und sektorenübergreifenden Qualitätssicherung

Abkürzung	Erläuterung
QFR-RL	Qualitätssicherungs-Richtlinie Früh- und Reifgeborene
QI	Qualitätsindikator
QS	Qualitätssicherung
QSKH-RL	Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern
RCT	Randomisierte kontrollierte Studien
SGB	Sozialgesetzbuch

Über die „Methodischen Grundlagen“

Das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG, im folgenden Text auch „Institut“) ist sowohl nach dem Gesetz – § 137a Fünftes Buch Sozialgesetzbuch (SGB V) – als auch nach seiner Satzung ein fachlich unabhängiges, wissenschaftliches Institut und hat zu gewährleisten, dass es seine Aufgaben auf Basis der maßgeblichen, international anerkannten Standards der Wissenschaften erfüllt. Hierzu veröffentlicht das IQTIG in seinen „Methodischen Grundlagen“ die wissenschaftlichen Methoden, 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. Die Anwendung dieser Methoden begründet die Empfehlungen an den Gemeinsamen Bundesausschuss (G-BA; siehe Abschnitt 1.3). Die „Methodischen Grundlagen“ werden fortlaufend weiterentwickelt und Stellungnahmeverfahren unterzogen, um sie den sich kontinuierlich ändernden Erfordernissen der Qualitätssicherung und der Entwicklung der Wissenschaft anzupassen.

Inhalte und Struktur

Die „Methodischen Grundlagen“ beinhalten diejenigen Methoden, die vom Institut zum Zeitpunkt der Erstellung angewandt werden, und bestehen aus drei Teilen:

- In Teil A werden grundlegende Begriffe, Konzepte und Vorgehensweisen der externen Qualitätssicherung und -förderung dargestellt.
- In Teil B werden die konkreten Prozesse der Entwicklung, Weiterentwicklung und Durchführung von QS-Verfahren erläutert.
- In Teil C werden einzelne methodische Elemente dargestellt, die an verschiedenen Stellen der Entwicklungs- und Durchführungsprozesse eine Rolle spielen, wie etwa Fokusgruppen und Auswertungsmethoden.

Im Rahmen von Beauftragungen neu entwickelte Methoden werden in den entsprechenden Abschlussberichten dargestellt und finden Eingang in die „Methodischen Grundlagen“ bei nachfolgenden Aktualisierungen. Beispiele hierfür sind:

- die Methodik für die Entwicklung von Qualitätsmessungen zu Zwecken der Krankenhausplanung und für Qualitätszu- und -abschläge der Vergütung
- die Methodik für die Entwicklung leistungserbringerübergreifender Qualitätsmessungen, z. B. für sektorenübergreifende Langzeit-Follow-up-Indikatoren, regionale Indikatoren oder für die Darstellung von Systemqualität
- Grundsätze für die Evaluation von QS-Verfahren

Was ist neu in der Version 1.1?

Im Vergleich zur Version 1.0 der „Methodischen Grundlagen“ wurde vor allem Teil A inhaltlich überarbeitet und im gesamten Dokument wurden redaktionelle Änderungen vorgenommen. Als größere Änderungen sind zu nennen:

- In Abschnitt 1.1 wird ein Rahmenkonzept für Qualität für die gesetzliche Qualitätssicherung bestehend aus sechs grundlegenden Dimensionen dargestellt.
- In Kapitel 2 wurde ein Abschnitt zum rechtlichen Rahmen der gesetzlichen Qualitätssicherung ergänzt (Abschnitt 2.2).
- Das Kapitel 3 zur Patientenzentrierung in der Qualitätssicherung wurde hinzugefügt.
- Die Darstellung der Methodik der Entwicklung von Qualitätsmodellen und ihrer Rolle für die Verfahrensentwicklung wurde überarbeitet (Abschnitt 5.2 und Abschnitt 6.1).
- Die Methodik der Entwicklung von Patientenbefragungen wurde um Ausführungen zu formativen Konstrukten bei der Itementwicklung ergänzt (Abschnitt 6.4.2).
- Die aktuelle Methodik zur Feststellung besonderen Handlungsbedarfs wurde aufgenommen (Abschnitt 7.4).
- Die Darstellung der biometrischen Methodik zur Auswertung von Follow-up-Indikatoren wurde hinzugefügt (Abschnitt 15.3).

Stellungnahmeverfahren

Die „Methodischen Grundlagen V1.1s“ wurden am 15. November 2018 auf der Website des Instituts veröffentlicht und einem öffentlichen Stellungnahmeverfahren unterzogen. Bis zum 15. Januar 2019 wurde allen interessierten Organisationen und Einzelpersonen die Möglichkeit gegeben, schriftlich Stellung zu den Methoden des Instituts nehmen. Zur Stellungnahme aufgefordert wurden insbesondere die Vertreterinnen und Vertreter der nach § 137a Abs. 7 SGB V zu beteiligenden Organisationen:

- 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

Auf Grundlage der Stellungnahmen wurden die „Methodischen Grundlagen“ überarbeitet und einschließlich entsprechender Würdigung der Stellungnahmen am 15. April 2019 als „Methodische Grundlagen V1.1“ auf der Website des IQTIG veröffentlicht.

Teil A: Grundlagen der Qualitätssicherung und -förderung

1 Qualität der Gesundheitsversorgung

Das Ziel von Gesundheitssystemen ist eine Gesundheitsversorgung der Bevölkerung, die zu guter Gesundheit führt, an den Erwartungen der Bevölkerung ausgerichtet ist und fair verteilt ist (WHO 2000: xi, sinngemäß auch Deutscher Ethikrat 2016: 37 f.). Aus diesen Zielen des Gesundheitswesens, die per se einen Wert haben und ohne weitere Begründung vorausgesetzt werden, leiten sich alle weiteren Anforderungen ab (WHO 2000: 24). Die Bevölkerung bzw. die Patientinnen und Patienten stehen damit im Zentrum der Bemühungen im Gesundheitssystem, und die Gesundheitsversorgung ist danach zu beurteilen, inwieweit sie diese übergeordneten Ziele und die daraus abgeleiteten Anforderungen erfüllt.

Das IQTIG definiert Qualität in der Gesundheitsversorgung daher wie folgt:

Qualität der Gesundheitsversorgung ist der Grad, in dem die Versorgung von Einzelpersonen und Populationen Anforderungen erfüllt, die patientenzentriert sind und mit professionellem Wissen übereinstimmen.

Diese Definition entspricht der Anwendung der allgemeinen Qualitätsdefinition der DIN-Norm EN ISO 9000:2015 „Qualitätsmanagementsysteme – Grundlagen und Begriffe“ auf die Gesundheitsversorgung. In dieser Norm wird Qualität definiert als der „Grad, in dem ein Satz inhärenter Merkmale [...] Anforderungen erfüllt“ (DIN EN ISO 9000 2015: 39). Zugleich steht diese Definition des IQTIG im Einklang mit der Definition von Qualität des Institute of Medicine¹, nach der „Versorgungsqualität der Grad ist, in dem Gesundheitsleistungen für Einzelpersonen und Populationen die Wahrscheinlichkeit erwünschter Gesundheitsergebnisse erhöhen und mit aktuellem professionellem Wissen übereinstimmen“ (Lohr 1990: 21). Das Institute of Medicine fordert dabei explizit die Berücksichtigung der Wünsche und Werte der Patientinnen und Patienten – „erwünschte Gesundheitsergebnisse“ meint also wie in der IQTIG-Definition die Erfüllung patientenzentrierter Anforderungen.

Dieses Qualitätsverständnis bildet den Ausgangspunkt für die Erfüllung aller Aufgaben des IQTIG. Das Verständnis und die Rolle von Patientenzentrierung in der Versorgung und für die Erfüllung der Aufgaben des IQTIG werden in Kapitel 3 ausführlich beschrieben.

1.1 Rahmenkonzept für Qualität

Für die Messung, Bewertung und Weiterentwicklung der Qualität im Gesundheitswesen ist es wichtig, sowohl für Leistungserbringer als auch für die Kostenträger und für die Patientinnen und Patienten deutlich zu machen, welche grundsätzlichen Anforderungen von der gesetzlichen Qualitätssicherung an die Gesundheitsversorgung gestellt werden. Beispielsweise sollte eine Behandlung einerseits wirksam sein und andererseits selbst möglichst wenig schädliche Wirkungen entfalten, also sicher sein. Gleichzeitig sollte eine Behandlung an den Bedürfnissen der Patientinnen und Patienten ausgerichtet und angemessen sein. Solche grundsätzlichen Anforderungen

¹ Das US-amerikanische Institute of Medicine der National Academies of Sciences, Engineering, and Medicine wurde 2016 umbenannt in Health and Medicine Division (HMD).

an die Gesundheitsversorgung werden häufig in Form von grundlegenden Qualitätsdimensionen in einem Rahmenkonzept (*conceptual framework*) für Qualität zusammengefasst. Es gibt eine Vielzahl von Rahmenkonzepten, die verschiedene Dimensionen² der Qualität von Gesundheitsversorgung und/oder von Gesundheitssystemen in unterschiedlichem Detailgrad zusammenfassen (Übersichten z. B. bei Arah et al. 2006, Ahluwalia et al. 2017, Klassen et al. 2010). Die Rahmenkonzepte weisen untereinander deutliche Ähnlichkeiten auf, haben aber auch je nach Zielsetzung unterschiedliche Schwerpunkte. Manche sind als Grundlage für die Qualitätsverbesserung im Gesundheitswesen gedacht (Commonwealth Fund 2006, Committee on Quality of Health Care in America/Institute of Medicine 2001: 39), für den internationalen Vergleich der Qualität der Gesundheitsversorgung (Arah et al. 2006), als Richtschnur für die Akkreditierung von Gesundheitseinrichtungen (z. B. Joint Commission 2016) oder für das interne Qualitätsmanagement von Einrichtungen (DIN EN 15224 2017). Die Schwerpunkte unterscheiden sich insbesondere danach, ob der Blick eher auf das Gesundheitssystem als Ganzes gerichtet wird oder auf die Versorgungsqualität einzelner Leistungserbringer (Wiig et al. 2014, Klassen et al. 2010). Gemeinsam ist allen Rahmenkonzepten, dass sie deutlich machen, dass Qualität multidimensional ist und nicht anhand weniger, isolierter Aspekte umfassend beurteilt werden kann. Das IQTIG verwendet ein Rahmenkonzept für Qualität, das sich an das Rahmenkonzept des Institute of Medicine (Committee on Quality of Health Care in America/Institute of Medicine 2001: 39 f.) anlehnt, da dieses ebenfalls auf Aktivitäten zur Qualitätsverbesserung ausgerichtet ist. Im Kontext der leistungserbringerbezogenen externen Qualitätssicherung im deutschen Gesundheitswesen ergeben sich allerdings einige besondere Schwerpunkte im Vergleich zum US-amerikanischen System. Beispielsweise machen die Herausforderungen der sektorenübergreifenden Versorgung einen besonderen Fokus auf die Koordination der Versorgung erforderlich. Das Rahmenkonzept wurde daher an solche Besonderheiten angepasst.

Das Rahmenkonzept richtet sich mit seinen Anforderungen an den Prinzipien der Patientenzentrierung aus, wie sie in Kapitel 3 erläutert sind. Dabei ist Patientenzentrierung nicht als eine Dimension des Rahmenkonzepts zu verstehen, sondern als übergreifendes Leitbild aller Dimensionen des Rahmenkonzepts. Folgende Dimensionen werden als grundlegende Anforderungen an die Gesundheitsversorgung angesehen, die bei einer umfassenden leistungserbringerbezogenen Qualitätsbeurteilung berücksichtigt werden sollten (in Klammern wird der entsprechende englische Fachbegriff genannt):

▪ **Wirksamkeit** (*effectiveness*)

Unter Wirksamkeit der Versorgung wird das Ausmaß verstanden, in dem die angestrebten Versorgungsergebnisse tatsächlich erreicht werden (Donabedian 2003: 5, Klassen et al. 2010).

Um diese Ergebnisse zu erreichen, müssen die Strukturen und Prozesse der Versorgung dem

² In der Literatur werden unterschiedliche Begriffe verwendet, wie „Qualitätsdimensionen“ (Arah et al. 2006, Wiig et al. 2014), „Komponenten der Qualität“ (Donabedian 2003: 3 ff., JCAHO 1990: 8), „grundlegende Qualitätsaspekte“ (DIN EN 15224 2017: 77 ff.), „Kernziele des Gesundheitssystems“ (Commonwealth Fund 2006) oder „Domänen“ (NQMC-Website der AHRQ [kein Datum]). Da keine sich gegenseitig ausschließenden Kategorien gemeint sind, sondern gleichzeitig zu berücksichtigende Facetten, wird in den „Methodischen Grundlagen“ von „Qualitätsdimensionen“ gesprochen. Der Begriff ist nicht mit den Messdimensionen nach Donabedian (Struktur, Prozess, Ergebnis) identisch, die oft ebenfalls Qualitätsdimensionen genannt werden (siehe Abschnitt 5.1).

aktuellen Wissensstand entsprechen und die Wahrscheinlichkeit angestrebter Ergebnisse erhöhen (DIN EN 15224 2017: 78, JCAHO 1990: 9, Committee on Quality of Health Care in America/Institute of Medicine 2001: 39), oder einfacher ausgedrückt, das, was getan wird, muss auf richtige Weise getan werden (JCAHO 1990: 9). Wirksamkeit der Versorgung zielt also unmittelbar auf das übergeordnete Ziel „gute Gesundheit“, d. h. auf den Nutzen für Patientinnen und Patienten einschließlich körperlichen, psychischen und emotionalen Wohlbefindens.

■ **Patientensicherheit** (*safety*)

Patientensicherheit ist die Abwesenheit unerwünschter, schädlicher Ereignisse im Rahmen der Versorgung (Kohn et al. 2000: 4, APS [kein Datum]). Das Ziel einer guten Gesundheit lässt sich nur dann erreichen, wenn die Gesundheitsversorgung nicht nur wirksam, sondern auch sicher ist. Die Versorgungsstrukturen und -prozesse müssen also so ausgerichtet sein, dass schädliche Ereignisse vermieden werden.

■ **Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten** (*responsiveness*)

Eine an den Patientinnen und Patienten ausgerichtete Versorgung berücksichtigt deren Bedürfnisse und Werte und stellt sicher, dass alle Versorgungsentscheidungen durch diese geleitet werden (Committee on Quality of Health Care in America/Institute of Medicine 2001: 40). Dazu gehören insbesondere der respekt- und würdevolle Umgang von Gesundheitsprofessionen mit den Patientinnen und Patienten, umfassende und bedarfsgerechte Information und Kommunikation, das Angebot zur aktiven Beteiligung an Versorgungsprozessen und Entscheidungen, die Berücksichtigung der Diversität von Patientinnen und Patienten und ihrer Lebenslagen und Lebensphasen sowie die Einbeziehung des sozialen Umfelds bei der Versorgungsgestaltung (siehe Kapitel 3). Diese Aspekte werden in der Literatur mit dem Begriff *responsiveness* (Arah et al. 2006, WHO 2000: 31) oder ähnlich auch *acceptability* (Maxwell 1992) beschrieben.³

■ **Rechtzeitigkeit und Verfügbarkeit** (*timeliness*)

Rechtzeitigkeit bezeichnet das Ausmaß, in dem Versorgungsmaßnahmen dann für die Patientinnen und Patienten zur Verfügung gestellt werden, wenn sie benötigt werden (JCAHO 1990: 10). Das Institute of Medicine beschreibt diese Anforderung konkret als „reducing waits and sometimes harmful delays“ (Committee on Quality of Health Care in America/Institute of Medicine 2001: 51). Die rechtzeitige Verfügbarkeit von Versorgungsmaßnahmen ist allerdings nicht nur ein zeitlicher Aspekt, sondern wird auch von weiteren Faktoren bestimmt, wie z. B. der geografischen Erreichbarkeit von Versorgungsangeboten oder organisatorischen Hindernissen. Diese Anforderungen an die Versorgung werden daher aus dem Blickwinkel der Patientinnen und Patienten auch als Zugänglichkeit (*accessibility*) (Arah et al. 2006, JCAHO 1990: 8, Klassen et al. 2010), Zugang (*access*) (Campbell et al. 2000) oder Verfügbarkeit (*availability*) (Klassen et al. 2010) bezeichnet. Anders als beispielsweise Arah et al. (2006) fasst das IQTIG Rechtzeitigkeit bzw. Verfügbarkeit von Versorgungsmaßnahmen als Teil der Qualität auf. Für die Zwecke der Qualitätssicherung ist jedoch in besonderem Maße zu beachten, wer Adressat

³ Die vom Institute of Medicine verwendete Bezeichnung *patient-centered* wird hier nicht übernommen, da Patientenzentrierung als übergeordnetes Leitbild der Gesundheitsversorgung verstanden wird, siehe dazu Kapitel 3.

der Qualitätsanforderungen ist. Rechtzeitigkeit und Verfügbarkeit sind manchmal durch die Leistungserbringer beeinflussbar (beispielsweise die Zeitdauer zwischen dem Entschluss zu einem Notfallkaiserschnitt und der Entbindung oder die Ermöglichung eines niedrighschwelligem Zugangs zu Versorgungsangeboten), während sie in anderen Fällen eher Systemqualität beschreiben (beispielweise die regionale Erreichbarkeit von Versorgungsangeboten). Während die leistungserbringerbezogenen Themen traditioneller Teil der externen vergleichenden Qualitätssicherung in Deutschland sind, rücken die Systemaspekte des „Zugangs zur Versorgung“ erst schrittweise in den Fokus der gesetzlichen Qualitätssicherung.

▪ **Angemessenheit** (*appropriateness*)

Eine angemessene Versorgung vermeidet Überversorgung und Unterversorgung, indem sie genau die richtigen Maßnahmen umfasst, also die, die den Patientenbedürfnissen und evidenzbasierten fachlichen Standards entsprechen (Klassen et al. 2010). Sie folgt damit den Grundsätzen der evidenzbasierten Medizin und verbindet die beste verfügbare Evidenz mit individueller klinischer Expertise und den Präferenzen der Patientinnen und Patienten (Sackett et al. 1996). Der Begriff Angemessenheit umfasst also insbesondere Aspekte der Indikationsqualität und Evidenzbasierung aus Sicht der in Gesundheitsberufen Arbeitenden (wie z. B. in der DIN EN 15224 2017: 78, und bei JCAHO 1990: 9) und hat gleichzeitig einen engen Bezug zur Qualitätsdimension „Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten“, da bei Entscheidungen für oder gegen diagnostische oder therapeutische Maßnahmen immer auch Präferenzen und das Selbstbestimmungsrecht der Patientinnen und Patienten zu berücksichtigen sind. Während andere Autoren Angemessenheit unter Wirksamkeit subsumieren (Arah et al. 2006, Committee on Quality of Health Care in America/Institute of Medicine 2001: 47), sollte dieses Thema aus Sicht des IQTIG für eine differenziertere Betrachtung bei einer Qualitätsmessung gesondert berücksichtigt werden.

▪ **Koordination und Kontinuität** (*coordination, continuity*)

Eine gute Versorgung ist nicht nur eine Anforderung an einzelne Leistungserbringer, sondern es kommt auch auf die effektive Kooperation zwischen verschiedenen Leistungserbringern bei der Behandlung von Patientinnen und Patienten an. Koordination beschreibt das Ausmaß, in dem die Versorgungsmaßnahmen für Patientinnen und Patienten aufeinander abgestimmt sind, sowohl zwischen den beteiligten Versorgern als auch in zeitlicher Hinsicht (Klassen et al. 2010). Sie ist eine Voraussetzung dafür, dass die Patientinnen und Patienten eine kontinuierliche Versorgung erfahren. In der Literatur wird daher der Begriff Kontinuität (*continuity*) teilweise bedeutungsgleich verwendet (JCAHO 1990: 9, DIN EN 15224 2017: 78). Koordination bzw. Kontinuität werden von manchen Autorinnen und Autoren als Element der Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten verstanden (Committee on Quality of Health Care in America/Institute of Medicine 2001: 49, Arah et al. 2006). Aufgrund der hohen Bedeutung dieses Themas für die sektorenübergreifende Qualitätssicherung wird es durch das IQTIG als gesonderte Qualitätsdimension berücksichtigt.

Die Qualitätsdimensionen für die leistungserbringerbezogene Qualitätssicherung des IQTIG-Rahmenkonzepts sind demnach im Überblick:

- **Wirksamkeit**
- **Patientensicherheit**
- **Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten**
- **Rechtzeitigkeit und Verfügbarkeit**
- **Angemessenheit**
- **Koordination und Kontinuität**

Das IQTIG versteht diese Qualitätsdimensionen als einander ergänzende grundlegende Anforderungen, die die Versorgung gleichzeitig erfüllen soll. Konkrete Qualitätsziele und Qualitätsindikatoren beziehen sich daher oft auf mehrere dieser Dimensionen gleichzeitig (JCAHO 1990: 10). Beispielsweise kann das Qualitätsziel „leitlinienkonforme Systemwahl“ bei Herzschrittmacherimplantation sowohl „Wirksamkeit“ als auch „Patientensicherheit“ zugeordnet werden, und das Qualitätsziel „kurze präoperative Verweildauer“ bei Oberschenkelbruch adressiert sowohl „Rechtzeitigkeit“ als auch „Patientensicherheit“ und „Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten“. Qualität der Gesundheitsversorgung wird demnach als mehrdimensionales Konstrukt verstanden.

Die Qualitätsdimensionen des Rahmenkonzepts können im Rahmen der Entwicklung und Weiterentwicklung von QS-Verfahren dem Abgleich dienen, ob ein QS-Verfahren (bzw. dessen Qualitätsmodell bzw. dessen Qualitätsindikatorenset; siehe Abschnitt 5.2) alle relevanten Dimensionen adressiert oder ob sich unerwünschte „blinde Flecke“ bei der Qualitätsdarstellung ergeben. Der Abgleich der Inhalte der QS-Verfahren mit den Qualitätsdimensionen kann vor allem bei der Entwicklung neuer QS-Verfahren, aber auch bei der Prüfung von Anpassungsbedarf laufender QS-Verfahren erfolgen.

Im Folgenden wird auf einige weitere, in der Literatur genannte Anforderungen eingegangen und begründet, weshalb sie vom IQTIG für die Zwecke der leistungserbringerbezogenen gesetzlichen Qualitätssicherung nicht gesondert berücksichtigt werden:

- *efficacy*: Unter *efficacy* wird im Gegensatz zu *effectiveness* (beides dt. „Wirksamkeit“) das Potenzial einer Versorgungsmaßnahme verstanden, wenn sie unter Idealbedingungen eingesetzt wird (Donabedian 2003: 4, ähnlich JCAHO 1990: 9), während *effectiveness* auf die Ergebnisse der Versorgungsmaßnahmen unter Realbedingungen abzielt. Aus Sicht der Qualitätssicherung ist die tatsächliche Wirksamkeit maßgeblich. *Efficacy* wird daher bei der Entwicklung von Qualitätsmessungen vom IQTIG nicht gesondert betrachtet (ähnlich auch Campbell et al. 2000).
- Legitimität (*legitimacy*): Donabedian (2003: 22 f.) versteht unter Legitimität, dass die Versorgung gesellschaftlichen Werten, Prinzipien und Regeln entspricht. Diese Anforderung überschreitet den Rahmen der auf die individuellen Patientinnen und Patienten zentrierten Anforderungen an die Versorgung und kann diesen in manchen Fällen widersprechen, etwa, weil nicht beliebig viele Ressourcen für eine Patientin oder einen Patienten aufgewendet werden können. Im Rahmen der externen Qualitätssicherung in Deutschland wird diesem Thema

dadurch Rechnung getragen, dass die Qualitätsanforderungen an die Leistungserbringer durch Beschlüsse des G-BA normiert und legitimiert werden (siehe Abschnitt 1.3).

- Einrichtungsinterne Anforderungen wie z. B. Führung, Infrastruktur, Finanzmanagement, Wissensmanagement, Innovationsmanagement, internes Qualitätsmanagement, Mitarbeiterorientierung und Arbeitssicherheit werden von manchen Rahmenkonzepten ebenfalls als Qualitätsdimensionen aufgeführt (Klassen et al. 2010). Bei diesen Themen handelt es sich um einrichtungsinterne Ziele und Prozesse des Qualitätsmanagements und der Unternehmensführung. Sie stellen zwar wichtige Voraussetzungen für eine hohe Qualität der Versorgung durch die Leistungserbringer dar, sind aber keine patientenzentrierten Versorgungsziele per se, sondern Mittel, um diese zu erreichen (siehe Abschnitt 2.3).

1.2 Qualität auf Systemebene

Qualität im Gesundheitswesen kann auf verschiedenen Ebenen betrachtet werden. Die gesetzliche, indikatorbasierte externe Qualitätssicherung ist auf die Betrachtung der Qualität der Versorgung auf Ebene einzelner Leistungserbringer ausgerichtet: Die Bewertung und Darstellung der Qualität erfolgt in erster Linie pro Leistungserbringer. Die Anforderungen an die Versorgung auf dieser Ebene (siehe Abschnitt 1.1) sind demnach Anforderungen an die Versorgung individueller Patientinnen und Patienten (auch Mikro-Ebene genannt). Auf übergeordneter Ebene kann die Qualität der Gesundheitsversorgung auch für die Gesamtpopulation betrachtet werden, die sogenannte Systemebene (Makro-Ebene). Qualität auf Systemebene wird hier definiert als die Qualität der Gesundheitsversorgung für die Gesamtpopulation und ist damit im Einklang mit der grundlegenden Definition von Qualität der Gesundheitsversorgung (siehe Abschnitt 1.1). Auch bei einer solchen Betrachtung von Versorgungsqualität für Populationen sind die Qualitätsdimensionen des Rahmenkonzepts (Abschnitt 1.1) relevante Beschreibungsdimensionen. Beispielsweise ist Überleben bzw. eine niedrige Sterblichkeit ein Ziel, das gleichermaßen für Einzelpersonen wie für Populationen gilt und das damit gleichermaßen ein Qualitätsziel für einzelne Leistungserbringer wie für das Gesundheitssystem als Ganzes darstellt. Qualität auf Systemebene ist also zum Teil bestimmt durch die Versorgungsqualität der Gesamtheit der Leistungserbringer. Im Vergleich zur Betrachtung auf Ebene einzelner Leistungserbringer ergeben sich auf Systemebene allerdings zusätzliche Anforderungen, wie beispielsweise die Verteilung von Ressourcen, da bei dieser Betrachtungsweise die Versorgungsqualität von Populationen adressiert wird (Campbell et al. 2000). So gehen beispielsweise aus individueller Sicht wünschenswerte, aber besonders teure Behandlungen zulasten der Ressourcen für andere Patientinnen und Patienten. In diesem Fall müssen die konkurrierenden Anforderungen gegeneinander abgewogen werden (Campbell et al. 2000, Berwick et al. 2008, Donabedian 2003: 23).

In US-amerikanischen Publikationen zu Kriterien für die Leistungsfähigkeit von Gesundheitssystemen finden die Kosten bzw. die Effizienz von Behandlungsmaßnahmen besondere Aufmerksamkeit (Commonwealth Fund 2006: 5, 12 ff., Berwick et al. 2008, Committee on Quality of Health Care in America/Institute of Medicine 2001: 52) und werden als eigene Qualitätsdimension verstanden. Bei der Beurteilung der Gesundheitsversorgung für Populationen kann die Wirtschaftlichkeit der versorgenden Einrichtungen als ein wichtiger Aspekt für ein Gesundheits-

wesen angesehen werden (Campbell et al. 2000). Auch Donabedian (2003: 11 ff.) nennt *optimality* als eine Komponente von Qualität und versteht darunter, dass eine bestimmte Verbesserung des Nutzens durch Versorgungsmaßnahmen an dem Punkt optimal ist, an dem diese Verbesserung mit dem geringsten zusätzlichen Ressourceneinsatz erzielt wurde. Die Bedeutung dieses Themas ist den begrenzten Ressourcen in jedem Gesundheitswesen geschuldet, die Effizienzüberlegungen auf Systemebene erforderlich machen (SVR-Gesundheit 2018: 45) und die dazu führen, dass ein Gesundheitswesen nicht maximale Versorgung für alle, sondern allenfalls „wesentliche Versorgung [...] für alle“ (sogenanter „new universalism“, WHO 2000: 15) leisten kann. Aus der Perspektive der individuellen Patientinnen und Patienten schlagen sich Kosten als deren Tragbarkeit (*affordability*) nieder. Da im deutschen System der gesetzlichen Krankenversicherung nach § 12 Abs. 1 SGB V eine Finanzierung aller notwendigen Versorgungsmaßnahmen durch die Krankenkassen übernommen werden soll, ist dieser Aspekt für die Erreichung der eingangs erwähnten übergeordneten Versorgungsziele (gute Gesundheit, Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten, Fairness) in Bezug auf die individuelle Behandlung von untergeordneter Bedeutung. Wie die Autoren des OECD-Modells (Arah et al. 2006) betrachtet das IQTIG die Kosten für Gesundheitsversorgung für die leistungserbringerbezogene externe Qualitätssicherung nicht als Teil der Qualität der Versorgung.

Auf Ebene der Versorgung von Populationen ist außerdem die Versorgungsgerechtigkeit als zusätzliche Anforderung zu berücksichtigen. Versorgungsgerechtigkeit (*equity*) meint, dass Versorgung nicht von persönlichen Patientenmerkmalen wie Geschlecht, ethnischer Abstammung, Wohnort und sozioökonomischem Status abhängig sein soll (Committee on Quality of Health Care in America/Institute of Medicine 2001: 53, Klassen et al. 2010), sondern nur von fachlichen Erwägungen und Patientenpräferenzen geleitet sein soll. Vom Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen wird hierfür auch der Begriff der bedarfsgerechten Versorgung verwendet (SVR-Gesundheit 2018: 49, ähnlich auch Deutscher Ethikrat 2016: 54 ff.). Im deutschen Gesundheitswesen soll diesem Ziel insbesondere durch das Solidaritätsprinzip (§ 1 und § 3 SGB V) Rechnung getragen werden. Unter der Annahme, dass sich die an der Versorgung Beteiligten wegen ihres Berufsethos um eine Gleichbehandlung aller Patientinnen und Patienten unter Berücksichtigung ihrer Bedarfe bemühen, wird die Versorgungsgerechtigkeit in erster Linie als Eigenschaft des Systems und nicht als Eigenschaft einzelner Leistungserbringer betrachtet. Versorgungsgerechtigkeit wird daher als eine Dimension von Systemqualität verstanden (Campbell et al. 2000).

Diese zusätzlichen Anforderungen „Kosten“ bzw. „Effizienz“ und „Versorgungsgerechtigkeit“ liegen größtenteils außerhalb des Einflussbereichs der Leistungserbringer und richten sich daher an Adressaten auf Ebene des Gesundheitssystems (Makro-Ebene oder Systemebene), beispielsweise an Einrichtungen der Gesundheitspolitik, Institutionen der Selbstverwaltung im Gesundheitswesen oder Berufsverbände.⁴ Systemqualität umfasst also nicht nur die Versorgungsqualität der Gesamtheit der Leistungserbringer, sondern auch den Grad der Erfüllung dieser zusätzlichen Anforderungen.

⁴ Bei Versorgungskonzepten nach dem Vorbild des *Triple Aim* des Institute for Healthcare Improvement (Berwick et al. 2008), bei denen eine Organisation die Gesamtverantwortung für die Versorgung einer

Die systembezogene Betrachtung von Qualität der Gesundheitsversorgung im Rahmen der externen Qualitätssicherung nach §§ 136 ff. SGB V ist bereits im SGB V angelegt. So ist beispielsweise in § 136d geregelt, dass der G-BA den Stand der Qualitätssicherung im Gesundheitswesen festzustellen hat. Dies beinhaltet eine leistungserbringerübergreifende Darstellung der Versorgungsqualität. In § 137a Abs. 10 ist weiterhin geregelt, dass wissenschaftliche Forschung auf Basis der QS-Daten erfolgen soll. Zum anderen lenken die Verfahren der sektorenübergreifenden Qualitätssicherung sowie die Einbeziehung von Follow-up-Erhebungen, Patientenbefragungen und Sozialdaten bei den Krankenkassen neben der QS-Dokumentation den Blick zunehmend auf leistungserbringerübergreifende Qualitätsaspekte. So wird z. B. die Versorgungsqualität von Patientinnen und Patienten mit chronischen Erkrankungen sowohl im Zeitverlauf als auch in der parallelen und abwechselnden Versorgung durch unterschiedliche Leistungserbringer zukünftig abbildbar werden. Insbesondere wird bei einer Versorgung, die mehrere Leistungserbringer über einen gewissen Zeitraum einschließt, die Frage nach der Verantwortlichkeit und dem Umgang mit Verbesserungsbedarfen in den Vordergrund rücken. Ein Konzept zur Messung und Darstellung von Versorgungsqualität, die nicht im Einflussbereich eines einzelnen Leistungserbringers liegt, wird in Zukunft vom IQTIG entwickelt.

1.3 Begründung der Qualitätsanforderungen

Qualität wurde eingangs als Grad der Erfüllung patientenzentrierter Anforderungen definiert. Welche Aspekte der Gesundheitsversorgung einer Messung und Qualitätsbewertung unterzogen werden sollten und welche Qualitätsanforderungen an die Gesundheitsversorgung gestellt werden, kann aus verschiedenen Perspektiven unterschiedlich beurteilt werden. Eine objektivierbare, Vergleiche ermöglichende Darstellung von Versorgungsqualität setzt daher einheitliche, begründete und nachvollziehbare Anforderungen an die Versorgung voraus. Da diese Anforderungen in Form von Qualitätsindikatoren in der externen gesetzlichen Qualitätssicherung Grundlage für Verbesserungsmaßnahmen und Steuerungsinstrumente im Gesundheitswesen sein sollen, müssen sie sie ihrerseits hohen Anforderungen genügen:

▪ Patientenzentrierung der Qualitätsziele

Die Anforderungen an die medizinische Versorgung müssen patientenzentriert sein (siehe Abschnitt 1.1 und Kapitel 3). Dies können z. B. Anforderungen sein, die als Ergebnisqualität den Patientennutzen direkt adressieren. Es können aber auch 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 gesellschaftliche Werte und Übereinkünfte Anforderungen vorgeben, die sich auf die an den Patientinnen und Patienten ausgerichtete Gestaltung der Versorgung beziehen, wie etwa die Erwartung partizipativer Entscheidungsfindung in der Versorgung.

Population inklusive der Verteilung der dafür benötigten Ressourcen hat, wären auch Leistungserbringer Adressaten von populationsbezogenen Qualitätsanforderungen. Bisher liegen solche Versorgungsansätze in Deutschland jedoch nur als Modellprojekte vor.

▪ Entwicklung und Durchführung gemäß wissenschaftlicher Methodik

Durch die transparente Anwendung einer wissenschaftlichen Methodik bei der Entwicklung und Durchführung der QS-Verfahren wird sichergestellt, dass die Instrumente sachgerecht sind. Die Instrumente müssen die in den Eignungskriterien festgehaltenen Anforderungen für Qualitätsmessungen erfüllen (siehe Kapitel 10). Beispielsweise muss es den Leistungserbringern prinzipiell möglich sein, die patientenbezogenen Ziele durch eine entsprechende Gestaltung ihrer Versorgung zu erreichen (Kriterien Beeinflussbarkeit und Angemessenheit des Referenzbereichs). Dadurch, dass die Entwicklung durch die Eignungskriterien geleitet wird, wird also sichergestellt, dass die Qualitätsanforderungen begründet sind.

Die gemäß diesen Anforderungen entwickelten Instrumente zur Messung und Bewertung der Qualität werden dem G-BA zur Beratung vorgelegt. Durch die Beratung und Beschlussfassung im G-BA wird ihre Anwendung in der externen gesetzlichen Qualitätssicherung formal legitimiert. Durch diesen letzten Schritt erlangen die Instrumente und die darin operationalisierten Qualitätsanforderungen normativen Charakter.

2 Qualitätssicherung, Qualitätsmanagement und Qualitätsverbesserung

Die Gesundheitsversorgung entspricht nicht immer den Anforderungen, die an sie gestellt werden. Dies wird auch im deutschen Gesundheitswesen beispielsweise an einer teilweise hohen Variation von Versorgungsergebnissen deutlich (z. B. Pross et al. 2017 oder Website Versorgungsatlas.de). Ein Teil solcher Unterschiede in der Versorgung ist dabei nicht auf Innovationen, Unterschiede in Patientenpräferenzen oder zufällige Einflüsse zurückzuführen, sondern stellt unerwünschte Variation dar (Appleby et al. 2011), d. h., er ist durch Qualitätsunterschiede bedingt. Qualitätssicherung und Qualitätsmanagement haben zum Ziel, diese unerwünschte Variation zu verringern und dadurch die Versorgungsqualität zu verbessern (Berwick 1991).

2.1 Der Begriff der Qualitätssicherung

Der Begriff der Qualitätssicherung ist historisch eng mit dem Begriff des Qualitätsmanagements verbunden, das im 20. Jahrhundert schwerpunktmäßig in Unternehmen des industriellen Wirtschaftssektors zunehmend Verbreitung fand. Die Orientierung an der Aufdeckung von Defiziten anhand von Qualitätskontrollen erklärt die frühe Prägung des Begriffs der „Qualitätssicherung“, die ein gefordertes Maß an Qualität sicherstellen soll.

In der Qualitätsmanagement-Normenreihe DIN 9000 ff. wird Qualitätssicherung als Teil des Qualitätsmanagements begriffen. Qualitätsmanagement umfasst danach mehrere Komponenten, wobei der Begriff „Qualitätssicherung“ den „Teil des Qualitätsmanagements, der auf das Erzeugen von Vertrauen darauf gerichtet ist, dass Qualitätsanforderungen erfüllt werden“, bezeichnet und von „Qualitätsplanung“, „Qualitätssteuerung“ und „Qualitätsverbesserung“ abgegrenzt wird (DIN EN ISO 9000 2015: 31). „Qualitätssicherung“ im Sinne der DIN wird als „Zusicherung“ verstanden (*quality assurance*) und umfasst die Prozesse, durch die Sicherheit darüber geschaffen werden soll, dass bzw. ob die Qualitätsanforderungen erfüllt werden. Qualitätssicherung nach dieser Definition meint also im Wesentlichen die Messung und Bewertung der Qualität und wird auch als Qualitätsdarlegung bezeichnet (Sens et al. 2018: 74 und 81).

Im Gesundheitswesen wird der Begriff Qualitätssicherung dagegen traditionell in einem umfassenderen Sinn verwendet und schließt die Maßnahmen zur Verbesserung der Qualität mit ein (Donabedian 2003: xxiv, Lohr 1990: 46). Der Begriff „Sicherung“ meint hier also auch „Sicherstellung“ von Qualität. Dieses Begriffsverständnis lässt sich ebenfalls in den rechtlichen Regelungen des SGB V zur Qualitätssicherung im Gesundheitswesen ablesen (deutlich z. B. in § 135b Abs. 1 und § 136a Abs. 2 SGB V), wenngleich die Gesetzestexte keine Nominaldefinition von Qualitätssicherung enthalten. In der Praxis der externen Qualitätssicherung im deutschen Gesundheitswesen hat sich das Verständnis des Begriffs darüber hinaus dahingehend erweitert, dass nicht nur die Qualitätsmessung und die Behebung von Qualitätsdefiziten das Ziel der QS-Verfahren ist, sondern auch die Förderung der Weiterentwicklung von Versorgungsqualität

(Wünschmann 2010: 23 f.). So sieht beispielsweise das Krankenhausstrukturgesetz nicht nur Abschlüsse 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.

Auch das IQTIG fasst den Begriff der Qualitätssicherung im Rahmen des SGB V daher weiter als die DIN EN ISO 9000 und versteht darunter die Prozesse und Maßnahmen, durch die die Qualität der Versorgung gewährleistet oder verbessert werden soll.

2.2 Rechtliche Grundlagen der gesetzlichen Qualitätssicherung

War die Umsetzung von Qualitätsstandards zunächst eine interne Aufgabe des ärztlichen und zahnärztlichen Berufsstandes, so haben sich in den letzten 15 bis 20 Jahren mehr und mehr gesetzliche und untergesetzliche Regelungen etabliert, die die Umsetzung von Qualitätsvorgaben und -maßnahmen im Gesundheitswesen steuern. Bereits in den allgemeinen Bestimmungen der §§ 2 und 70 SGB V werden Vorgaben zur Qualität der Gesundheitsversorgung von gesetzlich versicherten Patientinnen und Patienten gemacht. Der neunte Abschnitt des SGB V beschreibt nunmehr in den §§ 135 ff. die „Sicherung der Qualität der Leistungserbringung“. Dabei hat der G-BA die Aufgabe zugewiesen bekommen, in seinen Richtlinien nach den §§ 136 ff. SGB V Vorgaben zur Umsetzung der Qualitätssicherung in der vertrags(zahn)ärztlichen und stationären Versorgung zu machen. Diese Aufgabe hat er in mittlerweile über 20 Richtlinien umgesetzt. Zur Erarbeitung der fachlichen Grundlagen für seine Regelungen bedient sich der G-BA zweier Institute. Zum einen wurde 2004 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) nach § 139a SGB V gegründet, das u. a. für die Bewertung von Behandlungsmethoden sowie des Nutzens und der Kosten von Arzneimitteln zuständig ist. Zum anderen beauftragt der G-BA ein unabhängiges wissenschaftliches Institut mit der Entwicklung und Durchführung von einrichtungsübergreifenden QS-Verfahren. Nach der Beauftragung zunächst externer Institute gründete der G-BA im Jahre 2014 die Stiftung für Qualitätssicherung und Transparenz im Gesundheitswesen, die die Trägerin des gleichnamigen Instituts nach § 137a Abs. 1 SGB V (IQTIG) ist. Das IQTIG hat 2016 die Aufgaben nach § 137a SGB V übernommen.

Das Institut arbeitet im Auftrag des G-BA an Maßnahmen zur Qualitätssicherung und zur Darstellung der Versorgungsqualität im Gesundheitswesen. Es soll gemäß § 137a SGB V insbesondere beauftragt werden,

- für die Messung und Darstellung der Versorgungsqualität möglichst sektorenübergreifend abgestimmte risikoadjustierte Indikatoren und Instrumente einschließlich Module für ergänzende Patientenbefragungen zu entwickeln,
- die notwendige Dokumentation für die einrichtungsübergreifende Qualitätssicherung unter Berücksichtigung des Gebotes der Datensparsamkeit zu entwickeln,
- sich an der Durchführung der einrichtungsübergreifenden Qualitätssicherung zu beteiligen,
- die Ergebnisse der Qualitätssicherungsmaßnahmen in geeigneter Weise und in einer für die Allgemeinheit verständlichen Form zu veröffentlichen,
- auf der Grundlage geeigneter Daten, die in den Qualitätsberichten der Krankenhäuser veröffentlicht werden, einrichtungsbezogen vergleichende risikoadjustierte Übersichten über die

- Qualität in maßgeblichen Bereichen der stationären Versorgung zu erstellen und in einer für die Allgemeinheit verständlichen Form im Internet zu veröffentlichen,
- für die Weiterentwicklung der Qualitätssicherung zu ausgewählten Leistungen die Qualität der ambulanten und stationären Versorgung zusätzlich auf der Grundlage geeigneter Sozialdaten darzustellen, die dem Institut von den Krankenkassen nach § 299 Abs. 1a auf der Grundlage von Richtlinien und Beschlüssen des Gemeinsamen Bundesausschusses übermittelt werden, sowie
 - Kriterien zur Bewertung von Zertifikaten und Qualitätssiegeln, die in der ambulanten und stationären Versorgung verbreitet sind, zu entwickeln und anhand dieser Kriterien über die Aussagekraft dieser Zertifikate und Qualitätssiegel in einer für die Allgemeinheit verständlichen Form zu informieren.

Nach § 137a Abs. 4 Satz 2 SGB V kann auch das Bundesministerium für Gesundheit das Institut unmittelbar mit Untersuchungen und Handlungsempfehlungen zu den oben genannten Aufgaben für den G-BA beauftragen. Das IQTIG leitet seine Arbeitsergebnisse gem. § 137b SGB V dem G-BA als Empfehlungen zu. Dieser berücksichtigt sie im Rahmen seiner Aufgabenstellungen und entscheidet, ob und wie er sie in seinen Richtlinien und Beschlüssen umsetzt. Die Beschlüsse des G-BA sind nach § 91 Abs. 6 SGB V für die Träger des G-BA, deren Mitglieder und Mitgliedskassen sowie für die Versicherten und die Leistungserbringer verbindlich.

Die Aufträge zur Erarbeitung von Maßnahmen zur Qualitätssicherung und zur Darstellung der Versorgungsqualität im Gesundheitswesen beziehen sich dabei ausschließlich auf Leistungen der Gesetzlichen Krankenversicherung nach SGB V. Leistungen nach anderen Büchern des Sozialgesetzbuches wie z. B. Leistungen der Rentenversicherung (SGB VI) oder der Pflegeversicherung (SGB XI) sind nicht Inhalt dieser Aufträge.

2.3 Die Rolle der externen Qualitätssicherung für die Verbesserung der Versorgungsqualität

Mit der Münchner Perinatalstudie (1975–1977) wurde in Deutschland erstmals der Schritt von der Versorgungsforschung hin zur systematischen externen Qualitätssicherung 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 krankenhausbazogene, 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“ und „statistische Unterlagen zur Beantwortung gemeinsamer perinatologischer Fragestellungen bereitgestellt werden“ (Conrad 1977).

Das Verfahren umfasste die 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. Das Prinzip bestand also in externen Vergleichen der Versorgungsqualität zwischen Leistungserbringern mittels Qua-

litätsindikatoren. Dies gilt als die Geburtsstunde der externen Qualitätssicherung in Deutschland, der bald weitere Projekte in der stationären Versorgung folgten, z. B. in der Chirurgie, Anästhesiologie, und später auch in der ambulanten Versorgung.

Das Zusammenspiel zwischen Rückmeldung der Ergebnisse der externen Qualitätssicherung und der Verbesserung der Versorgungsqualität wird in Abbildung 1 dargestellt. Ziel der externen Qualitätssicherung ist es, die Erfüllung von Mindestanforderungen an die Versorgung, die nicht unterschritten werden dürfen, sicherzustellen und darüber hinaus Verbesserungen anzustoßen, die darauf zielen, die bestmögliche Qualität zu erreichen (siehe Abbildung 1). Medizinischer Fortschritt und Weiterentwicklung der Versorgung in all ihren Aspekten sorgen dafür, dass sich die Vorstellungen davon, was Mindestanforderungen sind und was die derzeit bestmögliche Qualität ist, kontinuierlich weiterentwickeln, z. B. durch neue Therapien, Medikamente, Versorgungsformen etc.

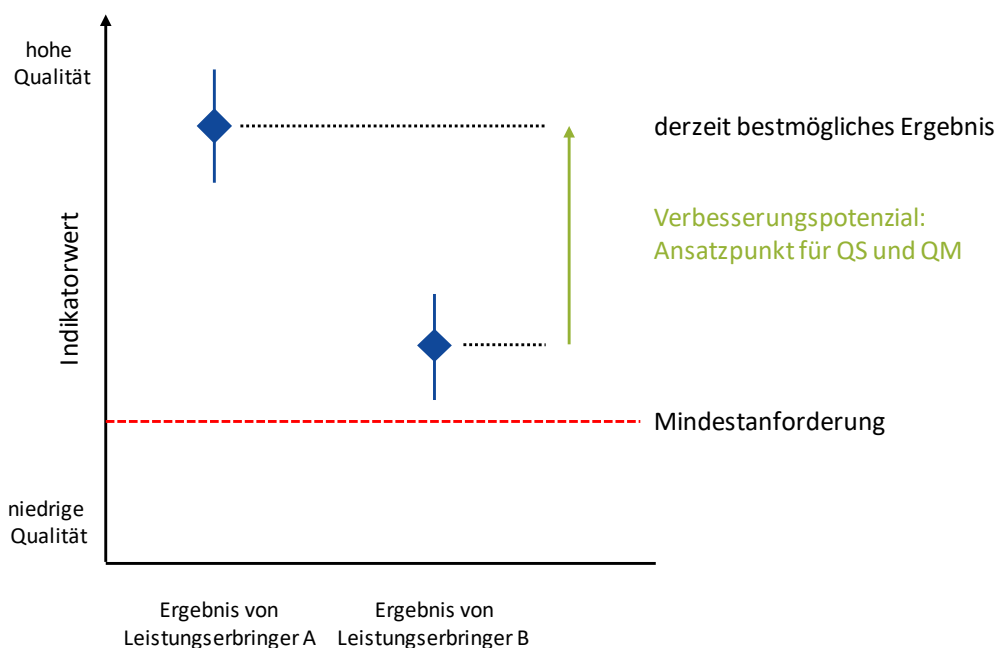


Abbildung 1: Verbesserungspotenzial ausgehend von der erreichten Qualität bei einem Leistungserbringer und der derzeit bestmöglichen Qualität als Ansatzpunkt von Qualitätssicherung und internem Qualitätsmanagement

In der Tradition der Münchner Perinatalstudie fördert die externe Qualitätssicherung die Qualität der Gesundheitsversorgung unter anderem dadurch, dass sie dem internen Qualitätsmanagement der Leistungserbringer aussagekräftige Messergebnisse aus gegebenenfalls adjustierten externen Vergleichen mit anderen Leistungserbringern oder mit Sollwerten (siehe „Mindestanforderung“ in Abbildung 1) zur Verfügung stellt (Geraedts 2009: 6). Mittels der Ergebnisse der externen Qualitätssicherung können die Leistungserbringer ihre Stärken und Verbesserungspotenziale besser einschätzen und die Erfolge ihres kontinuierlichen Verbesserungsprozesses kontrollieren. Für den Gebrauch der Messergebnisse im Rahmen des internen Qualitätsmanagements kann sich die Bewertung aber auch an ehrgeizigeren Zielen als den Referenzbereichen

der externen Qualitätssicherung orientieren oder sie kann die Verbesserung im Zeitverlauf beurteilen. Für das interne Qualitätsmanagement können daher auch Kennzahlen der externen QS-Verfahren genutzt werden, für die kein Referenzbereich festgelegt wurde.

Im Rahmen der externen Qualitätssicherung ist es jedoch wichtig, dass die Ergebnisse der Qualitätsmessungen bei verschiedenen Leistungserbringern anhand eines einheitlichen Maßstabs bewertet werden, damit vergleichende Aussagen zwischen Leistungserbringern sowie objektivierbare, durch Dritte nachvollziehbare Qualitätsaussagen möglich sind. Durch diese vergleichende Messung und Bewertung von Versorgungsqualität kann festgestellt werden, ob ein Leistungserbringer die an ihn gestellten Anforderungen (siehe Abschnitt 1.3) erreicht. Welcher Grad von Qualität erwartet werden kann, wird empirisch aus den Ergebnissen eines solchen Qualitätsvergleichs oder entsprechender Studien abgeleitet und bei manchen Themen, wie beispielsweise Fragen der partizipativen Entscheidungsfindung, auf Basis sozialer Normen festgelegt (Abschnitt 12.3).

Die externe Qualitätssicherung als Ganzes wird getragen durch die entsprechenden Aktivitäten der verschiedenen Akteure im Gesundheitswesen, wie dem G-BA, dem IQTIG, den beauftragten Stellen auf Landesebene und den Leistungserbringern. Die Verantwortung für die kontinuierliche Verbesserung der medizinischen Versorgungsqualität liegt dabei jedoch in erster Linie bei den Leistungserbringern. Denn die Versorgungsqualität auf Ebene der individuellen Patientenversorgung kann nur von den Leistungserbringern selbst verbessert werden und ihr Bekenntnis zur Qualität ist Voraussetzung für eine wirksame Qualitätsverbesserung (Donabedian 2003: xxx). Aufgabe des IQTIG ist dabei vor allem die unabhängige und transparente Messung und Darstellung von Versorgungsqualität in Kooperation mit den verschiedenen Partnern im Gesundheitswesen. Qualitätssicherung kann besonders dann gut funktionieren, wenn die Ergebnisse der externen Qualitätsvergleiche und die fachlichen Rückmeldungen für das interne Qualitätsmanagement nützlich sind und auch bestmöglich genutzt werden. Hierzu zählt z. B., dass zwischen der externen Qualitätssicherung und dem internen Qualitätsmanagement zeitnahe, inhaltlich relevante Schnittstellen existieren.

3 Patientenzentrierung in der Qualitätssicherung

Das Kapitel zielt darauf, das Verständnis des IQTIG von Patientenzentrierung und die damit verbundenen Implikationen für die Methoden und Inhalte der Verfahrensentwicklung, -weiterentwicklung, -pflege und -durchführung darzulegen. Darüber hinaus wird die angestrebte Information von Patientinnen und Patienten über die Arbeitsweise und -ergebnisse des IQTIG adressiert.

Der Begriff Patientenzentrierung ist national und international nicht einheitlich definiert. Darüber hinaus finden sich in der Literatur zahlreiche Begriffe, die eine hohe konzeptionelle Überschneidung aufweisen oder synonym verwendet werden (z. B. Patientenorientierung, Nutzerorientierung, *people-centeredness*; exemplarisch Gerlinger 2009). Das IQTIG greift hier auf die dahinterstehenden inhaltlichen Ansätze zurück, um das eigene Verständnis für den Kontext der externen Qualitätssicherung darzulegen. Es wird ausschließlich der Begriff Patientenzentrierung verwendet.

3.1 Verständnis von Patientenzentrierung

Das Ziel des Gesundheitssystems ist die Versorgung von Patientinnen und Patienten⁵ entsprechend ihrer Bedarfe und Bedürfnisse, basierend auf dem aktuellen wissenschaftlichen Kenntnisstand (Committee on Quality of Health Care in America/Institute of Medicine 2001, SVR-Gesundheit 2018). Um dieses Ziel zu erreichen, ist es erforderlich, im Sinne einer patientenzentrierten Versorgung, Patientinnen und Patienten in den Fokus der Bemühungen im Gesundheitssystem zu stellen (Battles 2006, Arah et al. 2006, Gerlinger 2009, Berwick 2002, Deutscher Ethikrat 2016): „Putting the patient first“, wie es in Forderungen zur Gestaltung einer qualitativ hochwertigen Gesundheitsversorgung pointiert formuliert wird (Mid Staffordshire NHS Foundation Trust Public Inquiry 2013: 1675). Dies beinhaltet, 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“ (Klemperer 2000: 15).⁶ Diese Forderung gilt zum einen für die direkte Interaktion zwischen Gesundheitsprofessionen und Patientinnen bzw. Patienten, die sogenannte Mikroebene der Gesundheitsversorgung (Committee on Quality of Health Care in America/Institute of Medicine 2001, Gerlinger 2009). Sie zielt zum anderen aber auch auf die Organisation und Ausrichtung von Institutionen und Einrichtungen im Gesundheitswesen, die sogenannte Meso- und Makroebene (Gerlinger 2009, Jo Delaney 2018, Greene et al. 2012, Steiger und Balog 2010). Um einer Über-, Unter- oder Fehlversorgung entgegenzusteuern, sollten dabei neben dem sogenannten objektiven Bedarf auch der sogenannte subjektive Bedarf sowie der sogenannte latente Bedarf⁷ von Patientinnen und Patienten Berücksichtigung finden, unter vordringlicher

⁵ Als Patientin bzw. Patient werden Menschen verstanden, die Gesundheitsleistungen in Anspruch nehmen.

⁶ Dieses Verständnis steht im Einklang mit international einflussreichen Betrachtungsweisen wie der des Institute of Medicine (Committee on Quality of Health Care in America/Institute of Medicine 2001).

⁷ Der latente Bedarf meint vor allem die Situation, „wenn eine Leistung medizinisch objektivierbar zwar sinnvoll und notwendig wäre, aber [...] der Patient darüber nicht informiert ist oder seinen Zustand falsch

Beachtung von deren Selbstbestimmungsrecht (vgl. SVR-Gesundheit 2018, SVR-Gesundheit 2014, SVR-Gesundheit 2002). Insgesamt folgt daraus, dass die Strukturen, Prozesse und Ergebnisse der Gesundheitsversorgung entsprechend den Anforderungen, die Patientinnen und Patienten bei der Nutzung des Gesundheitssystems stellen, kontinuierlich weiterentwickelt werden müssen (Klemperer 2000, Arah et al. 2006, Gerlinger 2009, Battles 2006). Um all diese Bemühungen tatsächlich praxiswirksam werden zu lassen, ist Patientenzentrierung auch als eine Frage der Organisationskultur zu verstehen (Moore et al. 2017, Bokhour et al. 2009, SVR-Gesundheit 2018).

Für das IQTIG ergeben sich aus dieser Betrachtung zwei Konsequenzen: Zum Ersten setzt sich das IQTIG – als Institution des Gesundheitswesens – das Ziel, sich selbst patientenzentriert auszurichten und dahingehend beständig weiterzuentwickeln. Zum Zweiten verfolgt das IQTIG mit seinen Arbeitsprodukten den Anspruch, so weit wie möglich eine patientenzentrierte Ausrichtung der Gesundheitsversorgung zu fördern.

Bisher mangelt es an einem Goldstandard, wie die Belange von Patientinnen und Patienten auf institutionalisierter Ebene in die Entwicklung, Weiterentwicklung, Pflege und Durchführung von QS-Verfahren einzubinden sind. Das IQTIG orientiert daher sein Verständnis von Patientenzentrierung in der Qualitätssicherung an der eingangs ausgeführten Begriffsbestimmung. Zugleich wird Patientenzentrierung als eine grundlegende Anforderung festgelegt, die durch QS-Verfahren bei der Messung und Darstellung von Qualität zu berücksichtigen ist (Kapitel 1).⁸ Ausgehend vom obigen Verständnis leitet das IQTIG konkrete Kernanforderungen an Patientenzentrierung ab, die als programmatisch für die Arbeit des IQTIG zu verstehen sind. Die Kernanforderungen unterstützen die Erfüllung des wissenschaftlichen Auftrags des IQTIG und konkretisieren das zugrundegelegte Verständnis von Patientenzentrierung. Die Möglichkeiten der Förderung von Patientenzentrierung in der Qualitätssicherung sind dabei immer innerhalb des vom Gesetzgeber beschriebenen Rahmens bzw. des gesetzlichen Auftrags des IQTIG zu verstehen (siehe Abschnitt 2.2).

Die Kernanforderungen an Patientenzentrierung lauten im Einzelnen (Committee on Quality of Health Care in America/Institute of Medicine 2001, Gerteis et al. 1993, Arah et al. 2006, Carinci et al. 2015, Battles 2006, Berwick 2002, Perfetto et al. 2017, SVR-Gesundheit 2018, Scholl et al. 2014):

1. Ausrichtung am Nutzen für Patientinnen und Patienten

- die Ausrichtung der Versorgungsstrukturen und -prozesse am Nutzen für Patientinnen und Patienten
- die Ausrichtung der Versorgungsergebnisse am Nutzen für Patientinnen und Patienten

einschätzt oder [...] die Nutzung der entsprechenden Gesundheitsleistung aus anderen Gründen, beispielsweise wegen mangelnder Angebotskapazitäten, nicht erfolgt“ (SVR-Gesundheit 2018: 50).

⁸ An dieser Stelle sei nochmal darauf verwiesen, dass Patientenzentrierung, im Sinne einer evidenzbasierten Gesundheitsversorgung, eingebettet in den aktuellen wissenschaftlichen Kenntnisstand betrachtet wird (siehe auch Kapitel 1).

2. Ausrichtung an der Sicherheit von Patientinnen und Patienten

- die Ausrichtung daran, dass Versorgungsstrukturen und -prozesse Patientinnen und Patienten nicht schaden
- die Ausrichtung daran, dass Versorgungsergebnisse Patientinnen und Patienten nicht schaden
- die Vermeidung von Über-, Unter- oder Fehlversorgung

3. Ausrichtung der Versorgungsgestaltung an Patientinnen und Patienten bzw. Responsiveness

- die Wahrnehmung der Patientinnen und Patienten als Individuum ausgehend von ihrer individuellen Lebenssituation (biopsychosoziale Perspektive)
- die Wahrnehmung der individuellen Bedarfe und Bedürfnisse, Präferenzen und Werte der Patientinnen und Patienten sowie die responsive Ausrichtung der gesamten Versorgungsgestaltung daran
- der Zugang zu einer angemessenen Versorgung
- eine koordinierte, kontinuierliche und integrierte Versorgungsgestaltung
- der respekt- und würdevolle Umgang von Gesundheitsprofessionen mit Patientinnen und Patienten ausgehend vom Angebot einer partnerschaftlichen Beziehungsgestaltung
- das Angebot von Information, Kommunikation, Edukation und Unterstützung, um Selbstbestimmung, informierte Entscheidungen, Teilhabe und Selbstmanagement zu ermöglichen (Empowerment)
- das Angebot zur aktiven Beteiligung an Versorgungsprozessen und Entscheidungen (Shared Decision Making) in selbstgewähltem Umfang
- die Förderung von körperlichem, psychischem und emotionalem Wohlbefinden
- die Einbeziehung des sozialen Umfelds (z. B. Familie)
- die Umsetzung und Wahrung von gesetzlich verankerten Rechten von Patientinnen und Patienten

4. Berücksichtigung der Diversität und Lebenslagen von Patientinnen und Patienten

- die Anerkennung und Berücksichtigung der Diversität, Lebenslagen und kulturellen Vielfalt von Patientinnen und Patienten einschließlich unterschiedlicher Lebensphasen, in denen Erkrankungen auftreten bzw. Behandlungen nötig werden
- die Anerkennung und Berücksichtigung der Diversität von Krankheitsverläufen und -phasen, die Patientinnen und Patienten erleben, einschließlich der Anforderungen von besonders vulnerablen Subpopulationen (z. B. Patientinnen und Patienten mit schwerwiegendem Erkrankungsverlauf)

5. Transparenz und Information für Patientinnen und Patienten

- die Bereitstellung von allgemeinverständlichen, relevanten und aktuellen Informationen über Prozesse
- die Bereitstellung von allgemeinverständlichen, relevanten und aktuellen Informationen über Ergebnisse
- die Verfügbarkeit von Ansprechpartnerinnen bzw. -partnern

6. Patientenbeteiligung und konsequente Einbindung der Patientenperspektive in Entwicklungsprozesse
 - die systematische Erfassung und Ausrichtung an der Patientenperspektive und -erfahrung
 - das Angebot strukturierter und zielgerichteter Beteiligungsmöglichkeiten für Patientinnen und Patienten
7. kontinuierliche Organisationsentwicklung zur responsiven Ausrichtung an Belangen von Patientinnen und Patienten
 - die langfristige, systematische Verankerung in der Organisationskultur sowie die kontinuierliche Weiterentwicklung
 - die langfristige, systematische Verankerung in den Arbeitsprozessen und -produkten sowie die kontinuierliche Weiterentwicklung

Für eine patientenzentrierte Ausrichtung der Gesundheitsversorgung in der Mikroebene, die über die Arbeitsprodukte des IQTIG gefördert werden kann, sind vor allem die Kernanforderungen 1 bis 4 maßgeblich. Hingegen berücksichtigt die patientenzentrierte Ausrichtung des IQTIG als Institution des Gesundheitswesens alle sieben aufgeführten Kernanforderungen, was sich in den nachfolgend ausgeführten Methoden, Inhalten sowie der Art und Weise der Information widerspiegelt.

3.2 Umsetzung in den Arbeitsprozessen und -produkten des IQTIG

Um die oben genannten, als programmatisch zu verstehenden Anforderungen umzusetzen, lassen sich drei zentrale Bereiche unterscheiden, an denen das IQTIG zur Förderung einer Patientenzentrierung in der Qualitätssicherung ansetzt:

1. die **Methoden**, die angewandt werden, um die inhaltliche Ausrichtung an den Belangen von Patientinnen und Patienten zu erhöhen und deren Perspektiven einzubinden
2. die **Inhalte**, die Belange von Patientinnen und Patienten adressieren können
3. die Art und Weise der **Information**, die Patientinnen und Patienten über die Arbeitsprozesse des IQTIG und die Ergebnisse der Qualitätssicherung erhalten

Diese drei Bereiche werden nachfolgend dargestellt. Eine Beschreibung der genauen Umsetzung findet sich in den entsprechenden Kapiteln.

3.2.1 Methoden zur Förderung der Patientenzentrierung

Das IQTIG setzt in der Verfahrensentwicklung, -weiterentwicklung, -pflege und -durchführung gezielt bestimmte Methoden ein, um die inhaltliche Ausrichtung an den Belangen von Patientinnen und Patienten zu erhöhen. Dies umfasst

- die Einbindung der Patientenperspektive sowie
- eine gezielte Adressierung patientenrelevanter Qualitätsaspekte, -merkmale und -indikatoren und
- die Einschätzung der patientenzentrierten Ausrichtung von Qualitätsaspekten, -merkmalen und -indikatoren.

Der Ansatz sieht sowohl vor, Patientinnen und Patienten direkt einzubeziehen, als auch im Sinne von Patientinnen und Patienten, also für sie, zu agieren (Gerlinger 2009).

Einbindung der Patientenperspektive

Die Einbindung der Patientenperspektive wird als nutzbringend erachtet, um z. B. spezifische Relevanzen, Bedürfnisse, Erfahrungen, Sichtweisen oder Erleben von den Betroffenen einer Krankheit oder Behandlung selbst zu erfahren. Auch ist die Patientenperspektive das Mittel der Wahl, um Versorgungsprozesse und -ergebnisse von Patientinnen und Patienten ausgehend nachzuvollziehen. Sie kann nicht durch Berichte von Gesundheitsprofessionen ersetzt werden. Die Einbindung der Patientenperspektive kann sowohl über (1) eine Analyse bereits vorliegender wissenschaftlicher Ergebnisse als auch (2) über eigene empirische Erhebung bzw. strukturierte Beteiligung geschehen.

Bei der **Analyse bereits vorliegender wissenschaftlicher Ergebnisse** werden Erkenntnisse zur Patientenperspektive generiert, ohne dass Patientinnen oder Patienten bzw. die Patientenvertreterinnen oder -vertreter der maßgeblichen Organisationen nach § 140f SGB V direkt beteiligt sind. Dies kann z. B. über Literaturanalysen zur Patientenperspektive erfolgen. Eine solche Vorgehensweise ermöglicht es, den aktuellen Forschungsstand zur Patientenperspektive zu einem Thema sehr breit zu erfassen sowie auch Unterthemen zu identifizieren, die dann in der weiteren Bearbeitung vertieft werden können. Sie kann aber auch den Mitarbeiterinnen und Mitarbeitern des IQTIG Informationen liefern, um sich in einem ersten Schritt mit der Situation betroffener Patientinnen und Patienten vertraut zu machen. Darüber hinaus können auf diese Weise mitunter Hinweise auf Perspektiven besonders vulnerabler Subgruppen (z. B. pflegebedürftige Menschen, Menschen mit terminalem Krankheitsverlauf) generiert werden. Diese können an eigenen empirischen Erhebungen des IQTIG bzw. strukturierten Beteiligungen in der Regel nicht selbst teilnehmen. Es besteht daher die Gefahr, dass ihre Perspektiven, Bedürfnisse und Bedarfe eher vernachlässigt werden.

Bei der **empirischen Erhebung bzw. strukturierten Beteiligung** kommen hingegen Methoden zum Einsatz, die Patientinnen und Patienten bzw. deren Vertreterinnen und Vertreter selbst in den Arbeitsprozess einbinden, z. B. in Fokusgruppen, Stellungnahmeverfahren oder auch als Teilnehmerinnen und Teilnehmer in Expertengruppen. Dabei ist im Wesentlichen zu unterscheiden zwischen dem direkten Einbezug der Patientenvertretung der maßgeblichen Organisationen nach § 140f SGB V (benannt über die Patientenvertretung des G-BA bzw. deren Mitgliedsorganisationen) und dem direkten Einbezug von betroffenen Patientinnen und Patienten, die in keinerlei Verbindung zur Qualitätssicherung oder zu Patientenverbänden stehen. Die Auswahl erfolgt jeweils auftragsspezifisch nach methodischen Überlegungen und dem angestrebten Erkenntnisinteresse, unter Berücksichtigung der gesetzlichen Regelungen. Hervorzuheben ist, dass die Einbindung von Patientinnen und Patienten bzw. deren Vertreterinnen und Vertretern kein Selbstzweck ist oder als eigenes Ergebnis per se betrachtet wird, sondern immer zielgerichtet erfolgt (Carman und Workman 2017).

Dabei werden die empirischen Methoden nach folgenden Grundprinzipien angewendet:

Fokusgruppen und Interviews zielen zumeist auf die Generierung von krankheits- und/oder behandlungsspezifischen Erkenntnissen über die Wissensordnung, Alltagserfahrungen, Perspektiven und Verständnisweisen von Menschen mit verschiedenen Bildungshintergründen, Krankheitserfahrungen, Geschlechtern etc. (siehe Kapitel 9). Hier strebt das IQTIG an, Patientinnen und Patienten, die selbst von einer Erkrankung bzw. Behandlung betroffen sind und projektspezifische weitere Kriterien (z. B. Alter, Geschlecht, Bildungshintergrund, Krankheitsdauer und -schwere) erfüllen, einzubeziehen. Dabei sollen, so weit wie möglich, die Diversität und unterschiedlichen Lebenslagen von Patientinnen und Patienten berücksichtigt werden. Die Rekrutierung erfolgt über unterschiedliche, möglichst breit gefächerte, Ansprachewege (z. B. Ärztinnen/Ärzte, Zeitungsannoncen).

Bedarf es einer Einordnung und Vertretung von patientenrelevanten Themen in übergeordneten Kontexten, z. B. in Expertengremien, bindet das IQTIG hingegen primär Patientenvertreterinnen und Patientenvertreter ein. Diese werden von der Patientenvertretung des G-BA bzw. deren Mitgliedsorganisationen (Vertretung der maßgeblichen Organisationen nach § 140f SGB V) entsandt. Darüber hinaus können sie zwecks Beratung zu spezifischen Themen, die für Patientinnen und Patienten von Belang sind, eingeladen werden, z. B. zum Thema Barrierefreiheit. Entsprechend dem projektspezifischen Erkenntnisinteresse zieht das IQTIG ggf. ergänzend auch Patientinnen und Patienten, die betroffen sind oder eine spezifische Funktion einnehmen (z. B. Krankenhausfürsprecher), hinzu. Der primäre Einbezug der Patientenvertretung des G-BA bzw. deren Mitgliedsorganisationen bleibt hiervon unberührt. Bei projektspezifischem Bedarf können auch die Patientenvertretung des G-BA bzw. deren Mitgliedsorganisationen unterstützend in die Rekrutierung von Patientinnen und Patienten für Fokusgruppen involviert werden. Die Entscheidung darüber obliegt dem IQTIG.

Eine weitere empirische Methode zur Erfassung der Patientenperspektive sind Patientenbefragungen (siehe Abschnitt 6.4). Sie kommen bei der Entwicklung und Umsetzung der QS-Verfahren als eigenständige Datenquelle in Betracht (siehe Kapitel 6). Ihre Entwicklung und Umsetzung wird vom G-BA beauftragt. Auch hier wird bei der Verfahrensentwicklung auf die beiden vorangehend beschriebenen Vorgehensweisen zur Erfassung der Patientenperspektive und der nachfolgend beschriebenen Adressierung und Einschätzung patientenrelevanter Qualitätsaspekte, -merkmale und -indikatoren zurückgegriffen (siehe Abschnitt 5.2).

Gezielte Adressierung patientenrelevanter Qualitätsaspekte, -merkmale und -indikatoren

Die Adressierung patientenrelevanter Qualitätsaspekte, -merkmale und -indikatoren kann hingegen auch durch Dritte (z. B. Wissenschaftlerinnen und Wissenschaftler) erarbeitet werden. Die Umsetzung erfolgt anhand der gezielten Herausarbeitung und Berücksichtigung der eingangs beschriebenen spezifischen Kernanforderungen, die als kennzeichnend für eine patientenzentrierte Ausrichtung gelten. Das bedeutet, dass in der Verfahrensentwicklung und -durchführung z. B. Themen wie Mortalität, Patientensicherheit, Schmerzmanagement, Aspekte der Lebensqualität und Entscheidungsbeteiligung bei der Analyse der Literatur besondere Beach-

tung finden. Diese Vorgehensweise zielt darauf, von Beginn eines Projekts an patientenrelevante Themen aufzugreifen, aufrechtzuerhalten und auch in den Phasen zu berücksichtigen, in denen Patientinnen und Patienten nicht direkt eingebunden werden können. Darüber hinaus werden auf diese Weise systematisch Themen berücksichtigt, bei denen von einer allgemein hohen Relevanz für einen Großteil der Patientinnen und Patienten ausgegangen werden kann (z. B. keine Schmerzen haben, bei einem elektiven Eingriff nicht versterben, deutliche Reduktion der Symptombelastung, keine Schädigung durch eine Behandlung erleiden).

Einschätzung der patientenzentrierten Ausrichtung von Qualitätsaspekten, -merkmalen und -indikatoren

Bei der Einschätzung der patientenzentrierten Ausrichtung von erarbeiteten Qualitätsaspekten und -merkmalen oder auch Qualitätsindikatoren wird abgewogen, ob die Kernanforderungen einer patientenzentrierten Qualitätssicherung berücksichtigt sind. Dies spiegelt sich beispielsweise auch in der Anwendung entsprechender Kriterien bei der Eignungsprüfung von Qualitätsaspekten, -merkmalen und -indikatoren wider (siehe Kapitel 10). Diese Herangehensweise kann förderlich sein, um den Anspruch einer patientenzentrierten Ausrichtung über den Projektverlauf hinweg aufrechtzuerhalten und kontinuierlich zu prüfen. Die Einschätzung kann durch Dritte (z. B. Wissenschaftlerinnen und Wissenschaftler) durchgeführt werden. Zusätzlich erfolgt zu definierten Zeitpunkten eine strukturierte Beteiligung von Patientinnen und Patienten bzw. Patientenvertreterinnen und Patientenvertretern bei der Einschätzung (z. B. in Expertengremien und in Stellungnahmeverfahren).

Einsatz von Methodenkombination

Alle drei Ansätze – die Einbindung der Patientenperspektive, die gezielte Adressierung patientenzentrierter Qualitätsaspekte, -merkmale und -indikatoren sowie die Einschätzung der patientenzentrierten Ausrichtung – können einander nicht ersetzen. Um die Stärken der Ansätze voll auszuschöpfen, kommen in der Regel alle drei zur Anwendung und werden miteinander verbunden. Beispielsweise kann über eine Recherche nach qualitativen Studien zur Patientenperspektive bezüglich einer bestimmten Prozedur ein umfänglicher, aktueller Stand der Forschung zu einem Thema erarbeitet werden. Darüber hinaus liegt ein besonderer Fokus auf Themen, die direkt auf die Förderung eines patientenrelevanten Nutzens und der Patientensicherheit sowie der Sicherung einer patientenzentrierten Versorgungsgestaltung zielen (z. B. in Leitlinien). Mittels Fokusgruppen mit Patientinnen und Patienten kann auf die gewonnenen Erkenntnisse aufgebaut werden und können diese um spezifische Informationen vertieft werden. Die Diskussion der abstrahierten Erkenntnisse bei der Konzeption der Qualitätsindikatoren erfolgt anschließend im Expertengremium mit Wissenschaftlerinnen und Wissenschaftlern sowie Fachexpertinnen bzw. -experten der Gesundheitsprofessionen unter Beteiligung von Patientenvertreterinnen und -vertretern.

3.2.2 Inhalte zur Förderung der Patientenzentrierung

Bei den Inhalten, die über QS-Verfahren adressiert werden, ist anzustreben, dass Themen fokussiert werden, die für Patientinnen und Patienten unmittelbar von Belang sind (siehe Abschnitt 10.2.1). Abgebildet werden diese über Qualitätsaspekte und -merkmale, die über Qualitätsindikatoren erfasst werden (siehe Abschnitt 5.2). Dies bedeutet, dass Struktur-, Prozess- und Ergebnisindikatoren einen patientenrelevanten Nutzen und/oder Schaden bzw. spezifische Themen adressieren sollen, die eine auf Patientinnen und Patienten ausgerichtete Versorgung kennzeichnen. Zu den unmittelbar patientenrelevanten Themen gehören Aspekte der Mortalität, Morbidität und Lebensqualität. Zu den patientenrelevanten Themen, die eine hohe Responsiveness bzw. Ausrichtung der interpersonellen, kontextuellen und organisatorischen Aspekte der Versorgungsgestaltung an Patientinnen und Patienten kennzeichnen, gehören z. B. Patienteninformation und -edukation, Berücksichtigung der Patientenpräferenzen und Mitbestimmung, körperliches Wohlbefinden und Schmerzmanagement, Versorgungskontinuität, Zugang zur Versorgung, psychische und emotionale Unterstützung, Einbeziehung der Familie. Mittelbar patientenrelevante Qualitätsindikatoren, die auf sogenannte Surrogate zurückgreifen, sind nur dann für eine patientenzentrierte Qualitätssicherung geeignet, wenn ein ausreichend beschriebener Zusammenhang mit einem dieser unmittelbar patientenrelevanten Ziele bzw. Themen vorliegt (siehe Abschnitt 10.2.2).

3.2.3 Informationen für Patientinnen und Patienten über die Arbeitsprozesse des IQTIG und Ergebnisse der Qualitätssicherung

Zu einer Ausrichtung als patientenzentrierte Institution gehört auch die transparente Bereitstellung von Informationen für Patientinnen und Patienten in einer allgemeinverständlichen Art und Weise. Das IQTIG hat nicht den Auftrag, Einzelpersonen zu beraten, sondern strebt an, Patientinnen und Patienten entsprechend seinem Auftrag nach § 137a SGB V transparent und verständlich über seine Arbeitsprozesse sowie die Ergebnisse der Qualitätssicherung zu informieren. Dies umfasst auch den kontinuierlichen Ausbau allgemeinverständlicher Berichtsformate. Die zentralen Abschlussberichte des IQTIG (ohne Anhänge) werden bereits weitestgehend barrierefrei veröffentlicht. Der jährliche Qualitätsreport ist auf der Website des IQTIG barrierefrei zugänglich. Perspektivisch wird auch „Information“ als eigenständiger Verwendungszweck von Qualitätsmessungen ausgebaut (siehe Kapitel 4). Darüber hinaus bietet das IQTIG eine zentrale Ansprechstelle per E-Mail und per Telefon sowohl für Patientinnen und Patienten als auch für Patientenvertreterinnen und -vertreter.

4 Verwendungszwecke von Qualitätsmessungen

Während alle QS-Verfahren die Verbesserung der Versorgungsqualität zum Ziel haben, können sich die Maßnahmen unterscheiden, mit denen dieses Ziel erreicht werden soll. Häufig wird die Verbesserung der Versorgungsqualität als letztendlich von den Leistungserbringern durchzuführende, interne Maßnahme unterschieden von externen Maßnahmen wie Auswahlentscheidungen durch Public Reporting, qualitätsorientierte Vergütung oder Zulassungsentscheidungen (Lohr 1990: 46 f., Davies et al. 2011, Jarvis 2000, Berwick et al. 2003, Remus und Fraser 2004, Damberg et al. 2011, Geraedts 2009: 6, Solberg et al. 1997). Die Verwendung von Qualitätsmessungen zur Steigerung der Versorgungsqualität durch externe Maßnahmen wird auch unter den Begriffen *accountability* (Solberg et al. 1997), Überwachungs- und Kontrollfunktion (Geraedts 2009: 6) oder *improvement through selection* (Berwick et al. 2003) subsumiert.

Die Nutzung von Qualitätsmessungen und deren Ergebnissen im Rahmen der gesetzlichen Qualitätssicherung fand bislang in Form der Qualitätsförderung durch Unterstützung des internen Qualitätsmanagements der Leistungserbringer und in Form der Veröffentlichung der Ergebnisse in den strukturierten Qualitätsberichten der Krankenhäuser statt. Der Gesetzgeber hat die Verwendungszwecke der Qualitätsförderung und der Information um selektivvertragliche Anreize, Krankenhausplanung und qualitätsabhängige Vergütung erweitert. Im Kontext der externen Qualitätssicherung nach SGB V unterscheidet das IQTIG folgende Verwendungszwecke von Qualitätsmessungen:

- Qualitätsförderung durch externen Vergleich (Verfahren der externen gesetzlichen Qualitätssicherung, auf Grundlage von § 136 Abs. 1 SGB V)
- Information (z. B. Patienteninformation über Versorgungsqualität einzelner Leistungserbringer, auf Grundlage von § 137a Abs. 3 SGB V)
- Selektivvertragliche Anreize (Qualitätsverträge nach § 110a SGB V)
- Planung der Versorgungsstrukturen (planungsrelevante Qualitätsindikatoren nach § 136c SGB V)
- Qualitätszu- und -abschläge (qualitätsabhängige Vergütung mit Zu- und Abschlägen nach § 136b SGB V)

Diese Verwendungszwecke unterscheiden sich im Wesentlichen in den Handlungsanschlüssen, für die die Qualitätsmessungen genutzt werden. Die Qualitätsmessungen selbst unterscheiden sich nicht in der zugrunde liegenden Definition von Versorgungsqualität (siehe Kapitel 1) oder in den grundlegenden Qualitätsdimensionen (siehe Abschnitt 1.1). Entsprechend den unterschiedlichen Verwendungszwecken müssen die Mess- und Bewertungsinstrumente (z. B. Referenzbereiche) gegebenenfalls unterschiedlichen Anforderungen genügen. Dies hat Auswirkungen auf die Entwicklung und Ausgestaltung der Verfahren (McGlynn 2003, Solberg et al. 1997).

Qualitätsförderung durch externen Vergleich

Bisher wurden die Qualitätsergebnisse der externen stationären Qualitätssicherung und der sektorenübergreifenden Qualitätssicherung hauptsächlich für die transparente Darstellung der Versorgungsqualität verwendet. Die Qualitätsergebnisse sollen dabei der Unterstützung des internen Qualitätsmanagements durch externen Vergleich dienen und so Verbesserungen der Versorgungsqualität initiieren (siehe Abschnitt 2.3). Die in diesem Rahmen möglichen Maßnahmen beinhalten unter anderem die Teilnahme an Fortbildungen, Peer Reviews, aber auch Vereinbarungen über die Implementation von Behandlungspfaden oder Handlungsempfehlungen anhand von Leitlinien.

Information

Unter Information wird die gezielte, adressatengerechte Bereitstellung von Qualitätsinformationen verstanden. Ein Beispiel hierfür sind Informationsangebote im Internet, anhand derer Patientinnen und Patienten Leistungserbringer hinsichtlich ihrer Versorgungsqualität vergleichen können (z. B. § 137a Abs. 3, Satz 2, Nr. 5 SGB V). Für gezielte Informationsangebote können auch Daten herangezogen werden, die nicht aus der gesetzlichen Qualitätssicherung stammen, aber z. B. für Patientinnen und Patienten bei der Auswahl von Versorgungseinrichtungen von Nutzen sein können. Darunter fallen beispielsweise Informationen zum Versorgungsangebot oder zur Erreichbarkeit einer Einrichtung.

Von der gezielten Information abzugrenzen ist die öffentliche Bereitstellung der Ergebnisse der gesetzlichen Qualitätssicherung ohne weitere ziel- und adressatengerechte Aufbereitung. Dies entspricht dem Gebot der Transparenz und ist ein grundlegendes Erfordernis für alle Qualitätsmessungen (siehe auch Kapitel 11).

Selektivvertragliche Anreize

Ein weiterer möglicher Verwendungszweck von Qualitätsmessungen ist die Förderung der Versorgungsqualität im Rahmen sogenannter Qualitätsverträge. Bei den Qualitätsverträgen nach § 110a SGB V handelt es sich um Selektivverträge, die von Krankenkassen mit ausgewählten Krankenhäusern in bisher vier ausgewählten Versorgungsbereichen⁹ geschlossen werden können. Im Rahmen der Erprobung von Qualitätsverträgen soll untersucht werden, inwieweit sich die stationäre Versorgungsqualität durch Anreize in Verbindung mit höheren Qualitätsanforderungen erreichen lässt. Die Wahl und die Bemessung der Anreize obliegen den vertragsschließenden Krankenkassen und Krankenhäusern. Qualitätsmessungen werden hier herangezogen, um das Erreichen der vereinbarten Qualitätsanforderungen zu überprüfen.

⁹ Der G-BA wählte 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.

Planung der Versorgungsstrukturen

Auch für die Planung von Versorgungsstrukturen im stationären Bereich können Qualitätsmessungen verwendet werden. Das Krankenhausfinanzierungsgesetz (KHG, Gesetz zur wirtschaftlichen Sicherung der Krankenhäuser und zur Regelung der Krankenhauspflegesätze) ermöglicht den Planungsbehörden der Bundesländer, Versorgungsqualität in ihrer Krankenhausplanung zu berücksichtigen. Zu diesem Zweck müssen sich die dafür herangezogenen Qualitätsindikatoren auf planungsrelevante Versorgungsaspekte beziehen und beispielsweise in der Lage sein, „in einem erheblichen Maß unzureichende Qualität“ (§ 8 KHG) als Grundlage für Interventionen der Planungsbehörden abzubilden.

Qualitätszu- und -abschläge

Qualitätsmessungen können auch zur Steuerung der Versorgungsqualität verwendet werden, indem sie bei der Vergütung in Form von qualitätsbezogenen Zu- oder Abschlägen berücksichtigt werden (§ 136b SGB V). Die Ausgestaltung der Zu- und Abschläge obliegt auch hier den Vertragspartnern. In § 8 Abs. 1b KHG wird darüber hinaus eine Verbindung zwischen qualitätsbezogenen Abschlägen und Krankenhausplanung insofern gezogen, als Krankenhäuser, für die in drei aufeinanderfolgenden Jahren Qualitätsabschläge erhoben wurden, ganz oder teilweise aus dem Krankenhausplan herauszunehmen sind.

Ausblick und Herausforderungen

Die Erweiterung der Verwendungszwecke von Qualitätsmessungen der externen Qualitätssicherung rückt zunehmend die Frage in den Mittelpunkt, wann welche Handlungsanschlüsse indiziert sind: Wann ist Versorgungsqualität planungsrelevant, wann abschlagspflichtig, wann zuschlagswürdig, wann förderungswürdig? Während bislang die Frage im Vordergrund stand, welche Versorgungsbereiche sich für die Entwicklung eines QS-Verfahrens zur Darstellung und Förderung der Versorgungsqualität eignen, stellt sich künftig zusätzlich die Frage, welche Steuerungsinstrumente für welche Qualitätsdefizite am geeignetsten erscheinen.

Da die Vorstellung dessen, was in einem bestimmten Themenbereich (z. B. Behandlung von Hüftgelenkerkrankungen) Versorgungsqualität darstellt, unabhängig von dem Verwendungszweck der Qualitätsmessung ist, sollten die verschiedenen Verwendungszwecke von Qualitätsergebnissen in einem Themenbereich als methodische Einheit betrachtet werden. Den methodischen Rahmen dafür stellt ein Qualitätsmodell für diesen Themenbereich dar (siehe Abschnitt 5.2), innerhalb dessen die Qualitätsmessungen mit ihren gegebenenfalls unterschiedlichen Handlungsanschlüssen aufeinander abgestimmt werden können.

Auch wenn die Vorstellung von Versorgungsqualität in einem Themenbereich unabhängig von der Verwendung der Qualitätsergebnisse ist, unterscheidet sich doch die Höhe der Qualitätsanforderungen zwischen ihnen. Die bisherige Qualitätsförderung differenziert bislang lediglich zwischen qualitativ unauffälligen und qualitativ auffälligen Ergebnissen, also zwischen guter und unzureichender Qualität. Das Krankenhausstrukturgesetz (KHSG, Gesetz zur Reform der Strukturen der Krankenhausversorgung) nennt dagegen beispielsweise im Rahmen von qualitätsbezogenen Zu- und Abschlägen außerordentlich gute und am anderen Ende der Skala unzureichende Qualität und impliziert damit eine dazwischenliegende Stufe der Qualität. Auch im

Rahmen der Krankenhausplanung mittels Qualitätsindikatoren findet sich eine Differenzierung der Versorgungsqualität: Minimalanforderungen bei Planungsentscheidungen, Ausschlusskriterien bei Interventionsentscheidungen und besonders gute Qualität bei Vergabeentscheidungen. Damit wird deutlich, dass zukünftig weitere, über die Unterscheidung von guter und unzureichender Qualität hinausgehende Differenzierungen notwendig werden. 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, hervorragende Qualität sichtbar zu machen und zu fördern. Damit geht die gesetzliche Qualitätssicherung über die Sicherung der Einhaltung von Mindestanforderungen hinaus.

5 Grundlegende Methodik der Qualitätsmessung

Um Versorgungsqualität zu verbessern und die Ziele der Gesundheitsversorgung zu erreichen, ist die Messung und Bewertung der Qualität der erste und grundlegende Schritt (Lohr 1990: 45 f., Donabedian 2003: xxvi, Berwick et al. 2003). Eine aussagekräftige Qualitätsmessung bildet die Grundlage für alle weiteren Verwendungszwecke.

5.1 Methoden zur Messung von Qualität

Die externe Qualitätssicherung hat zum Ziel, behandlungsfallübergreifende und vergleichende Aussagen über die Versorgungsqualität auf Ebene der Leistungserbringer und wenn möglich ergänzend auf Systemebene zu treffen (siehe Abschnitt 2.3). Dabei wird die Annahme geprüft, dass die Versorgung ein bestimmtes Qualitätsniveau erreicht. Zur Prüfung solcher Hypothesen eignen sich in erster Linie quantitative Analysemethoden (Döring und Bortz 2016: 184 f., Maats 2007: 293 f.), da sie eine hohe Objektivität und eine hohe Reliabilität der Messung ermöglichen. Dies ist insbesondere für steuernde Eingriffe im Rahmen der Krankenhausplanung oder für Vergütung auf Basis von Qualitätsmessung sowie für eine öffentliche Berichterstattung (Public Reporting) wichtig. Für die quantitative Analyse von Qualität müssen Merkmale der Versorgung messbar gemacht und mittels quantitativer Größen dargestellt werden (Operationalisierung; Meyer 2004: 28 und 51, Döring und Bortz 2016: 223 und 228, Schnell et al. 2013: 117 und 121) sowie mit den an sie gestellten Anforderungen verglichen werden. Diese qualitätsbezogenen Merkmale der Versorgung werden als Qualitätsmerkmale und die quantitativen Größen zu ihrer Messung und Darstellung werden als Qualitätsindikatoren bezeichnet.

Qualitätsindikatoren beschreiben den Grad der Erfüllung konkreter Qualitätsanforderungen. Bei Prozess- oder Ergebnisindikatoren (siehe unten) können durch die über die Einzelbeobachtung hinausgehende Betrachtungsweise Fehlschlüsse durch Zufallseinflüsse vermieden und Qualitätsprobleme dargestellt werden, die in Analysen von Einzelbeobachtungen u. U. nicht sichtbar werden (z. B. eine erhöhte Sterblichkeit, die bei Betrachtung weniger, plausibler Einzelfälle nicht auffällt). Das Bewertungskonzept eines Indikators (Kapitel 12) erlaubt dabei eine, meist statistische, Aussage über die Erfüllung der Qualitätsanforderungen (Soll-Ist-Abgleich), beispielsweise durch Vergleich des Indikatorwerts mit dem Referenzbereich. Während das Qualitätsziel eines Indikators die gewünschte Richtung der Indikatorergebnisse vorgibt, z. B. eine möglichst niedrige Komplikationsrate, definiert der Referenzbereich die konkret zu erfüllende Anforderung, z. B. dass eine Komplikationsrate unter 5 % erwartbarer Qualität entspricht. Ein Qualitätsindikator muss somit nach Definition des IQTIG die folgenden drei Komponenten umfassen:

- Qualitätsziel
- Messverfahren
- Bewertungskonzept

Messgrößen ohne Bewertungskonzept (d. h. ohne Referenzbereich und Methode zur Feststellung, ob dieser erreicht wurde) werden vom IQTIG nicht als Qualitätsindikatoren bezeichnet, da sie keine vergleichende Qualitätsbewertung erlauben.

Für die Beurteilung von Qualität mittels Qualitätsindikatoren werden dementsprechend Erhebungsmethoden mit möglichst hohem Grad an Standardisierung benötigt, die durch eine hohe Objektivität und Reliabilität die Grundlage für eine standardisierte, quantitative Bewertung der Ergebnisse legen. Darüber hinaus müssen die Erhebungsmethoden so ausgewählt werden, dass sie trotz der großen Zahl an Leistungserbringern und Behandlungsfällen praktikabel sind, also der Aufwand je Fall vertretbar ist. Dazu eignen sich in erster Linie standardisierte, schriftliche bzw. elektronische Befragungen, wie die fallbezogene und die einrichtungsbezogene QS-Dokumentation und die Patientenbefragung, sowie sekundäre Datenanalysen wie die Analyse von Sozialdaten, die die Krankenkassen erheben und bei denen es sich ebenfalls um in hohem Maße standardisierte Daten handelt.

Für manche Fragestellungen sind allerdings qualitative Analysemethoden geeigneter als quantitative Methoden. Sie sind meist dann angezeigt, wenn Sachverhalte erklärt oder Hypothesen generiert werden sollen (Döring und Bortz 2016: 184 f., Maats 2007: 293 f.). Qualitative Methoden können auch dann sinnvoll sein, wenn quantitative Bewertungen wegen einer geringen Zahl von Untersuchungseinheiten (z. B. weniger Behandlungsfälle oder sehr seltener Ereignisse) mit einer hohen statistischen Unsicherheit eingehen würden. In der medizinischen Qualitätssicherung gehen qualitative Untersuchungsmethoden oft von Detailanalysen einzelner Behandlungsfälle aus (kasuistische Analyse), beispielsweise mittels Dokumentenanalysen (Reviews von Patientenakten), Gruppeninterviews (Dialogverfahren) oder Kombinationen davon (z. B. bei Peer-Review-Verfahren). Ein Vorteil dieser Methoden ist auf der einen Seite, dass die besonderen Umstände des Einzelfalls berücksichtigt werden können, da potenziell alle Informationen aus der Behandlungsdokumentation für die Beurteilung zur Verfügung stehen. Da qualitative Analysemethoden mittels Induktion von den Einzelfällen auf zugrunde liegende Versorgungsprobleme verallgemeinern und rückschließen müssen (Maats 2007: 293), ergibt sich auf der anderen Seite der Nachteil, dass im Vergleich zu quantitativen Analysemethoden die Bewertungsergebnisse stärker von den impliziten, subjektiven Beurteilungskriterien der Untersuchenden abhängig sind (Donabedian 2003: 64 ff.) und dass sie durch anekdotische Evidenz, d. h. unsystematische Einflüsse, verzerrt werden können (Berwick 1991: 1218 f.). In der externen Qualitätssicherung werden daher qualitative Methoden meist eingesetzt, wenn die Ursachen für ein bestimmtes Indikatorergebnis oder wenn mögliche Maßnahmen zur Qualitätsverbesserung ermittelt werden sollen.

Nach Donabedian (2003: 47) kann unterschieden werden, ob Qualitätsbeurteilungen anhand von Angaben zu Strukturen, Prozessen oder Ergebnissen erfolgen. Die entsprechenden Qualitätsanforderungen werden oft auch den Begriffen Struktur-, Prozess- und Ergebnisqualität zugeordnet und letztere als Qualitätsdimensionen bezeichnet.¹⁰ Für die Verwendung von Indikatoren zur Qualitätsdarstellung ist allerdings weniger die formale Zuordnung eines Indikators zu Struktur-, Prozess- oder Ergebnisqualität von Bedeutung als vielmehr Fragen wie z. B. nach der

¹⁰ Donabedian stellt demgegenüber klar, dass Struktur, Prozess und Ergebnis keine Eigenschaften von Qualität sind: „[...] are not attributes of quality. They are only kinds of information one can obtain [...].“ (Donabedian 2003: 47)

Beeinflussbarkeit der Ergebnisse oder nach Belegen für die Patientenrelevanz von Qualitätszielen, die im Einzelfall geprüft werden müssen.¹¹

5.2 Qualitätsmodelle als Ausgangspunkt für die Verfahrensentwicklung

Qualität ist multidimensional und keine noch so umfassende Erhebung kann sämtliche Facetten der Qualität der Gesundheitsversorgung gleichzeitig abbilden. Eine Darstellung von Versorgungsqualität muss sich immer auf eine begrenzte Anzahl von Anforderungen an die Versorgung beschränken. Es muss eine Teilmenge von Anforderungen zielgerichtet und begründet ausgewählt werden, anhand derer die Qualität in dem interessierenden Versorgungsbereich (z. B. „Angemessene Behandlung von Hüftfrakturen“) angemessen abgebildet werden soll, z. B. indem zu jeder der grundlegenden Qualitätsdimensionen (siehe Abschnitt 1.1) entsprechende konkrete Anforderungen formuliert werden. Bei der Entwicklung von QS-Verfahren besteht der erste Schritt daher darin festzulegen, für welche Themen innerhalb eines Versorgungsbereichs Qualität dargestellt, d. h. in der Regel mittels Qualitätsindikatoren gemessen und bewertet werden soll (McGlynn und Asch 1998, Evans et al. 2009, Donabedian 2003: 29 ff.). Diese Themen werden in der Methodik des IQTIG in der Konzeptstudie für ein neues QS-Verfahren in Form von Qualitätsaspekten identifiziert und in einem sogenannten Qualitätsmodell für den spezifischen Versorgungsbereich zusammengefasst (siehe Abschnitt 6.1).

5.2.1 Qualitätsaspekte der Gesundheitsversorgung

Am Beginn der Entwicklung eines QS-Verfahrens steht die Beschreibung der Gesundheitsversorgung in einem ausgewählten Bereich (z. B. Herzkatheteruntersuchungen und PCI) im Hinblick auf die Qualität. Die Identifikation von patientenrelevanten Qualitätsaspekten stellt dabei den Ausgangspunkt dar. Qualitätsaspekte sind konkrete Themen der Versorgungspraxis, die einer oder mehrerer der allgemeinen Qualitätsdimensionen des Rahmenkonzepts (siehe Abschnitt 1.1) zugeordnet werden können und anhand derer sich die Qualität der Gesundheitsversorgung im ausgewählten Bereich beschreiben lässt.¹² Im oben eingeführten Beispiel könnten dies z. B. Indikationsstellung zur Herzkatheteruntersuchung, Information und Aufklärung vor elektiver Prozedur oder die Veränderung der Symptombelastung nach PCI sein.

Die Qualitätsaspekte werden auf Basis verschiedener Informationsquellen entwickelt. Bei der Themenauswahl für das spätere QS-Verfahren werden die Perspektive der Patientinnen und Patienten, die der Gesundheitsprofessionen und Erkenntnisse aus der wissenschaftlichen Literatur angemessen berücksichtigt (siehe Abschnitt 6.1). Die Hinweise aus diesen Informationsquellen werden mittels einer qualitativen Synthese zu voneinander abgrenzbaren Qualitätsaspekten zusammengefasst. Neben diesem induktiven Vorgehen zur Ermittlung von Qualitätsaspekten er-

¹¹ In diesem Zusammenhang wird auch oft die Indikationsqualität als eigene Qualitätsdimension diskutiert. Das IQTIG vertritt dabei die Auffassung, dass die Indikationsstellung durch die behandelnden Ärztinnen oder Ärzte gestaltbar ist, also einen Prozess darstellt. Sie thematisiert aber in besonderem Maße die Qualitätsdimension Angemessenheit (siehe Abschnitt 1.1).

¹² Der Begriff „Qualitätsaspekt“ bezeichnet also ein Thema der Versorgungspraxis, für das Qualität beschrieben wird. Es sind nicht die „grundlegenden Qualitätsaspekte“ nach der DIN EN 15224 (2017: 78 f.) gemeint.

folgt ein Abgleich mit den Qualitätsdimensionen des Rahmenkonzepts. Dabei wird geprüft, welche Qualitätsdimensionen durch die Qualitätsaspekte adressiert werden und ob es begründet ist, wenn eine oder gegebenenfalls mehrere Qualitätsdimensionen nicht adressiert werden. Wird eine umfassende Qualitätsbewertung für einen bestimmten Versorgungsbereich angestrebt, setzt dies voraus, dass alle relevanten Qualitätsdimensionen adressiert werden.

5.2.2 Auswahl der Qualitätsaspekte für ein Qualitätsmodell

Die für einen Versorgungsbereich identifizierten und anschließend nach bestimmten Kriterien ausgewählten Qualitätsaspekte werden in ihrer Gesamtheit als Qualitätsmodell bezeichnet. Folgende Kriterien werden für die Auswahl herangezogen (zu Details siehe Abschnitt 6.1.3):

- Es gibt Hinweise auf ein Verbesserungspotenzial der Versorgungsqualität.
- Es ist plausibel, dass sich die Versorgungsqualität durch Anstrengungen der Leistungserbringer positiv beeinflussen lässt.
- Die Versorgungsqualität lässt sich mit den Instrumenten der gesetzlichen Qualitätssicherung erfassen.
- Der Qualitätsaspekt liegt im Regelungsbereich des G-BA.

Das Qualitätsmodell zu einem Versorgungsbereich wird nicht immer Qualitätsaspekte aus allen Qualitätsdimensionen des Rahmenkonzepts enthalten, etwa weil eine bestimmte Qualitätsdimension bei diesem Thema mit den Mitteln der gesetzlichen Qualitätssicherung nicht abbildbar ist oder für das Thema bei dieser Dimension kein nennenswertes Verbesserungspotenzial gesehen wird. Durch Zuordnung der Qualitätsaspekte eines Qualitätsmodells zu den Qualitätsdimensionen des Rahmenkonzepts wird daher eine erste Einschätzung vorgenommen, ob auf Grundlage dieses Qualitätsmodells eine umfassende Qualitätsdarstellung gemäß dem Rahmenkonzept möglich ist.

Die in dem Qualitätsmodell zusammengefassten Qualitätsaspekte (QA 1 bis QA 3 in Abbildung 2) geben vor, für welche Themen in dem Versorgungsbereich in den nächsten Schritten der Verfahrensentwicklung Qualitätsindikatoren entwickelt werden sollen. Qualitätsaspekte, die die oben genannten Kriterien nicht erfüllen, werden nicht in das Qualitätsmodell aufgenommen (QA 4 und QA 5 in Abbildung 2). Auch wenn sich später für einen Aspekt kein geeigneter Indikator finden lässt, bleibt der Aspekt Teil des Qualitätsmodells (QA 3 und QI 3 in Abbildung 2). Dadurch wird sichtbar, welche Aspekte durch die Qualitätsindikatoren abgebildet werden und bei welchen Aspekten eine Darstellung der Versorgungsqualität derzeit nicht möglich ist. Dies ist für das Gesamtverständnis der Qualitätsergebnisse, ihrer Aussagekraft und ihrer Limitationen von Bedeutung.

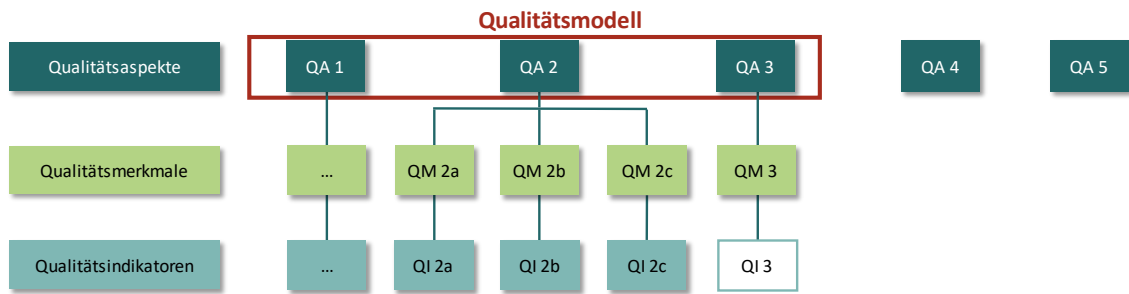


Abbildung 2: Qualitätsaspekte, Qualitätsmodell, Qualitätsmerkmale und Qualitätsindikatoren

Ein Qualitätsmodell ergibt ein Bild der Versorgung, das auf die mit den Instrumenten der Qualitätssicherung erfassbaren Qualitätsverbesserungspotenziale des jeweiligen Versorgungsbereichs fokussiert ist. Anders als das in Abschnitt 1.1 dargestellte allgemeine Rahmenkonzept für Qualität in der Gesundheitsversorgung, das zwischen verschiedenen grundlegenden Dimensionen der Versorgungsqualität unterscheidet, adressiert ein Qualitätsmodell in der IQTIG-Methodik spezifische Themen eines bestimmten Versorgungsbereichs. Das Qualitätsmodell ist damit eine wichtige Voraussetzung für eine zielgerichtete Entwicklung von Qualitätsindikatoren für einen Versorgungsbereich.

5.2.3 Qualitätsmerkmale als Zwischenschritt der Konkretisierung

Als Zwischenschritt auf dem Weg der Indikatorenentwicklung werden die in das Qualitätsmodell aufgenommenen Qualitätsaspekte thematisch differenziert in Qualitätsmerkmale (siehe Abbildung 2). Die Qualitätsmerkmale sind Konkretisierungen der Qualitätsaspekte, die mit Anforderungen verknüpft sind, passend zur Definition der DIN EN ISO 9000:2015, die unter einem Qualitätsmerkmal ein „inhärentes Merkmal eines Objekts, das sich auf eine Anforderung bezieht“ versteht, wobei „Objekt“ z. B. ein Produkt, ein Prozess oder eine Dienstleistung sein kann (DIN EN ISO 9000 2015: 53). So ließe sich im oben genannten Beispiel der Qualitätsaspekt „Information und Aufklärung vor elektiver Prozedur“ weiter thematisch differenzieren u. a. in die Merkmale „Information über alternative Behandlungsmöglichkeiten“ und „Information über Folgen der Behandlung“. Dies ist in Abbildung 2 anhand von Qualitätsaspekt 2 (QA 2) dargestellt, der durch drei Qualitätsmerkmale (QM 2a – QM 2c) repräsentiert ist. Diese Qualitätsmerkmale sind im Gegensatz zu den Qualitätsaspekten direkt mit Anforderungen an die Gesundheitsversorgung verbunden. Beispielsweise wird mit dem Merkmal „Information über alternative Behandlungsmöglichkeiten“ die Anforderung verbunden, dass alle betreffenden Patientinnen und Patienten des betrachteten Leistungserbringers vor einer PCI über alternative Behandlungsmöglichkeiten informiert sein sollten. Erst mit den Qualitätsmerkmalen werden konkrete Anforderungen an die medizinische Versorgung für einen spezifischen Aspekt gestellt, anhand deren Erfüllung die Versorgungsqualität bewertet werden kann. Diese Anforderungen werden als Qualitätsziele bezeichnet.

Auf Basis der Qualitätsmerkmale werden dann in einem weiteren Schritt Qualitätsindikatoren entwickelt, die das Merkmal und die Erfüllung der Anforderung messbar und bewertbar machen.

Im genannten Beispiel könnte das Merkmal über den Qualitätsindikator „Anteil von Patientinnen und Patienten, die über alternative Behandlungsmöglichkeiten informiert wurden“ gemessen und mit einem Referenzbereich (z. B. Anteil > 95 %¹³) als konkret zu erfüllende Anforderung bewertet werden. Tabelle 1 verdeutlicht die Unterschiede zwischen den verschiedenen methodischen Elementen der Qualitätsmessung anhand von Beispielen.

Tabelle 1: Fiktive Beispiele für die methodischen Elemente der Qualitätsmessung

Element	Beispiel 1	Beispiel 2
Qualitätsaspekt	Information und Aufklärung vor elektiver Prozedur (PCI) (Qualitätsdimensionen: Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten; Patientensicherheit)	Behandlungserfolg bei Patientinnen und Patienten mit Pneumonie (Qualitätsdimension: Wirksamkeit)
Qualitätsmerkmal	Information über alternative Behandlungsmöglichkeiten	Sterblichkeit im Krankenhaus bei Pneumonie
Qualitätsanforderung (Qualitätsziel)	Information aller Patientinnen und Patienten über alternative Behandlungsmöglichkeiten	möglichst geringe Sterblichkeit von Patientinnen und Patienten mit Pneumonie
Qualitätsindikator	Anteil von Patientinnen und Patienten, die über alternative Behandlungsmöglichkeiten informiert wurden	Anteil von Patientinnen und Patienten mit Pneumonie, die im Krankenhaus verstorben sind
Referenzbereich	Anteil von Patientinnen und Patienten, die über alternative Behandlungsmöglichkeiten informiert wurden, ist 95 %	Anteil von Patientinnen und Patienten mit Pneumonie, die im Krankenhaus verstorben sind, ist ≤ 10 %

Zu jedem Qualitätsaspekt ist in der Tabelle jeweils nur ein Qualitätsmerkmal als Beispiel genannt. Ein Qualitätsaspekt kann aber auch mehrere Qualitätsmerkmale umfassen.

5.2.4 Zusammenfassung

Das Qualitätsmodell ist ein zentrales Element in der Methodik des IQTIG zur Entwicklung von QS-Verfahren. Die umfassende Synthese der verfügbaren Informationen zur Qualität in einem Versorgungsbereich in Form von Qualitätsaspekten und deren kriteriengeleitete Zusammenfassung zu einem Qualitätsmodell bildet die Grundlage für die zielgerichtete Entwicklung von Qualitätsindikatoren. Daher ist die Entwicklung eines Qualitätsmodells immer verbindlich der Neuentwicklung eines QS-Verfahrens vorgeschaltet. Durch den späteren Abgleich zwischen den Qualitätsdimensionen des Rahmenkonzepts, den Qualitätsaspekten eines Qualitätsmodells und

¹³ Diese Prozentangabe ist ein fiktiver Wert und dient nur der Erläuterung des Beispiels.

dem entwickelten Indikatorenset ergibt sich die Möglichkeit einer Eignungsbeurteilung des Indikatorensets hinsichtlich der Abdeckung der aus Perspektive der Qualitätssicherung als relevant erachteten Qualitätsaspekte.

Mit Hinblick auf die verschiedenen Handlungsanschlüsse, für die die Qualitätsmessungen herangezogen werden können (siehe Kapitel 4), stellt ein für einen Versorgungsbereich einheitliches Qualitätsmodell den methodischen Rahmen dar: Auf Ebene des Qualitätsmodells werden nach den in Abschnitt 5.2.2 genannten Kriterien die grundsätzlich für die Qualitätssicherung relevanten Qualitätsaspekte zusammengefasst, unabhängig von möglichen Verwendungszwecken der Qualitätsergebnisse. Die Entwicklung eines QS-Verfahrens mit einem bestimmten Handlungsanschluss (z. B. qualitätsbezogene Zu- und Abschläge) erfolgt dann auf Ebene der Qualitätsindikatoren, die den spezifischen Anforderungen des Verwendungszwecks der Qualitätsmessungen genügen müssen. Dies kann prinzipiell dazu führen, dass für ein QS-Verfahren aus einem Qualitätsaspekt keine Qualitätsindikatoren abgeleitet werden, während für ein anderes QS-Verfahren im gleichen Versorgungsbereich aus diesem Qualitätsaspekt Qualitätsindikatoren abgeleitet werden. Durch den gemeinsamen Rahmen eines einheitlichen Qualitätsmodells für diesen Versorgungsbereich werden mögliche, durch unterschiedliche Anforderungen an die Indikatoren begründete Unterschiede in der Abdeckung der Qualitätsaspekte transparent.

Teil B: Entwicklung und Durchführung von Qualitätssicherungsverfahren

6 Neuentwicklung von Qualitätssicherungsverfahren

Die Entwicklung von Verfahren für die Qualitätssicherung nach §§ 136 ff. SGB V zur Messung und Bewertung der Versorgungsqualität mittels Qualitätsindikatoren ist eine Kernaufgabe des IQTIG. In diesem Kapitel wird die Methodik der Entwicklung solcher QS-Verfahren dargestellt. Die Methoden der diesen Entwicklungen zugrunde liegenden Elemente (Methoden der Stichprobenziehung, der Informationsrecherche, Auswertungsmethoden etc.) werden in Teil C beschrieben. Die folgenden Ausführungen (Abschnitt 6.1 bis 6.4) beziehen sich ausschließlich auf Verfahren zum Zweck der Qualitätsförderung durch externen Vergleich.

Alle Verfahrensentwicklungen richten sich an den Zielen der Qualitätsmessungen aus, die durch das Rahmenkonzept für Qualität (siehe Abschnitt 1.1) und die übergreifende Anforderung der Patientenzentrierung (siehe Kapitel 3) vorgegeben sind, und berücksichtigen dabei die Verwendungszwecke und Eignungskriterien für Qualitätsmessungen (siehe Kapitel 4 und 10). Die zu entwickelnden QS-Verfahren sollen also dort Maßnahmen anstoßen, wo Verbesserungspotenziale der patientenrelevanten Versorgungsqualität bestehen. Dazu müssen die Qualitätsziele und Verbesserungspotenziale bei der Entwicklung konkretisiert und präzise formuliert werden. Darüber hinaus orientiert sich die Entwicklung an folgenden Grundsätzen, die die spätere Umsetzung der QS-Verfahren betreffen:

- Die Verfahren sollen einrichtungsbezogene Auswertungen und Einrichtungsvergleiche ermöglichen.
- Die Verfahren sollen möglichst datensparsam sein.
- Die Verfahren sollen möglichst aufwandsarm sein.
- Bei Nutzung mehrerer Erfassungsinstrumente sollen diese aufeinander abgestimmt sein.
- Die Verfahren sollen möglichst alle Sektoren, Fachdisziplinen und Professionen einbeziehen, die für die Darstellung der Qualitätsthemen relevant sind.
- Die Logistik der Datenerhebung und der Datenschutz entsprechen den Regelungen nach § 299 SGB V.

Die Neuentwicklung von Qualitätsindikatoren für die sektorenübergreifende Qualitätssicherung nach §§ 136 ff. SGB V orientiert sich am Eckpunktepapier des Gemeinsamen Bundesausschusses. Der G-BA fordert hierin eine „präzise Formulierung der Ziele bzw. Qualitätsverbesserungsanliegen und Aufgaben des QS-Verfahrens vor der Entwicklung des Indikatorensets (vermutetes oder bekanntes Qualitätsdefizit), Identifikation von und Konzentration auf versorgungsrelevante Fragestellungen“ (G-BA [kein Datum]: 3 f.).

Für die Qualitätsmessung stehen bei Neuentwicklungen von QS-Verfahren derzeit vier Erhebungsinstrumente für drei verschiedene Datenquellen zur Verfügung:

▪ **Dokumentation der Leistungserbringer**

Bislang stützen sich QS-Verfahren vor allem auf die Dokumentation von Behandlungsfällen zu Zwecken der Qualitätssicherung durch die Leistungserbringer (fallbezogene QS-Dokumentation). Darüber hinaus können durch die Leistungserbringer auch Angaben auf Einrichtungsebene dokumentiert werden (einrichtungsbezogene QS-Dokumentation), z. B. Strukturangaben und interne Regelungen.

▪ **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 Versicherten der gesetzlichen Krankenversicherung (GKV-Versicherte). Sie stehen aufgrund der Struktur des Datenflusses erst mit Verzögerungen von bis zu 9 Monaten für die QS-Verfahren nach §§ 136 ff. SGB V zur Verfügung.

▪ **Befragung von Patientinnen und Patienten**

Nach §§ 136 ff. und § 299 SGB V sollen auch Patientenbefragungen für die Zwecke der gesetzlich verbindlichen, externen Qualitätssicherung eingesetzt werden.

Diese Datenquellen decken sowohl komplementäre als auch überlappende Informationen über die Versorgungsqualität ab. Die fallbezogene QS-Dokumentation macht die Qualität anhand durchgeführter Prozeduren und der erreichten Ergebnisse abbildbar. Die einrichtungsbezogene QS-Dokumentation dient in erster Linie der Erhebung von Strukturparametern, also ob beispielsweise Verfahrensregelungen zur Hygiene in einer Einrichtung vorliegen. Die Sozialdaten bei den Krankenkassen erlauben es darüber hinaus, den weiteren Behandlungsverlauf einschließlich Diagnosen und Prozeduren in die Qualitätsbewertungen einzubeziehen. Des Weiteren stellen die Patientinnen und Patienten selbst eine weitere Informationsquelle für die Qualität der medizinischen Versorgung dar, die durch Patientenbefragungen zugänglich gemacht werden kann. So können Patientinnen und Patienten beispielsweise Auskunft über Behandlungsabläufe, Behandlungsergebnisse und Rahmenbedingungen der Behandlung geben.

Die Erhebungsinstrumente unterscheiden sich im Hinblick auf den Erhebungsaufwand und an welcher Stelle dieser anfällt. So ist der Dokumentationsaufwand für die Leistungserbringer unter Umständen hoch, wenn in einem Verfahren viele fallbezogene Daten nur für die Qualitätssicherung erfasst werden sollen. Geringer bleibt der Aufwand in der Regel bei einer einrichtungsbezogenen QS-Dokumentation. Für die Nutzung der Sozialdaten bei den Krankenkassen fällt bei den Leistungserbringern dagegen kein zusätzlicher Aufwand an, da diese routinemäßig und unabhängig von der Qualitätssicherung erfasst und an die Krankenkassen weitergeleitet werden. Auch bei Patientenbefragungen fällt bei den Leistungserbringern nur ein geringer Aufwand an, da die Befragungen selbst nicht durch die Leistungserbringer durchgeführt werden. Aufwand

kann den Leistungserbringern jedoch bei der Weiterleitung der Patientenadressen sowie ergänzenden Informationen an die Versendestelle, die die Befragung durchführt, entstehen.

Die Entwicklungsschritte für QS-Verfahren, deren Ergebnisse für die Qualitätsförderung durch externen Vergleich verwendet werden (siehe Kapitel 4), sind in Abbildung 3 schematisch dargestellt.

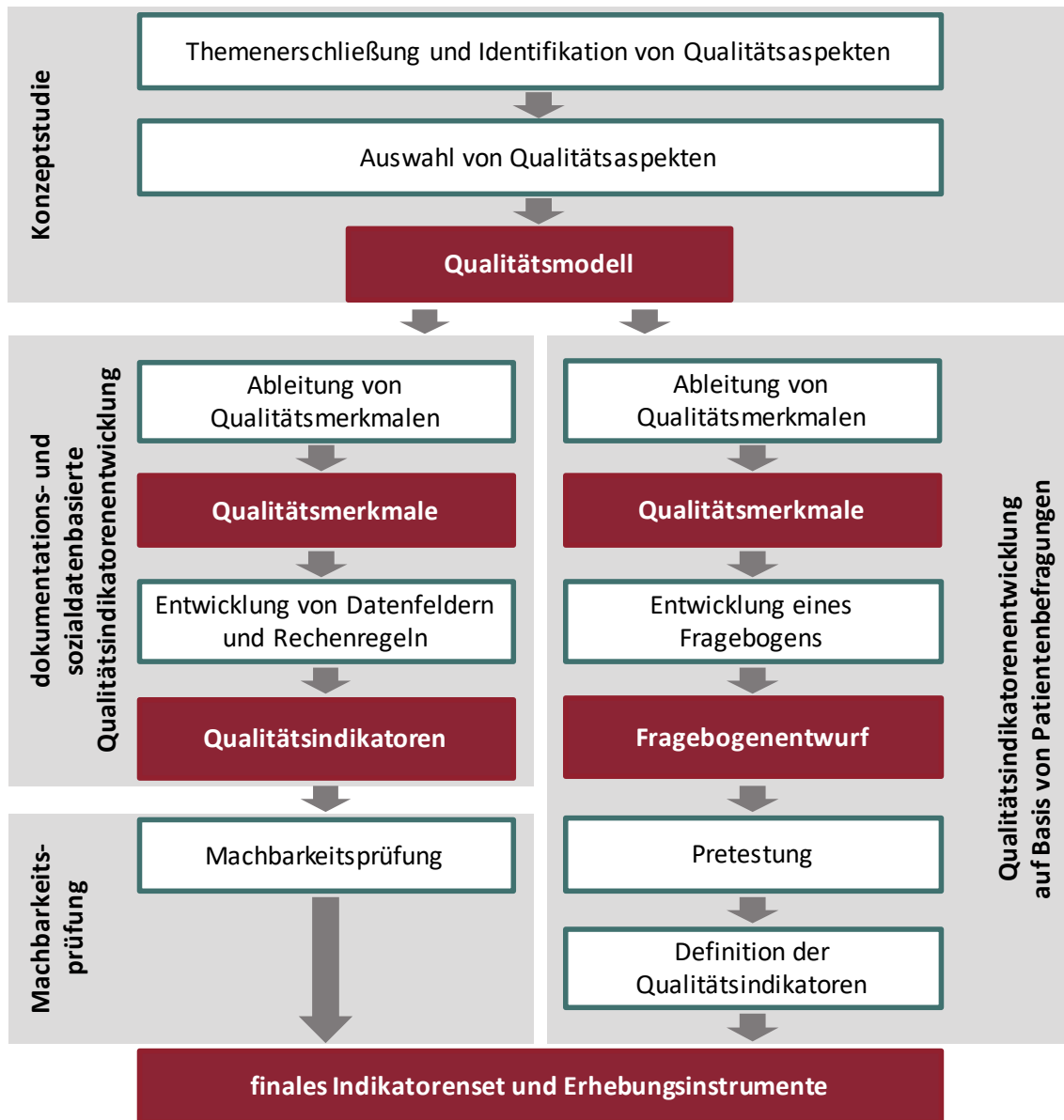


Abbildung 3: Schritte und Produkte der Entwicklung eines QS-Verfahrens

Die Entwicklung eines QS-Verfahrens beginnt mit einer Konzeptstudie, deren zentrales Ergebnis das Qualitätsmodell für diesen Themenbereich ist. Darauf basierend erfolgt die Entwicklung des eigentlichen QS-Verfahrens. Die Entwicklung von Qualitätsindikatoren kann dabei sowohl auf Basis der Leistungserbringerdokumentation sowie von Sozialdaten bei den Krankenkassen als auch auf Basis von Patientenbefragungen erfolgen. Ausgehend vom Qualitätsmodell werden zunächst die Qualitätsmerkmale abgeleitet. Die Entwicklung von Qualitätsindikatoren auf Basis

von Leistungserbringerdokumentationen und Sozialdaten bei den Krankenkassen einerseits und auf Basis von Patientenbefragungen andererseits erfolgt mit jeweils eigenen, den Erfordernissen entsprechenden Methoden. Zwischen beiden Entwicklungssträngen findet eine kontinuierliche Abstimmung statt, da die jeweiligen Indikatoren im Ergebnis ein gemeinsames, in sich konsistentes Indikatorenset bilden sollen.

Im Anschluss an die Entwicklung von Qualitätsindikatoren auf Basis der Leistungserbringerdokumentation und von Sozialdaten bei den Krankenkassen erfolgt eine Machbarkeitsprüfung, die die Erhebungsinstrumente und die Qualitätsindikatoren in einer probenhaften Anwendung prüft. Bei der Entwicklung einer Patientenbefragung ist die umfassende Pretestung fester Bestandteil der Fragebogenentwicklung. Die Erprobung der Erhebungsinstrumente kann bis in den Regelbetrieb des QS-Verfahrens (siehe Abschnitt 7.1) hineinreichen, sofern in der Richtlinie für das QS-Verfahren eine Erprobungsphase vorgesehen wird.

Am Ende aller in den folgenden Abschnitten beschriebenen Entwicklungen stehen Berichte mit Empfehlungen an den Auftraggeber, die den Entwicklungsprozess transparent machen. Im Rahmen der Entwicklung von QS-Verfahren wird ein Stellungnahmeverfahren mit den zu beteiligenden Organisationen nach § 137a Abs. 7 SGB V durchgeführt. Dieses erfolgt bei den dokumentations- und sozialdatenbasierten Indikatoren zu den komplett entwickelten Qualitätsindikatoren. Bei der Entwicklung einer Patientenbefragung dagegen findet das entsprechende Stellungnahmeverfahren nach der Entwicklung der Qualitätsmerkmale und vor der Entwicklung des Fragebogens und der entsprechenden Qualitätsindikatoren statt.

6.1 Konzeptstudie für Qualitätssicherungsverfahren

Die Entwicklung eines neuen QS-Verfahrens beginnt mit einer Konzeptstudie. Diese baut gegebenenfalls auf Informationen auf, die im G-BA-Verfahren zur Themenfindung und Priorisierung ermittelt wurden. Die Konzeptstudie hat die Aufgabe, eine Einschätzung zu folgenden Fragestellungen zu geben, die für die Beauftragung und Entwicklung eines QS-Verfahrens wichtig sind (G-BA [kein Datum]: 3 f.):

- Welche Hinweise auf Qualitätsdefizite gibt es und welche Potenziale der Qualitätsverbesserung bestehen?
- Was sind die Ziele und Qualitätsverbesserungsanliegen eines QS-Verfahrens und was sind die versorgungsrelevanten Fragestellungen?
- Kann Qualität in dem betreffenden Themenbereich gemessen werden und sind die benötigten Daten verfügbar?
- Kann das Thema mit den Instrumenten der gesetzlichen Qualitätssicherung adressiert werden?

Diese Fragen werden in der Konzeptstudie durch folgende Elemente adressiert:

- die Darstellung der aktuellen Versorgungspraxis für das beauftragte Thema einschließlich ihrer aktuellen Verbesserungsbedarfe und -potenziale
- die Entwicklung eines Qualitätsmodells für das beauftragte Thema
- die Identifizierung von geeigneten Datenquellen und Erhebungsinstrumenten

- eine Beurteilung, ob das geplante Verfahren in dieser Form grundsätzlich und angemessen aufwandsarm umsetzbar erscheint

Mittels der Konzeptstudie soll sichergestellt werden, dass ein geplantes QS-Verfahren einerseits ausreichend auf relevante Verbesserungspotenziale fokussiert und andererseits mögliche Umsetzungshürden frühzeitig erkannt werden und eine Entwicklung von unangemessen aufwendigen oder vor dem Hintergrund der Versorgungsrealität nicht durchführbaren Verfahrensansätzen vermieden wird. Die Konzeptstudie schafft damit die Grundlage für eine anschließende Entwicklung von Qualitätsindikatoren.

Die Entwicklung eines Qualitätsmodells gliedert sich in drei Schritte: die Themenerschließung der Versorgungspraxis, die Identifikation von Qualitätsaspekten und die Auswahl von Qualitätsaspekten für das Qualitätsmodell.

6.1.1 Themenerschließung

Zunächst wird eine qualitative und quantitative Analyse der aktuellen Versorgungspraxis anhand aller zur Verfügung stehenden Wissensquellen durchgeführt. Diese basiert auf einer umfassenden Literatur- und Leitlinienrecherche, einer Recherche der aktuellen Rahmenbedingungen (Gesetze, Richtlinien etc.), einer Internetsuche auf Websites relevanter Organisationen, leitfadengestützten Interviews mit Expertinnen und Experten und einer explorativen Sozialdatenanalyse. Dieses Vorgehen gewährleistet den Einbezug sowohl der wissenschaftlichen Perspektive auf das Versorgungsgeschehen als auch der Perspektive von Normen und Regelungen sowie der System- und Patientenperspektive. Der so geschaffene themenbezogene Wissensbestand dient der Beschreibung der Versorgungssituation, der inhaltlichen Vorbereitung der anschließenden Fokusgruppen und darüber hinaus der späteren Ableitung der Qualitätsaspekte.

Im Rahmen der Literaturrecherche wird nach wissenschaftlichen Erkenntnissen über den zu untersuchenden Versorgungsbereich gesucht (siehe Kapitel 8). Sie umfasst eine systematische Recherche von Leitlinien, systematischen Übersichtsarbeiten und Health-Technology-Assessment-Berichten (HTA-Berichten) sowie quantitativen und qualitativen Studien. Ergänzend erfolgt eine Handsuche z. B. in den Referenzlisten ausgewählter Schlüsselartikel. Dabei wird insbesondere nach Aussagen zur aktuellen Versorgungssituation mit Hinweisen auf Über-, Unter- und Fehlversorgung und zu Patientenerfahrungen in der Versorgung gesucht. Die recherchierten Quellen müssen dabei keinem definierten erforderlichen Evidenzgrad erfüllen, da die umfassende Erschließung der Versorgungspraxis im Fokus steht und nicht die Prüfung bestimmter einzelner Fragestellungen.

In Literatur und Internet wird zudem nach größeren regionalen bzw. überregionalen Qualitätsinitiativen und Versorgungsmodellen recherchiert. Die Ergebnisse dieser Projekte werden ergänzend genutzt, um tatsächliche Verbesserungspotenziale der Leistungserbringer abzuschätzen, um zu beurteilen, in welchem Verhältnis ein geplantes QS-Verfahren nach §§ 136 ff. SGB V zu bereits laufenden Initiativen der Qualitätsförderung stehen würde, und um berichtete Umsetzungsmöglichkeiten und -hürden dieser Projekte in die Abschätzung der grundsätzlichen Umsetzbarkeit eines QS-Verfahrens nach §§ 136 ff. SGB V einzubeziehen.

Leitfadengestützte Interviews werden mit Expertinnen und Experten durchgeführt, um Informationen zur Gesundheitsversorgung anhand von Erfahrungen aus der Versorgungstätigkeit oder aus wissenschaftlicher Tätigkeit zu erfassen. Dabei wird nicht nur der Ablauf der Versorgung thematisiert, sondern es werden auch übergreifende Fragestellungen wie beispielsweise Einflüsse von Vergütungsregelungen oder rechtlichen Rahmenbedingungen adressiert. Die Entwicklung der Interviewleitfäden und die Auswertungen der Interviews folgen der in Abschnitt 9.3.1 beschriebenen Methodik.

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. Gegenstand der explorativen Sozialdatenanalyse sind u. a. die Epidemiologie der Erkrankung, Abschätzungen der Zahl der Leistungserbringer und ihre Verteilung auf die einzelnen Sektoren sowie Abschätzungen der jährlichen Fallzahlen pro Leistungserbringer. Des Weiteren gibt die Sozialdatenanalyse wichtige Hinweise zu typischen Behandlungsverläufen, Sektorenübergängen, Verordnungen und Ergebnissen der Versorgung.

Die Versorgungspraxis wird auf Basis dieser Informationen anhand folgender Punkte dargestellt:

- Epidemiologie, Ätiologie und Risikofaktoren der Erkrankung, Komorbiditäten und Krankheitsverläufe
- Behandlungsabläufe, einschließlich Diagnostik, Behandlungs- und Versorgungszielen, Indikation, Therapie, Versorgung mit Heil- und Hilfsmitteln, sowie Häufigkeiten dieser Leistungen
- Versorgungssettings (stationär, ambulant, teilstationär und rehabilitativ), bestehende Versorgungsnetze, integrierte Versorgung, ambulante spezialfachärztliche Versorgung und Disease-Management-Programme
- an der Versorgung beteiligte Berufsgruppen
- Anreizsysteme (z. B. Vergütung)
- Behandlungsergebnisse (z. B. Symptomverlauf, Sterblichkeit, Lebensqualität)

Außerdem wird ein Versorgungspfad erstellt, der Schlüsselstrukturen und -prozesse der Versorgung sowie Schnittstellen und Beteiligte abbildet.

6.1.2 Identifikation von Qualitätsaspekten

Qualitätsaspekte sind konkrete Themen der Versorgungspraxis, anhand derer sich die Qualität der Gesundheitsversorgung beschreiben lässt (siehe Abschnitt 5.2). Ausgangspunkt für ihre Identifikation sind die Erkenntnisse und Informationen, die sich aus der oben beschriebenen Themenerschließung ergeben. Um auch das Erfahrungswissen aus der Versorgungspraxis umfassend zu berücksichtigen, werden außerdem Fokusgruppen mit Gesundheitsprofessionen durchgeführt. In die Fokusgruppen sollen ausschließlich aktive Praktikerinnen und Praktiker einbezogen werden, die am alltäglichen Behandlungsgeschehen beteiligt sind, um auch die Verbesserungsbedarfe und -potenziale in der Versorgung für die Patientinnen und Patienten aus Sicht

der Leistungserbringer zu eruieren. Die Teilnehmerinnen und Teilnehmer werden aus allen beteiligten Sektoren der Versorgung und aus allen relevanten Berufsgruppen rekrutiert.

Zur Ermittlung der Behandlungserfahrungen und des Behandlungserlebens von Patientinnen und Patienten werden Fokusgruppen mit Patientinnen und Patienten durchgeführt. So können qualitätsrelevante Themen oder wichtige Merkmale und Ereignisse einer Behandlung aus Sicht der Patientinnen und Patienten ermittelt werden. Die Durchführung der Fokusgruppen erfolgt nach der in Abschnitt 9.3.2 beschriebenen Methodik.

Die Qualitätsaspekte werden abgeleitet, indem die Informationen, die in der Themenerschließung und über die Fokusgruppen ermittelt wurden, aufgrund inhaltlicher Überlegungen zusammengefasst werden. Die Kategorien für diese Zusammenfassung orientieren sich dabei beispielsweise an Elementen des Versorgungspfads und den Qualitätsdimensionen des Rahmenkonzepts (siehe Abschnitt 1.1). So können z. B. verschiedene Gesichtspunkte der Aufklärung von Patientinnen und Patienten zu einem Qualitätsaspekt „Information und Aufklärung“ zusammengefasst werden. Der Abgleich der identifizierten Qualitätsaspekte mit den Dimensionen des Rahmenkonzepts kann außerdem Anlass zur Prüfung geben, ob die Nicht-Berücksichtigung einer oder mehrerer Qualitätsdimensionen sachgerecht ist.

Bei der Identifikation der Qualitätsaspekte werden die Eignungskriterien „Bedeutung für die Patientinnen und Patienten“ sowie „Zusammenhang mit einem unmittelbar patientenrelevantem Ziel“ berücksichtigt (siehe Abschnitt 10.2). Das bedeutet, es wird ermittelt, anhand welcher Themen der Versorgungspraxis sich die für die Patientinnen und Patienten relevante Qualität beschreiben lässt.

Der Versorgungspfad ermöglicht nicht nur eine strukturierte Beschreibung des gesamten Versorgungsgeschehens, sondern hilft auch, sowohl die identifizierten als auch die später selektierten Qualitätsaspekte zuzuordnen und ihre Verteilung über die Sektoren darzustellen (Großschädl et al. 2014). Auf diese Weise kann aufgezeigt werden, welche Bereiche des jeweiligen Versorgungsgeschehens im Rahmen der gesetzlichen Qualitätssicherung adressierbar sind.

6.1.3 Auswahl von Qualitätsaspekten für das Qualitätsmodell

In diesem Entwicklungsschritt werden aus den patientenrelevanten Qualitätsaspekten diejenigen für das Qualitätsmodell ausgewählt, die durch die QS-Verfahren adressiert werden sollen. Hierfür sind vier Kriterien maßgeblich:

- **Übereinstimmung mit dem Regelungsbereich des G-BA**

Der Qualitätsaspekt muss mit den verfahrensspezifischen Vorgaben der Beauftragung für das geplante QS-Verfahren übereinstimmen. Hierzu zählen die Begrenzung der G-BA-Richtlinien auf den Regelungsbereich des SGB V, d. h. der Ausschluss von Leistungen der Rehabilitation (SGB IX) und der Pflegeversicherung (SGB XI), aber auch andere formale oder inhaltliche Vorgaben, die der G-BA im Rahmen seines Themenfindungs- und Priorisierungsprozesses beschlossen hat. Im Bericht zur Konzeptstudie wird dargestellt, welche möglichen Einschränkungen für die Qualitätsdarlegung aus den Vorgaben der Beauftragung resultieren.

- **Potenzial zur Verbesserung**

Für die Qualitätsaspekte müssen außerdem Hinweise auf einen Verbesserungsbedarf für die Patientinnen und Patienten bzw. ein Verbesserungspotenzial durch die Leistungserbringer vorliegen (siehe Abschnitt 10.2). Die Hinweise auf Verbesserungsbedarfe und -potenziale ergeben sich aus der zusammenfassenden Analyse aller oben genannten Informationsquellen (siehe Abschnitt 6.1.1). Ausgewertet werden hierfür qualitative und quantitative Studien, beispielsweise zum Umsetzungsgrad von Leitlinienempfehlungen und zu Fragen der Über-, Unter- und Fehlversorgung. Die untersuchten Versorgungssettings müssen dabei mit dem deutschen Versorgungssystem vergleichbar sein; so werden beispielsweise bei internationalen Vergleichen nur solche zwischen Industrieländern einbezogen. Auch aus den Ergebnissen großer repräsentativer Erhebungen, z. B. aus Registern, und aus den Ergebnissen größerer einrichtungsübergreifender Qualitätsinitiativen und -projekte können anhand der Verteilung oder des Zeitverlaufs der Ergebnisse Verbesserungspotenziale abgeleitet werden. Verbesserungsbedarfe und -potenziale können auch aus der explorativen Sozialdatenanalyse oder den Analysen der Fokusgruppen mit Vertreterinnen und Vertretern von Gesundheitsprofessionen und mit Patientinnen und Patienten abgeleitet werden.

- **Beeinflussbarkeit durch den Leistungserbringer**

Es muss vor dem Hintergrund der verfügbaren Informationsbasis plausibel erscheinen, dass sich die Versorgungsqualität in diesem Qualitätsaspekt durch Anstrengungen der Leistungserbringer nach SGB V steigern lässt (siehe Abschnitt 10.2.3). Die Frage der Zuordnung einer einrichtungsbezogenen Verantwortlichkeit stellt sich dagegen erst auf der Ebene der Qualitätsmerkmale bzw. Qualitätsindikatoren.

- **Erfassbarkeit mit den Instrumenten der gesetzlichen Qualitätssicherung**

Schlussendlich müssen die Qualitätsaspekte mit einem Instrument der gesetzlichen Qualitätssicherung erfassbar sein. Dazu erfolgt eine Einschätzung, ob die bis dahin ausgewählten Qualitätsaspekte grundsätzlich praktikabel, d. h. grundsätzlich durch die gesetzliche Qualitätssicherung nach §§ 136 ff. SGB V erfassbar sind. Wenn für einen Qualitätsaspekt die Erfassung mittels einer zugänglichen Datenquelle und eines realisierbaren Erhebungsinstruments plausibel ist, werden jedem Qualitätsaspekt Erhebungsinstrumente zugeordnet. Dabei kann einem Aspekt durchaus auch mehr als ein Erhebungsinstrument zugeordnet werden, wenn damit unterschiedliche Teile des Aspekts betrachtet werden sollen. Doppelerhebungen sind im Sinne der Aufwandsangemessenheit jedoch zu vermeiden, bei mehreren gleichwertigen Erhebungsmöglichkeiten ist dem aufwandsärmeren Instrument der Vorzug zu geben. Die verschiedenen Charakteristika der Datenquellen und Erhebungsinstrumente werden bereits in diesem ersten Schritt der Neuentwicklung von QS-Verfahren berücksichtigt und es wird eine erste vorläufige Empfehlung bezüglich der geeigneten Datenquellen und Erhebungsinstrumente für das QS-Verfahren gegeben. Bei Bedarf und je nach Thematik können in der Verfahrensentwicklung aber auch weitere, neue Datenquellen im Rahmen der gesetzlichen Vorgaben berücksichtigt und die hierfür notwendigen Erhebungsinstrumente vorgeschlagen werden (z. B. Daten der klinischen Krebsregister). Für die Beurteilung wird außerdem abgeschätzt – soweit auf Basis öffentlich zugänglicher Dokumente möglich –, in welchem Umfang

Patientinnen und Patienten in Selektivverträgen zum jeweiligen Themengebiet versorgt werden und welche Erhebungsprobleme sich gegebenenfalls hieraus ergeben.

Die anhand der beschriebenen Kriterien ausgewählten Qualitätsaspekte werden zum Qualitätsmodell zusammengefasst. Sie werden im Versorgungspfad besonders gekennzeichnet, um die Eingrenzung der Betrachtung innerhalb des Themenbereichs explizit zu veranschaulichen. Zudem wird eine tabellarische Zuordnung der Qualitätsaspekte des Qualitätsmodells zu den Dimensionen des Rahmenkonzepts (siehe Abschnitt 1.1) vorgenommen (siehe Beispiel Tabelle 2). Dadurch wird eine erste Einschätzung ermöglicht, welche Qualitätsdimensionen durch ein späteres QS-Verfahren adressiert werden.

Tabelle 2: Beispiel für die Zuordnung der Qualitätsaspekte zu den Qualitätsdimensionen des IQTIG-Rahmenkonzepts

Qualitätsaspekt	Wirksamkeit	Patientensicherheit	Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten	Rechtzeitigkeit und Verfügbarkeit	Angemessenheit	Koordination und Kontinuität
Qualitätsaspekt 1	-	x	-	x	x	-
Qualitätsaspekt 2	-	x	x	-	-	-
...	x	x	x	-	x	x

Auf Basis der Ergebnisse einer Konzeptstudie soll eine wissenschaftlich fundierte Entscheidung über die Entwicklung eines QS-Verfahrens nach §§ 136 ff. SGB V getroffen werden. Im Abschlussbericht der Konzeptstudie wird eine Empfehlung ausgesprochen, die Entwicklung des Verfahrens weiterzuverfolgen oder nicht. Wird die Entwicklung von Qualitätsindikatoren für ein QS-Verfahren beauftragt, so wird dafür das in der Konzeptstudie entwickelte Qualitätsmodell verbindlich zugrunde gelegt.

6.2 Dokumentations- und sozialdatenbasierte Neuentwicklung von Qualitätsindikatoren

Die Grundlage für die Neuentwicklung von QS-Verfahren bildet das in der Konzeptstudie (siehe Abschnitt 6.4) formulierte Qualitätsmodell. Auf dieser Basis erfolgen die Entwicklung von Qualitätsindikatoren und die Erarbeitung von Empfehlungen zur Umsetzung eines künftigen QS-Verfahrens. Ziel der Indikatorenentwicklung ist ein auf die wesentlichen Verbesserungsbedarfe und -potenziale fokussiertes, aufwandsarmes und umsetzbares Indikatorenset, welches den Eignungskriterien für Qualitätsindikatoren (siehe Kapitel 10) entspricht. Ausgehend vom Qualitätsmodell (siehe Abschnitt 6.1) erfolgt der Prozess der Neuentwicklung von Qualitätsindikatoren durch das IQTIG in folgenden Entwicklungsschritten (Abbildung 4): Zunächst wird die Konkretisierung der Qualitätsaspekte zu Qualitätsmerkmalen vorgenommen, die anschließend zu Qualitätsindikatoren operationalisiert werden. Die Qualitätsindikatoren bilden nach abschließender Zusammenführung ein Set an Qualitätsindikatoren, das einem Stellungnahmeverfahren mit

den zu beteiligenden Organisationen nach § 137a Abs. 7 SGB V unterzogen wird. Der Ergebnisbericht an den G-BA enthält das empfohlene Indikatorenset.



Abbildung 4: Entwicklungsschritte vom Qualitätsmodell bis zum Qualitätsindikatorenset

Bei der Neuentwicklung von Qualitätsindikatoren auf Basis von fallbezogenen bzw. einrichtungsbezogenen QS-Dokumentationen und Sozialdaten bei den Krankenkassen wird über den gesamten Entwicklungsverlauf (siehe Abschnitte 6.2.1, 6.2.2 und 6.2.3) auch ein beratendes Expertengremium einbezogen. Hierfür sind regelhaft vier Expertentreffen vorgesehen.

Bei der Zusammensetzung des Expertengremiums wird insbesondere darauf geachtet, Personen aus unterschiedlichen Gesundheitsprofessionen einzubeziehen sowie dass in diesem beratenden Gremium nicht nur Personen mit Leitungsposition, sondern auch die mittlere Führungsebene, unter Berücksichtigung ausreichender Berufserfahrung und wissenschaftlicher Expertise, angemessen vertreten sind. Ebenso wird eine Ausgeglichenheit bezüglich der personellen Vertretung der Sektoren angestrebt. Darüber hinaus werden Patientenvertreterinnen und Patientenvertreter regelhaft in dieses Expertengremium einbezogen (siehe Kapitel 3).

Das Expertengremium hat bei allen Entwicklungsschritten eine beratende Funktion für das IQTIG (siehe Abschnitt 9.3.3). Die Empfehlungen des Gremiums werden vom IQTIG nach jedem Treffen reflektiert und berücksichtigt. Das Expertengremium ist somit eine unter verschiedenen Wissensquellen, die das IQTIG bei der Entwicklung von Qualitätsindikatoren nutzt, wobei die finalen Entscheidungen zu jedem Zeitpunkt der Entwicklung in der Verantwortung des IQTIG liegen.

6.2.1 Kriteriengeleitete Konkretisierung der Qualitätsaspekte zu Qualitätsmerkmalen

Für die Qualitätsaspekte des Qualitätsmodells werden im ersten Entwicklungsschritt patientenrelevante Qualitätsmerkmale der Versorgung identifiziert, die wichtige Teile des jeweiligen Aspekts abbilden. Diese Qualitätsmerkmale beziehen sich auf konkrete, begründete Qualitätsanforderungen an die Versorgungspraxis (Struktur, Prozess, Ergebnis), die im Verfahren der gesetzlichen Qualitätssicherung überprüft werden sollen. Dabei orientieren sich die Qualitätsmerkmale am Rahmenkonzept (siehe Abschnitt 1.1) und müssen insbesondere die Eignungskriterien des Qualitätsziels (siehe Abschnitt 10.2) erfüllen. Meist können mehrere patientenrelevante Merkmale je Qualitätsaspekt gefunden werden, wobei jedoch inhaltliche Überlappungen möglichst vermieden werden. So lassen sich beispielsweise für den Qualitätsaspekt „Komplikationen nach dem Eingriff“ eine Reihe möglicher Komplikationen (z. B. Blutung, Infektion) als konkretisierte Merkmale definieren. Jene Merkmale, die ein Eignungskriterium nicht erfüllen, werden nicht in die folgenden Entwicklungsschritte übernommen.

Entwicklung der abgeleiteten Qualitätsmerkmale entsprechend den Eignungskriterien „Bedeutung für die Patientinnen und Patienten“, „Zusammenhang mit einem unmittelbar patientenrelevanten Ziel“ und „Übereinstimmung mit einem Qualitätsaspekt“

Um relevante Abläufe oder Ereignisse der Patientenversorgung zu identifizieren, die für einen Qualitätsaspekt des Qualitätsmodells als Qualitätsmerkmal genutzt werden könnten, wird auf das Ergebnis der im Rahmen der Entwicklung des Qualitätsmodells recherchierten Leitlinien und/oder pflegerischen Expertenstandards zurückgegriffen. Stehen evidenzbasierte Leitlinien (S3, S2e) oder pflegerische Expertenstandards zur Verfügung, so bilden diese den Ausgangspunkt für die Ableitung von struktur- und prozessqualitätsbezogenen Qualitätsmerkmalen (Kötter et al. 2012, Nothacker et al. 2016), und sofern Core Clinical Outcomes in den Leitlinien enthalten sind, auch von ergebnisqualitätsbezogenen Qualitätsmerkmalen. Aus Leitlinien werden im Sinne fachlich begründeter Anforderungen nur Qualitätsmerkmale für jene Qualitätsanforderungen ausgewählt, die einen Empfehlungsgrad A oder B der deutschen AWMF-Leitlinien (Muche-Borowski et al. 2012) bzw. der Nationalen VersorgungsLeitlinien (BÄK et al. 2010) oder vergleichbare Empfehlungsgrade ausländischer Leitlinien aufweisen. Da pflegerische Expertenstandards in der Regel nur eine Gesamtempfehlung aussprechen, werden hieraus alle auf Qualitätsaspekte bezogenen Qualitätsanforderungen als Qualitätsmerkmal in die weitere Entwicklung aufgenommen. Bei Qualitätsmerkmalen aus ausländischen Leitlinien oder pflegerischen Expertenstandards wird zusätzlich geprüft, ob die Empfehlungen auf das deutsche Gesundheitssystem übertragbar sind.

Liegen keine evidenzbasierten Leitlinien oder pflegerische Expertenstandards zu einen Qualitätsaspekt vor, so erfolgt die Ableitung der Qualitätsmerkmale entlang des Versorgungspfads aus der besten verfügbaren Evidenz der eingeschlossenen, aktualisierten Wissensbestände aus der Konzeptstudie (siehe Abschnitt 6.1)

In der weiteren Entwicklung wird festgelegt, bei welchen Patienten- und Leistungserbringergruppen innerhalb der Grundpopulation des QS-Verfahrens das jeweilige Qualitätsmerkmal erhoben werden soll. Dann wird nochmals überprüft, ob das dem Qualitätsaspekt zugeordnete Erhebungsinstrument (siehe Abschnitt 6.1) zur Erfassung des jeweiligen Qualitätsmerkmals geeignet ist. Die Wahl der geeigneten Datenquellen und Erhebungsinstrumente für ein Qualitätsmerkmal orientiert sich daran, inwieweit valide und vollständige Daten zu erwarten sind bzw. ob der Aufwand zur Datenerhebung angemessen erscheint.

Prüfung der abgeleiteten Qualitätsmerkmale entsprechend den Eignungskriterien „Potenzial zur Verbesserung“, „Beeinflussbarkeit durch den Leistungserbringer“ „Zuschreibbarkeit der Verantwortung“ und „Berücksichtigung unerwünschter Wirkungen“

Die Prüfung der Qualitätsmerkmale anhand dieser Eignungskriterien durch das IQTIG erfolgt auf Basis der folgenden Wissensquellen:

- Ergebnisse der Literaturrecherche (aktualisiert)
- Ergebnisse der Sozialdatenanalyse (aktualisiert)
- Einschätzung des Expertengremiums

Hierfür werden die Literaturrecherchen und Sozialdatenanalysen aus der Konzeptstudie (siehe Abschnitt 6.1.1) aktualisiert und merkmalsbezogen ergänzt. Im Fokus der Ergänzung stehen die aktuelle Prävalenz und Inzidenz des Merkmals, die Risikofaktoren und der Verbesserungsbedarf aus Patientenperspektive sowie das Verbesserungspotenzial aus Leistungserbringerperspektive (siehe Abschnitt 6.1.3) ein. Auch konkrete Bezüge der Struktur- und Prozessmerkmale zu patientenrelevanten Outcomes sollen hier aufgezeigt werden.

Am Ende der Konkretisierung der Qualitätsmerkmale werden diese durch ein für die Neuentwicklung der Qualitätsindikatoren einberufenes, beratendes Expertengremium (siehe Abschnitt 9.3.3) hinsichtlich folgender Eignungskriterien bewertet (siehe Kapitel 10):

1. Verbesserungspotenzial für die Erfüllung der Merkmalsanforderung
2. Zuschreibbarkeit der Verantwortung der Leistungserbringer für die Ausprägung des Qualitätsmerkmals (schließt die Beeinflussbarkeit durch den Leistungserbringer ein)
3. Berücksichtigung von unerwünschten Wirkungen und Fehlanreizen

Ziel dieser Bewertung durch das Expertengremium ist es, eine Einschätzung der am besten geeigneten Qualitätsmerkmale für den betrachteten Versorgungsbereich zu erhalten. Die Erfassung der Expertenmeinung erfolgt dabei in Anlehnung an die RAND/UCLA-Appropriateness-Method (RAM) (Fitch et al. 2001, Anwendungen in der Qualitätsindikatoren-Entwicklung siehe z. B. bei CHP/PCOR/Battelle Memorial Institute 2011: 19 ff., Kötter et al. 2012, den Breejen et al. 2013, Ebert et al. 2017, Ingraham et al. 2017, Smits et al. 2017, Stegbauer et al. 2017, Ewald et al. 2018, Nag et al. 2018). Hierbei geben die Expertinnen und Experten sowohl in den schriftlichen Vorabbewertungen als auch in den persönlichen Treffen ihre Einschätzung merkmalsbezogen auf einer 9-stufigen Skala ab (Fitch et al. 2001). Nach einer ersten schriftlichen Vorabbewertung auf Basis der vorgelegten Qualitätsmerkmale findet in Kenntnis dieser Ergebnisse eine offene Diskussion der Qualitätsmerkmale durch die Expertinnen und Experten in der Sitzung statt, bevor eine zweite Bewertung erfolgt, die den Grad der Übereinstimmung zwischen den Mitgliedern des Expertengremiums ermittelt. Konsens des Expertengremiums, dass das jeweils beurteilte Kriterium erfüllt ist, wird vom IQTIG dann angenommen, wenn in der zweiten Bewertung mehr als 75 % der Expertinnen und Experten (ähnlich der Konsensstärke, die die AWMF (Muche-Borowski et al. 2012: 40) bei Abstimmungen über Leitlinienempfehlungen verwendet) einen Punktwert von 7 oder höher vergeben.

Außerdem sollen die Expertinnen und Experten Stellung nehmen, inwieweit mögliche unerwünschte Wirkungen oder Fehlanreize gegen eine Verwendung der Qualitätsmerkmale in der gesetzlichen Qualitätssicherung sprechen. Die Einschätzungen des Expertengremiums werden im Ergebnisbericht dokumentiert.

Die oben beschriebenen drei Wissensquellen werden für die Prüfung der Eignungskriterien berücksichtigt. Die Entscheidung, welche Qualitätsmerkmale in die weitere Indikatorenentwicklung eingehen, wird durch das IQTIG nach einer umfassenden Aufbereitung und Gegenüberstellung der Ergebnisse aus den oben genannten Wissensquellen getroffen.

Am Ende des Entwicklungsschritts der Konkretisierung liegen damit die Qualitätsmerkmale vor, die alle fünf Eignungskriterien des Qualitätsziels sowie die Eignungskriterien „Berücksichtigung

unerwünschter Wirkungen“ und „Zuschreibbarkeit der Verantwortung“ erfüllen (siehe Kapitel 10). Diese Qualitätsmerkmale werden im nächsten Entwicklungsschritt operationalisiert.

6.2.2 Operationalisierung der Qualitätsmerkmale

Um Qualitätsmerkmale messen und bewerten zu können, werden Qualitätsindikatoren als Messinstrumente einschließlich ihrer Bewertungskonzepte benötigt.

Zunächst ist zu entscheiden, welches Erhebungsinstrument am besten geeignet ist, ein Qualitätsmerkmal abzubilden. Im vorliegenden Kontext kommen entweder die fallbezogene bzw. die einrichtungsbezogene QS-Dokumentation oder Sozialdaten bei den Krankenkassen in Betracht. Entsprechend wird im Folgenden zunächst die Operationalisierung von Qualitätsmerkmalen mittels fall- oder einrichtungsbezogener QS-Dokumentation beschrieben. Danach wird auf die Operationalisierung von Qualitätsmerkmalen mittels Sozialdaten bei den Krankenkassen eingegangen. In beiden Fällen orientiert sich das Vorgehen an den Eignungskriterien des Messverfahrens und des Bewertungskonzepts (siehe Kapitel 10), insbesondere Objektivität, Reliabilität, Datenqualität, Validität und Praktikabilität.

Fall- oder einrichtungsbezogene QS-Dokumentation

Elementare Bestandteile einer QS-Dokumentation sind Datenfelder für einzelne Angaben wie z. B. Patientenmerkmale, Befunde, Untersuchungsergebnisse sowie die Regeln zur Verknüpfung der mithilfe dieser Datenfelder erfassten Daten (Rechenregeln).

Derzeit gebräuchlich sind z. B. Datenfelder zur Erfassung von Zahlen oder Zeichenketten wie etwa der Institutionskennzeichennummer (IK-Nummer) einer Einrichtung, der Körpergröße, der linksventrikulären Ejektionsfraktion (LVEF) oder von Kodierungen für Diagnosen oder Eingriffe. Ein weiterer, häufig verwendeter Typ von Datenfeld ermöglicht die Auswahl einer oder mehrerer Alternativen aus einer vorgegebenen Liste.

Grundlegende Anforderungen an die Gestaltung von Datenfeldern lassen sich aus der sozialwissenschaftlichen Literatur zur Formulierung von Fragebogenitems entnehmen (z. B. Bühner 2011: 133 ff.). Da die fall- und einrichtungsbezogene Dokumentation der gesetzlichen Qualitätssicherung grundsätzlich elektronisch erfolgt, sind als Kontextbedingungen auch Anforderungen an das Design von Nutzerschnittstellen für die Erfassung medizinischer Daten zu berücksichtigen (Nielsen 1994, Wilbanks und Moss 2018, Bühner 2011). Zu den zentralen Anforderungen zählt insbesondere, dass Beschriftungen für Auswahl- ebenso wie für (alpha-)numerische Eingabefelder möglichst präzise und verständlich gehalten sind; Gleiches gilt für die Bezeichnungen der Antwortalternativen in Auswahlfeldern. Die Alternativen eines Auswahlfeldes müssen den zu erfassenden Bereich vollständig abdecken und disjunkt sein (Jonkisz et al. 2012: 45 ff.).

Um die Objektivität eines Qualitätsindikators zu optimieren, sollten die erforderlichen Datenfelder vor allem manifeste Merkmale oder beim Leistungserbringer vorliegende Versorgungsdaten erfassen. Dies sind im Kontext der Qualitätssicherung z. B. direkt beobachtbare Patientenmerkmale sowie Befunde, die unmittelbar dem Krankenhausinformationssystem/Arztinformationssystem oder der Patientenakte entnommen werden können. „Bei einem **manifesten Merkmal**

bzw. einer **manifesten Variable** [...] sind die Ausprägungen direkt beobachtbar bzw. leicht feststellbar und die theoretische Bedeutung gilt als eindeutig und bekannt“ (Döring und Bortz 2016: 224, Hervorhebungen im Original).

Ebenfalls aus Gründen der Objektivität sind unzureichend standardisierte, globale Beurteilungen von Sachverhalten durch das dokumentierende Personal nach Möglichkeit zu vermeiden (z. B. die Einstufung des Grades einer Mitralklappeninsuffizienz als „gering“, „mittel“ oder „schwer“; angemessener wäre die Erfassung echokardiografischer Befunde und die Berechnung des Insuffizienzgrades anhand dieser Daten). Auf Einrichtungsebene wären entsprechend objektivierbare strukturelle Merkmale der Einrichtung oder Informationen zu Arbeitsabläufen zu bevorzugen.

Des Weiteren können bei der Operationalisierung von Qualitätsmerkmalen durch wissenschaftliche Fachgesellschaften konsenterte („Core Clinical Outcomes“; (Prinsen et al. 2016) oder durch Normgeber vorgegebene Endpunktdefinitionen von Nutzen sein.

Datenfelder sollten idealerweise ohne weitere Erläuterungen verständlich sein. Wenn jedoch ausführlichere Erklärungen erforderlich sind, können diese als zusätzliche Ausfüllhinweise angeboten werden. Ausfüllhinweise sollten nur wenn nötig eingesetzt werden und präzise und angemessen knapp formuliert sein.

Sozialdaten bei den Krankenkassen

Die Abbildbarkeit von Qualitätsmerkmalen durch die Sozialdaten bei den Krankenkassen wird im Rahmen einer Sozialdatenanalyse empirisch überprüft. Die Vorabberechnungen auf Grundlage vorliegender, anonymisierter Sozialdaten einer kooperierenden Krankenkasse ermöglichen eine erste Einschätzung, ob eine geeignete Operationalisierung von Qualitätsmerkmalen möglich ist. Analysiert wird unter anderem,

- ob die Grundgesamtheit der für das Qualitätsmerkmal relevanten Behandlungsfälle in den Daten korrekt abgebildet werden kann,
- ob die für das Qualitätsmerkmal relevanten Diagnosen/Prozeduren als Codes in den Sozialdaten definiert und geeignet sind, die für die Qualität relevanten Merkmalsausprägungen abzubilden (z. B. das Auftreten einer Komplikation) und
- ob Unterschiede im Kodierverhalten innerhalb der und/oder zwischen den verschiedenen Sektoren bestehen.

Die Operationalisierung eines Qualitätsmerkmals für das Erhebungsinstrument Sozialdaten bei den Krankenkassen erfordert:

1. die Zusammenstellung von Diagnose- oder Prozedurenkodens oder anderen zugänglichen kodierten Informationen, welche eine möglichst gut zutreffende Beschreibung des mit dem Qualitätsmerkmal thematisierten Versorgungsereignisses erlaubt
2. die Formulierung von Auswertungsalgorithmen (Rechenregeln), die festlegen, in welcher Kombination und (bei Follow-up-Indikatoren) wann die Codes dokumentiert worden sein sollen; die verbale Beschreibung eines Algorithmus könnte z. B. wie folgt lauten: „Innerhalb von 30 Tagen nach dem Indexeingriff wurde mindestens ein Kode aus der Diagnoseliste und mindestens ein Kode aus der Prozedurliste dokumentiert“

Nach der Zusammenstellung von Kodelisten und der Formulierung des Auswertungsalgorithmus können für neue Sozialdatenindikatoren Probeauswertungen von Sozialdaten durchgeführt werden zu:

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

Referenzbereiche, Risikofaktoren und Indizes

Zur Operationalisierung eines Qualitätsmerkmals zum Qualitätsindikator gehört auch die Festlegung eines (ggf. vorläufigen) Referenzbereichs (siehe Abschnitt 10.4.1). Es kann sich um einen verteilungsunabhängigen (festen) Referenzbereich oder um einen verteilungsabhängigen (perzentilbasierten) Referenzbereich, z. B. das 5. oder 95. Perzentil, handeln.

Sofern in den Literaturrecherchen zu einem Qualitätsmerkmal patientenbezogene Risikofaktoren, die das Ergebnis eines entsprechenden Indikators beeinflussen könnten, identifiziert wurden, werden diese in den Ergebnisbericht und in die Beschreibung des neu entwickelten Qualitätsindikators aufgenommen. Sie sollen zu einem späteren Zeitpunkt bei der Entwicklung eines Risikoadjustierungsmodells berücksichtigt werden.

Falls bereits zu diesem Entwicklungszeitpunkt vorgesehen ist, Qualitätsindizes auf der Basis der neuen Qualitätsindikatoren zu bilden, werden die Rationale der Indexbildung beschrieben und der Algorithmus zu seiner Berechnung formuliert (siehe Abschnitt 15.4.2).

Überprüfung der Operationalisierung

Im Anschluss an die Erstellung einer ersten Version der Qualitätsindikatoren werden diese vom IQTIG gemeinsam mit dem beratenden Expertengremium (siehe Abschnitt 9.3.3) weiter optimiert. Hierbei wird die vorliegende Operationalisierung überprüft. Im Einzelnen wird für sozialdatenbasierte Indikatoren die Korrektheit der vorgeschlagenen Codes (z. B. ICD-/OPS-/ATC-Kodes oder GOP) und für dokumentationsbasierte Indikatoren die Verständlichkeit der vorgesehenen Datenfelder untersucht, inklusive der Notwendigkeit von erläuternden Ausfüllhinweisen. Darüber hinaus werden die bei der Risikoadjustierung zu berücksichtigenden Faktoren thematisiert und die Angemessenheit der vorgeschlagenen Referenzbereiche für die Abbildung der vorgesehenen Bewertungskategorien mit den Expertinnen und Experten diskutiert. Bei der Verwendung von Sozialdaten bei den Krankenkassen können zudem noch externe Kodierexpertinnen und -experten zur Beantwortung spezieller Kodier- und Abrechnungsfragen einbezogen werden.

6.2.3 Zusammenführung zum Qualitätsindikatorenset

Gegenstand des letzten Entwicklungsschritts ist die Betrachtung der Qualitätsindikatoren insgesamt. Hierin wird das beratende Expertengremium einbezogen. Konkret werden folgende Fragen adressiert:

- Liegen Überschneidungen zwischen den Indikatoren vor? (Mehrfachbewertungen derselben Strukturen/Prozesse/Ergebnisse sind zu vermeiden)
- Sollen Qualitätsindikatoren zu Indizes zusammengeführt werden?
- 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?
- Werden konkrete Umsetzungshürden für das ausgestaltete QS-Verfahren erwartet und wenn ja, welche?

Die Einschätzungen des Expertengremiums werden im Ergebnisbericht dokumentiert. Gegebenenfalls werden unter Berücksichtigung dieser Einschätzungen Optimierungen an den Indikatoren vorgenommen. Danach werden die Qualitätsindikatoren zum vorläufigen Indikatorenset des IQTIG zusammengeführt. Darüber hinaus wird im Ergebnisbericht eine abschließende Würdigung des Indikatorensets durch das IQTIG durchgeführt, hinsichtlich folgender Fragen:

- Entspricht es dem intendierten Zweck des Verfahrens?
- Bildet es das Qualitätsmodell angemessen ab?

Aufgabe dieser Gesamtbetrachtung ist die Sicherstellung eines validen, ausgewogenen und fokussierten Sets an Qualitätsindikatoren für das QS-Verfahren. Zur Übersicht werden die Qualitätsindikatoren den sechs Qualitätsdimensionen des Rahmenkonzepts des IQTIG (siehe Abschnitt 1.1) und den Dimensionen der Struktur-/Prozess-/Ergebnisqualität zugeordnet. Das Ergebnis dieser Betrachtung kann schließlich als erste Einschätzung des übergeordneten Eignungskriteriums der „Brauchbarkeit“ des Qualitätsindikatorensets gelten.

6.2.4 Schriftliches Stellungnahmeverfahren und Ergebnisbericht

Nach Abschluss der Verfahrensentwicklung werden die Ergebnisse in einem Vorbericht für das schriftliche Stellungnahmeverfahren zusammengefasst. Dieser beschreibt den Entwicklungsprozess und das vorgesehene Qualitätsindikatorenset. Zusätzlich enthält der Vorbericht weitere Ausführungen und Informationen zur Umsetzung des vorgesehenen QS-Verfahrens:

- Auslösung der verwendeten Erhebungsinstrumente:
 - QS-Filter für die fall-/einrichtungsbezogene QS-Dokumentation
 - Patientenfilter für die Sozialdaten bei den Krankenkassen
- Details zur Datenerhebung:
 - Dokumentationsbögen für die fall- und/oder einrichtungsbezogene QS-Dokumentation, inkl. Ausfüllhinweisen

- Leistungs- und Medikationsfilter für die Sozialdaten bei den Krankenkassen
- Datenflüsse
- Erfassungszeiträume (Follow-up)
- Verfügbarkeit der Daten (Lieferfristen, Datenexporte)
- Überprüfung der Dokumentationspflicht (Sollstatistik)
- Anforderungen an eine Datenvalidierung
- Vollerhebung oder Stichprobenverfahren
- Auswertungskonzept:
 - Berichtswesen:
 - Berichtszeitpunkte
 - Inhalte für die Rückmeldeberichte an die beteiligten Leistungserbringer (Qualitätsindikatoren, Kennzahlen)
 - mögliche Risikofaktoren auf Basis der Literatur
 - Vorschlag eines Referenzbereichs (fest oder perzentilbasiert)
- möglicher Zeitplan bis zur Umsetzung des QS-Verfahrens in den Regelbetrieb
- Empfehlungen zu weiteren Schritten bis zum Regelbetrieb

Der Vorbericht geht sodann in ein schriftliches Stellungnahmeverfahren ein. Um Stellungnahme gebeten werden die bei der Entwicklung zu beteiligenden Organisationen nach § 137a Abs. 7 SGB V. Im Anschluss hieran finden weitere Anpassungen und Optimierungen der Entwicklungsinhalte durch das IQTIG statt, die in den nachfolgenden Ergebnisbericht eingearbeitet werden. Dieser wird gemeinsam mit dem ebenfalls angepassten Qualitätsindikatorenset als Empfehlung an den G-BA übersandt. In diesem Zuge erfolgt auch eine Würdigung der Stellungnahmen, die ebenfalls dem G-BA übermittelt wird.

6.3 Machbarkeitsprüfung für Dokumentation und Auswertung

Nach der dokumentations- und sozialdatenbasierten Neuentwicklung von Qualitätsindikatoren und ggf. auch nach der Weiterentwicklung eines bestehenden QS-Verfahrens folgt eine Machbarkeitsprüfung. Vor Beginn des Regelbetriebs werden so mit einer begrenzten Anzahl an Leistungserbringern, die von dem späteren QS-Verfahren betroffen sein werden, wichtige Komponenten des neuen oder weiterentwickelten QS-Verfahrens erprobt.

Die Machbarkeitsprüfung hat das Ziel, Dokumentationsprobleme, Fehler der Operationalisierung von Qualitätsmerkmalen oder der Definition des QS-Filters sowie Verständnisprobleme in Rückmeldungen von Ergebnissen an Leistungserbringer frühzeitig vor Aufnahme des Regelbetriebs eines QS-Verfahrens zu erkennen und zu beheben und dadurch die Qualitätsmessung hinsichtlich der Eignungskriterien des Messverfahrens (Objektivität, Reliabilität, Datenqualität, Validität und Praktikabilität, siehe Abschnitt 10.3) zu optimieren. Durch die Überprüfung und Optimierung zentraler Verfahrenskomponenten noch im Rahmen der Verfahrensentwicklung, d. h. vor Beginn des Regelbetriebs, ist die Machbarkeitsprüfung eine wichtige Voraussetzung für die erfolgreiche Aufnahme des Regelbetriebs.

Auch eine Arbeitsgruppe des Guidelines International Network (G-I-N) sieht Praxistests von Qualitätsindikatoren als wichtige Entwicklungsphase vor einer Entscheidung über ihre routinemäßige Anwendung. Allerdings weisen die Autorinnen und Autoren darauf hin, dass keine Standards hinsichtlich des Umfangs und des Vorgehens eines Praxistests vorliegen (Nothacker et al. 2016: 8).

6.3.1 Gegenstand der Machbarkeitsprüfung

Im Rahmen der Machbarkeitsprüfung werden folgende Verfahrenskomponenten erprobt:

1. die fallbezogene und/oder die einrichtungsbezogene QS-Dokumentation im Krankenhaus oder in der Praxis des ambulanten Leistungserbringers
2. der QS-Filter
3. die Qualitätsindikatoren, soweit sie auf der Basis von QS-Dokumentationsdaten berechnet werden können
4. Rückmeldeberichte an Leistungserbringer

Die Machbarkeitsprüfung des IQTIG ist damit einerseits keine umfassende Verfahrenserprobung wie von Campbell et al. (2011) vorgeschlagen, die für das britische Quality and Outcomes Framework (QOF) ein umfassendes Testprotokoll entwickelten, das u. a. einen Probetrieb mit einer Dauer von 6 Monaten vorsieht. Andererseits werden im Unterschied zum Vorgehen der Agency for Healthcare Research and Quality (AHRQ), die neu entwickelte Qualitätsindikatoren anhand bereits existierender Datensätze testet (CHP/PCOR/Battelle Memorial Institute 2011: 3f), durch das IQTIG neu entwickelte Qualitätsindikatoren auf der Basis neu erhobener Daten berechnet, wobei zudem auch die für ein neues QS-Verfahren entwickelte Dokumentation Gegenstand der Erprobung ist. Eine umfassendere Erprobung und Evaluation des neuen QS-Verfahrens insgesamt inklusive der erforderlichen Datenflüsse ist hingegen nicht Ziel der Machbarkeitsprüfung. Sie kann erst nach Aufnahme des Regelbetriebs erfolgen.

Sozialdatenbasierte Qualitätsindikatoren werden in einer Machbarkeitsprüfung nicht berücksichtigt. Für diesen Indikatortyp muss keine Dokumentation erprobt werden; die Rechenregeln dieser Indikatoren werden im Rahmen der Indikatorenentwicklung durch Probeauswertungen auf der Basis verfügbarer Sozialdaten überprüft (siehe Abschnitt 6.2).

Überprüfung und Optimierung der QS-Dokumentation

In der Machbarkeitsprüfung werden die Datenfeld- bzw. Schlüsselbeschriftungen und Ausfallhinweise sowie die Plausibilitätsregeln der fall- oder einrichtungsbezogenen Dokumentation überprüft.

Die Überprüfung einer Dokumentation erfolgt in zwei aufeinander aufbauenden Schritten:

1. Probedokumentation durch Leistungserbringer
2. Nachbefragung der Leistungserbringer zur Dokumentation

Aus den Ergebnissen von Probedokumentation und Nachbefragung werden Empfehlungen zur Optimierung der Dokumentation abgeleitet, die dem G-BA vorgelegt werden.

Überprüfung und Optimierung des QS-Filters

Die Ein- und Ausschlusskriterien für die zu dokumentierenden Fälle (QS-Filter) eines neuen QS-Verfahrens werden in der Machbarkeitsprüfung durch die teilnehmenden Leistungserbringer erstmals angewendet. Die Leistungserbringer werden um ihre Einschätzung gebeten, ob der QS-Filter aus ihrer Sicht Fälle ausschließt, die für die Qualitätssicherung dokumentiert werden müssten, oder ob umgekehrt Fälle ausgewählt werden, die zur Beurteilung der Versorgungsqualität im Bereich des neuen oder weiterentwickelten QS-Verfahrens nicht beitragen können.

Überprüfung und Optimierung der Qualitätsindikatoren

Es wird geprüft, ob die Qualitätsindikatoren entsprechend der im Rahmen der Indikatorentwicklung festgelegten Rechenregeln und des vorgelegten Dokumentationsentwurfs über alle Fälle ihrer jeweiligen Grundgesamtheit berechnet werden können. Es kann vorkommen, dass wegen eines hohen Anteils fehlender Werte in erforderlichen Datenfeldern ein erheblicher Teil der dokumentierten Behandlungsfälle aus der Berechnung ausgeschlossen werden muss. Für den späteren Regelbetrieb kann diesem Problem begegnet werden, indem Datenfelder der Dokumentation überarbeitet werden oder die Operationalisierung eines Indikators angepasst wird. Entsprechende Vorschläge werden im Abschlussbericht zur Machbarkeitsprüfung zusammengefasst.

Rückmeldeberichte für teilnehmende Leistungserbringer

Jeder an der Machbarkeitsprüfung teilnehmende Leistungserbringer erhält einen Rückmeldebericht mit Auswertungen zu den von ihm dokumentierten Fällen. Der Rückmeldebericht orientiert sich im Aufbau und Layout an den im Regelbetrieb der QS-Verfahren vom IQTIG bereitgestellten Rückmeldeberichten und enthält statistische Auswertungen zu den Qualitätsindikatoren und eine Basisauswertung. Ein Vergleich der Ergebnisse der eigenen Einrichtung mit Gesamtergebnissen (Benchmark) erfolgt jedoch im Rahmen einer Machbarkeitsprüfung nicht, da dieser aufgrund der fehlenden Repräsentativität der teilnehmenden Einrichtungen nicht sinnvoll ist. Die Leistungserbringer werden aufgefordert, Rückmeldungen zur Verständlichkeit und Nachvollziehbarkeit des Rückmeldeberichts zu geben. Es wird vom IQTIG geprüft, ob die Rückmeldungen Hinweise auf sinnvolle Ergänzungen der indikatorbezogenen Auswertungen enthalten, z. B. zusätzliche Kennzahlen.

6.3.2 Ablauf der Machbarkeitsprüfung

Ausgangspunkt einer Machbarkeitsprüfung ist ein Qualitätsindikatorensatz sowie der Entwurf eines Dokumentationsbogens zur Erfassung der erforderlichen Daten. Diese liegen nach Abschluss der Entwicklung der Qualitätsindikatoren für ein neues Verfahren vor. Abbildung 5 zeigt den Ablauf einer Machbarkeitsprüfung im Überblick. Die einzelnen Arbeitsschritte werden im Folgenden beschrieben.

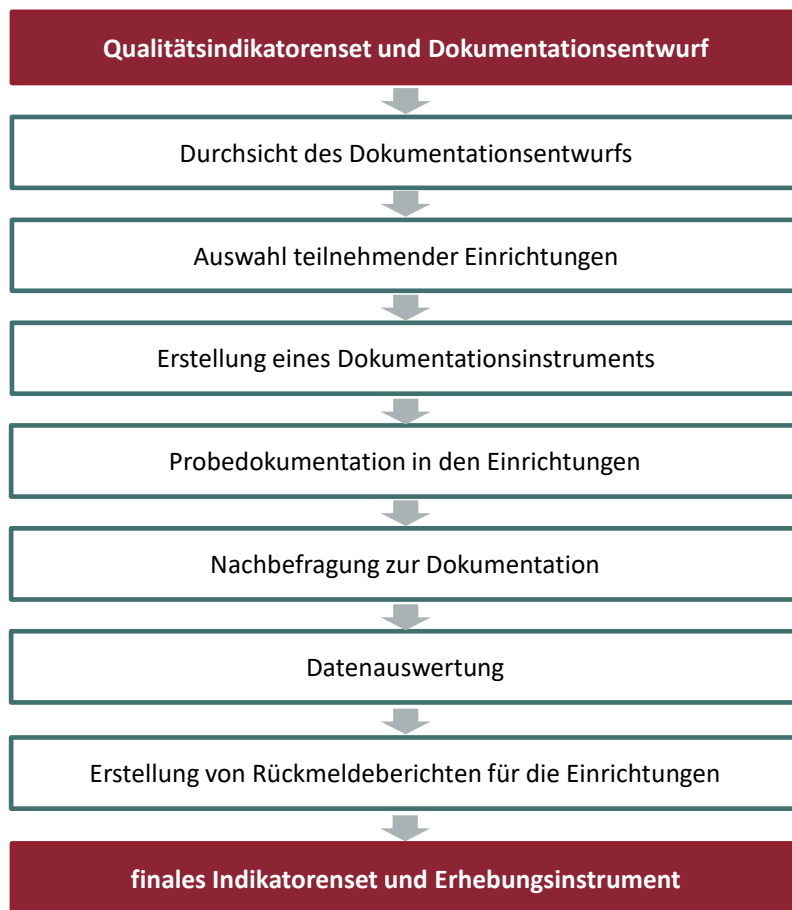


Abbildung 5: Ablauf einer Machbarkeitsprüfung

Durchsicht des Dokumentationsentwurfs

Im Vorfeld der Erstellung eines Datenerfassungsinstruments für den empirischen Test wird der nach den Entwicklungsschritten Konzeptstudie und Indikatorenentwicklung vorliegende Dokumentationsentwurf durchgesehen und bei Bedarf in Hinblick auf die Anforderungen einer nutzerfreundlichen EDV-gestützten Datenerfassung angepasst.

Es wird für jedes Datenfeld festgehalten, ob und inwieweit eine Nachbefragung beim Anwender der Dokumentation, d. h. bei den Leistungserbringern, erforderlich ist. Der entsprechende Klärungsbedarf ist gering, wenn Datenfelder aus Dokumentationsbögen anderer, bereits bestehender QS-Verfahren unverändert oder mit geringen Modifikationen übernommen werden. Hoher Klärungsbedarf besteht hingegen, wenn es sich um neue Datenfelder handelt oder wenn Datenfelder durch komplexe Plausibilitätsregeln verknüpft sind.

Dies kann zu Empfehlungen an die an der Machbarkeitsprüfung teilnehmenden Leistungserbringer führen, welche Behandlungsfälle ausgewählt werden sollen. Außerdem ergeben sich Auswirkungen auf den Leitfaden der Nachbefragung zur Dokumentation, da über Datenfelder mit hohem Klärungsbedarf mit den Dokumentierenden in jedem Fall gesprochen werden muss.

Mitwirkung der Leistungserbringer

Qualitätsförderung durch externen Vergleich ist ohne Akzeptanz aufseiten der betroffenen Leistungserbringer nicht sinnvoll denkbar. Insbesondere für QS-Verfahren, die zur Datenerfassung auf die Dokumentation von Behandlungsfällen oder Einrichtungsdaten durch Leistungserbringer angewiesen sind, ist die Mitwirkung der Leistungserbringer wichtig. Dasselbe gilt damit für die Erprobung entsprechender Datenerfassungsinstrumente, bei der niedergelassene Ärztinnen und Ärzte oder Krankenhäuser Behandlungsfälle aus ihrer Praxis oder Einrichtungsdaten dokumentieren. Das IQTIG sieht eine Machbarkeitsprüfung auch als Gelegenheit, den betreffenden Leistungserbringern Mitwirkungsmöglichkeiten zu eröffnen und von ihren Erfahrungen zu profitieren.

Auswahl teilnehmender Einrichtungen

Für die Durchführung einer Machbarkeitsprüfung ist das IQTIG auf die Unterstützung freiwillig teilnehmender ambulanter und stationärer Leistungserbringer angewiesen. Gegebenenfalls werden Fachgesellschaften und Interessenverbände über die geplante Machbarkeitsprüfung informiert und gebeten, unter ihren Mitgliedern für die Teilnahme an der Machbarkeitsprüfung zu werben. Eine repräsentative Zufallsauswahl aus allen Leistungserbringern und Behandlungsfällen ist unter dieser Voraussetzung nicht möglich.¹⁴ Dennoch muss eine möglichst hohe Generalisierbarkeit der Ergebnisse der Machbarkeitsprüfung für den späteren Regelbetrieb angestrebt werden. Zu diesem Zweck eignen sich Strategien einer gezielten Auswahl von Untersuchungseinheiten im Sinne eines *purposive sampling* (Creswell und Plano Clark 2018, Palinkas et al. 2015, Shadish et al. 2002), wie z. B. eine Auswahl mit dem Ziel maximaler Variation oder die Auswahl von besonders typischen Einrichtungen.

Für die Auswahl von Einrichtungen, die an der Machbarkeitsprüfung teilnehmen, wird eine Kombination beider Kriterien angestrebt, d. h., es werden sowohl die Variationsbreite wichtiger Unterscheidungsmerkmale als auch besonders häufige Ausprägungen („typische Leistungserbringer“) berücksichtigt. Wichtige Unterscheidungsmerkmale der Einrichtungen sind 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)
- regionale Verteilung: über ganzes Bundesgebiet, Land–Stadt

Für die Auswahl von Behandlungsfällen für die Falldokumentation sind zunächst die Ein- und Ausschlusskriterien des neuen QS-Verfahrens (QS-Filter) maßgeblich. Zusätzlich bittet das IQTIG

¹⁴ Anders wäre die Situation bei verpflichtender Teilnahme der Leistungserbringer. In diesem Fall ist neben einer Vollerhebung die Ziehung von repräsentativen Zufallsstichproben auf Einrichtungs- und Fallebene möglich.

die teilnehmenden Krankenhäuser und Arztpraxen, die beiden folgenden Vorgaben zu beachten:

1. Auswahl abgeschlossener Fälle aus einem zurückliegenden Zeitraum (z. B. Quartal oder Halbjahr): Die Fälle werden konsekutiv, d. h. in der Reihenfolge ihrer Entlassungs- oder Verlegungszeitpunkte bzw. bei ambulanten Einrichtungen in der Reihenfolge der letzten Quartalsabrechnungen oder entsprechend ihrer einrichtungsinternen fortlaufenden Nummerierung (Vorgangsnummer) dokumentiert.
2. Gegebenenfalls – abhängig vom QS-Verfahren und dem Indikatorenset sowie den Datenfeldern der zu erprobenden Dokumentation – werden die Leistungserbringer gebeten, erwartbar „schwierige“ Behandlungsfälle zu dokumentieren, um prüfen zu können, ob diese korrekt in der Dokumentation erfasst werden. Schwierig zu dokumentieren sind u. U. Behandlungsfälle mit ungewöhnlichem Behandlungsverlauf oder Komplikationen oder Behandlungsfälle, für deren adäquate Erfassung in der Dokumentation die Funktionsfähigkeit (auch komplexerer) Plausibilitätsregeln vorausgesetzt werden muss.

Was den Umfang der Datenerhebung angeht, werden folgende Punkte berücksichtigt:

- Es werden nach Möglichkeit etwa 20 bis 30 freiwillig teilnehmende Einrichtungen an einer Machbarkeitsprüfung beteiligt.
- Sofern eine einrichtungsbezogene Dokumentation erprobt wird, soll diese in jeder teilnehmenden Einrichtung durchgeführt werden; bei Erprobung der Falldokumentation sollen in jeder teilnehmenden Einrichtung ca. 15 bis 25 Falldokumentationen vorgenommen werden.

Erstellung des Dokumentationsinstruments und Probedokumentation in den Einrichtungen

Die Datenerfassung erfolgt computergestützt mithilfe von durch das IQTIG¹⁵ bereitgestellten ausfüllbaren PDF-Dateien oder einer Webapplikation. Während im Regelbetrieb der Qualitätssicherung eine Vielzahl von Datenerfassungsprogrammen unterschiedlicher Hersteller eingesetzt wird, kommt in den Machbarkeitsprüfungen des IQTIG ein einheitliches Erfassungsinstrument zum Einsatz, das neben den Minimalstandards der Basisspezifikation (IQTIG 2018b) die grundlegenden Anforderungen an die Dokumentation medizinischer Daten (Wilbanks und Moss 2018) sowie der Ergonomie und „Usability“ (Nielsen 1994, Jacobsen und Meyer 2017) erfüllt. Sofern die zu erprobende Dokumentation Ausfüllhinweise enthält, wird ein Layout gewählt, das diese automatisch, ohne zusätzliche Aktion der Nutzerin / des Nutzers, zusammen mit dem zugehörigen Datenfeld anzeigt. Die Probedokumentation in den Einrichtungen erfolgt mittels dieses Erfassungsinstruments.

Die computergestützte Datenerfassung bietet folgende Vorteile:

- Effiziente Datenübermittlung und -bereitstellung: Die Daten können aus ausgefüllten PDF-Dateien automatisiert ausgelesen oder im Falle einer Webapplikation unmittelbar in eine Datenbank des IQTIG übertragen werden.
- Möglichkeit der Implementation und Erprobung von Plausibilitätsregeln: Die Einhaltung der Wertebereiche numerischer Eingaben und die Regelkonformität von dokumentierten Codes

¹⁵ Denkbar wäre auch die Einbindung von kooperationsbereiten Softwareanbietern.

können überprüft werden. Zudem können Datenfelder ausgeblendet oder gesperrt werden, wenn diese aufgrund bereits dokumentierter Inhalte nicht ausgefüllt werden müssen.

Eine Verknüpfung des vom IQTIG für Machbarkeitsprüfungen bereitgestellten Erfassungsinstrumentes mit Klinik- oder Praxisinformationssystemen über definierte Schnittstellen ist derzeit nicht möglich. Deshalb können Daten wie ICD- oder OPS-Kodes nicht automatisch übernommen werden. Auch ein Abgleich des QS-Filters mit den in Einrichtungsinformationssystemen erfassten Behandlungsfällen ist nicht möglich. Die teilnehmenden Einrichtungen erhalten die ICD- und OPS-Listen des QS-Filters in Form von (elektronischen) Dokumenten und werden gebeten, ihre Fallauswahl manuell vorzunehmen. Das Vorgehen bei der Fallauswahl ist Gegenstand der späteren Nachbefragung.

Dennoch sind die Kodelisten des QS-Filters in das vom IQTIG bereitgestellte Erfassungsinstrument integriert. Sie werden bei der Dokumentation eines Behandlungsfalls verwendet, um zu prüfen, ob der Fall zur Grundgesamtheit des neuen QS-Verfahrens gehört. Wenn dies nicht zutrifft, wird ein Hinweis ausgegeben; die Dokumentation des Falls kann jedoch fortgesetzt werden. In der Nachbefragung wird dann thematisiert, warum der betreffende Fall aus Sicht der Einrichtung Gegenstand der Qualitätssicherung sein sollte.

Nachbefragung zur Dokumentation

Nachbefragungen vor Ort werden nach Abschluss der fall- oder einrichtungsbezogenen Dokumentation in den Einrichtungen mit den Personen durchgeführt, die an der Behandlung der Patientinnen und Patienten und der Dokumentation ihrer Behandlungsdaten beteiligt sind. Ziel ist die Erfassung der Perspektive der Anwenderinnen und Anwender. Die Dokumentierenden erhalten die Gelegenheit, über ihre Erfahrungen mit der Dokumentation und deren einzelnen Datenfeldern zu berichten.

Die im Rahmen einer Machbarkeitsprüfung durchgeführte Nachbefragung ist eine Form des halbstandardisierten Interviews (Döring und Bortz 2016).

In der Nachbefragung wird angestrebt, – abhängig von der Größe der Einrichtung – sowohl Ärztinnen und Ärzte als auch dokumentierendes Personal wie Study Nurses oder Dokumentarinnen und Dokumentare als Interviewpartnerinnen und -partner zu gewinnen. In den meisten Fällen steht pro Einrichtung mehr als eine Interviewpartnerin / ein Interviewpartner zur Verfügung, in einigen Fällen sind auch Qualitätsbeauftragte oder Medizincontrollerinnen und -controller anwesend.

Zu Beginn der Nachbefragung werden die Befragten gebeten zu berichten, was ihnen bei der Dokumentation positiv und was ihnen negativ aufgefallen ist.

Der überwiegende Teil der Fragen im Interview beziehen sich auf die Datenfelder der Dokumentation. Die Abfolge der Datenfelder kann als übergreifender Leitfaden des Interviews aufgefasst werden, wobei nicht alle Fragen zu jedem Datenfeld gestellt werden müssen.

Zu jedem Datenfeld sind die folgenden Fragen möglich:

- Ist das Feld – einschließlich Beschriftung und Ausfüllhinweis – verständlich?
- Ist die Angabe verfügbar?
- Wie hoch ist der Dokumentationsaufwand? Kann die Angabe aus dem Arztinformationssystem/Krankenhausinformationssystem übernommen werden?
- Zu welchen Dokumentationsfehlern kann es kommen?
- Wie kann der dokumentierte Eintrag in diesem Datenfeld interpretiert werden?

Die Fragen zu einem Datenfeld werden gestellt, sofern entsprechender Klärungsbedarf besteht. Für Datenfelder mit Antwortalternativen können zusätzlich die folgenden Nachfragen gestellt werden:

- Sind die Alternativen disjunkt?
- Decken die Alternativen alle möglichen Ausprägungen vollständig ab?
- Sind ggf. Mehrfachangaben sinnvoll?

Eine Themenliste mit allen Fragen zu den Datenfeldern wird den Interviewteilnehmerinnen und -teilnehmern ausgehändigt. Die Befragung erfolgt sukzessive zu allen Datenfeldern der Dokumentation, für die Klärungsbedarf besteht.

Nicht zu jedem Datenfeld erfolgt eine Nachbefragung. Zum einen erübrigen sich Nachfragen für Datenfelder, die eindeutig definierte Inhalte erfassen und/oder aus bereits existierenden Dokumentationen bekannt sind. Beispiele für Datenfelder dieser Art sind Geschlecht, Alter oder Körpergröße der Patientinnen und Patienten. Zum anderen erlauben es die zeitlichen und personellen Ressourcen einer Machbarkeitsprüfung bei umfangreicheren Dokumentationen nicht, jede Frage für jedes Datenfeld zu besprechen. Außerdem ist zu berücksichtigen, dass der zeitliche Aufwand für die freiwillig an der Nachbefragung teilnehmenden Ärztinnen / Ärzte und Dokumentarinnen / Dokumentare zumutbar bleiben muss. Diese Rahmenbedingungen erfordern eine vorrangige Behandlung derjenigen Datenfelder und Themen, für die erkennbarer Klärungsbedarf besteht. Durch die Interviewerinnen und Interviewer des IQTIG wird sichergestellt, dass diejenigen Themen und Datenfelder behandelt werden, welche bei der Durchsicht der Dokumentation vor Beginn der Machbarkeitsprüfung als klärungsbedürftig eingestuft wurden.

Im Anschluss an die Besprechung der Datenfelder werden den Interviewpartnerinnen und -partnern einige Fragen zur fall- oder einrichtungsbezogenen Dokumentation insgesamt und zum QS-Verfahren gestellt. Themen in diesem Teil der Nachbefragung sind u. a.:

- Sind in der Dokumentation aus Sicht der Befragten wichtige Punkte nicht oder nur unzureichend abgedeckt?
- Ist die Fallauswahl für den untersuchten Bereich der Qualitätssicherung adäquat? Falls die Einrichtung von den Vorgaben des QS-Filters abgewichen ist – was waren die Gründe?
- Werden die Krankheitsbilder sowie die Behandlungsmaßnahmen und ihre Folgen hinsichtlich der für das geplante QS-Verfahren relevanten Qualitätsaspekte in der Dokumentation adäquat abgebildet?
- Wie hoch ist der Zeitaufwand für die Dokumentation eines Behandlungsfalls oder für die Erfassung der einrichtungsbezogenen Daten?

- Welche Verbesserungen für ihre Einrichtung erwarten die Leistungserbringer von der Einführung des neuen QS-Verfahrens?

Der Zeitaufwand für eine Nachbefragung kann bis zu 4 Stunden betragen.

Die Antworten der Dokumentierenden auf die Fragen zu den einzelnen Datenfeldern und zur Dokumentation insgesamt liefern Hinweise zur korrekten Spezifikation der Datenfelder und der Rechenregeln der Indikatoren.

Datenauswertung

Die Daten aus Dokumentation und Nachbefragung werden auf zwei Ebenen ausgewertet. Zum einen wird für jedes Datenfeld eine deskriptiv-statistische Auswertung vorgenommen, zum anderen werden die Qualitätsindikatoren des künftigen QS-Verfahrens für die Gesamtstichprobe und einrichtungsbezogen berechnet. Für den Fall, dass Berechnungen für einen Indikator nicht möglich sind, wird analysiert, welche Ursachen eine erfolgreiche Berechnung verhindern. In diesem Fall können Änderungen von Datenfeldern und/oder der Rechenregeln vorgeschlagen und – sofern anhand der verfügbaren Daten möglich – bereits erprobt werden.

Bei einer ausreichenden Datenbasis werden für die Qualitätsindikatoren weitere Analysen und Auswertungen durchgeführt:

- Berechnung und grafische Darstellung der Kennwerte der Verteilung der einrichtungsbezogenen Indikatorenergebnisse
- probeweise Anwendung der in der Indikatorenentwicklung vorgeschlagenen Referenzbereiche

6.3.3 Ergebnisse der Machbarkeitsprüfung und Rückmeldeberichte

Nach Abschluss der Machbarkeitsprüfung wird dem G-BA ein Abschlussbericht übermittelt, der Vorgehen und Ergebnisse der Machbarkeitsprüfung beschreibt. Er enthält für jedes Datenfeld eine deskriptiv-statistische Auswertung, eine unkommentierte Zusammenfassung der Äußerungen der Nachbefragungsteilnehmerinnen und -teilnehmer zum Datenfeld sowie eine Beurteilung der Ergebnisse durch das IQTIG, die ggf. auch Vorschläge für eine Modifikation des Datenfelds enthält. Die Beurteilung ergibt sich aus einer Interpretation der Hinweise der Befragungsteilnehmerinnen und -teilnehmer vor dem Hintergrund der deskriptiv-statistischen Auswertung. Im Abschlussbericht werden außerdem die Ergebnisse zum QS-Filter und zu den Qualitätsindikatoren dargestellt, wobei – sofern möglich – auch Ergebnisse auf der Basis von eventuellen Änderungen der Rechenregeln berichtet werden, die sich aus den Proberechnungen und der Nachbefragung ergeben haben.

Sofern erforderlich und möglich werden bei der Ausarbeitung von Änderungsvorschlägen für QS-Filter, Dokumentation oder Rechenregeln Expertinnen und Experten, die an der Indikatorenentwicklung beteiligt waren, hinzugezogen. Falls die Machbarkeitsprüfung die Weiterentwicklung eines bestehenden QS-Verfahrens betrifft, werden bei Beratungsbedarf Mitglieder der zuständigen Expertengruppe auf Bundesebene konsultiert.

Änderungen des QS-Filters, der Dokumentation und der Rechenregeln der Qualitätsindikatoren, ggf. einschließlich der Streichung von Datenfeldern und Indikatoren werden den zuständigen Gremien des G-BA im Abschlussbericht zur Machbarkeitsprüfung als Vorschläge vorgelegt; eine Umsetzung der Änderungen erfolgt nach Zustimmung der G-BA-Gremien.

Der Abschlussbericht enthält darüber hinaus eine überarbeitete und optimierte Version der Dokumentation und der Rechenregeln der Indikatoren.

Die Einrichtungen, welche an der Machbarkeitsprüfung teilgenommen haben, erhalten Rückmeldeberichte mit einer Auswertung zu den jeweils dokumentierten Fällen; für den G-BA wird eine Musterauswertung bereitgestellt.

6.4 Entwicklung von Patientenbefragungen im Rahmen von Qualitätssicherungsverfahren

Patientenbefragungen haben sich 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, kann 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.

Die methodisch-wissenschaftlichen Grundprinzipien des IQTIG sowie die Rahmenbedingungen der gesetzlichen Qualitätssicherung nach SGB V stellen Anforderungen an die Entwicklung von QS-Verfahren (siehe Einleitung zu Kapitel 6), die damit auch für die Entwicklung einer Patientenbefragung im Kontext eines QS-Verfahrens gelten. Diese Anforderungen haben einen wesentlichen Einfluss auf die Methodik der Fragebogenentwicklung, die im Folgenden beschrieben wird.

6.4.1 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 (Williams 1994, Linder-Pelz 1982). Denn trotz tatsächlich gemessener Qualitätsunterschiede oder Unterschiede in der Outcomemessung zeigen Zufriedenheitsbefragungen oft keine Unterschiede, sondern eine gleichermaßen hohe Patientenzufriedenheit im Sinne von Deckeneffekten (Jacob und Bengel 2000, Heidegger et al. 2008). 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 PROM und PREM verfolgt werden. PROM wird als Oberbegriff für verschiedene Ansätze zur Erfassung von Behandlungsergebnissen verwendet, die von Patientinnen und Patienten berichtet werden, und bezieht sich häufig auf Symptome, Funktionsstatus, Aspekte der Lebensqualität und generelle Aspekte des Wohlbefindens (Patrick et al. 2008, Weldring und Smith 2013). 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 bestimmte Gesundheitsaspekte nur durch die Betroffenen selbst beurteilt werden können (z. B. Schmerzen). Mit dieser Befragungsmethode wird die Patientin bzw. der Patient in den Mittelpunkt gestellt, indem die tatsächlich erlebten Behandlungseffekte und Erfahrungen, die im Zusammenhang mit der Behandlung stehen, direkt erfasst 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. 1993, 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.

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 Kapitel 3 beschrieben, legt das IQTIG bei der Definition und Bewertung der Versorgungsqualität den Schwerpunkt auf die Patientenzentrierung. Dies schlägt sich unmittelbar in den Entwicklungsmethoden zur Patientenbefragung nieder und bedeutet, dass Patientinnen und Patienten nicht allein bezüglich bestimmter Situationen und Zustände Auskunft geben sollen. Vielmehr richten sich die zu erfragenden Inhalte nach den Bedürfnissen und der Bedeutung, die sie für die Patientinnen und Patienten haben. Die entsprechenden Methoden, die bei der Entwicklung von Patientenbefragungen Anwendung finden, werden im Folgenden dargestellt.

6.4.2 Methodische Schritte der Fragebogen- und Qualitätsindikatorenentwicklung

Die Fragebogenentwicklung geht von dem Qualitätsmodell aus, das im Rahmen einer vorangegangenen Konzeptstudie (siehe Abschnitt 6.1) erarbeitet wurde. Damit bestimmen die Qualitätsaspekte des Qualitätsmodells, die durch eine Patientenbefragung adressiert werden sollen, die inhaltlichen Schwerpunkte der Befragung. Diese werden in den weiteren Entwicklungsschritten anhand von Qualitätsmerkmalen konkretisiert. Die Qualitätsmerkmale bilden dann die Grundlage für die Operationalisierung der Befragungsinhalte anhand von Fragebogenitems. Anschließend erfolgen die Aggregation der Fragebogenitems und die Definition der Qualitätsindikatoren orientiert am Qualitätsmodell. Abbildung 6 gibt einen allgemeinen Überblick über die Entwicklungsschritte der Patientenbefragung.

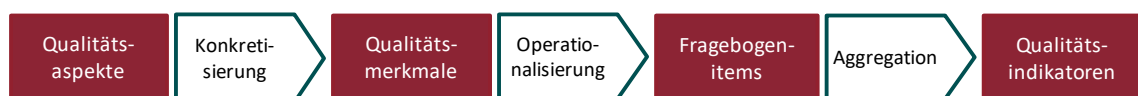


Abbildung 6: Schematische Darstellung der Entwicklungsschritte einer Patientenbefragung

Neben der Fragebogenentwicklung müssen für jedes Verfahren u. a. auch Fragebogenlogistik und Datenflüsse entwickelt werden, da sich Befragungszeiträume, Auslösekriterien und Rahmenbedingungen unterscheiden können. Bei der Entwicklung diesbezüglicher Empfehlungen werden sowohl methodische als auch organisatorische Anforderungen berücksichtigt.

Die im Regelfall auf eine Konzeptstudie bzw. auf dem Qualitätsmodell aufbauende Fragebogenentwicklung (Abbildung 7) beginnt zunächst mit der explorativen Phase der Entwicklung der Qualitätsmerkmale, wozu die Literaturrecherche und die Durchführung von Fokusgruppen und

ggf. leitfadengestützten Interviews zählen. Die auf dieser Basis entwickelten Qualitätsmerkmale werden im nächsten Schritt mit einem Expertengremium kritisch diskutiert und ggf. überarbeitet. Da die Qualitätsmerkmale als Konkretisierungen der Qualitätsaspekte die Basis für die Fragebogenentwicklung bilden, wird nach der Entwicklung der Qualitätsmerkmale ein schriftliches Stellungnahmeverfahren durchgeführt. Dabei erhalten die nach § 137a Abs. 7 SGB V zu beteiligenden Organisationen die Möglichkeit, die Entwicklungen bis zur Themensetzung der Fragebogenentwicklung zu kommentieren. Zu diesem Zeitpunkt können Anregungen in die weiteren Entwicklungen einfließen, d. h., bevor mit der konkreten Fragebogenentwicklung und -testung begonnen wird. Änderungen nach Fertigstellung der Fragebögen erfordern in der Regel erneute Testungen, was erhebliche Auswirkungen auf die zeitlichen Planungen der Entwicklung der Patientenbefragungen hat.

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 der Fragebogen nochmals getestet. Abschließend werden auf Grundlage des validierten Fragebogens Qualitätsindikatoren definiert. Diese Entwicklungsschritte werden in den folgenden Abschnitten jeweils genauer erläutert.

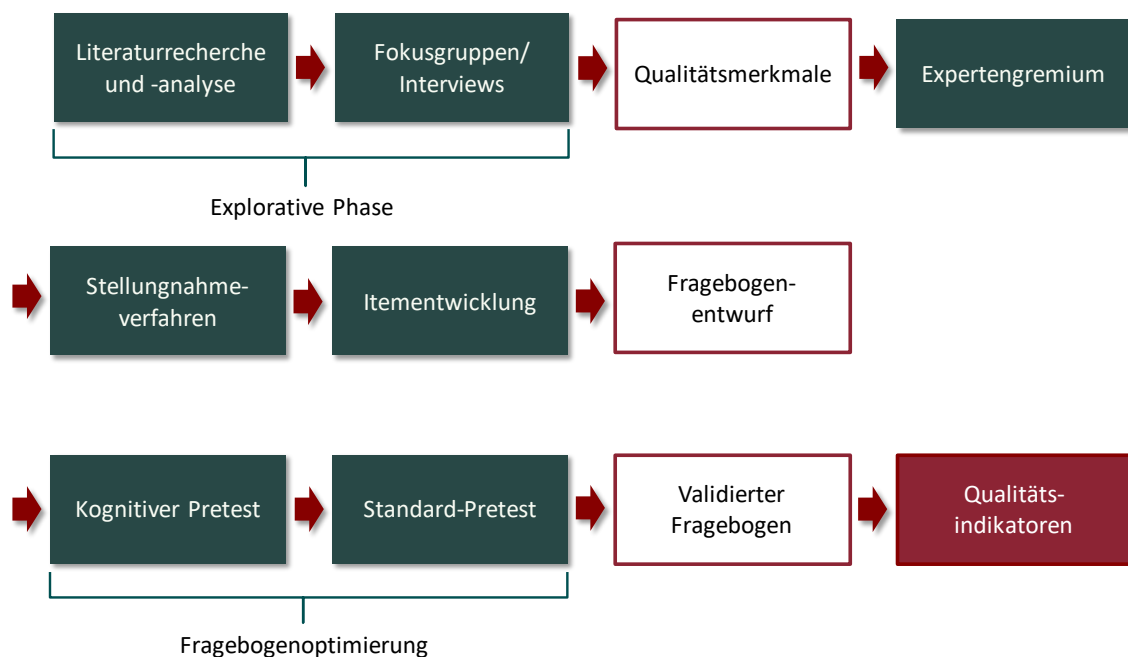


Abbildung 7: Methodische Schritte und Zwischenprodukte der Fragebogenentwicklung

Eine Patientenbefragung kann auch als Ergänzung zu einem QS-Verfahren beauftragt werden. Wenn in diesen Fällen kein Qualitätsmodell vorliegt, das den Entwicklungen der Patientenbefragung zugrunde gelegt werden kann, werden im Rahmen der Entwicklung zur Patientenbefragung ausschließlich die Qualitätsaspekte erarbeitet, die mittels Patientenbefragung adressierbar sind, bevor hierzu die jeweiligen Qualitätsmerkmale entwickelt werden.

Literaturrecherche und -analyse

Die Entwicklung einer Patientenbefragung beginnt mit einer Literaturrecherche und -analyse, die auf die Erkenntnisse aus der Literaturrecherche der Konzeptstudie aufbaut (siehe Kapitel 8 und Abschnitt 6.1). Die Literaturrecherche ist eine Wissensquelle zur Identifizierung, Analyse und Beschreibung der für die Patientenbefragung relevanten Qualitätsmerkmale. Außerdem dienen die Ergebnisse der Generierung von Themen für die Diskussionsleitfäden der Fokusgruppen bzw. der Interviewleitfäden und geben auch Anregungen und Hinweise für die spätere Entwicklung der Fragebogenitems.

Für die Literaturrecherche und -analyse bestimmen die Fragestellungen, die sich aus der Zielsetzung der Patientenbefragung ergeben, die Auswahl der Literatur- und Studientypen. Für alle Zielsetzungen werden vorab die inhaltlichen Ein- und Ausschlusskriterien (z. B. Population, Indikation, Intervention, Sektor, Endpunkte), die methodisch-formalen Ein- und Ausschlusskriterien (z. B. Studien- bzw. Leitlinientyp, erforderliche Studienqualität, Sprache, Publikationszeitraum) sowie die Datenquellen (z. B. bibliografische Datenbanken wie MEDLINE/EMBASE, Leitliniendatenbanken, Websiterecherche) definiert. Um sicherzustellen, dass alle relevanten Informationen berücksichtigt werden, wird neben der Recherche in bibliografischen Datenbanken auch eine ergänzende Handsuche in Referenzlisten von Publikationen sowie auf Websites relevanter Organisationen vorgenommen (siehe Kapitel 8).

Die Literaturrecherche umfasst neben dem Rechercheteil auch immer eine Analyse der gesichteten Literatur. Im Kontext der Patientenbefragung bedeutet Literaturanalyse nicht, dass Evidenz für Zusammenhänge zwischen patientenzentrierten Prozessen, wie etwa partizipativer Entscheidungsfindung, und klinischen Ergebnissen vorliegen muss. Gerade für die Patientenzentrierung einer Behandlung, die sich primär auf Prozesse im Behandlungsgeschehen bezieht, spielen Daten zu deren klinischer Effektivität eine untergeordnete Rolle. Anforderungen der Patientenzentrierung, wie die Ausrichtung der Versorgungsgestaltung an Patientinnen und Patienten in der Behandlung, gehen einher mit einem veränderten Rollenverständnis der Patientinnen und Patienten – weg von passiven Behandlungsempfängerinnen und -empfängern hin zu aktiven Beteiligten des Behandlungsgeschehens (siehe Kapitel 3). Die damit verbundene Anerkennung der Patientenbedürfnisse, die Einbindung in Entscheidungsprozesse und die Informiertheit über die Behandlung haben bereits Eingang in die Gesetzgebung gefunden. Folglich begründet sich Patientenzentrierung auch aus gesellschaftlichen Werten, deren Auswirkungen – beispielweise auf die Therapietreue, Lebensqualität oder Ängste – in einzelnen Bereichen untersucht wurden (z. B. Härter et al. 2005, Scheibler et al. 2003, Haynes et al. 1996).

Diesem Umstand wird bei der Literaturanalyse 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 10). Weiterhin liegt bereits bei der Analyse der Literatur im Kontext der Patientenbefragung der Fokus auf der Beurteilbarkeit durch die Patientinnen und Patienten. Das bedeutet, dass die recherchierten Inhalte auch unter dem Gesichtspunkt analysiert werden, inwiefern sie von Patientinnen und Patienten erlebt werden können, sodass diese im Rahmen einer Befragung darüber berichten können.

Durch die Literaturanalyse wird Transparenz darüber hergestellt, mit welcher Deutlichkeit die recherchierten Quellen Hinweise auf konkrete patientenrelevante Qualitätsmerkmale geben.

Wird eine Patientenbefragung im Rahmen eines bereits bestehenden QS-Verfahrens nachentwickelt, ohne dass bereits ein verfahrensübergreifendes Qualitätsmodell besteht, ist die Literaturrecherche und -analyse eine Grundlage für die Definition der relevanten Qualitätsaspekte des Qualitätsmodells sowie der Qualitätsmerkmale der Patientenbefragung.

Fokusgruppen und Interviews

Fokusgruppen und Interviews bilden gemeinsam mit den Erkenntnissen aus der Literaturrecherche die Grundlage für die Ableitung der Qualitätsmerkmale zur Fragebogenentwicklung. Dabei wird den Fokusgruppen (siehe Abschnitt 9.3.2) 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 (Lamnek 2010, Flick 2016). Das Ziel der Fokusgruppen liegt dementsprechend in der Vertiefung und Konkretisierung der Qualitätsaspekte, die über eine Patientenbefragung adressiert werden sollen, sodass die Themensetzungen des späteren Fragebogens genauer bestimmt und Qualitätsmerkmale als Ausgangspunkt für 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 Vertreterinnen und Vertretern von Gesundheitsprofessionen, 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 Konzeptstudie 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 Vertreterinnen und Vertretern der Gesundheitsprofessionen sollen diese Merkmale aus ihrem Erfahrungshorizont heraus ergänzen.

Bei der Zusammensetzung der Fokusgruppen wird darauf geachtet, dass die Charakteristiken der zu adressierenden Zielgruppen in ihren wesentlichen Ausprägungen vertreten sind.

Im Falle von Rekrutierungsschwierigkeiten (z. B. Schwierigkeiten, einen gemeinsamen Termin zu finden) werden alternativ persönliche leitfadengestützte Interviews durchgeführt.

Die Analyse der Ergebnisse der Fokusgruppen und Interviews wird in den Abschnitten 9.3.2 und 9.3.1 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 bilden im Kontext der Patientenbefragung die Basis für die Itemgenerierung. 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 überprüft, ob die Qualitätsmerkmale die folgenden Eignungskriterien (siehe Kapitel 10) erfüllen:

- Bedeutung für die Patientinnen und Patienten
- Zusammenhang mit einem unmittelbar patientenrelevanten Ziel
- Potenzial zur Verbesserung
- Übereinstimmung mit einem Qualitätsaspekt
- Beeinflussbarkeit durch den Leistungserbringer
- Zuschreibbarkeit der Verantwortung

Für eine Diskussion der Qualitätsmerkmale hinsichtlich dieser Kriterien werden externe Fachexpertinnen und -experten in Form eines Expertengremiums einbezogen. Ziel der Expertenconsultation ist zunächst die Darlegung der möglicherweise unterschiedlichen Ansichten und Einschätzungen der Expertinnen und Experten hinsichtlich der o. g. Kriterien (Niederberger und Wassermann 2015, Krick 2013). 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 9.3.3 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 vorgelegt. Es erfolgt eine schriftliche Einschätzung eines jeden Qualitätsmerkmals anhand der sechs o.g. Eignungskriterien. Die Ergebnisse dieser postalischen Vorabbefragung dienen als Moderationsinstrument für die zweite Phase, in der die aufbereiteten Antworten bei einem gemeinsamen Treffen der

Expertinnen und Experten diskutiert werden. Die protokollierten Aussagen der Expertinnen und Experten werden zusammenfassend aufbereitet und in den weiteren Entwicklungsprozess der Qualitätsmerkmale und Fragebogenitems einbezogen.

Itementwicklung

Grundlage der Itementwicklung bilden die auf Basis der Qualitätsaspekte des Qualitätsmodells abgeleiteten Qualitätsmerkmale. 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.¹⁶

Bei der Entwicklung von Items kann entweder ein reflektives oder formatives Konstrukt zugrunde gelegt werden (Bollen und Diamantopoulos 2017, Bühner 2011, Edwards und Bagozzi 2000). Reflektive Konstrukte entsprechen dem Konzept der klassischen Testtheorie, die voraussetzt, dass einer Messung mittels mehrerer Items immer ein eindimensionales Konstrukt zugrunde liegt (Bollen und Diamantopoulos 2017). Bei einem reflektiven Konstrukt existiert das zugrunde liegende Konstrukt unabhängig von den Items, mit denen es gemessen wird. So existiert beispielsweise die Persönlichkeitsfacette Extraversion unabhängig von den spezifischen Items, mit denen sie erfasst werden soll – die Items *reflektieren* das Konstrukt. Dies wird in Abbildung 8 durch die Pfeilrichtung von dem Konstrukt (η) hin zu den gemessenen Variablen (x_1 bis x_3) symbolisiert. Formative Konstrukte hingegen werden durch die Items definiert. Beispielsweise wird der sozioökonomische Status einer Person als die Kombination von schulischer und beruflicher Bildung, beruflicher Stellung und Einkommen definiert (Pollak 2018, Hoffmeyer-Zlotnik et al. 2010: 23 ff.). Allerdings existiert er nicht unabhängig von seinen konstituierenden Variablen – diese *formen* das Konstrukt. Während bei reflektiven Konstrukten die Itemausprägungen kausal durch das zugrunde liegende Konstrukt beeinflusst werden, ist die Kausalrichtung bei formativen Konstrukten umgekehrt. Dies wird in Abbildung 8 durch die Pfeilrichtung von den gemessenen Variablen (x_1 bis x_3) hin zum Konstrukt (η) symbolisiert.

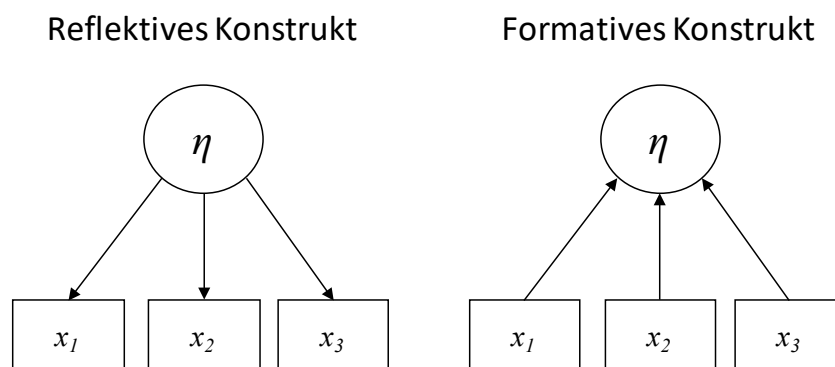


Abbildung 8: Formative und reflektive Konstrukte unterscheiden sich in der angenommenen Kausalrichtung, entweder vom Konstrukt (η) hin zu den gemessenen Variablen (x_1 bis x_3) im Fall reflektiver Konstrukte oder von den gemessenen Variablen hin zum Konstrukt im Fall formativer Konstrukte.

¹⁶ Eine Ausnahme stellen offene Fragen dar, die im vorliegenden Kontext allerdings eine untergeordnete Rolle spielen.

Eine ähnliche Unterscheidung wurde von Feinstein (1987) vorgeschlagen, der zwischen „clinimetrics“ und „psychometrics“ differenziert. „Clinimetrics“ bezeichnet in dieser Terminologie die Wissenschaft von der Messung klinischer Phänomene, wie etwa von Symptomen mittels Indizes, wohingegen „psychometrics“ die Herangehensweise der klassischen Testtheorie bei der Skalenkonstruktion beschreibt. Als Beispiel für „clinimetrics“ wird häufig der Apgar-Score angeführt, der recht unterschiedliche Aspekte der Gesundheit eines Neugeborenen in einem Index zusammenfasst (Apgar 1953).

Die Annahme darüber, ob das zugrunde liegende Konstrukt formativer oder reflektiver Natur ist, hat bedeutende Konsequenzen für die Auswahl der Variablen bzw. Items, die das Konstrukt abbilden sollen. Bei reflektiven Konstrukten sollten die Variablen möglichst hoch korreliert sein, da sie alle dasselbe Konstrukt abbilden sollen. Die Variablen sind dann prinzipiell austauschbar, ohne dass sich die Bedeutung des Konstrukts ändert (Bollen und Lennox 1991). Bei reflektiven Konstrukten werden die Items daher mittels (konfirmatorischer) Faktorenanalyse ausgewählt und die Itemanzahl wird so auf diejenigen Items reduziert, die nach den Kriterien der klassischen Testtheorie das Konstrukt optimal abbilden (Bühner 2011, Raykov und Marcoulides 2011). Die endgültige Auswahl der Items wird durch ihre statistische Konformität mit dem Modell der klassischen Testtheorie in Form von hohen Faktorladungen, unkorrelierten Fehlertermen, hohen Interitemkorrelationen, hoher interner Konsistenz etc. begründet (Diamantopoulos und Winklhofer 2001).

Bei formativen Konstrukten ändert sich jedoch die Bedeutung des Konstrukts, wenn seine konstituierenden Variablen ausgetauscht werden (Bollen und Diamantopoulos 2017, Bollen und Lennox 1991, Bühner 2011, Diamantopoulos und Winklhofer 2001), da das Konstrukt erst durch die ausgewählten Variablen definiert wird. Die Auswahl der Variablen erfolgt primär nach inhaltlichen Gesichtspunkten. Die Variablen müssen daher auch nicht unbedingt korreliert sein und müssen deshalb auch nicht durch einen Faktor in einer Faktorenanalyse beschreibbar sein. Daher kann bei formativen Konstrukten Cronbachs α auch niedrig ausfallen (Streiner 2003). Bei formativen Konstrukten ist das Ziel häufig eher eine möglichst breite Auswahl von Items, um das Konstrukt in seiner gesamten Breite abzubilden (Diamantopoulos und Winklhofer 2001). So ist beispielsweise die Auswahl der Variablen, die in den Apgar-Score eingehen, klinisch-inhaltlich begründet und nicht durch statistische Kriterien, wie etwa die Korrelation der Variablen untereinander (Streiner 2003).

Formative Konstrukte können weiter unterteilt werden in solche, die eine „konzeptuelle Einheit“ des zugrunde liegenden Konstrukts annehmen, und solche, die diese Annahme nicht treffen (Bollen und Diamantopoulos 2017). Unter konzeptueller Einheit wird das Ausmaß verstanden, zu dem die konstituierenden Variablen (x_1 bis x_n) des Konstrukts (η) die inhaltlich selbe Dimension abbilden. Ein Beispiel für ein formatives Konstrukt, das konzeptuelle Einheit nach dieser Nomenklatur aufweist, ist die Messung von Lebenszufriedenheit mittels Fragen zur Zufriedenheit mit der Wohnsituation, dem Hobby, der Familie, dem Freundeskreis und der Gesundheit (Bollen und Ting 2000). Lebenszufriedenheit als Konstrukt wird hier kausal bedingt durch die genannten Faktoren und nicht umgekehrt, d. h., die einzelnen Faktoren ändern sich nicht in Abhängigkeit von der Lebenszufriedenheit (Kausalrichtung). Alle Faktoren haben gemeinsam, dass

sie unterschiedliche Facetten desselben Konstrukts abbilden, und weisen damit konzeptuelle Einheit auf. Ein Gegenbeispiel für ein formatives Konstrukt ohne konzeptuelle Einheit ist sozio-ökonomischer Status. Die Variablen schulische und berufliche Bildung, berufliche Stellung und Einkommen bilden keine konzeptuelle Einheit – sie wurden aus inhaltlichen Gründen zusammengestellt.

Für die Methodik der Itementwicklung für die Patientenbefragung wurde aus mehreren Gründen ein formativer Ansatz gewählt. Die Items werden aus den Qualitätsmerkmalen so abgeleitet, dass sie jeweils qualitätsrelevante Ereignisse oder Situationen aus dem Themenbereich des Qualitätsmerkmals abbilden. Dabei erfasst jedes Item für sich genommen einen qualitätsrelevanten Sachverhalt eines Qualitätsmerkmals und ist daher nicht durch ein anderes Item ersetzbar. Die Auswahl der Items begründet sich also inhaltlich aus den Qualitätsmerkmalen und den ihnen zugrunde liegenden Datenquellen. Trotzdem bilden die Items eines Qualitätsmerkmals eine konzeptuelle Einheit ab – das jeweilige aus dem Qualitätsaspekt abgeleitete Qualitätsmerkmal.

Da die Patientenbefragung im Schwerpunkt einen faktenorientierten Ansatz verfolgt, werden die entsprechenden Qualitätsmerkmale in Form von PREMS und PROMS operationalisiert. Dies stellt bestimmte Anforderungen an die Frageformulierung und kommt folglich auch in den Antwortkategorien zum Ausdruck.

Neben weiteren, allgemeingültigen Prinzipien der Itemformulierung (z. B. Überschneidungsfreiheit, keine doppelten Stimuli, eindeutige Begrifflichkeiten)¹⁷ werden an die Itementwicklung 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 durch entsprechende Formulierungen ausgeschlossen oder mindestens kontrolliert werden (siehe Pretestung).

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 (Porst 2014, Hippler et al. 1991).

¹⁷ Eine gute Übersicht der allgemeinen Regeln der Itemformulierung findet sich u. a. bei Porst (2014).

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, Items zur Stichprobenbeschreibung).

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 (Lenzner et al. 2015, Prüfer und Rexroth 1996, Bradburn et al. 2004, Porst 2014). 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-Pretests¹⁸, die in einem zweistufigen Testverfahren bei der Fragebogenentwicklung im Rahmen der QS-Verfahren eingesetzt werden.

Kognitiver Pretest

In der ersten Testphase wird jeder Fragebogen einem kognitiven Pretest unterzogen. 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. Die kognitiven Pretests beziehen sich auf die Prüfung von Einzelitems (Prüfer und Rexroth 2005, Willis 2005, Willis 2004, Lenzner et al. 2015, Porst 2014).

¹⁸ In der Literatur werden für „Standard-Pretest“ häufig auch die Bezeichnungen „Standardbeobachtungspretest“ oder „klassischer Pretest“ verwendet.

Zu den am häufigsten verwendeten Techniken zählen:

- Nachfragetechniken (*Probing*), bei denen Begriffe, Fragentexte oder gegebene Antworten mittels einer oder mehrerer Zusatzfragen (*probes*) hinterfragt werden
- 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 pro Fragebogen 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 der kognitiven Pretests sind halbstandardisierte Testleitfäden, in denen die zu testenden Fragen zusammen mit der entsprechenden Nachfrage und Zielrichtung der intendierten Testung aufgeführt sind. Im offenen Teil werden die Kommentare der Befragten und die Erkenntnisse der Testleiterin bzw. des Testleiters eingetragen. Die Testleitfäden sind spezifisch auf die jeweiligen Fragebögen angepasst (D'Ardenne 2015, Lenzner et al. 2015). Diese Vorgehensweise gewährleistet einen einheitlichen Ablauf der Pretestungen, insbesondere im Hinblick auf

- einen gleichen Einstieg und Abschluss des Interviews,
- eine vorgegebene Reihenfolge der zu testenden Items und Antwortoptionen und
- eine zielgerichtete Testung, indem im Testleitfaden die einzelne zu testende Frage immer zusammen mit dem vorgegebenen Ziel und der entsprechenden Nachfragetechnik aufgeführt wird.

Innerhalb der standardisierten Vorgehensweise lässt der Testleitfaden eine Offenheit für neue Themen und Gesichtspunkte, die durch das vertiefte Nachfragen auftreten können.

In jedem kognitiven Pretest werden zwei verschiedene Testansätze parallel eingesetzt. In einem Testansatz werden gezielt einzelne Fragen bzw. Itembatterien getestet und den Testpersonen vorgelegt (Lenzner et al. 2015). In einem zweiten Testansatz steht der komplette Fragebogen im Vordergrund. Hier werden die Testpersonen gebeten, den Fragebogen zunächst vollständig auszufüllen. Nach der Beantwortung des gesamten Fragebogens schließen sich Interviews zur Gesamteinschätzung des Fragebogens sowie vertiefende Nachfragen zu einzelnen Items an. Zusätzlich werden spontane Eindrücke und bedeutsame nonverbale Signale erfasst. Ziel dieses Testansatzes ist es, einen Überblick über generelle Schwierigkeiten bei der Beantwortung einschließlich der Beantwortungsdauer zu erhalten. Durch den kombinierten Einsatz der zwei Testansätze erfolgt eine intensive Prüfung der Fragebögen.

Alle Interviews werden mit Einverständnis der Testpersonen aufgezeichnet und im Anschluss transkribiert und analysiert. Auf Basis der Ergebnisse des kognitiven Pretests wird der erste Fragebogenentwurf modifiziert.

Generell legt das IQTIG bei der Fragebogenentwicklung einen großen Schwerpunkt auf die kognitive Pretestung, da das richtige Fragenverständnis die Grundlage für valide Angaben bildet, die zu keinem späteren Zeitpunkt so transparent sichergestellt werden kann.

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. Im Rahmen des Standard-Pretests wird der Fragebogen unter möglichst ähnlichen Bedingungen, wie sie im späteren Regelbetrieb des QS-Verfahrens erwartet werden, eingesetzt. Hierzu wird der Fragebogen erstmals in einer Stichprobe eingesetzt, die sich aus Patientinnen und Patienten mehrerer Leistungserbringer zusammensetzt. 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 angewandt. Außerdem werden auf Basis der dort gewonnenen Daten die Indikatoren berechnet, die zuvor auf Grundlage inhaltlicher Überlegungen konzipiert wurden.

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.

Konstruktion der Qualitätsindikatoren

Das zugrunde liegende Qualitätsmodell einschließlich der Qualitätsmerkmale stellt nicht nur den Rahmen für die Fragebogenentwicklung dar, sondern es strukturiert auch die Entwicklung der Qualitätsindikatoren. Damit sich die theoretische Struktur des Qualitätsmodells in der Struktur der Qualitätsindikatoren widerspiegelt, werden die Items des Fragebogens zu Indizes aggregiert. Um die geschachtelte Struktur von Qualitätsmerkmalen in Qualitätsaspekten in der Struktur der Qualitätsindikatoren widerzuspiegeln, werden die Indizes durch zwei Aggregationsschritte gebildet (siehe Abbildung 9). Im ersten Schritt erfolgt eine Aggregation der Items (I) eines Merkmals (Merkmalsebene) und im zweiten Schritt eine Aggregation der Werte aus dem ersten Schritt zu einem Qualitätsindikator. Abgeleitet aus den Qualitätsaspekten und Qualitätsmerkmalen wird für jeden Qualitätsindikator ein möglichst konkretes Qualitätsziel und ein Referenzbereich definiert (siehe Kapitel 12).

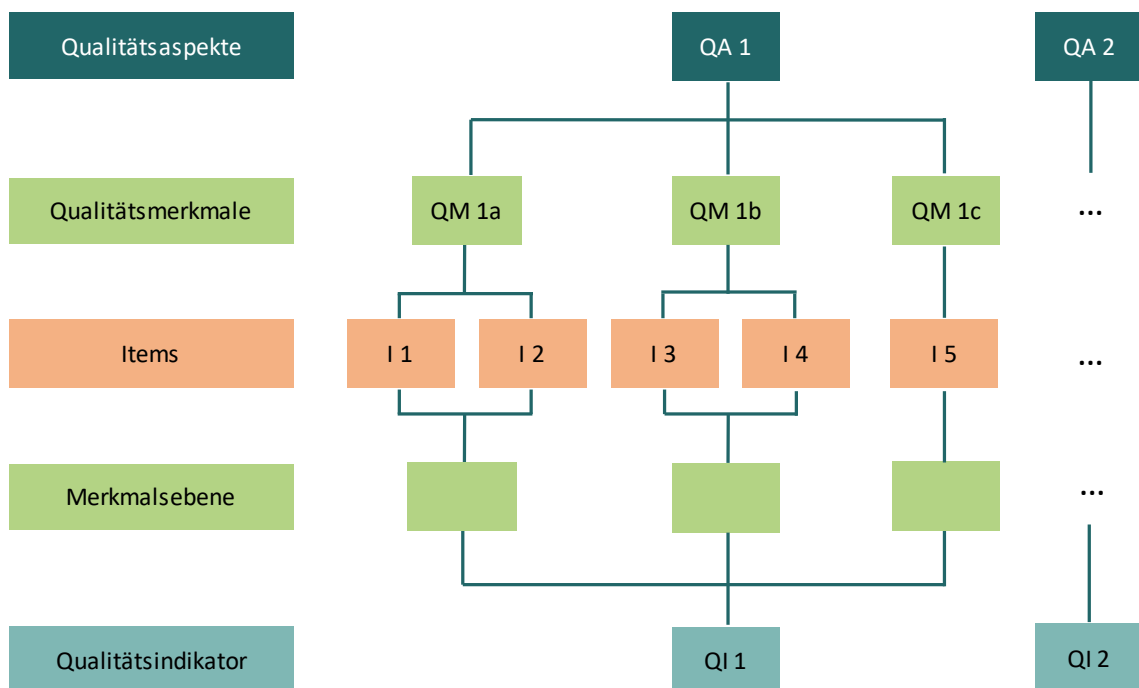


Abbildung 9: Hierarchische Struktur von Qualitätsaspekten, Qualitätsmerkmalen und Items und hierarchische Aggregation zu Qualitätsindikatoren

6.4.3 Eignungskriterien für Qualitätsindikatoren auf Basis von Patientenbefragungen

Die Entwicklung von Qualitätsindikatoren auf Basis von Patientenbefragungen orientiert sich an den Eignungskriterien für Qualitätsmessungen (siehe Kapitel 10). Diese werden als Leitfaden über den gesamten Entwicklungsprozess herangezogen und deren Erfüllung wird auf Basis unterschiedlicher Maßnahmen und Wissensbestände sichergestellt. In Tabelle 3 werden die Eignungskriterien für Qualitätsmessungen und die jeweiligen Maßnahmen und Wissensbestände, die für deren Sicherstellung herangezogen werden, gegenübergestellt.

Tabelle 3: Eignungskriterien für Qualitätsindikatoren und dazugehörige Maßnahmen und Informationsquellen der Patientenbefragung

Eignungskriterium	Maßnahmen und Wissensbestände
Qualitätsziel des Indikators	
Bedeutung für die Patientinnen und Patienten	Literaturrecherche insbesondere zur Patientenperspektive, Fokusgruppen insbesondere mit Patientinnen und Patienten, Expertengremium
Zusammenhang mit einem unmittelbar patientenrelevantem Ziel	Literaturrecherche insbesondere zur Patientenperspektive, Fokusgruppen insbesondere mit Patientinnen und Patienten, Expertengremium
Potenzial zur Verbesserung	Literaturrecherche, Fokusgruppen, Expertengremium
Übereinstimmung mit Qualitätsaspekt	Expertengremium
Beeinflussbarkeit durch den Leistungserbringer	Literaturrecherche, Expertengremium
Übergreifende Kriterien	
Brauchbarkeit	Zusammenschau der entwickelten Indikatoren vor dem Hintergrund der Zielsetzung des betreffenden QS-Verfahrens
Berücksichtigung unerwünschter Wirkungen	Literaturrecherche, Expertengremium
Zuschreibung der Verantwortung	Literaturrecherche, Expertengremium
Messverfahren	
Objektivität der Messung	faktenorientierter Befragungsansatz, kognitiver Pretest, Standard-Pretest
Datenqualität	kognitiver Pretest, Standard-Pretest
Reliabilität der Messung	Standard-Pretest
Validität der Messung	Expertengremium, kognitiver Pretest, Standard-Pretest
Praktikabilität der Messung	kognitiver Pretest, Standard-Pretest
Bewertungskonzept	
Angemessenheit des Referenzbereichs	Literaturrecherche, Indikator Konstruktion (unter Berücksichtigung von Antwortkategorien, Anzahl der Items und Auswertungskonzept)

Eignungskriterium	Maßnahmen und Wissensbestände
Klassifikationsgüte	Indikatorkonstruktion und Auswertungsmethodik
Angemessenheit der Risikoadjustierung	Literaturrecherche, Daten des Standard-Pretests (Weiterentwicklung im Regelbetrieb)

Für Fragebogeninstrumente sind einige Eignungskriterien des Messverfahrens besonders zu berücksichtigen. Daher werden diese Eignungskriterien und ihre Sicherstellung im Folgenden ausführlicher erläutert.

Objektivität der Messung

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 der Messung

Reliabilität stellt ein Maß der klassischen Testtheorie für die Präzision einer Messung dar (z. B. Moosbrugger und Kelava 2012, Bühner 2011).¹⁹ 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.

¹⁹ Da die klassische Testtheorie Intervallskalenniveau voraussetzt, beschränkt sich diese Betrachtung auf intervallskalierte Messungen.

Reliabilität ist formal definiert als der Anteil an der Varianz (σ^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 zustande.

Im Kontext der Qualitätssicherung ist das Ziel der Patientenbefragungen vorrangig ein Vergleich der Indikatorwerte der Leistungserbringer mit einem fixen Wert (z. B. einer Referenzbereichsgrenze; siehe Kapitel 12). Dazu werden in der Regel die Angaben der befragten Patientinnen und Patienten eines Leistungserbringers aggregiert, um einen Indikatorwert pro Leistungserbringer zu berechnen. Um diesen Indikatorwert wird ein Vertrauensbereich berechnet, der die Unsicherheit aufgrund der Stichprobenziehung und die Unsicherheit aufgrund nicht perfekter Reliabilität quantifiziert.

Die Reliabilität der Messung spiegelt sich somit in der Breite des Vertrauensbereichs um die Indikatorwerte der Leistungserbringer wider: Je größer die Variabilität der Patientenangaben innerhalb der Gruppe behandelter Patientinnen und Patienten eines Leistungserbringers ist, desto breiter wird der Vertrauensbereich um den Indikatorwert. Die Variabilität der Patientenangaben innerhalb eines Leistungserbringers hängt ihrerseits von der Reliabilität der Messung auf Patientenebene ab: Je geringer diese ist, desto größer wird die Variabilität der Patientenangaben eines Leistungserbringers und desto breiter wird der Vertrauensbereich. Die Breite des Vertrauensbereichs ist demnach ein Maß für die Präzision der Schätzung des 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 Vertrauensbereichs eingeht, wird dieser mit steigender Stichprobengröße schmaler. Daher steigt die Präzision der Schätzung des zugrunde liegenden Indikatorwerts eines Leistungserbringers 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 schon bei niedrigerer Reliabilität auf Patientenebene möglich (s. a. Frost et al. 2007).

Validität der Messung

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 6.4.2):

- 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 inhaltvaliden Itempool zu generieren.

Nebengütekriterien

Für die Patientenbefragung spielen auch die sogenannten Nebengütekriterien (Übersicht z. B. bei Döring und Bortz 2016: 449) Testökonomie, Nützlichkeit und Zumutbarkeit eine entscheidende Rolle. Es werden Fragebögen entwickelt, die für den jeweiligen Verwendungszweck der Qualitätsmessung nützlich sind und keine darüberhinausgehenden Aspekte beinhalten (Nützlichkeit). Ebenso 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. Schließlich sollen die für die Patientenbefragung benötigten Ressourcen im Verhältnis zum erwarteten Nutzen stehen (Ökonomie). Die einzelnen methodischen Entwicklungsschritte werden durch diese wie durch die anderen Kriterien geleitet und u. a. im Rahmen der zweistufigen Pretestung überprüft (siehe Abschnitt 6.4.2 ab S. 85).

6.4.4 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 15.2 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 eine zumutbare Anzahl an Items zur Risikoadjustierung in einem Fragebogen begrenzt sein muss 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. Die Erstellung des statistischen Modells zur Risikoadjustierung erfolgt auf Basis der Daten des ersten Jahres des jeweiligen Regelbetriebs.

7 Durchführung, Pflege und Weiterentwicklung von Qualitätssicherungsverfahren

Die gesetzlichen Aufgaben des IQTIG schließen neben Entwicklungsaufgaben auch die Beteiligung an der Umsetzung der vom G-BA beschlossenen QS-Verfahren ein (§ 137a Abs. 3 Satz 2 Nr. 3 SGB V). Vor dem Hintergrund des stetigen medizinischen und technischen Fortschritts sowie der sich ändernden Anforderungen der Gesellschaft ist es notwendig, die QS-Verfahren im Regelbetrieb nicht nur richtlinienkonform durchzuführen, sondern sie auch kontinuierlich auf Anpassungsbedarf zu prüfen und ggf. notwendigen Änderungsbedarf umzusetzen oder zur Umsetzung vorzuschlagen. Letzteres geschieht im Rahmen der Verfahrenspflege, die daher einerseits eng mit den Prozessen der Verfahrensdurchführung verbunden ist, aber andererseits einen von der Verfahrensdurchführung getrennten inhaltlichen Anspruch hat. Verfahrensweiterentwicklung beschreibt dagegen die von Durchführung und Pflege unabhängige Entwicklung umfangreicherer Änderungen eines QS-Verfahrens (z. B. Erweiterungen eines QS-Verfahrens um eine neue Datenquelle), die in Form von Empfehlungen an den G-BA in einem Abschlussbericht zusammengefasst werden und deren Umsetzung vom G-BA beschlossen werden muss. Abhängig vom Inhalt der jeweiligen Weiterentwicklung wird das IQTIG die Durchführung einer Machbarkeitsprüfung (siehe Abschnitt 6.3) empfehlen, wie sie auch bei der Entwicklung neuer Verfahren erfolgt.

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.

7.1 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 (länderbezogene) 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, Auswertungen, Rückmeldeberichte, Bewertung von Ergebnissen, Klärung von rechnerischen Auffälligkeiten mit den Leistungserbringern, Vereinbarung und Durchführung weiterer Maßnahmen). Dies betrifft gegenwärtig insbesondere QS-Verfahren mit wenigen beteiligten Leistungserbringern (z. B. derzeit die herzchirurgischen und transplantationsmedizinischen Verfahren). 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 mög-

lich. Darüber hinaus könnten die Beratungen bei sehr wenigen Leistungserbringern je Bundesland faktisch nicht pseudonymisiert durchgeführt werden, und ein Benchmarking der Leistungserbringer auf Bundeslandebene wäre ebenfalls nicht sinnvoll durchführbar. Bei indirekten QS-Verfahren werden im Gegensatz dazu wesentliche Anteile von den jeweils zuständigen Stellen auf der Ebene der Bundesländer durchgeführt. Dies beinhaltet insbesondere die Klärung von rechnerischen Auffälligkeiten mit den jeweils betroffenen Leistungserbringern.

Die Verfahrensdurchführung umfasst nicht die Entwicklung inhaltlicher Änderungsvorschläge für die QS-Verfahren. Vielmehr fasst sie die Prozesse zusammen, die das IQTIG in enger Umsetzung der Vorgaben des G-BA ohne nennenswerten inhaltlichen Entscheidungsspielraum umsetzt, das betrifft insbesondere IT-technische Verfahrensbestandteile wie die Datenannahme, die Bildung eines Bundesdatenpools, die Anwendung der finalisierten Rechenregeln, die Erstellung der unterschiedlichen Auswertungen, aber auch die Klärung von rechnerischen Auffälligkeiten und ggf. auch die Vereinbarung von weitergehenden Maßnahmen mit den betroffenen Leistungserbringern. Auch die Verfahrensdurchführung wendet regelhaft methodische Instrumente an, die vom IQTIG als unabhängigem wissenschaftlichen Institut für ihren jeweiligen Zweck entwickelt wurden. Beispiele hierfür sind neben dem Strukturierten Dialog in den direkten Qualitätssicherungsverfahren die Bewertung der Bundesergebnisse unter Beteiligung der Bundesfachgruppen bzw. der Sektorenübergreifenden Expertengremien auf Bundesebene (Expertengruppen des IQTIG) hinsichtlich der Feststellung eines besonderen Handlungsbedarfs (siehe Abschnitt 7.4) oder die Bewertung der Qualitätsindikatoren hinsichtlich ihrer Eignung zur standortbezogenen Veröffentlichung ihrer Ergebnisse (siehe Kapitel 11). Auch die Anpassung der QS-Verfahren an geänderte Vorgaben in den Richtlinien des G-BA ist in der Regel Gegenstand der Verfahrensdurchführung.

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

7.2 Verfahrenspflege

Der Begriff „Verfahrenspflege“ bezeichnet die Gesamtheit aller Prozesse, die das IQTIG mit dem Ziel durchführt, inhaltlichen Anpassungsbedarf der QS-Verfahren im Regelbetrieb festzustellen und die nötigen Änderungen dem G-BA vorzuschlagen. Anpassungsbedarf entsteht insbesondere durch neue 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 Umsetzung der Anpassungen in der Verfahrenspflege wird in der Regel im Jahresrhythmus durchgeführt und durch den G-BA beschlossen.

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.

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 wird durch die Umsetzung dieser Vorschläge jedoch nicht verändert, d. h., im Rahmen der Verfahrenspflege wird das IQTIG keine Qualitätsindikatoren zu Qualitätsaspekten vorschlagen, die in dem bestehenden QS-Verfahren bisher 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 G-BA abgestimmt und von diesem beschlossen.

Im Gegensatz zur Verfahrensweiterentwicklung beruht die Verfahrenspflege nicht auf einem gesonderten 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 konkret auf die einzelnen im Regelbetrieb durchgeführten QS-Verfahren beziehen. Da die Verfahrenspflege insbesondere auf der fachlichen Bewertung der betrachteten Qualitätsaspekte beruht, unterliegt sie insgesamt der Methodik des IQTIG als fachlich unabhängiges, wissenschaftliches Institut. Weiterentwicklungsbedarf im Sinne von Abschnitt 7.3 wird ebenfalls im Rahmen der Verfahrenspflege gesammelt und im Rahmen von Weiterentwicklungsberichten vom IQTIG zusammengefasst und an den G-BA berichtet.

7.2.1 Informationsquellen

Die Verfahrenspflege nutzt eine Reihe unterschiedlicher Informationsquellen, um Hinweise auf Veränderungen bzw. Verbesserungen der QS-Verfahren zu erlangen und diese in die Verfahren einzupflegen. Bei diesen Informationsquellen kann es sich bspw. um folgende handeln:

- Hinweise von Leistungserbringern, für die Qualitätssicherung beauftragten Stellen auf der Landesebene, Fachgesellschaften oder 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)
- Auswirkungen von Änderungen an den Qualitätsindikatoren
- Auswirkungen von Änderungen an der Datenerfassung einschließlich der eingesetzten QS-Filter
- Ergebnisse des Strukturierten Dialogs bzw. des Stellungnahmeverfahrens (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 Leitlinienaktualisierungen sowie neuen Forschungsergebnissen)
- Ergebnisse von Sonderauswertungen
- Ergebnisse der Prüfung der Qualitätsindikatoren auf Erfüllung der Eignungskriterien
- Hinweise der Trägerorganisationen in den Arbeitsgruppen und Gremien des G-BA

7.2.2 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 Expertengremien im Rahmen der Verfahrensneuentwicklung vergleichbar sind, nämlich eine fachliche Bewertung der Ergebnisse der QS-Verfahren im Regelbetrieb mit dem Ziel der Identifizierung von inhaltlichem Änderungsbedarf. Sie werden vom IQTIG zu allen aus Sicht des Instituts wesentlichen Informationen, die im Verfahrensverlauf gesammelt und erarbeitet wurden, 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 insbesondere die festgelegten Grundsätze des IQTIG (siehe Kapitel 9) maßgeblich. Erreicht werden soll, dass sowohl die von den jeweiligen QS-Verfahren betroffenen medizinischen Fachgebiete (z. B. Kardiologie) als auch die Fachvertreterinnen und Fachvertreter der jeweils einbezogenen Bereiche der Versorgung (z. B. stationäre Versorgung, vertragsärztliche Versorgung) hierbei repräsentiert werden. Bei der Zusammensetzung wird darüber hinaus berücksichtigt, dass die Expertise von Patientinnen und Patienten sowie von Vertreterinnen und Vertretern der Pflege und falls notwendig auch weiterer beteiligter Gesundheitsfachberufe oder weiterer beteiligter Gruppen eingebunden wird.

Durch die Einbindung eines breiten Wissens- und Erfahrungsspektrums soll eine fundierte medizinisch-fachliche Beratung des IQTIG für 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 seiner unabhängigen fachlichen Beratung weitere Expertinnen und Experten hinzuzuziehen. Die Zusammensetzung der einzelnen Expertengruppen wird auf der Website des Instituts veröffentlicht.

7.2.3 Handlungsoptionen

Sofern im Rahmen der Verfahrenspflege Änderungsbedarf festgestellt wird, werden die notwendigen Anpassungen im Einklang mit den Vorgaben des G-BA und unter Beachtung 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 Anpassungsvorschlägen hinsichtlich folgender Verfahrensbestandteile:

- Qualitätsindikatoren (z. B. Beschreibungstexte, Rechenregeln, Risikoadjustierung, Referenzbereiche, Literaturangaben, Vergleichbarkeit zum Vorjahr, Streichen von Qualitätsindikatoren, Entwicklung neuer Qualitätsindikatoren)
- Instrumente zur Datenerfassung (z. B. QS-Filter, Datenfeldbezeichnungen, Ausgestaltung einzelner Datenfelder [z. B. Zahl und Art der Ausprägungen eines Merkmals], Ausfüllhinweise, Plausibilitätsregeln, Streichen von Datenfeldern, Einführung neuer Datenfelder)
- Maßnahmen zur Datenvalidierung

Möglich sind auch weitergehende Empfehlungen, beispielsweise:

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

Anpassung der Qualitätsindikatoren

Die in der Qualitätsindikatorendatenbank 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 Prüfungen zur Einhaltung der Eignungskriterien für Qualitätsindikatoren (siehe Kapitel 10), Anpassungsbedarf, der im Rahmen der Auswertungen identifiziert wurde, Hinweise auf Optimierungsbedarf aus biometrischer Sicht (insbesondere Eindeutigkeit, Verständlichkeit und Übersichtlichkeit der Rechenregeln) sowie ggf. festgestellter Anpassungsbedarf für die angewandten Risikoadjustierungsmodelle.

Für alle im Rahmen des Regelbetriebs erhobenen Qualitätsindikatoren erfolgt eine wiederkehrende 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 8 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 10 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 Einführung eines neu entwickelten Qualitätsindikators empfohlen wird.

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 Qualitätsindikatorensatz 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 erfolgt in diesen Fällen grundsätzlich anhand der in Kapitel 10 dargestellten Eignungskriterien. Die Einbeziehung der medizinisch-fachlichen Expertise erfolgt nach den in Kapitel 9 dargestellten Prinzipien 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

Grundsätzlich sind derzeit in der externen gesetzlich verpflichtenden Qualitätssicherung folgende Instrumente zur Datenerfassung verfügbar:

- fallbezogene QS-Dokumentation der Leistungserbringer
- Abrechnungsdaten (Beispiel: Risikostatistik für das QS-Verfahren *Pflege: Dekubitusprophylaxe* in den Krankenhäusern)
- einrichtungsbezogene QS-Dokumentation der Leistungserbringer (Beispiel: Einrichtungsbefragung im QS-Verfahren *Vermeidung nosokomialer Infektionen – postoperative Wundinfektionen*)
- Sozialdaten bei den Krankenkassen

In absehbarer Zukunft wird als neues Erfassungsinstrument die Patientenbefragung hinzukommen. Perspektivisch bestehen voraussichtlich auch Möglichkeiten zur Nutzung der Daten aus der gesetzlichen Krebsregistrierung sowie des Transplantationsregisters und des geplanten Implantatregisters.

Alle derzeit genutzten Verfahren zur Datenerfassung beruhen auf vom IQTIG erstellten Spezifikationen, die wo nötig durch Softwarehersteller in ein IT-Produkt umgesetzt und dann z. B. 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 Prozedurenkodes 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 Qualität der erfassten Informationen und damit 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 Prüfung der Qualität 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 14 dargelegt.

7.3 Verfahrensweiterentwicklung

Ziel einer Verfahrensweiterentwicklung ist die Erarbeitung von Empfehlungen zur systematischen Umsetzung von umfangreichem Änderungsbedarf im Bereich des jeweiligen QS-Verfahrens im Regelbetrieb. Der Änderungsbedarf kann dabei insbesondere

- organisatorischer Art sein, z. B. wenn bereits bestehende QS-Verfahren hinsichtlich Datenerfassung und Auswertung zusammengelegt oder getrennt werden sollen und in diesem Zusammenhang die Qualitätsindikatoren hinsichtlich ihrer Evidenzgrundlage und der Grundgesamtheiten überprüft und ggf. überarbeitet werden müssen,
- in der Hinzunahme einer neuen Datenquelle einschließlich der Neuentwicklung von Indikatoren bestehen, ohne dass sich dadurch die betrachteten Qualitätsaspekte verändern, z. B. durch die Entwicklung von Follow-up-Indikatoren durch Nutzung von Sozialdaten bei den Krankenkassen, oder
- in der Ergänzung eines isolierten neuen Qualitätsaspekts in ein bestehendes QS-Verfahren bestehen, solange sich das Qualitätsmodell dieses Verfahrens dadurch nicht grundlegend ändert, z. B. durch die zusätzliche Betrachtung einer weiteren Eingriffsart in einem QS-Verfahren.

In Abgrenzung zur Verfahrenspflege beruht die Verfahrensweiterentwicklung auf einem Weiterentwicklungsauftrag des G-BA, der im Benehmen mit dem IQTIG erstellt wird und die Zielsetzung der angestrebten Weiterentwicklung sowie wesentliche zu berücksichtigende Inhalte festlegt. In jedem Fall wird die Bearbeitung eines Weiterentwicklungsauftrags durch Vorlage eines Abschlussberichts abgeschlossen, der die erarbeiteten Inhalte und Empfehlungen zur Umsetzung für den G-BA zusammenfasst. Die Zielsetzung ist damit prinzipiell identisch mit dem Ziel der Verfahrenspflege. Der Unterschied liegt in der Größenordnung der angestrebten Änderungen, der angewendeten Methodik, die sich von der Verfahrenspflege unterscheidet und grundsätzlich der Methodik zur Neuentwicklung von Verfahren entspricht, der Bearbeitung des Auftrags in einem zeitlich begrenzten Projekt einschließlich Berichterstattung sowie – soweit beauftragt – der Durchführung einer Machbarkeitsprüfung.

Im Gegensatz zur Verfahrenspflege kann im Rahmen einer Verfahrensweiterentwicklung damit auch das dem QS-Verfahren zugrunde liegende Qualitätsmodell in Teilen geändert werden. Im Unterschied zur Neuentwicklung eines QS-Verfahrens wird das Qualitätsmodell jedoch lediglich ergänzt und nicht grundsätzlich neu entwickelt. Es erfolgt also nicht immer eine systematische Überprüfung der im bereits bestehenden QS-Verfahren betrachteten Qualitätsaspekte auf Vollständigkeit und Aktualität, betrachtet wird nur der beauftragte Ausschnitt. Die Verfahrensweiterentwicklung soll gewährleisten, dass tiefgreifende Veränderungen an bestehenden QS-Verfahren, die den Umfang der Verfahrenspflege überschreiten, strukturiert umgesetzt werden können.

7.3.1 Anlässe für Verfahrensweiterentwicklungen

Weiterentwicklungen für bestehende QS-Verfahren werden in aller Regel vom IQTIG angestoßen. Grundlage hierfür können beispielsweise Anpassungsbedarfe aus der Verfahrenspflege sein. Zunächst wird geprüft, ob ein festgestellter 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 jedem Fall werden zunächst Empfehlungen zu Eckpunkten der angestrebten Anpassungen erarbeitet und im Bericht zur Weiterentwicklung der externen Qualitätssicherung des IQTIG dargelegt. Diese Empfehlungen legen insbesondere dar, welche Ziele angestrebt werden sollen sowie – soweit bereits konkretisierbar – welche Entwicklungsschritte im Rahmen der Umsetzung voraussichtlich notwendig sein werden. Mit dem Bericht zur Weiterentwicklung der externen Qualitätssicherung informiert das IQTIG den G-BA über den Sachstand, der dann zu entscheiden hat, ob er einen Auftrag zur Weiter- bzw. Neuentwicklung erteilen will.

7.3.2 Durchführung von Verfahrensweiterentwicklungen

Das Vorgehen entspricht grundsätzlich den Vorgaben für Verfahrensneuentwicklungen (siehe Kapitel 6).

Die wissenschaftlich-fachliche Expertise wird über die Einbindung der jeweils zuständigen Expertengruppe auf Bundesebene in die Entwicklungsarbeit gewährleistet. Die damit verbundenen Aufgaben entsprechen denen der Expertengremien bei der Verfahrensneuentwicklung. 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 6.3) notwendig ist.

7.4 Feststellung von besonderem Handlungsbedarf

7.4.1 Zweck der Einstufung des besonderen Handlungsbedarfs

Die Ergebnisse der externen Qualitätssicherung erlauben über die Qualitätsdarstellung für einzelne Leistungserbringer hinaus auch eine Gesamtbetrachtung der Versorgungsqualität im jeweiligen Leistungsbereich. Im Rahmen dieser übergreifenden Bewertung der Indikatorergebnisse nimmt das IQTIG jährlich mit Unterstützung der Expertengruppen eine Beurteilung hinsichtlich eines besonderen Handlungsbedarfs vor. Damit soll auf spezifische Qualitätsdefizite in der Versorgung aufmerksam gemacht werden, die das Qualitätsziel eines Indikators betreffen und deren Ausmaß über ein Qualitätsdefizit einzelner Einrichtungen hinausgeht. Die Feststellung eines besonderen Handlungsbedarfs hat zum Ziel, Maßnahmen anzustoßen, die über die einrichtungsbezogenen Regelungen der externen Qualitätssicherung hinaus für eine Verbesserung erforderlich sind.

7.4.2 Definition von besonderem Handlungsbedarf

Während ein Qualitätsindikator die Erfüllung von Qualitätszielen auf Einrichtungsebene abbildet, beschreibt die Einstufung des besonderen Handlungsbedarfs die Erfüllung dieser Qualitätsziele übergreifend über alle Einrichtungen, also auf Systemebene. Zeigt sich dabei ein ausgeprägtes oder persistierendes Qualitätsdefizit, wird dies als „besonderer Handlungsbedarf“ für diesen Qualitätsindikator bezeichnet. Liegt kein solches besonderes Qualitätsdefizit vor, wird für den Qualitätsindikator „kein besonderer Handlungsbedarf“ festgestellt. Die Einstufung erfolgt also in eine von zwei möglichen Kategorien.

Die Beurteilung, inwieweit die Qualitätsziele eines Indikators auf Systemebene erfüllt werden, erfolgt aus zwei Perspektiven: Zum einen wird untersucht, ob von einem angemessenen Anteil der Einrichtungen die Ziele des Qualitätsindikators erreicht werden, zum anderen, ob für die Patientenpopulation als Ganzes die Versorgungsziele des Qualitätsindikators angemessen erreicht werden.

Ein Qualitätsdefizit wird auch dann als Hinweis auf einen besonderen Handlungsbedarf verstanden, wenn es zwar nicht die Kriterien für eine besondere Ausprägung erfüllt, aber zeitlich persistiert. Dies folgt daraus, dass Qualitätsindikatoren in der externen Qualitätssicherung dann eingesetzt werden, wenn von einem nennenswerten Verbesserungspotenzial in der Gesundheitsversorgung ausgegangen werden kann (siehe Abschnitt 10.2.3). Bleiben bei einem Indikator erwartbare Verbesserungen der Versorgungsqualität aus, weist dies auf ein persistierendes Qualitätsdefizit hin, das Aktivitäten erfordert, die über die Maßnahmen der externen Qualitätssicherung hinausgehen.

Die Einstufung in „besonderer Handlungsbedarf“ oder „kein besonderer Handlungsbedarf“ wird für jeden Qualitätsindikator einzeln vorgenommen, da ein klarer Bezug zwischen der Feststellung eines Qualitätsdefizits und den daraus abzuleitenden Maßnahmen möglich sein soll.

Die Einstufung des besonderen Handlungsbedarfs erfolgt auf Grundlage der Ergebnisse der Qualitätsmessung der externen Qualitätssicherung, d. h. anhand der Indikatorergebnisse sowie ggf. anhand zugehöriger Sonderauswertungen und qualitativer Rückmeldungen von den an der externen Qualitätssicherung Beteiligten. Informationen aus anderen Quellen können zwar zur Interpretation ergänzend herangezogen werden, fließen aber nicht unmittelbar in die Einstufung ein.

Tabelle 4 zeigt den Vergleich zwischen einrichtungsbezogener und -übergreifender Qualitätsdarstellung in der Übersicht; die dort genannten Punkte werden in den folgenden Abschnitten erläutert.

Tabelle 4: Qualitätsdarstellung auf Einrichtungs- und auf Systemebene im Vergleich

	Einrichtungsebene	Systemebene
Ziel der Qualitätsdarstellung	Hinweise auf Verbesserungsbedarf in einzelnen Einrichtungen	Hinweise auf einrichtungsübergreifenden Verbesserungsbedarf
Aussage in Bezug auf die Qualität	Ausmaß der Erfüllung der Qualitätsanforderungen durch die einzelne Einrichtung	<ul style="list-style-type: none"> ▪ Ausmaß der Erfüllung der Qualitätsanforderungen durch die Gesamtheit der Einrichtungen ▪ Ausmaß der Erfüllung der Versorgungsziele für die Gesamtheit der Patientinnen und Patienten
Ergebnis der Einstufung	auffällige vs. nicht auffällige Einrichtung (gemäß QSKH-RL, DeQS-RL oder plan. QI-RL)	besonderer vs. kein besonderer Handlungsbedarf
Konsequenzen aus der Einstufung	Maßnahmen nach G-BA-Richtlinien (z. B. Strukturierter Dialog, Stellungnahmeverfahren)	Veränderungen auf Systemebene (z. B. mittels G-BA-Richtlinien, Fördermaßnahmen zur Implementation von Leitlinien)

QSKH-RL = Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern²⁰; DeQS-RL = Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung²¹; plan. QI-RL = Richtlinie zu planungsrelevanten Qualitätsindikatoren²²

7.4.3 Vorgehen zur Einstufung des besonderen Handlungsbedarfs

Um die Einstufung des besonderen Handlungsbedarfs für alle Indikatoren hinweg einheitlich vorzunehmen, werden Leitkriterien auf die Ergebnisse der Qualitätsindikatoren angewandt. Diese weisen in besonderem Maße auf ein ausgeprägtes oder persistierendes Qualitätsdefizit hin. Zusätzlich zur Anwendung dieser Leitkriterien erfolgt eine Beratung durch die Expertengruppen auf Bundesebene der jeweiligen QS-Verfahren. Eine Einstufung, die von den Hinweisen aus den Leitkriterien abweicht, ist dabei möglich. In diesem Fall werden die besonderen Gründe für die Einstufung in den Berichten des IQTIG dargestellt.

²⁰ Richtlinie des Gemeinsamen Bundesausschusses gemäß § 136 Abs. 1 SGB V i. V. m. § 135a SGB V über Maßnahmen der Qualitätssicherung für nach § 108 SGB V zugelassene Krankenhäuser. In der Fassung vom 15. August 2006, zuletzt geändert am 21. Juni 2018, in Kraft getreten am 1. Januar 2019. URL: <https://www.g-ba.de/informationen/richtlinien/38/> (abgerufen am 01.04.2019).

²¹ Richtlinie des Gemeinsamen Bundesausschusses nach § 92 Abs. 1 Satz 2 Nr. 13 i. V. m. § 136 Abs. 1 Nr. 1 SGB V über die einrichtungs- und sektorenübergreifenden Maßnahmen der Qualitätssicherung. In der Fassung vom 19. April 2010, zuletzt geändert am 21. Juni 2018, in Kraft getreten am 1. Januar 2019. URL: <https://www.g-ba.de/informationen/richtlinien/72/> (abgerufen am 01.04.2019).

²² Richtlinie zu planungsrelevanten Qualitätsindikatoren gemäß § 136 Absatz 1 SGB V i. V. m. § 136c Absatz 1 und Absatz 2 SGB V. In der Fassung vom 15. Dezember 2016, zuletzt geändert am 18. Januar 2018, in Kraft getreten am 12. Mai 2018. URL: <https://www.g-ba.de/informationen/richtlinien/91/> (abgerufen am 11.04.2019).

7.4.4 Leitkriterien

Folgende Leitkriterien gelten für einen Indikator als Hinweis auf besonderen Handlungsbedarf:

1. ein hoher Anteil von Leistungserbringern, die die Qualitätsanforderungen nicht erfüllen:
 - a. ein Anteil von > 10 % der Einrichtungen, deren Indikatorwert statistisch signifikant außerhalb des Referenzbereichs liegt, sofern es mindestens 5 Einrichtungen sind, oder
 - b. ein Anteil von > 5 % der Einrichtungen, deren Versorgungsqualität in dem Indikator im Rahmen des Strukturierten Dialogs bzw. Stellungnahmeverfahrens als qualitativ auffällig eingestuft wurde, sofern es mindestens 5 Einrichtungen sind, oder
2. ein Bundesergebnis bei Berechnung für alle Behandlungsfälle der Patientinnen und Patienten (Bundeswert), das außerhalb des Bereichs liegt, der für die Gesamtpopulation erreichbar und daher zu fordern wäre (Bundesreferenzbereich)

Ein Leitkriterium weist dann auf einen besonderen Handlungsbedarf hin, wenn die jeweilige Prüfgröße (z. B. der Anteil der Einrichtungen, deren Indikatorwert statistisch signifikant außerhalb des Referenzbereichs liegt) den festgelegten Grenzwert (10 % bzw. 5 % bzw. den Bundesreferenzbereich) überschreitet oder wenn der Zeitverlauf der Prüfgröße auf eine Persistenz des Qualitätsdefizits hinweist (siehe Abschnitt „Beurteilung von Zeitverläufen bei den Leitkriterien“). Bereits die Erfüllung eines dieser Leitkriterien ist hinreichend, um auf besonderen Handlungsbedarf für den betreffenden Indikator hinzuweisen. Die Kriterien werden im Folgenden näher erläutert.

Kriterium 1: hoher Anteil von Leistungserbringern, die die Qualitätsanforderungen nicht erfüllen

Dieses Kriterium operationalisiert das Ausmaß der Erfüllung der Qualitätsanforderungen anhand der Verbreitung von Qualitätsdefiziten bei den Einrichtungen. Da die Beurteilungen von Einrichtungen anhand des Indikatorwerts und anhand des Strukturierten Dialogs bzw. Stellungnahmeverfahrens nicht immer übereinstimmen, werden zwei Prüfgrößen unabhängig voneinander eingesetzt: zum einen der Anteil der Einrichtungen, deren Indikatorwert statistisch signifikant vom Referenzbereich abweicht, zum anderen der Anteil der Einrichtungen, die im Strukturierten Dialog bzw. Stellungnahmeverfahren als qualitativ auffällig bzw. als qualitativ unzureichend in diesem Indikator²³ bewertet wurden, jeweils bezogen auf alle Einrichtungen, die mindestens eine Patientin oder einen Patienten versorgt haben, die/der unter die Definition des Qualitätsindikators fällt.

Die Feststellung von besonderem Handlungsbedarf soll Handlungsempfehlungen und Maßnahmen (siehe Abschnitt 7.4.7) nicht aufgrund zufallsbedingter oder unsicherer Ergebnisse auslösen. Aus diesem Grund wird bei der Beurteilung anhand der rechnerischen Indikatorergebnisse statistische Unsicherheit berücksichtigt. Dies erfolgt wie bei der Berechnung statistisch signifikanter Abweichungen vom Referenzbereich für planungsrelevante Qualitätsindikatoren

²³ Gemeint ist die Beurteilung als auffällige Versorgungsqualität, also derzeit Schlüsselwerte A nach den Regelungen des G-BA zum Qualitätsbericht der Krankenhäuser (Qb-R). Dokumentationsqualität soll nicht durch die Einstufung des besonderen Handlungsbedarfs dargestellt werden.

(einseitiger exakter Binomial- bzw. Poisson-Test mit mid-p-Werten, mit einem Signifikanzniveau von 5 %, vgl. IQTIG 2016).

Welchen Grenzwert der Anteil auffälliger Einrichtungen überschreiten sollte, um auf ein besonderes Qualitätsdefizit hinzuweisen, ist eine Setzung und kann nicht methodisch hergeleitet werden. Für die Beurteilung anhand der Indikatorwerte wird der Grenzwert auf 10 % der Einrichtungen unter der Vorstellung festgelegt, dass Qualitätsdefizite in mehr als jeder zehnten Einrichtung in besonderem Maße Aktivitäten zur Verbesserung erfordern. Dieser Grenzwert berücksichtigt außerdem, dass es auch zufallsbedingt – selbst wenn keinerlei Qualitätsdefizite auf Einrichtungsebene vorliegen – in bis zu 5 % der Einrichtungen zu einer statistisch signifikanten Abweichung vom Referenzbereich kommen kann, wenn diese Abweichung jeweils auf einem Signifikanzniveau von 5 % geprüft wird.

Die Setzung von 5 % als Grenzwert für die Beurteilung anhand der Ergebnisse des Strukturierten Dialogs bzw. Stellungnahmeverfahrens folgt der Überlegung, dass im Strukturierten Dialog bzw. Stellungnahmeverfahren eine Bewertung als qualitativ auffällig bzw. als qualitativ unzureichend nach deutlich strengeren Kriterien erfolgt als anhand der Indikatorwerte und dass daher bereits ein niedrigerer Anteil von Einrichtungen mit Qualitätsproblemen Anlass für einen Hinweis auf besonderen Handlungsbedarf sein sollte. Bei der Beurteilung werden qualitativ auffällige Einrichtungen nach QSKH-RL und DeQS-RL einerseits und qualitativ unzureichende Einrichtungen nach plan. QI-RL andererseits zusammengezählt, da das Einstufungsergebnis der Einrichtung in beiden Fällen die Nichterfüllung erwartbarer Qualitätsanforderungen kenntlich macht.

Aus Kriterium 1 wird nur dann ein Hinweis auf einen besonderen Handlungsbedarf abgeleitet, wenn die jeweils beurteilte Prüfgröße mindestens 5 auffällige Einrichtungen umfasst. Bei weniger als 5 auffälligen Einrichtungen ergibt sich allein anhand der Indikatorergebnisse kein ausreichender Anhaltspunkt, dass Maßnahmen erforderlich sind, die über die einrichtungsbezogenen Regelungen der externen Qualitätssicherung hinausgehen.

Kriterium 2: der Bundeswert liegt außerhalb des Bundesreferenzbereichs

Dieses Kriterium operationalisiert das Ausmaß der Erfüllung der Versorgungsziele bezogen auf die Gesamtheit der betreffenden Patientinnen und Patienten. Der Bundeswert ist der Wert, der bei Berechnung des Indikators auf Grundlage aller Behandlungsfälle bundesweit resultiert. Für die Beurteilung ist maßgeblich, ob der Bundeswert in einem tolerablen Bereich liegt, d. h. ob der Unterschied zwischen dem tatsächlich gemessenen Bundeswert und dem bei Einhaltung aller Qualitätsanforderungen erreichbaren Bundeswert nicht ein bestimmtes Maß überschreitet.

Dieses Vorgehen folgt dem gleichen Prinzip wie die Bewertung der Qualitätsindikatorergebnisse für Einrichtungen mittels Referenzbereichen (siehe Kapitel 12). Auch die bisher in der externen Qualitätssicherung festgelegten Referenzbereiche berücksichtigen meist, dass weitere Maßnahmen (z. B. Strukturierter Dialog) nur dann erfolgen sollen, wenn der Indikatorwert um ein bestimmtes, nicht mehr tolerables Ausmaß von dem für die Patientenpopulation der Einrichtung erreichbaren Wert abweicht. Die Referenzbereiche sind allerdings auf die Beurteilung einzelner Einrichtungen ausgelegt, und die darin einkalkulierten tolerablen Abweichungen sind in der Regel, insbesondere für Ergebnisindikatoren, zu groß für die Anwendung auf Bundesebene. Da die

Auslösung des Strukturierten Dialogs nach QSKH-RL zufällige Einflüsse nicht berücksichtigt und Risikoadjustierungen stets gewissen Limitationen unterliegen, müssen Referenzbereiche für einzelne Einrichtungen eine entsprechend große Toleranz gewährleisten, die für die Gesamtpopulation nicht in diesem Ausmaß erforderlich ist. Für die Beurteilung auf Bundesebene müssen daher eigene Bundesreferenzbereiche entwickelt werden. Da dies für die Qualitätsindikatoren der externen Qualitätssicherung bisher noch nicht systematisch erfolgt ist, werden anhand eines einzelnen Bundeswerts, z. B. nur des letzten Erfassungsjahres, derzeit keine Hinweise auf besonderen Handlungsbedarf generiert. Dies erfolgt, sobald das IQTIG angemessene Bundesreferenzbereiche für die Indikatoren ermittelt hat. Das Kriterium „der Bundeswert liegt außerhalb des Bundesreferenzbereichs“ kommt also derzeit nicht zur Anwendung.

Beurteilung von Zeitverläufen bei den Leitkriterien

Die drei Prüfgrößen der oben genannten Leitkriterien, d. h. der Anteil der Einrichtungen mit statistisch signifikanter Abweichung vom Referenzwert, der Anteil der Einrichtungen mit qualitativer Auffälligkeit nach dem Strukturierten Dialog bzw. Stellungnahmeverfahren und der Bundeswert über alle Behandlungsfälle, werden nicht nur für den jüngsten Erfassungszeitpunkt mit einem Grenzwert verglichen, sondern zusätzlich im Zeitverlauf betrachtet, um auch bei weniger stark ausgeprägten, aber zeitlich persistierenden Qualitätsdefiziten auf besonderen Handlungsbedarf hinzuweisen.

Liegt zum Zeitpunkt der Beurteilung weiterhin eine verbesserungsbedürftige Qualität vor und zeigt der Verlauf der letzten drei Jahre oder länger eindeutig keine Veränderung oder einen eindeutig negativen Trend der Versorgungsqualität, so wird dies als Hinweis auf ein persistierendes Qualitätsdefizit gewertet. Voraussetzung für diese Bewertung ist, dass ein Vergleich der Werte angemessen ist, d. h. dass der Verlauf plausibel auf Unterschiede der Versorgungsqualität und nicht z. B. auf geändertes Dokumentationsverhalten zurückzuführen ist.²⁴ Die Beurteilung, auf welche Ursachen der beobachtete Zeitverlauf der Prüfgrößen zurückzuführen ist und ob weiterhin eine verbesserungsbedürftige Qualität vorliegt, erfolgt unter Zusammenschau aller Informationen durch das IQTIG mit Beratung durch die Fachexpertinnen und -experten für die jeweiligen Verfahren.

Anwendung der Leitkriterien in Abhängigkeit vom Bewertungskonzept der Qualitätsindikatoren

Ob alle Leitkriterien für besonderen Handlungsbedarf bei einem Qualitätsindikator sinnvoll angewandt werden können, hängt auch von der Art des Bewertungskonzepts des Indikators ab (siehe Kapitel 12). Für verschiedene Arten von Bewertungskonzepten wird daher unterschiedlich vorgegangen.

²⁴ Ein detailliertes Konzept, wann Vergleiche von Werten verschiedener Zeitpunkte angemessen sind („Vergleichbarkeit“), wird vom IQTIG derzeit entwickelt.

Qualitätsindikatoren mit festen Referenzbereichen

Bei Indikatoren mit festen Referenzbereichen sind die Qualitätsanforderungen eindeutig und unabhängig von der Verteilung der Einrichtungsergebnisse definiert. Bei diesen Indikatoren werden daher alle Leitkriterien angewandt.

Qualitätsindikatoren mit verteilungsabhängigen Referenzbereichen

Verteilungsabhängige (z. B. perzentilbasierte) Referenzbereiche werden dann für einen Qualitätsindikator genutzt, wenn es nicht möglich ist zu definieren, bei welchem Wert des Indikators die Schwelle zu unzureichender Qualität zu setzen ist. Per Definition können Indikatoren mit verteilungsabhängigen Referenzbereichen daher nur relative Aussagen über Qualität treffen. Bei diesen Indikatoren kann ein ausgeprägtes Qualitätsdefizit im Sinne einer verbreiteten unzureichenden Erfüllung von Mindeststandards nicht anhand des Anteils der Einrichtungen, deren Ergebnis statistisch signifikant vom Referenzbereich abweicht (Leitkriterium 1a), abgeleitet werden.

Sentinel-Event-Indikatoren

Die Sentinel-Event-Indikatoren der externen Qualitätssicherung werden in der Regel als Auslöser oder Vorfilter für weitere Beurteilungen eingesetzt (siehe Abschnitt 12.5). Ihr Indikatorwert erlaubt keine hinreichende Bewertung der Versorgungsqualität, sondern eine sinnvolle Qualitätsbewertung einzelner Einrichtungen ist erst durch die ergänzende Beurteilung z. B. im Strukturierten Dialog bzw. Stellungnahmeverfahren möglich. Für die Sentinel-Event-Indikatoren der externen Qualitätssicherung können daher aus dem Anteil rechnerisch auffälliger Einrichtungen (Leitkriterium 1a) keine Hinweise auf einen besonderen Handlungsbedarf abgeleitet werden.

Risikoadjustierte Qualitätsindikatoren

Bei Qualitätsindikatoren mit Risikoadjustierung muss berücksichtigt werden, dass der Bundeswert und dessen Zeitverlauf ggf. durch die Risikoadjustierung beeinflusst werden. Beispielsweise wird für Indikatoren, die das Verhältnis der beobachteten Rate (*observed*, O) eines interessierenden Ereignisses zur erwarteten Rate (*expected*, E) darstellen, das Bundesergebnis in der Regel unter Bezug auf das Vorjahr risikoadjustiert, sodass das bundesweite O/E-Verhältnis nur Veränderungen der Versorgungsqualität zum Vorjahr anzeigt, also in der Regel nahe 1 liegt. Bei jährlicher Anpassung der Risikoadjustierung ist daher ein Vergleich der Bundeswerte mehrerer Erfassungsjahre nicht sinnvoll. Die Zeitverläufe der Bundeswerte werden dementsprechend für risikoadjustierte Indikatoren nicht als Kriterium für besonderen Handlungsbedarf verwendet. Ein detailliertes Konzept, wann Vergleiche über mehrere Zeitpunkte angemessen sind und unter welchen Bedingungen diese auch für risikoadjustierte Indikatoren möglich sind, wird vom IQTIG entwickelt.

Tabelle 5 zeigt in der Übersicht, welche Leitkriterien in Abhängigkeit vom Bewertungskonzept des Indikators angewandt werden.

Tabelle 5: Anwendung der Leitkriterien 1a, 1b und 2 und der Zeitverläufe bei Indikatoren mit verschiedenen Bewertungskonzepten

	Indikator mit festem Referenzbereich		Indikator mit verteilungsabhängigem Referenzbereich		Sentinel-Event-Indikator
	rohe Rate	O / E	rohe Rate	O / E	
Leitkriterium 1a (Indikatorwerte)	ja	ja	nein	nein	nein
Zeitverlauf zu 1a	ja	ja	nein	nein	nein
Leitkriterium 1b (Einstufungen im Strukturierten Dialog bzw. Stellungnahmeverfahren)	ja	ja	ja	ja	ja
Zeitverlauf zu 1b	ja	ja	ja	ja	ja
Leitkriterium 2 (Bundeswert)	ja (zukünftig)	ja (zukünftig)	ja (zukünftig)	ja (zukünftig)	ja (zukünftig)
Zeitverlauf zu 2	ja	nein	ja	nein	ja

ja = Kriterium wird angewandt; nein = Kriterium wird nicht angewandt; rohe Rate = nicht risikoadjustierter Indikator; O / E = risikoadjustierter Indikator

7.4.5 Qualitätsindikatoren, für die keine Aussage zu besonderem Handlungsbedarf getroffen wird

Für manche Indikatoren ist die Einstufung, ob ein besonderer Handlungsbedarf vorliegt oder nicht, nicht sinnvoll anwendbar, da sie keine eindeutige und sichere Qualitätsaussage treffen. Für folgende Gruppen von Indikatoren wird daher keine Einstufung vorgenommen.

Qualitätsindikatoren in der Erprobungsphase

Für neu entwickelte bzw. erheblich geänderte Qualitätsindikatoren wird während ihrer Erprobungsphase (mindestens das erste Anwendungsjahr) keine Einstufung bezüglich besonderen Handlungsbedarfs vorgenommen. Insbesondere soll bekannt sein, ob sich aus dem Strukturierten Dialog bzw. Stellungnahmeverfahren Hinweise hinsichtlich der Validität des Indikators und der Angemessenheit des Referenzbereichs ergeben, bevor für einen Indikator geprüft wird, ob ein besonderes Qualitätsdefizit besteht.

Qualitätsindikatoren mit Bezug zur Vollständigkeit der Dokumentation

Bei speziellen Qualitätsindikatoren wird das Fehlen einer Information mit einer bestimmten Merkmalsausprägung gleichgesetzt. Beispiele hierfür sind Indikatoren, bei denen, wenn der Überlebensstatus der Patientin oder des Patienten nicht per Dokumentation übermittelt wurde oder der Überlebensstatus dem Krankenhaus selbst nicht bekannt ist, dies in der Auswertung als Versterben der Patientin oder des Patienten interpretiert wird (sogenannte Worst-Case-Analyse der Überlebensrate). Diese Indikatoren wurden eingeführt, um den Anreiz zu einer vollständigen Nachsorge und einer vollständigen Dokumentation zu erhöhen. Das bedeutet jedoch,

dass mehrere Aspekte in den Indikatorwert einfließen (hier: Überlebensrate, Nachsorgequalität, Dokumentationsqualität). Die genauen Ursachen für einen auffälligen Indikatorwert können im Strukturierten Dialog bzw. Stellungnahmeverfahren mit einzelnen Einrichtungen untersucht werden. Da jedoch bei einrichtungsübergreifender Betrachtung die Art des Qualitätsdefizits (z. B. Versorgungsproblem vs. Dokumentationsproblem) nicht eindeutig anhand der Indikatorergebnisse benannt werden kann, erfolgt für solche Indikatoren keine Einstufung des besonderen Handlungsbedarfs.

7.4.6 Prozess der Einstufung des besonderen Handlungsbedarfs

Die Einstufung der Ergebnisse der Qualitätsindikatoren bezüglich besonderen Handlungsbedarfs wird jährlich durchgeführt und erfolgt in drei Schritten:

- Im ersten Schritt werden die Leitkriterien angewandt, die eine Einstufung als besonderer Handlungsbedarf in besonderem Maße nahelegen und die dem IQTIG sowie den beratenden Expertinnen und Experten als Richtschnur bei der Entscheidung über die Einstufung dienen.
- Im zweiten Schritt werden die Qualitätsindikatoren, für die sich anhand der Leitkriterien Hinweise auf einen besonderen Handlungsbedarf ergeben, in den Expertengruppen auf Bundesebene der jeweiligen QS-Verfahren diskutiert. Darüber hinaus werden Qualitätsindikatoren diskutiert, bei denen zusätzliche Informationen auf besonderen Handlungsbedarf hinweisen. Solche zusätzlichen Informationen können z. B. Rückmeldungen aus den Expertengruppen auf Landesebene oder Sonderauswertungen zu regionalen Unterschieden oder zu besonderen Patienten- oder Einrichtungsgruppen sein. Ergebnis dieser Beratung ist eine Empfehlung, ob für den jeweiligen Qualitätsindikator unter Zusammenschau aller Informationen ein besonderer Handlungsbedarf konstatiert werden soll.
- Die Entscheidung über die endgültige Einstufung trifft das IQTIG im dritten und letzten Schritt. Dabei werden die Einstufungsergebnisse nochmals daraufhin geprüft, ob methodisch korrekt vorgegangen wurde, ob gegebenenfalls ein Abweichen von dem Ergebnis der Leitkriterien angemessen begründet wurde und ob vergleichbare Konstellationen bei verschiedenen Indikatoren auf gleiche Weise eingestuft wurden.

Während des Prozesses wird auch untersucht, ob ergänzend zur Feststellung eines besonderen oder persistierenden Qualitätsdefizits Handlungsempfehlungen zu dessen Behebung gegeben werden können. Mögliche Empfehlungen können beispielsweise sein:

- verstärkte Kommunikation der Gültigkeit der Qualitätsanforderungen und der Evidenzgrundlage der Indikatoren
- Schulungsmaßnahmen für Einrichtungen oder Personengruppen, z. B. durch die Ärztekammern
- Beschluss von Strukturvorgaben durch den G-BA
- Vorschlag der Berücksichtigung des Themas in regulierenden QS-Verfahren oder durch die Planungsbehörden der Länder
- Optimierung von Versorgungspfaden, Entwicklung von Versorgungsmodellen

7.4.7 Konsequenzen aus der Einstufung des besonderen Handlungsbedarfs

Das IQTIG informiert den G-BA und die Öffentlichkeit jährlich über die Einstufungsergebnisse je Qualitätsindikator, beispielsweise im Qualitätsreport und einem Bericht an den G-BA zu den Qualitätsindikatoren mit besonderem Handlungsbedarf. Die Einstufung durch das IQTIG hat primär informativen Charakter, d. h., es sollen besondere Versorgungsprobleme dargestellt werden und Institutionen des Gesundheitswesens auf besonderen Handlungsbedarf außerhalb der einrichtungsbezogenen externen Qualitätssicherung aufmerksam gemacht werden. Falls sich aus den Indikatorergebnissen, aus Sonderauswertungen oder aus der Beratung durch die Fachexpertinnen und -experten konkrete Handlungsempfehlungen ableiten lassen (siehe Abschnitt 7.4.6), übermittelt das IQTIG diese Empfehlungen ebenfalls an den G-BA und an die entsprechenden Adressaten.

Unabhängig vom Einstufungsergebnis zum besonderen Handlungsbedarf können sich bei der Beurteilung der Indikatorergebnisse Hinweise auf Anpassungs- oder Weiterentwicklungsbedarf der Qualitätsindikatoren bzw. der QS-Verfahren, auf erforderliche Maßnahmen zur Steigerung der Datenqualität oder auf das Fehlen eines relevanten Verbesserungspotenzials für einzelne Qualitätsindikatoren ergeben. Solche Hinweise werden vom IQTIG im Rahmen der Verfahrensdurchführung und -pflege berücksichtigt.

Teil C: Methodische Elemente

8 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 QS-Verfahren zu identifizieren. Beispiele hierfür sind die Literaturrecherchen im Rahmen von Konzeptstudien bzw. der Entwicklung von Qualitätsmodellen, im Rahmen von Fragebogenentwicklungen sowie bei der Verfahrenspflege.

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

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

Wichtige Prinzipien sind dabei die Transparenz und die 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 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 8.1) 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 Fragestellungen zu Qualitätsaspekten bei der Entwicklung des Qualitätsmodells sowie Fragen nach Versorgungsprozessen oder Therapien im Rahmen der Themenerschließung 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 Recherchen nach evidenzbasierten Leitlinien und systematischen Reviews im Rahmen der Verfahrenspflege.

8.1 Systematische Recherche

Bei der systematischen Recherche bezieht sich der Begriff „systematisch“ 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 inklusive der 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). Je nach Fragestellung des Projekts werden auch Kriterien für die Übertragbarkeit der Inhalte auf den deutschen Versorgungskontext definiert (Beispiel: Nur Publikationen aus Industrienationen, die wie Deutschland zum WHO Stratum A der Staateneinteilung des Weltgesundheitsberichts 2003 der World Health Organization (WHO) gehören, werden eingeschlossen (WHO 2003)).

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

8.1.1 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 angewandt. 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).

8.1.2 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 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 eine 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 wird immer in der Cochrane Library und nach HTA-Berichten über die Website des Centre for Reviews and Dissemination (CRD) recherchiert. Ebenso erfolgt eine Suche in den bibliografischen Datenbanken MEDLINE (via Ovid) und Embase (via Elsevier) und abhängig von der Fragestellung ergänzend in weiteren Datenbanken wie z. B. CINAHL (via EBSCO) und PsycINFO (via Ovid).

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

Recherche von Leitlinien

Für die Leitliniensuche werden Leitliniendatenbanken wie das „Portal der wissenschaftlichen Medizin“ der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften

(AWMF)²⁵ und die „International Guideline Library“ des Guideline International Network (G-I-N)²⁶ genutzt. Zudem findet – abhängig von der Thematik – eine Suche in weiteren Datenbanken von nationalen und/oder internationalen Leitlinienerstellern sowie Websites der Fachgesellschaften statt.

8.1.3 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 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. spezifische Websites oder bibliografische Datenbanken) für die Recherche hinzugezogen werden, wird themenspezifisch entschieden.

Recherche nach Versorgungsdaten

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

8.1.4 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 via Ovid entwickelt. Der Inhalt und die Syntax werden regelhaft von einer zweiten Person überprüft. Anschließend erfolgt die Übertragung entsprechend den spezifischen Suchmöglichkeiten auf die anderen genutzten Datenbanken wie Embase (via Elsevier), CINAHL (via EBSCO) etc. Die Syntax wird von einer zweiten Person geprüft. Ausgehend von der strukturierten Fragestellung (siehe Abschnitt 8.1.1) und ggf. von 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.

²⁵ <http://www.awmf.org/leitlinien/aktuelle-leitlinien.html> (abgerufen am 04.04.2019)

²⁶ <https://www.g-i-n.net/> (abgerufen am 04.04.2019)

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 diejenigen, die themenrelevant sind, zu identifizieren. Die themenspezifische Recherche in weiteren Informationsquellen erfolgt entsprechend den Anforderungen der Informationsquelle.

Die Suchstrategie wird für jede Datenbank bzw. Informationsquelle tabellarisch dokumentiert, inklusive der Suchoberfläche, des Suchzeitraums sowie sprachlicher und anderer Einschränkungen bzw. Filter.

8.1.5 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 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).

8.2 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 (Fletcher und Fletcher 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-2-Instrument (A Measurement Tool to Assess systematic Reviews) (Shea et al. 2017) 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 das AGREE-II-Instrument (Appraisal of Guidelines for Research & Evaluation) (AGREE Next Steps Consortium 2014, AGREE Next Steps Consortium 2017). 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 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. Bei Unstimmigkeiten von mehr als zwei Punkten werden die jeweiligen Items von den bewertenden Personen diskutiert und final konsentiert.

8.3 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 angewandter 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 2019: 2. Kap. 3. Abs. § 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], Schünemann (2009)). Bei dieser Form der Evidenzklassifizierung ist die Qualität der Evidenz der Gradmesser für das Vertrauen, dass ein ermittelter Effekt korrekt ist (Schünemann 2009).

8.3.1 Einzelstudien

Das Institut orientiert sich bei der zusammenfassenden Evidenzbewertung von Einzelstudien 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 2017).

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 6: 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 7: 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
mäßig	Nicht-RCT mit Kontrolle für Störfaktoren und großen, präzisen Schätzungen des Effekts oder RCT ohne gravierende Mängel, die zu Verzerrungen führen könnten, aber mit entweder indirekten Nachweisen oder ungenauen Schätzungen des Effekts
niedrig	RCT mit Mängeln, die zu Verzerrungen führen oder Nicht-RCT mit kleinen oder unpräzisen Schätzungen des Effekts oder ohne Kontrolle von Störfaktoren
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 8: 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	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; wenn nur eine Studie vorliegt, dann sollte die Schätzung des Nutzens den potenziellen Schaden für Patientinnen und Patienten überwiegen (nur eine vorliegende Studie kann nicht die Bewertung „hoch“ für Konsistenz erhalten)
niedrig	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 oder breite Konfidenzintervalle erschweren die Schätzung des Nutzens; wenn nur eine Studie vorliegt, dann überwiegt die Schätzung des Nutzens nicht den potenziellen Schaden für die Patientinnen und Patienten
unzureichend	keine Bewertung der Höhe und Richtung von Nutzen und Schaden für Patientinnen und Patienten möglich

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.

8.3.2 Systematische Übersichtsarbeiten

Bei einer systematischen Übersichtsarbeit werden zu einer strukturierten, präzisen Frage alle verfügbaren Einzelstudien 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-2- 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.

8.3.3 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.

8.4 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.

9 Einbeziehung externen Wissens und externer Erfahrung

9.1 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.

Mit der 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

Das IQTIG bezieht die Empfehlungen und Beratungsergebnisse der Expertinnen und Experten in seine Entscheidungsfindung ein und berücksichtigt sie in seinen Berichten und Empfehlungen, wobei das IQTIG als fachlich unabhängiges, wissenschaftliches Institut für die Inhalte und die ordnungsgemäße Wahrnehmung seiner Aufgaben letztverantwortlich bleibt.

9.2 Beratungszwecke bei der Einbindung externen Wissens

Bei der Einbindung von Personen mit ergänzendem Wissen und Erfahrung werden in Anlehnung an Niederberger und Wassermann (2015: 37 ff.) 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 angewandt, die eine für die Fragestellung sachgerechte Zusammensetzung der Gruppe sicherstellen sollen (siehe Abschnitt 9.3.3).

- **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 (siehe Abschnitt 9.3.3).

In Tabelle 9 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 9: Beratungszwecke der Einbindung externen Wissens mit Beispielen für die Einsatzbereiche und für die verwendeten Methoden

	Ermittlung oder Prüfung von Fakten	Ermittlung unterschiedlicher Perspektiven	Bewertung diskussionsbedürftiger Sachverhalte
Form der Einbindung	einzelne Personen	mehrere Personen	mehrere Personen als Gruppe
Beispiele Einsatzbereich	<ul style="list-style-type: none"> ▪ Erläuterung eines Therapieverfahrens ▪ Klärung eines Datenfelds 	<ul style="list-style-type: none"> ▪ Themenerschließung ▪ Ermittlung von Fragebogen-dimensionen ▪ Interpretation von Indikator-ergebnissen 	<ul style="list-style-type: none"> ▪ Beurteilung von Qualitätsindikatoren ▪ Beratung zum besonderen Handlungsbedarf
Beispiele Methode der Einbindung	<ul style="list-style-type: none"> ▪ Einzelbefragung (Hausärztin) ▪ Einzelbefragung (Medizin-Controller) 	<ul style="list-style-type: none"> ▪ Scoping Workshop ▪ Fokusgruppe ▪ Bundesfachgruppe ▪ qualitatives Einzel-interview 	<ul style="list-style-type: none"> ▪ Expertengremium im Rahmen der Verfahrensentwicklung ▪ Bundesfachgruppe

9.3 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 9.3.1 und 9.3.2).
- 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 9.3.3 beschrieben.

9.3.1 Leitfadengestützte Interviews

Das leitfadengestützte Interview ist eine qualitative Forschungsmethode, die regelhaft im Rahmen der Entwicklungsarbeiten eingesetzt wird. Ziel von leitfadengestützten Interviews ist der Erkenntnisgewinn durch die Befragung bestimmter Vertreterinnen und Vertreter einer Zielgruppe, die im Fokus der Untersuchung steht. 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 5.2.1): 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 6.4).

Die leitfadengestützten Interviews werden für diese Zwecke meist in Kombination mit Fokusgruppen eingesetzt (siehe Abschnitt 9.3.2). 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.²⁷ Zur Unterstützung der qualitativen Analyse wird eine Software eingesetzt, anhand derer die Kernaussagen abgeleitet und bei mehreren Interviews zusammengeführt werden.

9.3.2 Fokusgruppen

Fokusgruppen gelten als eine Methode der qualitativen Sozialforschung, bei der eine Gruppe moderiert und fokussiert zu bestimmten, vorgegebenen Themen diskutiert (Tausch und Menold 2015, Kühn und Koschel 2011). 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

²⁷ Die Teilnehmerinnen und Teilnehmer werden vorab darüber informiert und geben ihr Einverständnis. Die Aufzeichnungen werden nach der Transkription gelöscht.

Unterschied zu Einzelinterviews Gedanken, Erlebnisse, Einstellungen, Wahrnehmungen, Bedürfnisse und Orientierungen der Gruppe in einer größeren Breite transparent. Eine 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 auf ein breiteres Fundament gestellt. Folglich helfen Fokusgruppen, die Forschungsfrage bzw. den Untersuchungsgegenstand besser zu verstehen. Sie können als einzelne qualitative Methode (*stand alone*) genauso wie in Kombination mit anderen qualitativen oder quantitativen Methoden (Kühn und Koschel 2011) angewandt werden.

Grundsätzlich sollen durch den Einsatz von Fokusgruppen wesentliche Meinungsbilder von Repräsentanten des Untersuchungsgegenstandes, d. h. Patientinnen und Patienten bzw. Vertreterinnen und Vertreter der Gesundheitsberufe, gewonnen werden. Bei Fokusgruppen als eine qualitative Methode der Sozialforschung steht ein ausgewogenes Diskussionsbild im Vordergrund, was dadurch sichergestellt wird, dass die adressierte Zielgruppe durch charakteristische Patientinnen und Patienten bzw. Gesundheitsprofessionen vertreten ist (Krueger und Casey 2015, Dreher und Dreher 1982). So zielt die Methode darauf ab, Erkenntnisse von einer gezielt zusammengestellten, der Forschungsfrage folgenden Auswahl an bestimmten Personen zu gewinnen und nicht auf Basis einer repräsentativen Stichprobe (Nyumba et al. 2018, Scheuren 2004).

Neben der Zusammensetzung hängt die Qualität der Fokusgruppenergebnisse auch von der Anzahl der Gruppen ab. Das bedeutet, dass so viele Gruppen durchgeführt werden sollen, dass der benötigte Erkenntnisgewinn aller Voraussicht nach erreicht wird (Krueger und Casey 2015, Dreher und Dreher 1982). Vor diesem Hintergrund und unter Berücksichtigung der Praktikabilität (Kosten-Nutzen/Erkenntnisgewinn) werden die Teilnehmerinnen und Teilnehmer der Fokusgruppen hinsichtlich ihrer Charakteristika und Anzahl zusammengestellt.

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

Fokusgruppen werden zum einen bei der Erstellung eines Qualitätsmodells zur Erschließung von Qualitätsaspekten durchgeführt (siehe Abschnitt 6.1). Zum anderen werden Fokusgruppen bei der Entwicklung von Patientenbefragungen zur Ableitung von Qualitätsmerkmalen eingesetzt (siehe Abschnitt 6.4).

Um dem explorativen Charakter des Einsatzbereichs gerecht zu werden, erfolgt die Moderation der Fokusgruppen anhand eines Leitfadens, der im Sinne eines Topic Guides erstellt wird. Das heißt, es werden lediglich bestimmte Themen vorgegeben, die frei in der Gruppe diskutiert werden und durch die Teilnehmerinnen und Teilnehmer um eigene Themen ergänzt werden können.

Am Ende einer jeden Gruppendiskussion füllen alle Teilnehmerinnen und Teilnehmer auf freiwilliger Basis einen Kurzfragebogen aus, in dem soziodemografische Merkmale wie Alter und

Geschlecht sowie weitere auswertungsrelevante Angaben erfragt werden, die je nach Einsatz und Zielgruppe variieren können (z. B. Angaben zur Gesundheit, zur medizinischen Versorgung bzw. bei den Fokusgruppen mit medizinischem Fachpersonal Angaben zur Funktion u. ä.) Diese Angaben werden im Rahmen der Auswertung des Diskussionsmaterials zur genaueren Beschreibung der Gruppenzusammensetzung herangezogen.

Von allen Fokusgruppendifkussionen werden digitale Tonaufnahmen erstellt, die im Anschluss transkribiert und anonymisiert werden.²⁸ Ein Co-Moderator bzw. einer Co-Moderatorin beobachtet zusätzlich die Diskussion und protokolliert wesentliche Erkenntnisse und Aussagen des Diskussionsgeschehens, die die späteren Auswertungen unterstützen.

Die Analyse der anonymisierten Transkripte erfolgt gruppenübergreifend in Anlehnung an die zusammenfassende Inhaltsanalyse und die inhaltlich-strukturierende Inhaltsanalyse (Kuckartz 2014, Mayring 2015). Mithilfe der schriftlichen Aufzeichnungen werden die Transkripte mit Blick auf die Hauptfragestellung der Fokusgruppen gruppenübergreifend ausgewertet. Dabei werden in einem ersten Schritt deduktiv auf Basis des Topic Guides zentrale Diskussionsthemen erfasst, die in einem Kategoriensystem abgebildet werden. In einem zweiten induktiven Auswertungsschritt werden die Themen anhand von (Sub-)Kategorien ausdifferenziert, anhand derer sich zentrale Aussagen mit Ankerbeispielen ableiten lassen. Um die Qualität der Analyse zu gewährleisten, wird die Datenanalyse und -interpretation durch mindestens zwei Personen unabhängig voneinander durchgeführt. Mögliche Unstimmigkeiten werden gemeinsam diskutiert und konsentiert. Für die qualitative Analyse wird eine Software eingesetzt.

9.3.3 Einbindung externer Expertengruppen

Begriffsdefinition „Expertin/Experte“

Für den Begriff „Expertin/Experte“ gibt es keine allgemein konsentierte, 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 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, Datenschutzbeauf-

²⁸ 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.

tragte, 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 herangezogen. 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
- Beratung zum besonderen Handlungsbedarf auf Bundesebene
- Beratung zu möglichem Anpassungs- und Entwicklungsbedarf der Qualitätsindikatoren
- Präsentation der Verfahrensergebnisse in Fachkreisen und der 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“. Im vorliegenden Abschnitt wird auf die übergreifenden Aspekte und deren methodische Begründung eingegangen.

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 9.2). 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 Experten-Gruppe). 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 können es Erfahrungen mit der infrage stehenden Erkrankung oder medizinischen Versorgung sein. Bei Expertinnen und Experten aus anderen Institutionen und Fachgebieten können beispielsweise bestimmte Kenntnisse 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 Interessen, 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 Expertinnen und Experten auch nur zulasten der Fachlichkeit gelingen würde (Roters in: Kasseler Kommentar, § 137a, Rdnr. 36).

Es wird geprüft, ob die angegebenen Interessen zu relevanten Interessenkonflikten für die vorgesehenen Beratungsthemen führen 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 Interessenkonflikten erfolgt durch die Selbstauskunft der als Expertinnen und Experten einzubeziehenden Personen. Hierzu verwendet das Institut ein Formblatt zur Offenlegung von Interessen, wie es bereits seit vielen Jahren bei der Entwicklung und Durchführung von QS-Verfahren 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 Expertinnen und Experten sowie eine Zusammenfassung der Angaben zu den 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 Zusammensetzung 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 9.2). Diese Bewertungen sollen so weit 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 9.3.3 ab S. 134).

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, Stelfox und Straus 2013, Campbell et al. 2004), 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 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 Neu-berufung 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 (Boukadedid et al. 2011, Mullen 2003, Waggoner et al. 2016, Murphy et al. 1998). 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 eine anonyme Abgabe von Bewertungen durch die Gruppenmitglieder gruppenspezifisch oder psychologisch bedingte Einflüsse auf das Bewertungsergebnis verringern kann (Fitzpatrick 1989). Der Bewertungsprozess kann daher beispielsweise in Form eines Delphi-Verfahrens (Mullen 2003) durchgeführt werden, mit der Modifikation, dass postalische und persönliche Bewertungsrounds 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 Bewertungsrounds 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.

10 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 diesen Zweck geeignet sind. Dieses Kapitel gibt einen Überblick über die grundsätzlichen Kriterien, anhand derer die Eignung der Qualitätsindikatoren für die Qualitätsmessung in der Gesundheitsversorgung beurteilt wird. Das konkrete Vorgehen bei der Entwicklung von Qualitätsmessungen ist in Kapitel 6 beschrieben.

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. Reiter et al. 2007, AQUA 2015, NQF 2015, Mainz 2003, de Koning 2007, Geraedts et al. 2003, Jones et al. 2014, 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 et al. 2007). 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“ (Jones et al. 2014, AQUA 2015), 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 Autorinnen und 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, Geraedts et al. 2017: e103). Es wurde daher unter Berücksichtigung der Literatur eine eigene Systematik für die Eignungsbeurteilung von Qualitätsindikatoren bzw. Qualitätsmessungen entwickelt.

10.1 Grundsätze zur Eignungsbeurteilung

Ob ein Qualitätsindikator für ein bestimmtes QS-Verfahren geeignet ist, bestimmt sich danach, inwieweit er dem Verwendungszweck der Qualitätsmessung (Qualitätsförderung, Information, selektivvertragliche Anreize, Planung der Versorgungsstrukturen, Qualitätszu- und -abschläge; siehe Kapitel 4) dient. Die Eignung von Qualitätsindikatoren ist nicht nur von der Konstruktion

des Indikators (Qualitätsmerkmal, Rechenregeln etc.), sondern auch von der konkreten Datengrundlage, der betrachteten Population, den Rahmenbedingungen der Qualitätsmessung und -bewertung sowie der Verwendung der Qualitätsmessung abhängig. So können etwa potenzielle unerwünschte Wirkungen bei Verfahren zur Qualitätsförderung durch externen Vergleich und bei Verfahren mit Qualitätszu- und -abschlägen unterschiedlich sein. Ebenso erfordern Indikatoren, die zur Qualitätsförderung durch externen Vergleich eingesetzt werden andere Bewertungskonzepte als Indikatoren, die beispielsweise zwischen guter Qualität und außerordentlich guter Qualität unterscheiden sollen. 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. Ein Qualitätsindikator kann folglich für unterschiedliche Anwendungszwecke auch unterschiedlich gut geeignet sein (Meyer 2004: 24).

Durch die Eignungsbeurteilung der Qualitätsindikatoren soll für alle Komponenten eines Qualitätsindikators (Qualitätsziel, Messverfahren, Bewertungskonzept; siehe Abschnitt 5.1) eine Aussage ermöglicht werden, inwieweit diese für den Zweck eines bestimmten QS-Verfahrens als geeignet eingeschätzt werden.

Die Anwendung der Eignungskriterien erfolgt je nach Anwendungszeitpunkt auf verschiedene Weisen:

- prospektiv als Leitfaden für die Neuentwicklung von Qualitätsmessungen; da die Neuentwicklung in mehreren Schritten von den Qualitätsaspekten als theoretisches Konzept über die konkreten Qualitätsmerkmale bis hin zu den Qualitätsindikatoren als konkrete Messinstrumente erfolgt, beziehen sich einige der in den folgenden Abschnitten dargestellten Eignungskriterien nicht nur auf den Qualitätsindikator im engeren Sinne, sondern auch auf Qualitätsaspekte oder Qualitätsmerkmale; aus Gründen der Lesbarkeit wird in der folgenden Darstellung vereinfachend von Eignung bzw. Eignungskriterien der Qualitätsindikatoren gesprochen
- 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; dabei ist die Eignungsbeurteilung nur dann adäquat möglich, wenn nicht allein einzelne Qualitätsindikatoren, sondern auch deren Auswahl, Zusammenspiel und Verwendung im Set betrachtet werden

Bei der Beurteilung der Eignung von Qualitätsindikatoren unterscheidet das IQTIG vier Kategorien, wobei drei davon den Komponenten der Qualitätsindikatoren (Qualitätsziel, Messverfahren, Bewertungskonzept; siehe Abschnitt 5.1) entsprechen und die vierte Kategorie Eignungskriterien enthält, die sich komponentenübergreifend auf den Qualitätsindikator als Ganzes beziehen (siehe Abbildung 10).

Eignung			
des Qualitätsziels	des Messverfahrens	des Bewertungskonzepts	übergreifende Kriterien
Bedeutung für die Patientinnen und Patienten	Objektivität der Messung	Angemessenheit d. Referenzbereichs	Brauchbarkeit
Zusammenhang mit patientenrelevantem Ziel	Datenqualität	Klassifikationsgüte	Berücksichtigung unerwünschter Wirkungen
Potenzial zur Verbesserung	Reliabilität der Messung	Angemessenheit d. Risikoadjustierung	Zuschreibbarkeit der Verantwortung
Übereinstimmung mit Qualitätsaspekt	Validität der Messung		
Beeinflussbarkeit	Praktikabilität der Messung		

Abbildung 10: Überblick über die Eignungskriterien für Qualitätsindikatoren

Die Einschätzung hinsichtlich der Erfüllung jedes Eignungskriteriums durch einen Qualitätsindikator erfolgt grundsätzlich durch das Institut in Form qualitativer Beurteilung, Einordnung und Zusammenschau der verfügbaren Informationen. Denn mit Ausnahmen der Reliabilität der Messung und der Klassifikationsgüte kann der Erfüllungsgrad der Kriterien nicht sinnvoll in Form einer Kennzahl quantifiziert werden. Beispielsweise lässt sich das Ausmaß, zu dem Indikatorwerte durch die Leistungserbringer beeinflussbar sind, nicht quantifizieren. Die Beurteilung, inwieweit ein Kriterium auch im Vergleich zu anderen Indikatoren angemessen erfüllt ist, erfolgt daher durch eine qualitative Einschätzung unter Zuhilfenahme ergänzender Informationen wie Literatur und externer Expertise. Für die Einholung externer Expertise können Gruppenverfahren nach Kapitel 9 zum Einsatz kommen. Da es sich bei den Eignungskriterien größtenteils um qualitative Kriterien handelt, kann ihre Erfüllung nicht, wie etwa bei quantitativen Größen, über das Erreichen eines Schwellenwerts definiert werden. Ein einzelnes Eignungskriterium wird vielmehr als angemessen erfüllt angesehen, wenn in der Gesamtbetrachtung der Informationen dazu die positiven Belege und Hinweise gegenüber möglichen gegenteiligen Hinweisen überwiegen.

Die Gesamtbeurteilung der Indikatoreignung erfolgt in Form einer qualitativen Abwägung der Erfüllung der Einzelkriterien, die in Form eines Indikatorprofils dargestellt werden. Die Auswahl von Indikatoren erfolgt auf der Basis möglichst reichhaltiger Informationen (z. B. Literatur, Expertengespräche) durch Abwägung von Vor- und Nachteilen, die für oder gegen einen Indikator sprechen, und lässt sich nicht sinnvoll durch einen festen Algorithmus treffen (Meyer 2004: 40). Aufgrund der qualitativen Natur der Eignungskriterien wird auch von einer numerischen Ge-

wichtung der Kriterien untereinander abgesehen. Transparenz über das entwickelte QS-Verfahren und die Indikatoren entsteht somit durch die nachvollziehbare Darlegung der Gründe, weshalb das IQTIG einen Indikator für geeignet hält.

Im Folgenden werden die Eignungskriterien in abstrakter Form als grundlegende Anforderungen an Qualitätsmessungen beschrieben. Die Operationalisierung dieser Kriterien erfolgt je nach Anwendungszeitpunkt (in der Entwicklung von Qualitätsmessungen vs. für bestehende Qualitätsindikatoren) unterschiedlich und ist nicht Gegenstand dieses Kapitels.

10.2 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.

10.2.1 Bedeutung für die Patientinnen und Patienten

Anhand dieses Kriteriums wird beurteilt, wie bedeutend ein Qualitätsaspekt, ein Qualitätsmerkmal bzw. das Qualitätsziel eines Indikators für die Patientinnen und Patienten ist. Qualitätsaspekte, -merkmale und Indikatorziele sind dann patientenrelevant, wenn sie mindestens einer Qualitätsdimension des Rahmenkonzepts (Abschnitt 1.1) zuordenbar sind und sich unmittelbar oder mittelbar auf Endpunkte beziehen, die für Patientinnen und Patienten erfahrbar sind. Beispiele hierfür sind:

- Sterblichkeit
- Krankheitsbeschwerden und -folgen
- gesundheitsbezogene Lebensqualität
- Behandlung der Patientinnen und Patienten mit Respekt
- emotionale Unterstützung
- Information der Patientinnen und Patienten
- Partizipation der Patientinnen und Patienten

Anzustreben ist, dass das jeweils adressierte Qualitätsthema für die Erfüllung der Bedarfe und Bedürfnisse von Patientinnen und Patienten durch die Gesundheitsversorgung eine hohe Bedeutung besitzt. Das Ausmaß der Bedeutung, die ein Qualitätsaspekt oder -merkmal für die Patientinnen und Patienten hat, hängt beispielsweise von Ausprägung und Dauer einer Erkrankung ab, von dem Ausmaß der Verkürzung oder Verlängerung von Lebensdauer oder Krankheitsdauer, von dem Risiko für das Eintreten einer Behandlungskomplikation oder von der Dringlichkeit einer Versorgungsmaßnahme (z. B. Erhalt eines Spenderorgans zur Transplantation). Indikatorziele, die nicht unmittelbar patientenrelevant in diesem Sinn sind, eignen sich nur für die Qualitätssicherung, wenn der Zusammenhang mit einem unmittelbar patientenrelevanten Ziel ausreichend belegt ist (siehe Abschnitt 10.2.2).

Welche Aspekte Bedeutung für Patientinnen und Patienten haben, muss in jedem Themenbereich gezielt ermittelt werden (Mühlhauser und Müller 2009: 36 f.). In den Entwicklungsprojekten des IQTIG sind die Informationsgrundlage hierfür insbesondere die Ergebnisse der Literaturrecherche und -analyse, eigener qualitativer Erhebungen wie Fokusgruppen mit Patientinnen und Patienten sowie mit Vertreterinnen und Vertretern von Gesundheitsprofessionen und ergänzend Hinweise von Expertinnen und Experten einschließlich der Patientenvertretung. Die Beurteilung der Bedeutung für die Patientinnen und Patienten erfolgt zum einen in der Konzeptstudie bezogen auf die ermittelten Qualitätsaspekte, zum anderen im Prozess der Indikatorenentwicklung bezogen auf die Qualitätsmerkmale. Die Notwendigkeit einer mehrmaligen Beurteilung ergibt sich aus der zunehmenden Konkretheit in jedem Entwicklungsschritt. Es soll sichergestellt werden, dass das konkrete Qualitätsmerkmal und der darauf basierende Indikator (z. B. „Anteil von Patientinnen und Patienten mit 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. „Dokumentation von Komplikationen in der Patientenakte“) 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.

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 Konstrukt durch diesen Indikator gemessen werden soll, und dieses einer entsprechenden Beurteilung unterzogen werden.

10.2.2 Zusammenhang mit einem unmittelbar patientenrelevanten Ziel

So wie bei der Gesundheitsversorgung zwischen einem unmittelbaren und einem mittelbaren Nutzen für die Patientinnen und Patienten unterschieden werden kann (Porzsolt und Gaus 1993), so gibt es auch unmittelbar und mittelbar für Patientinnen und Patienten relevante Qualitätsmerkmale und Qualitätsziele eines Indikators (siehe Tabelle 10). Unmittelbar patientenrelevante Indikatorziele liegen bei Qualitätsindikatoren vor, die ein Qualitätsmerkmal messen, das gemäß dem Eignungskriterium „Bedeutung für die Patientinnen und Patienten“ *per se* Bedeutung für die Patientinnen und Patienten hat und das in der Regel unmittelbar von diesen berichtet oder erfahren werden kann, wie z. B. Sterblichkeit, Komplikationen einer Operation oder Aufklärung über Behandlungsalternativen. Unmittelbar patientenrelevante Qualitätsmerkmale basieren auf ethischen Überlegungen und sozialem Konsens. Im Sinne einer patientenzentrierten Qualitätssicherung (siehe Kapitel 3) müssen geeignete Qualitätsmessungen entweder ein unmittelbar patientenrelevantes Qualitätsziel adressieren oder es muss ein überzeugender Nachweis vorliegen für den Zusammenhang des mittelbar relevanten Qualitätsziels mit einem für die Patientinnen und Patienten unmittelbar relevanten Qualitätsziel. 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 patientenrelevanten Ergebnisses erhöhen.

Tabelle 10: Beispiele für unmittelbar und mittelbar patientenrelevante Indikatorziele

	unmittelbar patientenrelevantes Indikatorziel	mittelbar patientenrelevantes Indikatorziel
Strukturen	barrierefreier Zugang	Desinfektionsmittelspender in Patientenzimmer
Prozesse	partizipative Entscheidungsfindung	Antibiotikaprophylaxe bei Operation
Ergebnisse	niedrige Sterblichkeit	CRP-Verringerung 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.²⁹ 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. Falls keine ausreichende externe Evidenz für diesen Zusammenhang vorliegt, kann auf eine konsentiertere 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 (siehe Grundsätze zur Eignungsbeurteilung in Abschnitt 10.1). Die Bewertung von Qualität und Konsistenz der Evidenz wird in der Beschreibung der Indikatoren dokumentiert.

10.2.3 Potenzial zur Verbesserung

Die Wichtigkeit einer Qualitätsmessung bemisst sich neben der Bedeutung für die Patientinnen und Patienten auch nach den möglichen Effekten.³⁰ 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 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 Patientinnen und Patienten haben, aber deren Messung keine nützlichen Zusatzinformationen bietet, da hohe Versorgungsqualität über alle Leistungserbringer hinweg vorliegt. Umgekehrt kann Verbesserungspotenzial bei Merkmalen bestehen, die aus Patientensicht wenig bedeutsam sind. Aus Aufwand-Nutzen-Gründen ist die Einführung oder

²⁹ Für Struktur- und Prozessindikatoren entspricht dieses Eignungskriterium dem im QUALIFY-Instrument als Indikatorevidenz bezeichneten Kriterium (Reiter et al. 2007).

³⁰ Es wird also – anders als z. B. im AQUA-Verfahren (AQUA 2015) und deutlicher als im QUALIFY-Instrument (Reiter et al. 2007) – bei der Beurteilung zwischen Patientenrelevanz und Verbesserungspotenzial getrennt.

Beibehaltung von Indikatoren zu vermeiden, deren Nützlichkeit begrenzt ist oder deren Ergebnisse eine Sättigungsgrenze erreicht haben (Meyer 2004: 39, NQF 2015: 44 ff.). 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 Hinweisen auf

- eine hohe Variation in der Versorgung,
- Qualitätsdefizite oder Fehlentwicklungen oder
- Über-, Unter- oder Fehlversorgung.

Ist ein Verbesserungspotenzial gegeben, so kann dessen Ausmaß beispielsweise anhand der Zahl betroffener Patientinnen und Patienten, anhand der Kosten für nicht indizierte Behandlungen, die dann für andere Maßnahmen nicht zur Verfügung stehen, oder anhand der Zahl der Leistungserbringer mit Verbesserungsbedarf näher eingeschätzt werden.

10.2.4 Übereinstimmung mit einem Qualitätsaspekt

In der Entwicklung, Pflege und Weiterentwicklung von QS-Verfahren werden aus den Qualitätsaspekten des Qualitätsmodells Qualitätsmerkmale abgeleitet und in Form von Qualitätsindikatoren operationalisiert (siehe Kapitel 6 bzw. Abschnitt 7.2). Dabei soll sichergestellt werden, dass das mögliche Qualitätsmerkmal (z. B. „Transfusionspflichtige postoperative Blutung“) in Bezug zu einem Qualitätsaspekt des Qualitätsmodells (z. B. „Komplikationen der Behandlung“) steht. So soll die Entwicklung von Indikatoren vermieden werden, die Aspekte adressieren, die nicht Teil des Qualitätsmodells sind. Gleichzeitig soll damit sichergestellt werden, dass jeder Qualitätsaspekt des Qualitätsmodells durch mindestens einen Qualitätsindikator abgebildet wird.³¹ Die Beurteilung der Übereinstimmung erfolgt durch sachlogische Überlegungen zum Versorgungsgeschehen.

10.2.5 Beeinflussbarkeit durch den Leistungserbringer

In der externen Qualitätssicherung werden Anforderungen an die Leistungserbringer formuliert und anhand der Qualitätsmerkmale und Qualitätsindikatoren dargestellt (siehe Abschnitt 5.2). Ein Indikator beschreibt jedoch nur dann leistungserbringerbezogene Qualität, wenn die Beeinflussung des von ihm gemessenen Qualitätsmerkmals auch innerhalb der Handlungsmöglichkeiten des Leistungserbringers liegt.

³¹ Das Ausmaß, in dem alle Qualitätsaspekte des Qualitätsmodells durch das Qualitätsindikatorenset eines QS-Verfahrens abgebildet werden, kann als Inhaltsvalidität des Indikatorensets für die Abbildung des Qualitätsmodells verstanden werden.

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 patientenseitigen und sonstigen Faktoren sowie von Wechselwirkungen zwischen diesen Faktoren (siehe auch Abschnitt 15.1). Der Anteil des Einflusses, den die Leistungserbringer auf die Indikatorwerte haben, wird als Beeinflussbarkeit der Indikатораusprägung bezeichnet und ist bei verschiedenen Indikatoren unterschiedlich groß. Sowohl bei Ergebnisindikatoren (z. B. die Sterblichkeit von Organempfängerinnen und -empfängern während des Krankenhausaufenthalts im Rahmen der Transplantation) als auch bei einigen Prozessindikatoren (z. B. die Eingriffsdauer beim Wechsel eines Herzschrittmacher- oder Defibrillatoraggregats) haben patientenseitige Faktoren einen relevanten Einfluss auf den Indikatorwert (Abbildung 11). Für andere Prozess- und Ergebnisindikatoren (z. B. Eingehen auf Informationsbedürfnis der Patientinnen und Patienten, Tod bei elektiver Operation ohne Vorerkrankungen) kann dagegen bei ausreichend spezifischer Formulierung des Indikators unterstellt werden, dass patientenseitige 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. Antibiotikaprophylaxe), bei denen der Indikatorwert in stärkerem Maße sowohl von patientenseitigen oder sonstigen Faktoren als auch vom Handeln des Leistungserbringers beeinflusst ist.

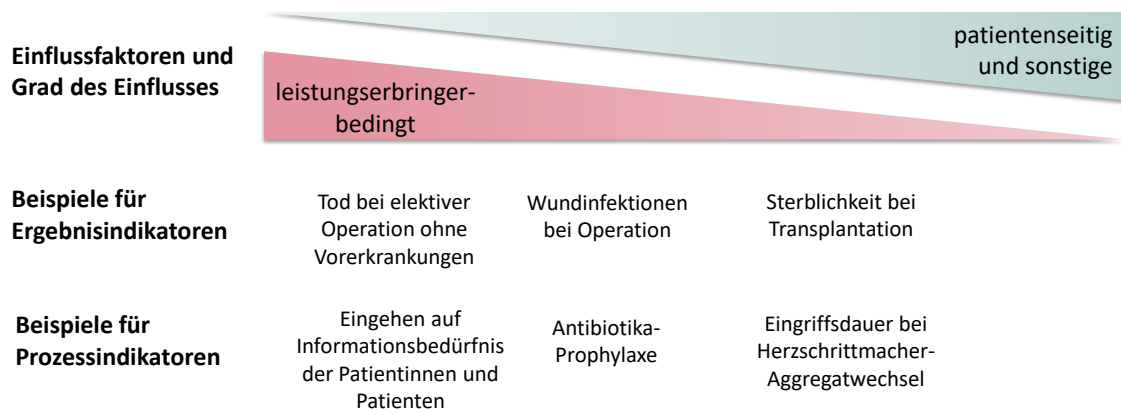


Abbildung 11: Schematische Darstellung des Ausmaßes der leistungserbringerbedingten und nicht leistungserbringerbedingten Einflüsse auf den Indikatorwert anhand von Beispielen fallbezogener Qualitätsindikatoren

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 12). Konkrete Einflüsse von Rahmenbedingungen auf die Erfüllbarkeit von Qualitätsanforderungen werden ebenfalls bei der Festlegung des Bewertungskonzepts berücksichtigt und durch das Kriterium „Angemessenheit des Referenzbereichs“ adressiert (siehe Abschnitt 10.4.1).

10.3 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.

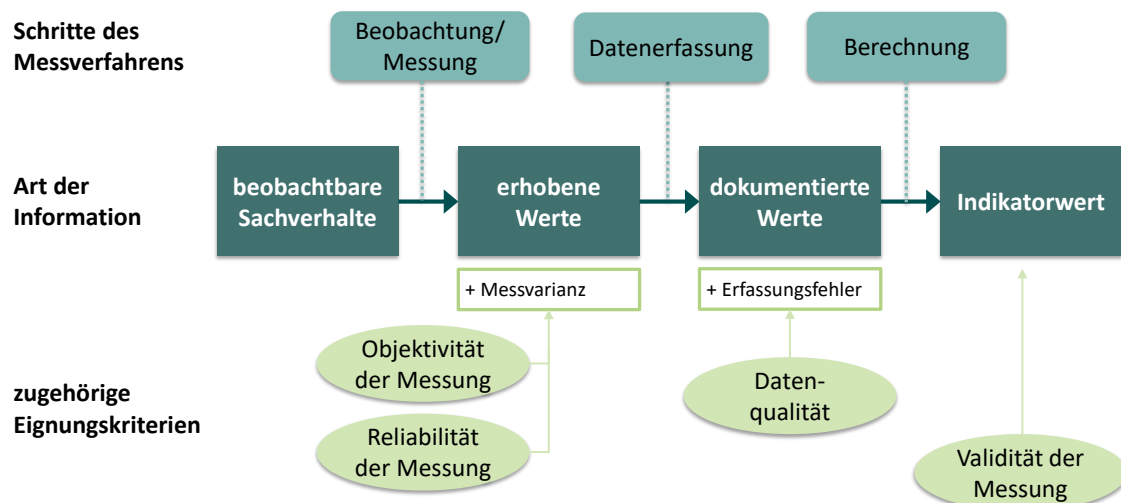


Abbildung 12: Schritte und zugehörige Eignungskriterien bei der Ermittlung und Aufarbeitung von Informationen für die Berechnung von Indikatorwerten

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 12). Der erste Schritt besteht in der Beobachtung eines Sachverhalts und Erhebung eines Wertes für diesen Sachverhalt auf Patientenebene, wie z. B. in Form einer Bestimmung des Blutdrucks oder einer Befragung nach Schmerzen. Dieser Vorgang der Zuordnung von Zahlenwerten 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 Abrech-

nungszwecken und nachfolgender Übermittlung von den Krankenkassen an die Auswertungsstelle für die Qualitätssicherung.³² Bei Patientenbefragungen für die Qualitätssicherung fallen Messung auf Patientenebene und Datenerfassung zusammen, d. h., beides erfolgt gleichzeitig beim Ausfüllen des Fragebogens. Der dritte Schritt besteht schließlich in der Berechnung des Indikatorwerts gemäß den Festlegungen zu den verwendeten Datenfeldern, Dateninhalten, Rechenregeln und Risikoadjustierungsmethoden, kurz als Berechnungsspezifikation des Indikators bezeichnet.

Bei den Schritten des Messverfahrens einschließlich der Berechnung des Indikatorwerts existieren 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.

10.3.1 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 Befragung von Einrichtungen 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 6.4). 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 6.4.1), und eindeutige Spezifikationen zum Erhebungsprozess einschließlich leicht verständlicher und eindeutiger Ausfüllhinweise erstellt.

³² Genau genommen besteht hier also die Datenerfassung ihrerseits aus zwei Schritten.

10.3.2 Datenqualität

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 Datenqualität 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 Datenqualität sind damit Spezifikationen der Instrumente zur Datenerfassung (QS-Filter, QS-Dokumentationssoftware, Sozialdatenspezifikation, Patientenfragebogen) und der zugehörigen Datenflüsse, die eindeutig sind und die eine Überprüfung der Vorgaben ermöglichen.

Formal kann die Datenqualität als Teilaspekt der Durchführungsobjektivität auf Ebene der dokumentierenden Einrichtungen aufgefasst werden: Einschränkungen der Datenqualität 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 14).

10.3.3 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 (Bühner 2011: 60, Schermelleh-Engel und Werner 2012: 120, Meyer 2004: 28, Adams et al. 2010). 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 eine Patientin / ein Patient verstorben ist. Hier erscheint es nicht sinnvoll anzunehmen, dass es zu Messfehlern kommt, und eine solche Trennung von wahren und beobachtetem Wert vorzunehmen.³³ 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

³³ 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 „Datenqualität“ adressiert wird.

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 12.4), kann die Reliabilität, auch auf Leistungserbringerebene, nur geschätzt werden. Entsprechende Verfahren sind Gegenstand zukünftiger Entwicklungen.

10.3.4 Validität der Messung

Indikatoren machen theoretische Konzepte durch beobachtbare Sachverhalte messbar (Meyer 2004: 28 und 51, Döring und Bortz 2016: 223 und 228, Schnell et al. 2013: 117 und 221). Damit 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 diese unterschiedlichen Facetten des Validitätsbegriffs mittels verschiedener Eignungskriterien, um sie transparenter zu machen. Validität meint hier also kein Gesamturteil über die Eignung eines Qualitätsindikators. Ein Gesamturteil über die Eignung eines Indikators erlaubt in diesem Konzept vielmehr die Gesamtbetrachtung *aller* Eignungskriterien. Das Eignungskriterium „Validität der Messung“ wird in Anlehnung an Schnell et al. (2013: 144), Hartig et al. (2012: 144), Reiter et al. (2007) definiert als das Ausmaß, in dem ein Qualitätsindikator das, was er abbilden soll, tatsächlich erfasst. Die Messung, die einem Qualitätsindikator zugrunde liegt, ist demnach umso valider, je eher sie das interessierende Merkmal abbildet, das durch die Indikatorbezeichnung beschrieben wird. Beispielsweise ist der Indikator „Sterblichkeit während des Krankenhausaufenthalts 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 dadurch bestimmt, dass die gewählte Indikatorbezeichnung, die ausgewählten Datenfelder und die Rechenregeln angemessen und korrekt spezifiziert sind.

Die Validität der Messung wird bei der Neuentwicklung von Qualitätsindikatoren und bei der Überprüfung von Qualitätsindikatoren im Regelbetrieb in erster Linie mittels sachlogischer Überlegungen eingeschätzt. Dabei werden Literatur, externe Expertinnen und Experten sowie eigene Analysen und Erfahrung zu den verfügbaren Datenquellen einbezogen, um die am besten

geeigneten Datenfelder und operationale Definitionen 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 6.4.3 verwiesen.

10.3.5 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 Qualitätsindikators 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 12). 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 in der Regel bereits im Rahmen der Patientenbehandlung erhoben und in der Patientenakte dokumentiert wurden. Der Aufwand entsteht bei den Datensätzen der QS-Dokumentation also vor allem 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.

10.4 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 12). Dieses Bewertungskonzept beinhaltet eine umschriebene Menge an Kategorien, die unterschiedliche Qualitätsniveaus repräsentieren (z. B. außerordentlich gute Qualität, unauffällige Qualität, unzureichende Qualität), numerische Grenzwerte für diese Kategorien und eine Zuordnungsregel, nach der die Indikatorergebnisse den Kategorien zugeordnet

werden. Die Eignung des Bewertungskonzepts eines Indikators im Rahmen eines QS-Verfahrens lässt sich zusammenfassend anhand der drei nachfolgend beschriebenen Kriterien beurteilen.

10.4.1 Angemessenheit des Referenzbereichs

Damit Indikatorwerte sinnvoll interpretiert und mit Handlungsempfehlungen verknüpft werden können, ist nicht nur die korrekte Messung, sondern gleichermaßen auch eine angemessene Wahl des Referenzbereichs wichtig, mit dem der Indikatorwert verglichen wird (Goodwin 1996). Dazu ist es erforderlich, dass die Kategorien, in die ein Indikatorwert klassifiziert wird, klar definiert werden (Berk 1996). Die Angemessenheit eines Referenzbereichs bemisst sich danach, ob er das gewünschte Niveau der Versorgungsqualität, z. B. unzureichende Qualität, repräsentiert.

Dazu wird zusätzlich zur grundsätzlichen Beeinflussbarkeit des Indikatorwerts durch die Leistungserbringer (siehe Abschnitt 10.2) geprüft, ob die konkrete Zielvorgabe des Referenzbereichs durch einen Leistungserbringer (wenigstens unter optimalen Bedingungen) überhaupt erreichbar ist und damit begründet gefordert werden kann. Zu dieser Prüfung gehört auch eine Einschätzung, wie viele Leistungserbringer die Anforderung voraussichtlich erreichen (Norcini und Shea 1997, Berk 1996) und ob es plausibel ist, wenn die Zielvorgabe für das jeweilige Qualitätsniveau von sehr vielen oder sehr wenigen Leistungserbringern erreicht wird. Stünde schon vor der Messung fest, dass keines der Ergebnisse im Referenzbereich liegen würde, würde sich die Klassifikation mittels des Indikators erübrigen. 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 auch bei optimaler Qualität nicht erreicht werden kann. Somit könnte der Indikator nicht zwischen Leistungserbringern mit erwartbarer Qualität und solchen unterhalb erwartbarer Qualität differenzieren und wäre für diese Klassifikation wegen fehlenden Informationsgehalts ungeeignet.

Unterschiedliche Verwendungszwecke können unterschiedliche Qualitätsniveaus adressieren und damit unterschiedliche Referenzbereiche erfordern. Beispielsweise ist es sinnvoll, den Indikatorwert eines Leistungserbringers in Verfahren zur Qualitätsförderung mittels externen Vergleichs anders zu bewerten als hinsichtlich der Frage, ob auf Grundlage des Indikatorwerts ein qualitätsbezogener Vergütungszuschlag gezahlt werden soll. Je nach Zielsetzung und Kontext des QS-Verfahrens können Indikatorwerte zudem für andere Zwecke als den Vergleich mit einem Referenzbereich genutzt werden, z. B. für den Vergleich der Leistungserbringer untereinander. Für diese Zwecke sind die Lage und die Erreichbarkeit des Referenzbereichs nicht von Belang. Details zu den Methoden zur Festlegung von Bewertungskonzepten und Referenzbereichen sind in Kapitel 12 dargestellt.

10.4.2 Klassifikationsgüte

Für Prozess- und Ergebnisindikatoren, bei denen die Fallzahl des Leistungserbringers bei der Bewertung des gemessenen Indikatorwerts zu berücksichtigen ist (siehe Abschnitt 15.1), 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 12.4). 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.³⁴

10.4.3 Angemessenheit der Risikoadjustierung

Damit die Klassifikation der Indikatorwerte verschiedener Leistungserbringer (z. B. der Vergleich mit einem Referenzbereich) sowie der Vergleich der Indikatorwerte zwischen verschiedenen 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 15.2). Das Kriterium beschreibt somit das Ausmaß, in dem patientenseitige Einflussfaktoren auf die Indikatorwerte durch Risikoadjustierung kontrolliert sind.

10.5 Übergreifende Eignungskriterien

In die vierte Kategorie von Eignungskriterien für Qualitätsindikatoren fallen Anforderungen, die nicht nur auf eine einzelne Komponente des Indikators (Qualitätsziel, Messverfahren, Bewertungskonzept) bezogen sind.

10.5.1 Brauchbarkeit

Dieses Kriterium fasst eine Reihe von Aspekten zusammen, die beschreiben, wie gut die Ergebnisse der Qualitätsmessung in der praktischen Anwendung für die Ableitung von Entscheidungen genutzt werden können (McGlynn 2003, NQF 2015: 20), vorausgesetzt, dass Qualitätsziel, Messverfahren und Bewertungskonzept geeignet sind.³⁵ Je nach Verwendungszweck (siehe Kapitel 4) können für die Brauchbarkeit der Indikatoren verschiedene Aspekte von Bedeutung sein. Für die Qualitätsförderung mittels Feedback ist wichtig, wie verständlich die Aussage des Indikators für die Leistungserbringer ist und wie gut sich aus den Indikatorergebnissen Hinweise auf konkrete Maßnahmen für das interne Qualitätsmanagement ableiten lassen. Bei Qualitätsmessungen

³⁴ 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.

³⁵ Das IQTIG unterscheidet bei der Beurteilung des potenziellen Nutzens einer Qualitätsmessung zwischen der Bedeutung des Qualitätsziels, dem Verbesserungspotenzial in Bezug auf dieses Qualitätsziel und der Brauchbarkeit für die konkrete Anwendung.

zum Zweck der Information von Patientinnen und Patienten bzw. der Öffentlichkeit soll dagegen vor allem beurteilt werden, ob die berichteten Indikatorwerte eine Information zur Verfügung stellen, die hilfreich ist und beispielsweise die Auswahl eines Leistungserbringers durch Patientinnen und Patienten unterstützen kann. Für jeden Verwendungszweck sollte der Prozess der Qualitätsmessung daraufhin beurteilt werden, wie zeitnah die Indikatorwerte als Grundlage für Handlungsempfehlungen zur Verfügung stehen. Beispielsweise könnten zwei Indikatoren A und B die Sterblichkeit während des Krankenhausaufenthalts 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 Beurteilung der Brauchbarkeit erfolgt nicht nur für einzelne Indikatoren, sondern teilweise auch in Zusammenschau der Indikatoren eines QS-Verfahrens (d. h. für ein Indikatorenset). 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, also z. B. alle Qualitätsdimensionen des Rahmenkonzepts (siehe Abschnitt 1.1) berücksichtigt.

10.5.2 Berücksichtigung unerwünschter Wirkungen

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 für Komplikationen und ggf. für auffällige Indikatorwerte einhergeht, nicht zu behandeln (Patientenselektion). Die Prüfung dieses Kriteriums dient dazu, geeignete Maßnahmen zum Ausgleich solcher unerwünschten Effekte 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 Expertinnen und Experten. In laufenden QS-Verfahren kommen Rückmeldungen zum Verfahren sowie ggf. eigene Datenauswertungen als weitere Informationsquelle hinzu.

10.5.3 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 Transplantatiere“ 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 im Rahmen der leistungserbringerbezogenen Qualitätssicherung die Verantwortung für einen Indikatorwert einem Leistungserbringer zugewiesen werden kann, ist also nicht nur die Beeinflussbarkeit des Indikatorwerts Voraussetzung (siehe Abschnitt 10.2.5), sondern die Beeinflussung des Indikatorwerts muss im Wesentlichen diesem Leistungserbringer zugeschrieben werden können. Dieses Kriterium ist immer dann relevant, wenn die auf die Qualitätsmessung folgenden Maßnahmen (z. B. gezielte Patienteninformation, Qualitätszu- und -abschläge) auf einzelne konkret benennbare Leistungserbringer ausgerichtet sind. Qualitätsindikatoren mit auf mehrere Leistungserbringer verteilter Verantwortung für die Versorgung können dann eingesetzt werden, wenn eine Gruppe von beteiligten Leistungserbringern gemeinschaftlicher Adressat von Qualitätssicherungsmaßnahmen sein soll (z. B. Accountable Care Organizations in den USA) oder wenn Versorgungsqualität auf regionaler Ebene dargestellt und gefördert werden soll.

11 Veröffentlichung von Indikator- und Kennzahlergebnissen

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 sowie der Ergebnisse weiterer Kennzahlen der einzelnen Leistungserbringer. Im Folgenden wird hierfür auch der international gebräuchliche Begriff „Public Reporting“, d. h. öffentliche Berichterstattung, 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)³⁶. 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 und welche zusätzlichen Kennzahlergebnisse 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 den Beitragszahlerinnen und -zahlern nachkommen (Schwenk und Schmidt-Kaehler 2016, Murphy und Scheu 2010) und verdeutlichen, dass sie eine qualitativ hochwertige und effiziente Versorgung anstreben. In diesem Sinne sieht das Institut eine Veröffentlichung von leistungserbringerbezogenen Indikator- und Kennzahlergebnissen 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 selbstbestimmte Information durch Patientinnen und Patienten 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

³⁶ 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 in der Neufassung vom 16. Mai 2013, zuletzt geändert am 22. November 2018, in Kraft getreten am 30. November 2018. URL: <https://www.g-ba.de/informationen/richtlinien/39/> (abgerufen am 01.04.2019).

sie beratenden Personen sowie die Motivation der Leistungserbringer zur Qualitätsverbesserung (Totten et al. 2012, Schwenk und Schmidt-Kaehler 2016, Geraedts et al. 2009, Berwick et al. 2003). Das Institut versteht die Information von Patientinnen und Patienten als eigenständigen Verwendungszweck von Qualitätsmessungen und strebt an, zukünftig explizit bedarfs- und nutzerorientierte Qualitätsinformationen zu entwickeln, beispielsweise für das Internet-Qualitätsportal des G-BA nach § 137a Absatz 3 Satz 2 Nr. 5 SGB V.

Kriterien zur Veröffentlichung von Qualitätsindikatoren

Systematische Literaturübersichten zeigen, dass es nur wenige Studien zu Wirkungen von Public Reporting gibt und diese eine unterschiedliche Qualität aufweisen (Totten et al. 2012, Fung et al. 2008). 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 Qualitätsziel, Messung und Bewertungskonzept der Qualitätsindikatoren mit hinreichender Sicherheit für die Qualitätsmessung geeignet sind (siehe Kapitel 10). 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 Indikatorergebnisse zu dienen, werden daher nicht vorgenommen. Alle veröffentlichten Informationen sollen stattdessen möglichst allgemeinverständlich dargestellt werden. Eine ergänzende Abschätzung der Verständlichkeit und Nutzerorientierung von Qualitätsinformationen wird aber zukünftig angestrebt, wenn Qualitätsaussagen eigens für den Zweck gezielter Patienteninformation entwickelt oder aufbereitet werden sollen.

Das Institut empfiehlt daher, grundsätzlich alle Qualitätsindikatorergebnisse 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 erhebliche Bedenken bestehen, werden auch nicht zur leistungserbringerbezogenen Veröffentlichung der Ergebnisse empfohlen.

Veröffentlichung der Ergebnisse neu entwickelter Qualitätsindikatoren

Für Qualitätsindikatoren, die neu entwickelt werden, wird die Eignung für das jeweilige QS-Verfahren durch die Berücksichtigung der Eignungskriterien (siehe Kapitel 10) sichergestellt. Teilweise können die Eignungskriterien jedoch erst anhand der Erfahrung aus den ersten Jahren des

Regelbetriebs 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 Aussagekraft des Indikators nicht ausgeschlossen werden können. In beiden Fällen gelten diese Indikatoren als „neu entwickelt“ und ihre Ergebnisse werden vorerst nicht zur Veröffentlichung empfohlen. Bei der wiederkehrenden Beurteilung der Indikatoren im Rahmen der Verfahrenspflege (siehe Abschnitt 7.2) wird individuell geprüft, ob die Eignungskriterien hinreichend sicher als erfüllt gelten und die Indikatorergebnisse damit auch zur Veröffentlichung empfohlen werden können.

Veröffentlichung begleitender Informationen zu Qualitätsindikatoren

Als begleitende Informationen zu den Indikatorergebnissen können nach Beratung mit den jeweiligen fachlichen Expertinnen und Experten ergänzende Texte veröffentlicht werden. Diese erläutern methodische Limitationen und inhaltliche 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.

Veröffentlichung von Kennzahlen

Zusätzlich zu Qualitätsindikatoren werden in der externen Qualitätssicherung weitere Kennzahlen verwendet, die die Versorgung beschreiben, die aber – weil für sie beispielsweise aus methodischen und inhaltlichen Gründen kein Referenzbereich festgelegt ist – zur Bewertung von Versorgungsqualität nicht geeignet sind. Gleichwohl kann es sinnvoll sein, diese Kennzahlen zusammen mit den Indikatorergebnissen zu veröffentlichen, da sie z. B. die Verständlichkeit der Indikatorergebnisse deutlich erhöhen. Weiterhin sind manche Kennzahlen per se für die Information der Öffentlichkeit zur Versorgungsqualität einer Einrichtung informativ. Daher sollen (in den Qualitätsberichten der Krankenhäuser und den Bundes- und Landesauswertungen) nicht nur die Ergebnisse von Qualitätsindikatoren, sondern auch weitere, wichtige Kennzahlen ohne Indikatorfunktion dargestellt werden, die auf den für die Qualitätssicherung erhobenen Daten basieren. Hierzu dient die Festlegung verschiedener Kennzahltypen mit unterschiedlichen Zwecken und Eigenschaften.

Folgende Kennzahltypen decken derzeit das Spektrum erkennbarer Erfordernisse ab:

▪ Kalkulatorische Kennzahlen

Bei der Berechnung eines Indikatorwerts gibt es assoziierte Kennzahlen, die direkt mit dieser Berechnung zusammenhängen und deren Darstellung das Indikatorergebnis besser verstehen lässt. Kalkulatorische Kennzahlen können z. B. die Zwischenergebnisse eines komplexeren Qualitätsindikators, d. h. die Einzelergebnisse eines Index, sein. Weitere Beispiele für kalkulatorische Kennzahlen sind – bei risikoadjustierten Qualitätsindikatoren mit dem Verhältnis O/E – die Anzahl der Fälle n in der Grundgesamtheit, die Anzahl der beobachteten Fälle O , die Anzahl der erwarteten Fälle E sowie die risikoadjustierte Rate der Einrichtung i : $R_i = O_i / E_i$ * (Gesamtrate über alle Fälle hinweg). Mit der Zunahme komplexerer Kalkulationen

zur Ermittlung eines Indikatorwerts wird die Sinnhaftigkeit und Notwendigkeit der ergänzenden Darstellung kalkulatorischer Kennzahlen wachsen, z. B. bei der Darstellung von komplexen Indexergebnissen oder Ergebnissen der Ereigniszeitanalyse.

▪ **Ergänzende Kennzahlen**

Zum Verständnis eines Indikatorwerts kann es hilfreich sein zu analysieren, wie sich dieser zusammensetzt, z. B. welche Ergebnisse in bestimmten Subgruppen der Grundgesamtheit erzielt wurden. So kann es z. B. von Interesse sein, bei einer erhöhten Rate an Wundinfektionen zu betrachten, ob diese Wundinfektionen bei Patientinnen und Patienten ohne Antibiose oder bei bestimmten Behandlungsstrategien gehäuft auftreten oder ob Patientinnen und Patienten bestimmter Risikogruppen häufiger betroffen sind als andere. Solche Analysen können helfen, die Ursachen auffälliger Indikatorwerte zu finden. Im Gegensatz zu den kalkulatorischen Kennzahlen ist die Rechenregel der ergänzenden Kennzahlen nicht bereits in der Rechenregel des Qualitätsindikators enthalten.

▪ **Transparenzkennzahlen**

Transparenzkennzahlen sind für die Beschreibung des Versorgungsgeschehens von Nutzen. Sie erhöhen die Transparenz als zusätzliche Information ohne direkten Bezug zu einem Qualitätsindikator. Künftig könnten z. B. strukturelle Parameter wie Personalausstattung als Transparenzkennzahlen eingesetzt werden.

▪ **Verfahrensspezifische Kennzahlen**

In regulierenden Verfahren der Qualitätssicherung wie z. B. qualitätsorientierte Zu- und -abschläge oder planungsrelevante Qualitätsindikatoren fallen Informationen an, die zwar keine Qualitätsindikatorergebnisse sind, aber zur Ergebnisdarstellung und zur Nachvollziehbarkeit der getroffenen Entscheidungen und Vereinbarungen benötigt werden. So spielt z. B. die wiederholte statistische Auffälligkeit bei planungsrelevanten Qualitätsindikatoren eine wichtige Rolle.

Diese Kennzahltypen erweitern den Informationsgehalt in Auswertungen und Berichten auf Bundes- und Landesebene, sie sind jedoch nicht als eine Beschränkung der Auswertungen auf nur diese Kennzahlen und Qualitätsindikatoren zu verstehen. Für die Einleitung des Strukturier-ten Dialogs bzw. des Stellungnahmeverfahrens haben die Kennzahltypen keine unmittelbare Bedeutung.

Das IQTIG schlägt Kennzahlen der erläuterten Typen als berichts- oder veröffentlichungspflichtig vor. Diese zusätzlichen Kennzahlen sind auf das Notwendige zu beschränken, um eine übersichtliche und prägnante Berichterstattung zu gewährleisten. Dabei gibt es immer einen Ermessensspielraum, der formal nicht geregelt werden kann. Wesentlich für die Entscheidung, berichts- oder veröffentlichungspflichtige Kennzahlen einzusetzen, ist, ob ihre Darstellung in Auswertungen und Qualitätsberichten die Transparenz des Versorgungsgeschehens erhöhen kann.

12 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 10), 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 15). 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.

12.1 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 13).

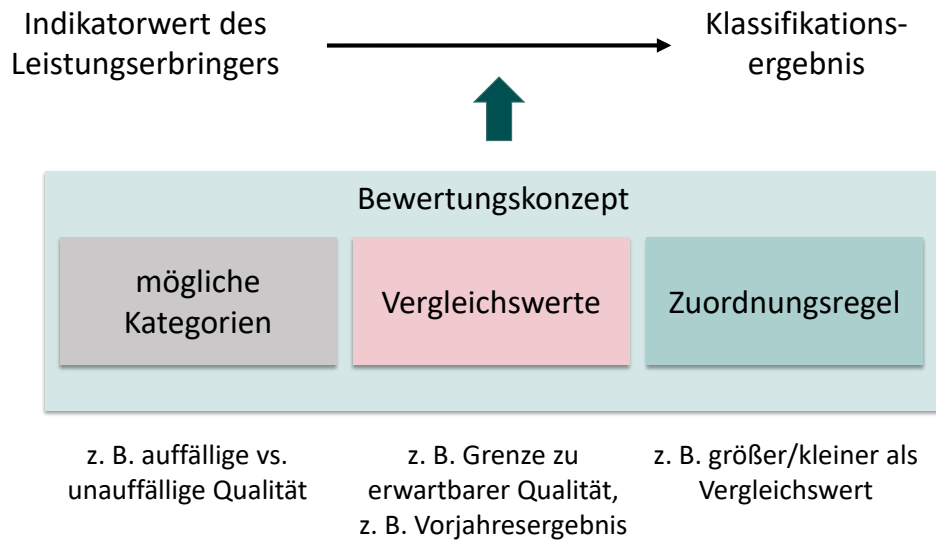


Abbildung 13: Funktion und Bestandteile des Bewertungskonzepts von Qualitätsindikatoren

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“³⁷ Qualität als Grundlage für die Entscheidung, ob ein Strukturierter Dialog bzw. ein Stellungnahmeverfahren als Qualitätssicherungsmaßnahme durchgeführt werden soll oder nicht. Die Konzepte für andere Kategorien (z. B. „außerordentlich gute“ und „unzureichende“ Qualität), wie sie etwa im Rahmen von qualitätsbezogenen Zu- oder Abschlägen benötigt werden, sind derzeit noch nicht in Anwendung (siehe Kapitel 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.).

³⁷ Während in der QSKH-RL der Begriff „rechnerische Auffälligkeit“ Verwendung findet, wird in der DeQS-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 durch externen Vergleich wird in der Regel nur ein einziger Vergleichswert definiert, der die Schwelle zu erwartbarer Qualität im Sinne eines einzuhaltenden Standards beschreiben soll. In diesem Fall gibt es genau einen Referenzbereich je Indikator, der zwischen dem Vergleichswert (einschließlich) und dem Ende der Indikatorkala mit guter Qualität liegt.³⁸ Zukünftige QS-Verfahren werden auch mehrere 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.

12.2 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 (siehe Kapitel 4). Beispielsweise kann es für die Qualitätsförderung durch externen Vergleich 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 Verwendungszweck eingesetzt werden.

Die Spezifikation des konkreten Referenzbereichs und der Zuordnungsregel erfolgt dagegen für jeden Qualitätsindikator spezifisch unter Berücksichtigung der Besonderheiten des Indikators

³⁸ Auch andere Vergleichswerte und Referenzbereiche können sinnvoll sein, wie etwa Vorjahreswerte, wenn auf Veränderungen im Zeitverlauf reagiert werden soll.

(z. B. Messdimension nach Donabedian (2003: 47) und Grad der Beeinflussbarkeit durch die 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 bisherigen Verfahren nach der DeQS-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.

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 7.2).

12.3 Festlegung von Referenzbereichen

Die Referenzbereiche der Qualitätsindikatoren in den Verfahren der Qualitätssicherung nach §§ 136 ff. SGB V mit dem Zweck der Qualitätsförderung durch externen Vergleich 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 10.2.5).

12.3.1 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 7.2) 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 10.2.2), 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?

12.3.2 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 7.2) 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 12.3.1) 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 8): 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 %).

12.4 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 15.1). 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 unbekanntes 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 10.2.5) 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 11 stellt die Entscheidungssituation bei der Feststellung, ob der zugrunde liegende Indikatorwert eines Leistungserbringers innerhalb eines Referenzbereichs liegt, dar.

Tabelle 11: 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 11). 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.³⁹

Das Instrument zur Optimierung von Sensitivität und Spezifität in diesem Kontext ist die Teststärkeanalyse (im Rahmen frequentistischer Statistik). Abbildung 14 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-

³⁹ 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.

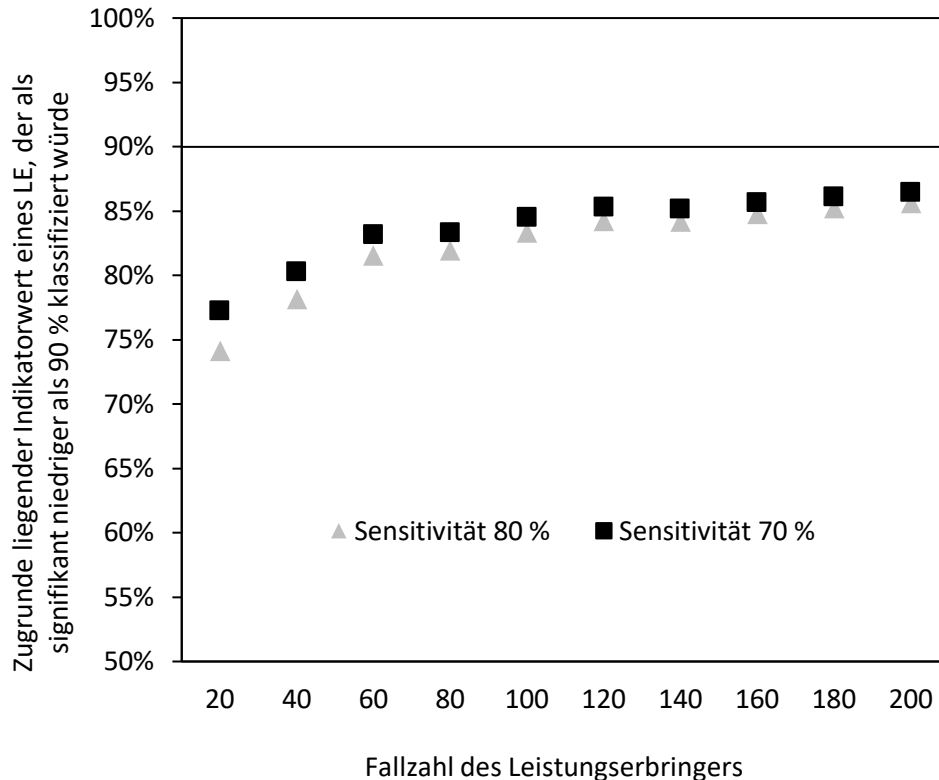


Abbildung 14: Beispielhafte Darstellung der zugrunde liegenden Indikatorwerte (y-Achse), die mit einer Spezifität von 85 % in Abhängigkeit von Fallzahl (x-Achse) und Sensitivität (80 % vs. 70 %) mit einem einseitigen exakten Binomialtest als signifikant verschieden von einem Wert von 90 % klassifiziert werden

Lesebeispiel: Bei einem Leistungserbringer 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 (siehe Kapitel 4) 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.

12.5 Sentinel Events

Sentinel Events sind eine besondere Form von Bewertungskonzept für Prozess- und Ergebnisindikatoren. Dabei wird der Indikatorwert eines Leistungserbringers bereits bei einem einzigen aufgetretenen Ereignis, das der Merkmalsdefinition des Indikators entspricht, in die Kategorie „auffällig“ klassifiziert.⁴⁰ Fallbezogene Einflussfaktoren wie patientenseitige Risikofaktoren und stochastische Einflüsse (siehe Abschnitt 15.1) werden bei diesem Bewertungskonzept nicht berücksichtigt.

Die Begründung für die Wahl von Sentinel Events als Bewertungskonzept ist in der Literatur uneinheitlich. Die Joint Commission beispielsweise bezeichnet Sentinel Events als schwerwiegende, häufig vermeidbare Ereignisse (JCAHO 1989: 332) und versteht diese in neueren Publikationen in erster Linie als Patientensicherheitsereignisse (Joint Commission 2017: 1), also als unerwünschte Ereignisse, die eher auf der Behandlung als auf der Erkrankung beruhen (vgl. Glossar des APS [kein Datum]). Diese Sicht hebt die Vermeidbarkeit der Ereignisse als definierendes Merkmal hervor (ähnlich: Donabedian 2003: 102 f., dort als *clinical aberrance monitoring* bezeichnet, und McLoughlin et al. 2006) und setzt das Auftreten von Sentinel Events mit gravierenden Qualitätsmängeln gleich (Mainz 2003, IGZ 2016). Beispiele für solche Sentinel Events sind Transfusionsreaktionen durch ABO-Inkompatibilität oder das Zurücklassen von Fremdmaterial im Körper bei Eingriffen (Joint Commission 2017: 2). Die Nicht-Berücksichtigung patientenseitiger Risikofaktoren sowie stochastischer Einflüsse ist bei dieser Betrachtungsweise durch die Annahme gerechtfertigt, dass der Indikatorwert weitestgehend durch die Einrichtung beeinflussbar sei und fallbezogene Einflussfaktoren vernachlässigbar seien (vgl. Abschnitt 10.2.5).

In der externen Qualitätssicherung in Deutschland werden Sentinel-Event-Bewertungskonzepte dagegen in erster Linie für Indikatoren zu seltenen Ereignissen genutzt, ohne dass eine weitgehende Beeinflussbarkeit des Indikatorwertes bzw. eine Vermeidbarkeit der Ereignisse durch die Leistungserbringer unterstellt werden kann. Beispiele für solche Indikatoren sind die Sterblichkeit beim Aggregatwechsel von Schrittmachern und die Müttersterblichkeit im Zusammenhang mit der Geburt. Der Verzicht auf die Berücksichtigung patientenseitiger Risikofaktoren begründet sich bei diesen Indikatoren dadurch, dass voraussichtlich nur wenige Ereignisse auftreten, sodass die Erhebung der Risikofaktoren bei allen potenziell gefährdeten Patientinnen und Patienten einen unverhältnismäßig großen Aufwand bedeuten würde. Die Indikatorwerte von Sentinel-Event-Indikatoren erlauben daher ohne ergänzende Beurteilung keine Aussage über die Versorgungsqualität, sondern werden lediglich als Auslöser für nachfolgende, genauere Unter-

⁴⁰ Formal entspricht ein Sentinel Event einem Referenzwert von 0 % und einem einfachen mathematischen Größer-Vergleich als Zuordnungsregel.

suchungen verwendet, beispielsweise den Strukturierten Dialog. Dieses Verständnis von Sentinel Events ist der Trigger-Methodik (Resar et al. 2003, Geraedts et al. 2017: e110, Hibbert et al. 2016) ähnlich, bei der Ereignisse, die häufig mit Patientenschaden korrelieren, in der Behandlungsdokumentation gesucht und z. B. als Auslöser für ausführlichere Aktenreviews genutzt werden. Ziel ist in beiden Fällen, eine Vorauswahl zu treffen, auf deren Grundlage gezielter nach Qualitätsproblemen gesucht werden kann.

13 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 14), 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 Zweck des Verfahrens 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.

13.1 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 15), 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 2010). 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 (Schnell et al. 2013, Lohr 2010). 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 (Schnell et al. 2013, Lohr 2010). 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 von 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). 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 6).

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 14)

13.2 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 werden im Folgenden dargelegt.

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 dem einmaligen Auftreten eines Ereignisses (z. B. Versterben nach Hüft-Endoprothesen-Erstimplantation bei geringer Sterbewahrscheinlichkeit) in jedem Fall nachgegangen werden soll, sind Stichprobenverfahren nicht geeignet, da diese nicht gewährleisten können, jedes Auftreten eines sogenannten Sentinel Events zu erfassen.

13.3 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 2010). 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 15). 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.

14 Datenvalidierung von QS-Dokumentationsdaten und Strukturierter Dialog

Für quantitative Messungen von Versorgungsqualität mithilfe von Qualitätsindikatoren stellt die Qualität der zugrunde liegenden Daten eine wichtige Voraussetzung dar (siehe Abschnitt 10.3.2). Das vorliegende Kapitel beschreibt, mit welchem Verfahren die Qualität 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 Datenqualität 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-RL⁴¹ beschrieben und begründet. Für die sektorenübergreifenden Verfahren nach der Qesü-RL⁴² und die damit verbundenen neuen Datenquellen (einrichtungsbezogene QS-Dokumentation, Patientenbefragung, Sozialdaten bei den Krankenkassen) wurde ein Konzept für die Prüfung der Datenqualität durch das IQTIG entwickelt, das nach Abschluss der Beratungen in den zuständigen Gremien 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 Datenqualität 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 (Ehling et al. 2007, Nonnemacher et al. 2014, Chen et al. 2014, Arts et al. 2002).

⁴¹ Richtlinie über Maßnahmen zur Qualitätssicherung der Versorgung von Früh- und Reifgeborenen gemäß § 136 Absatz 1 Nummer 2 SGB V in Verbindung mit § 92 Abs. 1 Satz 2 Nr. 13 SGB V (Qualitätssicherungs-Richtlinie Früh- und Reifgeborene).

⁴² Richtlinie des Gemeinsamen Bundesausschusses nach § 92 Abs. 1 Satz 2 Nr. 13 i. V. m. § 136 Abs. 1 Nr. 1 SGB V über die einrichtungs- und sektorenübergreifenden Maßnahmen der Qualitätssicherung. In der Fassung vom 19. April 2010, zuletzt geändert am 21. Juni 2018, in Kraft getreten am 1. Januar 2019. URL: <https://www.g-ba.de/informationen/richtlinien/72/> (abgerufen am 11.04.2019).

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

- **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 Prüfungen 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 (Schubert et al. 2010, Nonnemacher et al. 2014). 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 eine Kombination aus interner und externer Validierung angewandt.

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

- technische Prüfung bei Dateneingabe und -annahme (interne Validierung)
- statistische Prüfung der erhobenen Daten nach Datenannahme (interne Validierung)
- Datenabgleich mit Referenzquelle nach Datenannahme (externe Validierung)

14.1 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

sen 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.

14.2 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.

14.3 Datenabgleich mit Referenzquelle nach Datenannahme

Als Referenzquelle zur Validierung von Gesundheitsdaten werden verschiedene Datenquellen verwendet (ärztliche Originaldokumentationen, Befragungsdaten, Abrechnungsdaten, amtliche Statistiken oder Registerdaten) (Horenkamp-Sonntag et al. 2014, Benchimol et al. 2011). 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 13). 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 Datenqualität 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 Datenqualität und des Handlungsbedarfs) und der Ebene der geprüften Leistungserbringer (leistungserbringerbezogene Bewertung der Datenqualität und des Handlungsbedarfs). Zur Bewertung der Datenqualität reicht die Übereinstimmungsrate alleine nicht aus, es müssen weitere Informationen hinzugezogen werden:

Verfahrensbezogene Bewertung der Datenqualität und des Handlungsbedarfs

1. Anzahl/Anteil der Leistungserbringer mit Abweichungen
2. Relevanz der Abweichungen für die Berechnung der Qualitätsindikatoren und Risikoadjustierung (Unter-, Über-, Fehldokumentation)
3. 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 7.2).

Leistungserbringerbezogene Bewertung der Datenqualität und des Handlungsbedarfs

1. Anzahl/Anteil der Datenfelder mit Abweichungen
2. Art und Auswirkung der Abweichungen auf die rechnerischen Ergebnisse der Qualitätsindikatoren und Risikoadjustierung (Unter-, Über-, Fehldokumentation)
3. Beeinflussbarkeit der Dokumentationsmängel durch Maßnahmen des internen Struktur- und Prozessmanagements der Leistungserbringer
4. 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:

1. Sterbefälle, die nur in den §21-Daten dokumentiert sind
2. Sterbefälle, die nur in den QS-Daten dokumentiert sind
3. 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 Krankenhausinformationssystem) geprüft und ggf.

nachdokumentiert. Der Abgleich wurde erstmalig für die Daten der Erfassungsjahre 2010 bis 2014 durchgeführt.

14.4 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 Problembhebung

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, indem 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.

15 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 auffällig⁴³ ist (siehe Kapitel 12), 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 15.2), der Auswertung von Follow-up-Indikatoren (Abschnitt 15.3) und Indexbildung (Abschnitt 15.4) 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).

15.1 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 Leistungserbringerergebnisse mit einem Referenzbereich (siehe Kapitel 12). Daher fokussiert die folgende Darstellung auf die Einflussgrößen, die bei einer Darstellung der Indikatorergebnisse von Leistungserbringern relevant sind.

⁴³ Während in der QSKH-RL der Begriff „rechnerische Auffälligkeit“ Verwendung findet, wird in der DeQS-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 und 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 (Patientenmix) 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 15.2) dienen dazu, Indikatormessergebnisse um solche systematischen, patientenseitigen Einflüsse so weit wie 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 in der Regel 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.

15.2 Risikoadjustierung

Unter dem Begriff der 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 es zu entsprechenden Unterschieden in den beobachteten Indikatorwerten der Leistungserbringer kommen, 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 vom IQTIG 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 Auswertungskonzepts anhand einer ausführlichen Literaturrecherche und unter Einbeziehung der fachlich-inhaltlichen Expertise der entsprechenden Expertengremien. 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 können 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 (vgl. z. B. Keiding und Clayton 2014).

15.2.1 Stratifizierung

Bei einer Risikoadjustierung mittels Stratifizierung wird die dem Qualitätsindikator zugrunde 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 dargestellt (Tabelle 12).

Tabelle 12: Im Krankenhaus verstorbene Patientinnen und Patienten mit isoliertem koronarchirurgischem Eingriff geschichtet nach Dringlichkeit des Eingriffs im Jahr 2015

Dringlichkeit des Eingriffs	verstorbene Patientinnen und Patienten / Gesamtanzahl	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 Sterblichkeitsrate bei dieser Einrichtung zu erwarten, da bei Notfalleingriffen das Risiko entsprechend höher ist. Die erhöhte Sterblichkeitsrate 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 Stratifizierungsvariablen 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 12) 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“ ($\leq 30\%$; $31\text{--}50\%$; $> 50\%$) und das Geschlecht der Patientinnen und Patienten als Risikofaktor zu berücksichtigen, steigt die Anzahl der möglichen Strata auf 12 ($2 \times 3 \times 2$). Die Fallzahl der Patientinnen und Patienten in den einzelnen Strata kann mit steigender Anzahl an Risikofaktoren so weit 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 mmHg oder von diastolisch ≤ 60 mmHg sowie ein Alter ≥ 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 im Nachgang zur Überprüfung einer Übertragbarkeit für die Messung des jeweiligen Qualitätsindikators genutzt werden.

15.2.2 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 E (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 stehen ein erhöhter Aufwand im Rahmen der Verfahrenspflege (siehe Abschnitt 7.2), 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.

15.2.3 Zusammenfassung und Ausblick

In der vorliegenden Fassung der „Methodischen Grundlagen“ werden lediglich Grundbegriffe der zum Zeitpunkt der Erstellung eingesetzten Methoden der Risikoadjustierung dargestellt. Wichtige Aspekte der gegenwärtig bereits verwendeten Verfahren werden 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. Allerdings müssen die Risikofaktoren auch tatsächlich im Rahmen der Verfahren der Qualitätssicherung nach §§ 136 ff. SGB V messbar sein. Die Variablenselektion erfolgt aus einer Kombination von inhaltlichen Überlegungen und datengestützten Modellselektionsverfahren.

Zu einem besseren Verständnis der Abhängigkeitsstrukturen zwischen Risikofaktoren und Indikator wird sowohl 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 und dem damit verbundenen erhöhten Aufwand bei der Dokumentation durch die Leistungserbringer.

Darüber hinaus trägt Risikoadjustierung auch dazu bei, 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.

15.3 Auswertung von Follow-up-Indikatoren

Als Follow-up-Indikatoren werden im Rahmen der externen Qualitätssicherung Qualitätsindikatoren bezeichnet, deren Auswertung sich auf Ereignisse (z. B. Revisionseingriffe oder Tod einer Patientin / eines Patienten) innerhalb eines bestimmten Beobachtungszeitraums im Anschluss an ein interessierendes Ereignis (z. B. eine Organtransplantation) bezieht. Üblicherweise werden aus diesen Daten Rückschlüsse auf die Qualität des interessierenden Ereignisses gezogen. Beispielsweise ist das Qualitätsziel für die in Anlage 3 der QSKH-RL definierten Follow-up-Indikatoren das möglichst seltene Auftreten von Revisionen nach Hüftendoprothesen-, Knieendoprothesen- und Herzschrittmacherimplantation innerhalb eines definierten Zeitraums nach einem interessierenden Ereignis (dem Ersteingriff).

15.3.1 Aktives und passives Follow-up

In der externen Qualitätssicherung lassen sich zwei Arten von Follow-up-Indikatoren unterscheiden: jene mit aktivem Follow-up, bei denen Informationen zu Ereignissen im Follow-up-Zeitraum zu jedem Fall aktiv von den Leistungserbringern gemeldet werden; sowie jene mit passivem Follow-up, bei denen Fälle nicht aktiv nachverfolgt, sondern Folgeereignisse im Follow-up-Zeitraum erst durch eine Verknüpfung mit Daten zum interessierenden Ereignis (z. B. über ein Patientenpseudonym) identifiziert werden. Weitere Möglichkeiten des Follow-up stellen die Verknüpfung von Daten aus den QS-Verfahren *Geburtshilfe* und *Neonatologie* sowie sozialdatengestützte Follow-up-Indikatoren dar.

Die beiden Arten von Follow-up-Indikatoren unterscheiden sich im Zeitpunkt, wann die Follow-up-Informationen für eine Auswertung zur Verfügung stehen. Bei Indikatoren mit aktivem Follow-up liegen die relevanten Informationen zu Ereignissen im Beobachtungszeitraum erst nach dessen Ende vor, da diese erst dann von den Leistungserbringern übermittelt werden. Daher ist hier eine Auswertung oft erst nach Ende des Beobachtungszeitraums möglich. Bei Indikatoren mit passivem Follow-up können dagegen Informationen zum Verlauf im Beobachtungszeitraum unter Umständen schon vor Ende des Beobachtungszeitraums komplett vorliegen. Im Folgenden werden neben der Darstellungsweise von Follow-up-Daten (Abschnitt 15.3.2) die methodischen Grundlagen für zwei Auswertungsstrategien von Follow-up-Indikatoren beschrieben (Kohorten- und Periodensichtweise, siehe Abschnitt 15.3.3). Diese unterscheiden sich hinsichtlich der Wahl der Grundgesamtheit sowie der statischen Auswertungsmethodik und auch in Bezug auf die Beobachtungszeiträume. Beide Strategien können prinzipiell bei der Auswertung von

Indikatoren mit passivem Follow-up angewandt werden. Indikatoren mit aktivem Follow-up sind dagegen prinzipiell auf eine Auswertungsstrategie (Kohortensichtweise) begrenzt.

Follow-up-Indikatoren erfordern schon deshalb eine eigene Auswertungsstrategie, da sich die zu einem Indikator erhobenen Daten nicht zwangsläufig wie bei den übrigen Indikatoren auf ein kalendarisches Jahr beschränken. Bei einem Indikator mit einem Beobachtungszeitraum von einem Jahr nach Ersteingriff (interessierendes Ereignis) ist beispielsweise eine Auswertung von Informationen aus zwei Erfassungsjahren notwendig, da der Beobachtungszeitraum von einem Jahr nach Ersteingriff nicht im Kalenderjahr des Ersteingriffs beendet sein kann. Bei diesen Indikatoren gehen also auch Informationen ein, welche aus dem kalendarischen Folgejahr des Ersteingriffs stammen, weshalb für die Auswertung die Betrachtung lediglich eines Erfassungsjahres nicht ausreicht. Das IQTIG fasst im Folgenden „Erfassungsjahr“ als äquivalent zu „Kalenderjahr“ auf.

15.3.2 Lexis-Diagramme

Die Daten eines Follow-up-Indikators lassen sich mittels eines sogenannten Lexis-Diagramms (Keiding 1990) illustrieren. Abbildung 15 zeigt beispielhaft die beobachteten Behandlungsverläufe eines Follow-up-Indikators mit einem Beobachtungszeitraum von einem Jahr in einem solchen Lexis-Diagramm. Auf der x-Achse ist die Kalenderzeit und auf der y-Achse die Beobachtungszeit bzw. Follow-up-Zeit seit dem interessierenden Ereignis (in Tagen) dargestellt. In diesem Kontext wird die zeitliche Dimension eines Behandlungsverlaufs durch Erst- und Folgeingriff definiert; die Verbindungslinie zwischen den Zeitpunkten dieser Ereignisse, auch Lebenslinie genannt, stellt den Behandlungsverlauf im Koordinatensystem dar und hat, bei gleicher Zeiteinheit von x- und y-Achse (wie z. B. in Abbildung 15), einen Winkel von 45 Grad. In diesem Beispiel wird dabei angenommen, dass die Datenerhebung nur bis zum 31. Dezember 2017 erfolgt ist, d. h., dass keine Behandlungsverläufe nach diesem Zeitpunkt beobachtet wurden (Rechtszensierung). Aus der Abbildung wird deutlich, dass für Patientinnen und Patienten mit interessierendem Ereignis innerhalb des Erfassungsjahres 2017 kein vollständiger Beobachtungszeitraum vorliegt. Die Rechtszensierung der Beobachtungszeit muss in der Auswertung berücksichtigt werden, sofern diese Patientinnen und Patienten in die Grundgesamtheit eingeschlossen werden. Methoden aus der Ereigniszeitanalyse, die in IQTIG (2017) – und detaillierter beispielsweise in Klein und Moeschberger (2003) – beschrieben sind, erlauben eine Berücksichtigung unterschiedlich langer Beobachtungszeiten (durch Rechtszensierung, Linkstrunkierung oder konkurrierende Ereignisse) und stellen daher einen geeigneten Rahmen für die statistische Auswertungsmethodik von Follow-up-Indikatoren dar.

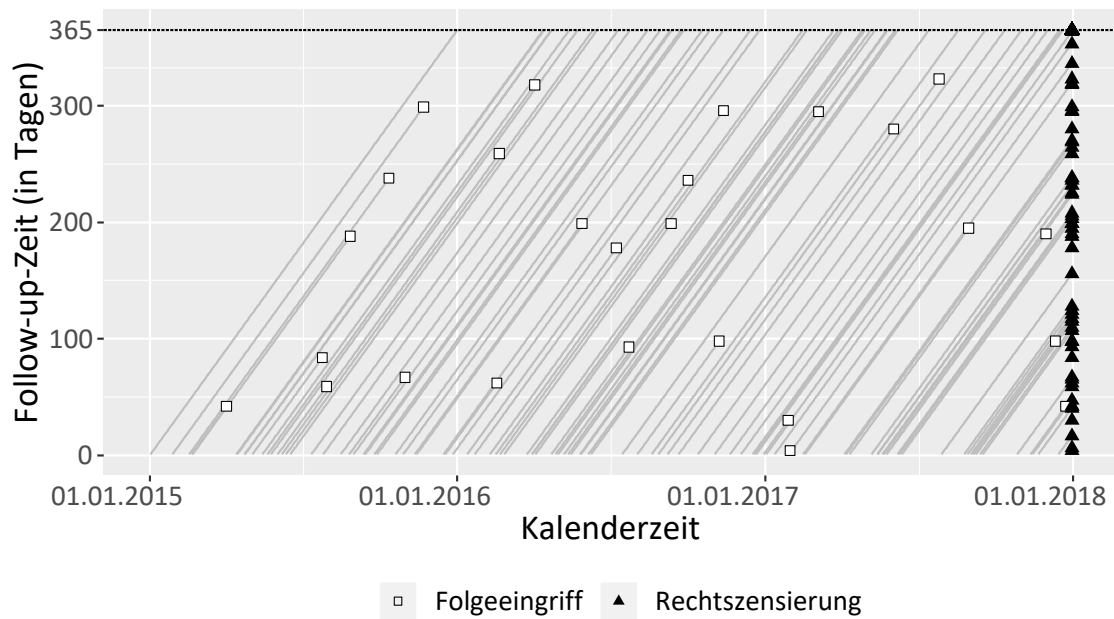


Abbildung 15: Beispielhafte Darstellung der Daten eines 1-Jahres-Follow-up-Indikators mit 100 Fällen in Form von Lebenslinien in einem Lexis-Diagramm

15.3.3 Auswertung nach Kohortensichtweise und nach Periodensichtweise

Im Folgenden werden die einmalige Auswertung des gesamten Beobachtungszeitraums nach dessen vollständiger Beobachtung (siehe Abbildung 16: Kohortensichtweise) sowie die Auswertung der Daten, die im aktuellen Erfassungsjahr beobachtet wurden (siehe Abbildung 16: Periodensichtweise), dargestellt.

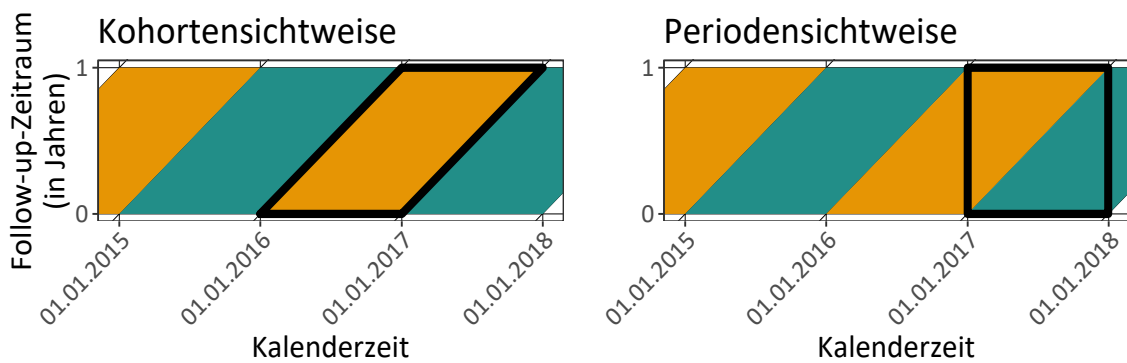


Abbildung 16: Lexis-Diagramme für einen 1-Jahres-Follow-up-Indikator in Kohortensichtweise und in Periodensichtweise

Die x-Achsen in Abbildung 16 stellen die kalendarische Zeit und die y-Achsen die vergangene Zeit nach dem interessierenden Ereignis (Beobachtungszeitraum) dar. Prinzipiell gibt es drei Betrachtungsweisen der beobachteten Zeiträume.⁴⁴ Die Lexis-Diagramme verdeutlichen die Kohorten- und die Periodensichtweise. Im Rahmen der Qualitätssicherung sind hauptsächlich diese relevant; die möglichen Unterschiede des Risikos für das Eintreten eines Ereignisses in den verschiedenen Follow-up-Strata kann als patientenseitige Eigenschaft gesehen werden und wird daher immer gleich in der Berechnung der Maßzahl des Qualitätsindikators berücksichtigt werden. Dies entspricht im Allgemeinen einer Risikoadjustierung für die Beobachtungszeit.

Die Kohorten- und Periodensichtweisen unterscheiden sich darin, dass die Grundgesamtheit bei der Kohortensichtweise auf interessierende Ereignisse, d. h. die Ersteingriffe, innerhalb eines Erfassungsjahres beschränkt wird, während bei der Periodensichtweise alle Ersteingriffe innerhalb des Beobachtungszeitraums eingeschlossen werden. In Abbildung 16 sind die beiden Grundgesamtheiten der Sichtweisen jeweils als schwarz umrandete Flächen dargestellt, alle Verläufe, die in dieser Fläche liegen, werden analysiert.

Bei Indikatoren, bei denen die Auswertungsstrategie basierend auf der Kohortensichtweise gewählt wird, werden in der Regel alle interessierenden Ereignisse des aktuellsten Erfassungsjahres in die Analyse eingeschlossen, für die der definierte Beobachtungszeitraum komplett beobachtet wurde. Diese Strategie hat den Vorteil, dass alle interessierenden Ereignisse aus einem einzigen Erfassungsjahr stammen, was der aktuellen Verfahrensweise des Strukturierten Dialogs entspricht. Der wesentliche Nachteil ist, dass der Beobachtungszeitraum zum Zeitpunkt der Auswertung bereits vollständig beobachtet sein muss. Indikatoren mit langen Beobachtungszeiträumen können bei dieser Auswertungsstrategie daher erst mit erheblicher Verzögerung ausgewertet und deren Ergebnisse an die Leistungserbringer zurückgemeldet werden. Im Gegensatz dazu erlaubt eine Auswertungsstrategie basierend auf der Periodensichtweise die Berücksichtigung des ganzen aktuellen Erfassungsjahres als Beobachtungszeitraum. Dies hat den Vorteil, dass alle im aktuellen Erfassungsjahr erhobenen Daten ohne zeitliche Verzögerung in die Auswertung einfließen. Der wesentliche Nachteil ist, dass dadurch betrachtete Ersteingriffe aus mehreren Erfassungsjahren in einer Auswertung zusammengefasst werden und sich eventuelle Qualitätsdefizite nicht immer eindeutig einem Erfassungsjahr zuordnen lassen.

Bei Indikatoren mit aktivem Follow-up werden Informationen zum Beobachtungszeitraum erst nach dessen Ablauf von den Leistungserbringern übermittelt. Daher ist eine periodenweise Auswertungsstrategie, in die auch unvollständige Beobachtungszeiträumen eingehen, für diese Indikatoren nicht sinnvoll, und es kommt nur die Auswertung nach der Kohortensichtweise infrage. Für Indikatoren mit passivem Follow-up sind dagegen beide Auswertungsstrategien prinzipiell möglich. Die Wahl der Auswertungsstrategie wird in Abhängigkeit der interessierenden Grundgesamtheit und des Beobachtungszeitraums indikatorspezifisch getroffen, da Follow-up-Indikatoren jeweils verschiedene Qualitätsziele haben können und damit auch unterschiedliche Auswertungsstrategien für unterschiedliche Indikatoren sinnvoll sein können.

⁴⁴ In der Krebsepidemiologie ist diese Analyseart unter dem Begriff Age-Period-Cohort-Modelling bekannt (Carstensen 2007).

Unabhängig von der Wahl der Strategie ist bei der Auswertung von Follow-up-Indikatoren zu berücksichtigen, dass sich die Beobachtungszeiträume zwischen einzelnen Fällen unterscheiden können, beispielsweise durch das Versterben von Patientinnen und Patienten oder andere sogenannte konkurrierende Ereignisse, die dazu führen, dass der Beobachtungszeitraum für einen Fall nicht vollständig beobachtet werden kann. Statistische Methoden aus dem Kontext der Ereigniszeitanalysen erlauben die Berücksichtigung unterschiedlich langer Beobachtungszeiträume und bieten daher generell einen geeigneten statistischen Rahmen für die Auswertung von Follow-up-Indikatoren. Die Methoden der Ereigniszeitanalyse sind jedoch hinsichtlich ihrer zugrunde liegenden Annahmen und Berechnungen in der Regel komplexer als die bisher zur Anwendung kommenden Methoden für die Auswertung von Qualitätsindikatoren. Ob die Wahl von komplexeren statistischen Methoden für die Auswertung gerechtfertigt ist, wird letztendlich immer indikatorspezifisch in Abwägung von notwendiger Komplexität und erwünschter Transparenz beurteilt. Details der statistischen Methodik aus der Ereigniszeitanalyse, die bei risikoadjustierten und nicht risikoadjustierten Follow-up-Indikatoren Anwendung finden kann, sind in IQTIG (2017) und in IQTIG (2018a) beschrieben.

15.4 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 15.4.1) und die Indexbildung (siehe Abschnitt 15.4.2), 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 zugunsten 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 bei jeder Aggregation eine übersichtlichere Darstellung einen Informationsverlust mit sich bringt. 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.

15.4.1 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 Dialogs 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 Dialogs 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 15.4.2). 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 15.2). 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.

15.4.2 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, Schwartz 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, Schwartz 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.

Indexkonstruktion

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 („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 10). 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 z. B. 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 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

Siehe Datenqualität

Datenqualität

Die Datenqualität beschreibt, ob die zur Berechnung des Indikatorwerts verwendeten Daten den Vorgaben des Indikators hinsichtlich Vollzähligkeit, Vollständigkeit und Dateninhalten entsprechen.

Erfassungsjahr

Das Erfassungsjahr bezeichnet das Jahr, in dem die Daten für die externe Qualitätssicherung erhoben werden.

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.

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 Qualitätsmerkmale 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.

QS-Filter

Summe der Einschluss- und Ausschlussbedingungen, die definieren, welche Behandlungsfälle in der externen Qualitätssicherung dokumentationspflichtig sind.

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 handelt es sich bei Qualitätsaspekten in der Methodik des IQTIG nicht um generelle Kategorien, sondern um spezifisch für jedes Verfahren abgeleitete Themen.⁴⁵

⁴⁵ Der Begriff ist nicht gleichbedeutend mit den elf „grundlegenden Qualitätsaspekten“ der DIN EN 15224, die nach der Methodik des IQTIG Qualitätsdimensionen entsprechen.

Qualitätsdimension

Grundlegende Kategorie von Anforderungen an die Gesundheitsversorgung. Das IQTIG unterscheidet in seinem Rahmenkonzept für Qualität sechs Qualitätsdimensionen (Wirksamkeit; Patientensicherheit; Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten; Rechtzeitigkeit und Verfügbarkeit; Angemessenheit; Koordination und Kontinuität).

Qualitätsindikator

Qualitätsindikatoren sind quantitative Größen, die Aussagen über die Erfüllung konkreter Qualitätsanforderungen mittels eines Messverfahrens und eines Bewertungskonzepts ermöglichen. Qualitätsindikatoren umfassen die drei Komponenten Qualitätsziel, Messverfahren und Bewertungskonzept.

Qualitätsmerkmal

Eigenschaft der Versorgung (z. B. Erlangen von Gehfähigkeit, Information über alternative Behandlungsmöglichkeiten), die mit Anforderungen verknüpft ist.

Qualitätsmessung

Empirische Beschreibung der Erfüllung von Anforderungen anhand von Qualitätsindikatoren.

Qualitätsmodell

Die Summe der zur Beschreibung der Qualität eines Versorgungsbereichs ausgewählten Qualitätsaspekte (siehe Qualitätsaspekte). Für jeden Versorgungsbereich, in dem ein neues QS-Verfahren entworfen wird, wird ein individuelles Qualitätsmodell entwickelt, welches als Basis für die Entwicklung von Qualitätsindikatoren dient.

Qualitätssicherung

Unter Qualitätssicherung im Gesundheitswesen werden hier alle Prozesse und Maßnahmen verstanden, durch die die Qualität der Versorgung gewährleistet oder verbessert werden soll.

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 im Wesentlichen die Festlegung der einbezogenen Leistungen oder Leistungsbereiche, Qualitätsindikatoren (mit Qualitätszielen, Mess- und Auswertungsmethoden sowie Regeln zur Bewer-

tung), 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

Anforderung an Strukturen, Prozesse oder Ergebnisse einer medizinischen Versorgungsleistung. Ein Qualitätsziel gibt die gewünschte Richtung der Ergebnisse eines Qualitätsindikators vor (z. B. „Die Sterblichkeit nach elektiver Hüftendoprothesenversorgung soll möglichst niedrig sein“).

Referenzbereich

Der Bereich auf der Messskala eines Qualitätsindikators, der ein bestimmtes Qualitätsniveau repräsentiert. Ein Referenzbereich stellt die konkret zu erfüllende Anforderung für ein Qualitätsmerkmal dar (z. B. „Eine Komplikationsrate unter 5 % entspricht erwartbarer Qualität“).

Reliabilität

Die Zuverlässigkeit einer Messung. Sie ist definiert als der Anteil der Varianz an der Gesamtvarianz einer Messreihe, 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. Landesgeschäftsstelle für Qualitätssicherung, Landesarbeitsgemeinschaft, 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.

Verwendungszweck von Qualitätsmessungen

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 durch externen Vergleich (wie Verfahren nach der QSKH-RL und DeQS-RL), der Information (z. B. von Patientinnen und Patienten) und der Planung der Versorgungsstrukturen (z. B. planungsrelevante Qualitätsindikatoren für die Krankenhausplanung) vor.

Versorgungsbereich

Bereich thematisch eng miteinander verbundener Anteile der Gesundheitsversorgung, z. B. Versorgung von Patientinnen und Patienten mit Diabetes mellitus.

Versorgungspfad

Eine Methode, die den Behandlungsprozess für eine definierte Gruppe von Patientinnen und Patienten in einem definierten Zeitraum strukturiert und dazu die Ziele der Behandlung, ihre Kernelemente, die erforderliche Kommunikation und Koordination sowie die Erreichung der Ergebnisse beschreibt. Der idealtypische Ablauf der Versorgung von Patientinnen und Patienten kann anhand eines Versorgungspfades dargestellt werden.

Versorgungsqualität

Grad, in dem die Gesundheitsversorgung von Einzelpersonen und Populationen Anforderungen erfüllt, die patientenzentriert sind und mit professionellem Wissen übereinstimmen.

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 (abgerufen am: 18.08.2017).
- AGREE Next Steps Consortium (2014): Appraisal of Guidelines for Research & Evaluation II. AGREE II Instrument. Deutsche Version. AGREE Research Trust. URL: http://www.agreetrust.org/wp-content/uploads/2014/03/AGREE_II_German-Version.pdf (abgerufen am: 30.01.2017).
- AGREE Next Steps Consortium (2017): Appraisal of Guidelines for Research & Evaluation II. AGREE II Instrument. Update: December 2017. AGREE Research Trust. URL: <https://www.agreetrust.org/wp-content/uploads/2017/12/AGREE-II-Users-Manual-and-23-item-Instrument-2009-Update-2017.pdf> (abgerufen am: 23.10.2018).
- Ahluwalia, SC; Damberg, CL; Silverman, M; Motala, A; Shekelle, PG (2017): What Defines a High-Performing Health Care Delivery System: A Systematic Review. *Joint Commission Journal on Quality and Patient Safety* 43(9): 450-459. DOI: 10.1016/j.jcjq.2017.03.010.
- 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).
- Apgar, V (1953): A Proposal for a New Method of Evaluation of the Newborn Infant. *Current Researches in Anesthesia and Analgesia* 32(4): 260-267.
- Appleby, J; Raleigh, VS; Frosini, F; Bevan, G; Gao, H; Lyscom, T (2011): Variations in Health Care. The good, the bad and the inexplicable. London: King's Fund. ISBN: 978-1-85717-641-8. URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/Variations-in-health-care-good-bad-inexplicable-report-The-Kings-Fund-April-2011.pdf (abgerufen am: 20.06.2018).
- APS [Aktionsbündnis Patientensicherheit] ([kein Datum]): Glossar. Berlin: APS. URL: <http://www.aps-ev.de/glossar/> (abgerufen am: 20.06.2018).
- 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; Shwartz, 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.
- BÄK [Bundesärztekammer]; KBV [Kassenärztliche Bundesvereinigung]; AWMF [Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften]; ÄZQ [Ärztliches Zentrum für Qualität in der Medizin] (2010): Programm für Nationale VersorgungsLeitlinien. Methoden-Report. 4. Auflage. [Stand:] 30.07.2010. Berlin: BÄK [u. a.]. DOI: 10.6101/AZQ/000061.
- 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.
- Battles, JB (2006): Quality and safety by design. *Quality and Safety in Health Care* 15(Suppl. 1): i1-i3. DOI: 10.1136/qshc.2006.020347.
- 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.
- Berk, RA (1996): Standard Setting: The Next Generation (Where Few Psychometricians Have Gone Before!). *Applied Measurement in Education* 9(3): 215-225. DOI: 10.1207/s15324818ame0903_2.
- Berwick, DM (1991): Controlling Variation in Health Care: A Consultation from Walter Shewhart. *Medical Care* 29(12): 1212-1225.
- Berwick, DM (2002): A User's Manual For The IOM's 'Quality Chasm' Report. Patients' experiences should be the fundamental source of the definition of "quality.". *Health Affairs* 21(3): 80-90. DOI: 10.1377/hlthaff.21.3.80.
- Berwick, DM; James, B; Coye, MJ (2003): Connections Between Quality Measurement and Improvement. *Medical Care* 41(1 Suppl.): I-30-I-38. DOI: 10.1097/00005650-200301001-00004.
- Berwick, DM; Nolan, TW; Whittington, J (2008): The Triple Aim: Care, Health, and Cost. *Health Affairs* 27(3): 759-769. DOI: 10.1377/hlthaff.27.3.759.
- Bokhour, BG; Pugh, MJ; Rao, JK; Avetisyan, R; Berlowitz, DR; Kazis, LE (2009): Improving Methods for Measuring Quality of Care. A Patient-Centered Approach in Chronic Disease. *Medical Care Research and Review* 66(2): 147-166. DOI: 10.1177/1077558708327174.
- Bollen, K; Lennox, R (1991): Conventional Wisdom on Measurement: A Structural Equation Perspective. *Psychological Bulletin* 110(2): 305-314. DOI: 10.1037/0033-2909.110.2.305.

- Bollen, KA; Ting, K-f (2000): A Tetrad Test for Causal Indicators. *Psychological Methods* 5(1): 3-22. DOI: 10.1037/1082-989X.5.1.3.
- Bollen, KA; Diamantopoulos, A (2017): In Defense of Causal-Formative Indicators: A Minority Report. *Psychological Methods* 22(3): 581-596. DOI: 10.1037/met0000056.
- 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). URL: 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.
- 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; Roland, MO; Buetow, SA (2000): Defining quality of care. *Social Science & Medicine* 51(11): 1611-1625. DOI: 10.1016/S0277-9536(00)00057-5.
- 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 and Safety in Health Care* 13(6): 428-434. DOI: 10.1136/qshc.2003.007815.
- Campbell, SM; Kontopantelis, E; Hannon, K; Burke, M; Barber, A; Lester, HE (2011): Framework and indicator testing protocol for developing and piloting quality indicators for the UK quality and outcomes framework. *BMC Family Practice* 12: 85. DOI: 10.1186/1471-2296-12-85.
- Carinci, F; Van Gool, K; Mainz, J; Veillard, J; Pichora, EC; Januel, JM; et al. (2015): Towards actionable international comparisons of health system performance: expert revision of the OECD framework and quality indicators. *International Journal for Quality in Health Care* 27(2): 137-146. DOI: 10.1093/intqhc/mzv004.
- Carman, KL; Workman, TA (2017): Engaging patients and consumers in research evidence: Applying the conceptual model of patient and family engagement. *Patient Education and Counseling* 100(1): 25-29. DOI: 10.1016/j.pec.2016.07.009.
- Carstensen, B (2007): Age-period-cohort models for the Lexis diagram. *Statistics in Medicine* 26(15): 3018-3045. DOI: 10.1002/sim.2764.

- 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 (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: 978-0-309-07280-9.
- 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/aug/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).
- Creswell, JW; Plano Clark, VL (2018): Designing and Conducting Mixed Methods Research. Third Edition. Los Angeles [u. a.]: SAGE Publications. ISBN: 9781506386621.
- D'Ardenne, J (2015): Developing Interview Protocols. Chapter 5. In: Collins, D: *Cognitive Interviewing Practice*. Los Angeles [u. a.]: Sage Publications, 101-125. ISBN: 978-1-4462-5600-5.
- Damberg, CL; Sorbero, ME; Lovejoy, SL; Lauderdale, K; Wertheimer, S; Smith, A; et al. (2011): An Evaluation of the Use of Performance Measures in Health Care. Santa Monica, US-CA: RAND Health. No. HHSM-500-2009-00010C. URL: https://www.rand.org/content/dam/rand/pubs/technical_reports/2011/RAND_TR1148.pdf (abgerufen am: 03.07.2018).

- Davies, S; McDonald, KM; Schmidt, E; Schultz, E; Geppert, J; Romano, PS (2011): Expanding the Uses of AHRQ's Prevention Quality Indicators: Validity From the Clinician Perspective. *Medical Care* 49(8): 679-685. DOI: 10.1097/MLR.0b013e3182159e65.
- de Koning, J; Smulders, A; Klazinga, N (2007): 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 (2007): 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).
- den Breejen, EME; Nelen, WLDM; Schol, SFE; Kremer, JAM; Hermens, RPMG (2013): Development of guideline-based indicators for patient-centredness in fertility care: what patients add. *Human Reproduction* 28(4): 987-996. DOI: 10.1093/humrep/det010.
- Deutscher Ethikrat (2016): Patientenwohl als ethischer Maßstab für das Krankenhaus. Stellungnahme. [Stand:] 05.04.2016. Berlin: Deutscher Ethikrat. ISBN: 978-3-941957-71-8. URL: <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-patientenwohl-als-ethischer-massstab-fuer-das-krankenhaus.pdf> (abgerufen am: 17.07.2018).
- Diamantopoulos, A; Winklhofer, HM (2001): Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research* 38(2): 269-277. DOI: 10.1509/jmkr.38.2.269.18845.
- 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 (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: MWV [Medizinisch Wissenschaftliche Verlagsgesellschaft], 162-208. ISBN: 978-3-941468-71-9.
- Ebert, ST; Pittet, V; Cornuz, J; Senn, N (2017): Development of a monitoring instrument to assess the performance of the Swiss primary care system. *BMC Health Services Research* 17: 789. DOI: 10.1186/s12913-017-2696-z.
- Edwards, JR; Bagozzi, RP (2000): On the Nature and Direction of Relationships Between Constructs and Measures. *Psychological Methods* 5(2): 155-174. DOI: 10.1037//1082-989X.5.2.155.
- 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-ngaf/Eurostat-HANDBOOK%20ON%20DATA%20QUALITY%20ASSESSMENT%20METHODS%20AND%20TOOLS%20%20I.pdf> (abgerufen am: 13.11.2018).
- Ennker, J; Albert, A; Thanner, M; Nagel, E; Rosendahl, U; Ennker, I (2009): Transparenz medizinischer Qualitätsdaten. Ethische Bringschuld oder Instrument zur Schaffung unlauterer Wettbewerbsvorteile? *DMW – Deutsche Medizinische Wochenschrift* 134(Suppl. 6): S188-S189. DOI: 10.1055/s-0029-1241904.
- Evans, SM; Lowinger, JS; Sprivulis, PC; Copnell, B; Cameron, PA (2009): Prioritizing quality indicator development across the healthcare system: identifying what to measure. *Internal Medicine Journal* 39(10): 648-654. DOI: 10.1111/j.1445-5994.2008.01733.x.
- Ewald, DA; Huss, G; Auras, S; Caceres, JR-C; Hadjipanayis, A; Geraedts, M (2018): Development of a core set of quality indicators for paediatric primary care practices in Europe, COSI-PPC-EU. *European Journal of Pediatrics* 177(6): 921-933. DOI: 10.1007/s00431-018-3140-z.
- 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.
- Feinstein, AR (1987): Clinimetric perspectives. *Journal of Chronic Diseases* 40(6): 635-640. DOI: 10.1016/0021-9681(87)90027-0.

- Fitch, K; Bernstein, SJ; Aguilar, 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 (abgerufen am: 30.01.2017).
- Fitzpatrick, AR (1989): Social Influences in Standard Setting: The Effects of Social Interaction on Group Judgments. *Review of Educational Research* 59(3): 315-328. DOI: 10.3102/00346543059003315.
- 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 (abgerufen am: 30.01.2017).
- G-BA [Gemeinsamer Bundesausschuss] (2019): 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 16. August 2018, veröffentlicht im Bundesanzeiger BAnz AT 05.03.2019 B2, in Kraft getreten am 6. März 2019. Stand: 15.03.2019. Berlin: G-BA. URL: https://www.g-ba.de/downloads/62-492-1777/VerfO_2018-08-16_iK-2019-03-06.pdf (abgerufen am: 11.04.2019).
- G-BA [Gemeinsamer Bundesausschuss] ([kein Datum]): Eckpunkte zu einem gemeinsamen Verständnis und Handlungsempfehlungen zur sektorenübergreifenden Qualitätssicherung. Berlin: G-BA. URL: <https://www.g-ba.de/downloads/17-98-3536/Eckpunkte-Handlungsempfehlungen-sQS.pdf> (abgerufen am: 12.11.2018).
- 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 (2009): Einsatz von Qualitätsindikatoren. Kapitel 3. In: ÄZQ [Ärztliches Zentrum für Qualität in der Medizin]: *Programm für Nationale VersorgungsLeitlinien von BÄK, KBV und AWMF. Qualitätsindikatoren – Manual für Autoren.* (ÄZQ Schriftenreihe, Band 36). Neukirchen: Make a Book, 5-7. ISBN: 978-3940218-551. URL: <https://www.aezq.de/mdb/edocs/pdf/schriftenreihe/schriftenreihe36.pdf> (abgerufen am: 04.07.2018).
- 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.
- Geraedts, M; Drösler, SE; Döbler, K; Eberlein-Gonska, M; Heller, G; Kuske, S; et al. (2017): DNVF-Memorandum III „Methoden für die Versorgungsforschung“, Teil 3: Methoden der Qualitäts- und Patientensicherheitsforschung. *Gesundheitswesen* 79(10): e95-e124. DOI: 10.1055/s-0043-112431.
- Gerlinger, T (2009): Nutzerorientierung im Gesundheitswesen – Probleme und Perspektiven. In: Mozygemba, K; Mümken, S; Krause, U; Zündel, M; Rehm, M; Höfling-Engels, N; et al.: *Nutzerorientierung – ein Fremdwort in der Gesundheitssicherung?* Bern: Hans Huber, 17-29. ISBN: 978-3-456-84597-5.
- Gerteis, M; Edgman-Levitan, S; Daley, J; Delbanco, T (1993): Through the patient's eyes. Understanding and promoting patient-centered care. San Francisco: Jossey-Bass. ISBN: 978-0-7879-6220-3.
- 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.
- Goodwin, LD (1996): Focus on Quantitative Methods: Determining Cut-Off Scores. *Research in Nursing & Health* 19(3): 249-256. DOI: 10.1002/(SICI)1098-240X(199606)19:3<249::AID-NUR8>3.0.CO;2-K.
- Greene, SM; Tuzzio, L; Cherkin, D (2012): A Framework for Making Patient-Centered Care Front and Center. *Permanente Journal* 16(3): 49-53. DOI: 10.7812/TPP/12-025.
- Großschädl, F; Freidl, W; Stolz, E; Rásky, É (2014): Versorgungspfade. Fokus: Sektorenübergreifende Instrumente. Endbericht. [Stand:] Dezember 2014. Graz: Medizinische Universität Graz. URL: <http://www.hauptverband.at/cdscontent/load?contentid=10008.616665&version=1426857734> (abgerufen am: 12.11.2018).
- 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/howto5mh.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]; HHS.gov [U.S. Department of Health and Human Services], FDA [Food and Drug Administration], CBER [Center for Biologics Evaluation and Research]; HHS.gov [U.S. Department of Health and Human Services], FDA [Food and Drug Administration], 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 [u. a.]. URL: <http://www.fda.gov/downloads/Drugs/.../Guidances/UCM193282.pdf> (abgerufen am: 30.01.2017).
- Hibbert, PD; Molloy, CJ; Hooper, TD; Wiles, LK; Runciman, WB; Lachman, P; et al. (2016): The application of the Global Trigger Tool: a systematic review. *International Journal for Quality in Health Care* 28(6): 640-649. DOI: 10.1093/intqhc/mzw115.
- 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 (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.
- Hoffmeyer-Zlotnik, JHP; Hanefeld, U; Herter-Eschweiler, R; Mohr, S (2010): Statistik und Wissenschaft. Demographische Standards. 5. überarbeitete und erweiterte Auflage. [Stand:] Dezember 2010. (Statistik und Wissenschaft, Bd. 17/2010). Wiesbaden: Destatis [Statistisches Bundesamt]. ISBN: 978-3-8246-0904-8. URL: https://www.destatis.de/GPStatistik/servlets/MCRFileNodeServlet/DEMonografie_derivate_00000210/10308171090_04.pdf (abgerufen am: 08.04.2019).

- 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.
- IGZ [Inspectie voor de Gezondheidszorg] (2016): Brochure voor zorgaanbieders: Calamiteiten melden aan de IGZ. [Stand:] 13.12.2016. Den Haag: Ministerie van Volksgezondheid, Welzijn en Sport. URL: <https://www.rijksoverheid.nl/documenten/brochures/2016/12/13/brochure-voor-zorgaanbieders-calamiteiten-melden-aan-de-igz> [Download] (abgerufen am: 21.08.2018).
- Ingraham, A; Nathens, A; Peitzman, A; Bode, A; Dorlac, G; Dorlac, W; et al. (2017): Assessment of emergency general surgery care based on formally developed quality indicators. *Surgery* 162(2): 397-407. DOI: 10.1016/j.surg.2017.03.025.
- 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 (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 (abgerufen am: 18.08.2017).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018a): Biometrische Methodik für die Follow-up-Indikatoren nach QSKH-RL für die Bundesauswertung des Auswertungsjahres 2017. Stand: 23.10.2018. Berlin: IQTIG. URL: https://iqtig.org/downloads/berichte/2017/IQTIG_Biometrische-Methodik-FU-Indikatoren_BuAW-AJ-2017_2018-10-23.pdf (abgerufen am: 12.11.2018).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018b): Technische Dokumentation zur Basisspezifikation für Leistungserbringer. Erfassungsjahr 2019. Stand: 31.07.2018. Berlin: IQTIG. URL: https://iqtig.org/downloads/spezifikation/2019/v01/2019_TechDok_LE_XML_V01.pdf (abgerufen am: 13.08.2018).

- Jacob, G; Bengel, J (2000): Das Konstrukt Patientenzufriedenheit: Eine kritische Bestandsaufnahme. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie* 48(3): 280-301.
- Jacobsen, J; Meyer, L (2017): ISO 9241 – Ergonomie der Mensch-System-Interaktion. Abschnitt 4.1. In: *Praxisbuch Usability und UX. Was jeder wissen sollte, der Websites und Apps entwickelt*. Bonn: Rheinwerk, 61-62. ISBN: 978-3-8362-4423-7.
- Jarvis, H (2000): Differentiating among Research, Evaluation and Measures to Assure Quality. *Healthcare Management Forum* 13(4): 34-36. DOI: 10.1016/S0840-4704(10)60793-4.
- JCAHO [Joint Commission on Accreditation of Healthcare Organizations] (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.
- Jo Delaney, L (2018): Patient-centred care as an approach to improving health care in Australia. *Collegian* 25(1): 119-123. DOI: 10.1016/j.colegn.2017.02.005.
- Joint Commission (2016): Standards for Behavioral Health Care 2016. Oak Brook, US-IL: Joint Commission Resources. ISBN: 978-1-59940-912-2.
- Joint Commission (2017): Sentinel Events (SE) [Hospitals]. [Stand:] 01.07.2017. Oakbrook Terrace, US-IL: Joint Commission. URL: https://www.jointcommission.org/assets/1/6/CAMH_SE_0717.pdf (abgerufen am: 21.08.2018).
- 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. *EMA – Emergency Medicine Australasia* 26(2): 113-124. DOI: 10.1111/1742-6723.12195.
- Jonkisz, E; Moosbrugger, H; Brandt, H (2012): Planung und Entwicklung von Tests und Fragebogen. Kapitel 3. In: Moosbrugger, H; Kelava, A; Hrsg.: *Testtheorie und Fragebogenkonstruktion*. 2. Auflage. Berlin [u. a.]: Springer, 27-74. ISBN: 978-3-642-20071-7.
- 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 (1990): Statistical inference in the Lexis diagram. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* 332(1627): 487-509. DOI: 10.1098/rsta.1990.0128.
- 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.
- Klassen, A; Miller, A; Anderson, N; Shen, J; Schiariti, V; O'Donnell, M (2010): Performance measurement and improvement frameworks in health, education and social services systems: a systematic review. *International Journal for Quality in Health Care* 22(1): 44-69. DOI: 10.1093/intqhc/mzp057.
- Klein, JP; Moeschberger, ML (2003): Survival Analysis: Techniques for Censored and Truncated Data. Second Edition. New York [u. a.]: Springer. ISBN: 978-1-4419-2985-3.
- 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://freidok.uni-freiburg.de/data/1505> [Download] (abgerufen am: 09.04.2019).
- 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.
- Kohn, LT; Corrigan, JM; Donaldson, MS (2000): To Err Is Human: Building a Safer Health System. Washington, D.C.: National Academy Press. ISBN: 978-0-309-26174-6.
- Kötter, T; Blozik, E; Scherer, M (2012): Methods for the guideline-based development of quality indicators – a systematic review. *Implementation Science* 7(21). DOI: 10.1186/1748-5908-7-21.
- 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.
- 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.

- Kuckartz, U (2014): Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. 2., durchgesehene Auflage. Weinheim [u. a.]: Beltz Jventa. ISBN: 978-3-7799-2922-2.
- 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, KN (1990): Medicare: A Strategy for Quality Assurance, Volume I. Washington, D.C.: National Academies Press. ISBN: 978-0-309-04230-7. URL: https://www.nap.edu/login.php?record_id=1547&page=https%3A%2F%2Fwww.nap.edu%2Fdownload%2F1547 [Download nach Registrierung] (abgerufen am: 21.06.2018).
- Lohr, SL (2010): Sampling: Design and Analysis. 2. International Edition. Boston, US-MA: Brooks/Cole. ISBN: 978-0-495-11084-2.
- Maats, P (2007): Einführung in das Datenmanagement und die Datenauswertung. Kapitel 9. In: Stockmann, R: *Handbuch zur Evaluation. Ein praktische Handlungsanleitung*. (Sozialwissenschaftliche Evaluationsforschung, Band 6). Münster [u. a.]: Waxmann, 278-313. ISBN: 978-3-8309-1766-3.
- 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.

- Maxwell, RJ (1992): Dimensions of quality revisited: from thought to action. *Quality in Health Care* 1(3): 171-177. DOI: 10.1136/qshc.1.3.171.
- Mayring, P (2015): Qualitative Inhaltsanalyse. Grundlagen und Techniken. 12., überarbeitete Auflage. Weinheim [u. a.]: Beltz Verlag. ISBN: 978-3-407-25730-7.
- McGlynn, EA; Asch, SM (1998): Developing a Clinical Performance Measure. *American Journal of Preventive Medicine* 14(3): 14-21. DOI: 10.1016/S0749-3797(97)00032-9.
- McGlynn, EA (2003): Selecting Common Measures of Quality and System Performance. *Medical Care* 41(1): I-39-I-47.
- McLoughlin, V; Millar, J; Mattke, S; Franca, M; Jonsson, PM; Somekh, D; et al. (2006): Selecting indicators for patient safety at the health system level in OECD countries. *International Journal for Quality in Health Care* 18(Suppl. 1): 14-20. DOI: 10.1093/intqhc/mzl030.
- 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. 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 (abgerufen am: 30.01.2017).
- Meyer, W (2004): Indikatorenentwicklung: eine praxisorientierte Einführung. 2. Auflage. (CEval-Arbeitspapier, Nr. 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.
- Mid Staffordshire NHS Foundation Trust Public Inquiry (2013): Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry. Volume 3: Present and future Annexes. [Stand:] February 2013. London: TSO [The Stationery Office]. ISBN: 9780102981469. URL: <http://www.midstaffpublicinquiry.com/sites/default/files/report/Volume%203.pdf> (abgerufen am: 23.07.2018).
- Moore, L; Britten, N; Lydahl, D; Naldemirci, Ö; Elam, M; Wolf, A (2017): Barriers and facilitators to the implementation of person-centred care in different healthcare contexts. *Scandinavian Journal of Caring Sciences* 31(4): 662-673. DOI: 10.1111/scs.12376.
- 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 (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: MWV [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.
- Nag, N; Millar, J; Davis, ID; Costello, S; Duthie, JB; Mark, S; et al. (2018): Development of Indicators to Assess Quality of Care for Prostate Cancer. *European Urology Focus* 4: 57-63. DOI: 10.1016/j.euf.2016.01.016.
- Niederberger, M; Wassermann, S (2015): Methoden der Experten- und Stakeholdereinbindung in der sozialwissenschaftlichen Forschung. Wiesbaden: Springer VS. ISBN: 978-3-658-01686-9.
- Nielsen, J (1994): Enhancing the Explanatory Power of Usability Heuristics. In: Adelson, B; Dumais, S; Olson, J: *CHI '94 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM Conference on Human Factors in Computing Systems. 24-28.04.1994. Boston, US-MA*. New York: ACM [Association for Computing Machinery], 152-158. ISBN: 0-89791-650-6.
- 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.
- Norcini, JJ; Shea, JA (1997): The Credibility and Comparability of Standards. *Applied Measurement in Education* 10(1): 39-59. DOI: 10.1207/s15324818ame1001_3.
- Nothacker, M; Stokes, T; Shaw, B; Lindsay, P; Sipilä, R; Follmann, M; et al. (2016): Reporting standards for guideline-based performance measures. *Implementation Science* 11: 6. DOI: 10.1186/s13012-015-0369-z.

- 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).
- NQF [National Quality Forum] (2017): Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement. Effective August 2017. Washington, D.C.: NQF. URL: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards/2017_Measure_Evaluation_Criteria.aspx (abgerufen am: 16.10.2018).
- Nyumba, TO; Wilson, K; Derrick, CJ; Mukherjee, N (2018): The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution* 9(1): 20-32. DOI: 10.1111/2041-210X.12860.
- 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.
- Palinkas, LA; Horwitz, SM; Green, CA; Wisdom, JP; Duan, N; Hoagwood, K (2015): Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research* 42(5): 533-544. DOI: 10.1007/s10488-013-0528-y.
- 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.
- Patrick, DL; Guyatt, GH; Acquadro, C (2008): Patient-reported outcomes. Chapter 17. In: Higgins, JPT; Green, S: *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley-Blackwell, 531-545. ISBN: 978-0-470-69951-5.
- Perfetto, EM; Oehrlein, EM; Boutin, M; Reid, S; Gascho, E (2017): Value to Whom? The Patient Voice in the Value Discussion. *Value in Health* 20(2): 286-291. DOI: 10.1016/j.jval.2016.11.014.
- 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: MWV [Medizinisch Wissenschaftliche Verlagsgesellschaft]. ISBN: 978-3-941468-71-9.
- Pollak, R (2018): Status, sozialer. In: Kopp, J; Steinbach, A; Hrsg.: *Grundbegriffe der Soziologie*. Wiesbaden: Springer VS, 433-435. ISBN: 978-3-658-20977-3.
- Porst, R (2014): Fragebogen. Ein Arbeitsbuch. 4. erweiterte Auflage. Wiesbaden: Springer VS. ISBN: 978-3-658-02117-7.

- Porzsolt, F; Gaus, W (1993): Wirksamkeit und Nutzen medizinischer Maßnahmen: Ein Beitrag zur Optimierung des Gesundheitssystems. *Kliniker* 12(22): 522-528.
- Prinsen, CAC; Vohra, S; Rose, MR; Boers, M; Tugwell, P; Clarke, M; et al. (2016): How to select outcome measurement instruments for outcomes included in a “Core Outcome Set” – a practical guideline. *Trials* 17(1): 449. DOI: 10.1186/s13063-016-1555-2.
- Pross, C; Busse, R; Geissler, A (2017): Hospital quality variation matters – A time-trend and cross-section analysis of outcomes in German hospitals from 2006 to 2014. *Health Policy* 121(8): 842-852. DOI: 10.1016/j.healthpol.2017.06.009.
- 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 (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 (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]. [Zugesandt von G-BA auf Nachfrage].
- Remus, D; Fraser, I (2004): Guidance for Using the AHRQ Quality Indicators for Hospital-level Public Reporting or Payment. [Stand:] August 2004. (AHRQ Publication, No. 04-0086-EF). Rockville, US-MD: AHRQ [Agency for Healthcare Research and Quality]. URL: https://www.leg.state.nv.us/74th/Interim_Agendas_Minutes_Exhibits/Exhibits/HealthCare/E030608J-6.pdf (abgerufen am: 03.07.2018).
- Resar, RK; Rozich, JD; Classen, D (2003): Methodology and rationale for the measurement of harm with trigger tools. *Quality and Safety in Health Care* 12(Suppl. 2): ii39-ii45. DOI: 10.1136/qhc.12.suppl_2.ii39.
- 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, DL; Rosenberg, WM; Gray, JA; Haynes, RB; Richardson, WS (1996): Evidence based medicine: what it is and what it isn't. *BMJ – British Medical Journal* 312: 71-72. DOI: 10.1136/bmj.312.7023.71.

- Sackett, DL; Rosenberg, WMC; Gray, JAM; Hanyes, 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. 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 (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.
- Scheuren, F (2004): What is a Survey? Alexandria, US-VA: ASA [American Statistical Association]. URL: https://www.whatisasurvey.info/downloads/pamphlet_current.pdf (abgerufen am: 08.11.2018).
- 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.
- Scholl, I; Zill, JM; Härter, M; Dirmaier, J (2014): An Integrative Model of Patient-Centeredness – A Systematic Review and Concept Analysis. *PLoS One* 9(9): e107828. DOI: 10.1371/journal.pone.0107828.
- 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 (abgerufen am: 30.01.2017).
- Sens, B; Pietsch, B; Fischer, B; Hart, D; Kahla-Witzsch, H; von Friedrichs, V; et al. (2018): Begriffe und Konzepte des Qualitätsmanagements – 4. Auflage. *GMS Medizinische Informatik, Biometrie und Epidemiologie* 14(1). Doc04. DOI: 10.3205/mibe000182.
- Shadish, WR; Cook, TD; Campbell, DT (2002): *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, US-MA [u. a.]: Houghton. ISBN: 0-395-61556-9.
- Shea, BJ; Reeves, BC; Wells, G; Thuku, M; Hamel, C; Moran, J; et al. (2017): AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ – British Medical Journal* 358: j4008. DOI: 10.1136/bmj.j4008.
- 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.
- Smits, KPJ; Sidorenkov, G; Kleefstra, N; Bouma, M; Meulepas, M; Voorham, J; et al. (2017): Development and validation of prescribing quality indicators for patients with type 2 diabetes. *International Journal of Clinical Practice* 71(1): e12922. DOI: 10.1111/ijcp.12922.
- Solberg, LI; Mosser, G; McDonald, S (1997): The Three Faces of Performance Measurement: Improvement, Accountability, and Research. *Joint Commission – Journal on Quality Improvement* 23(3): 135-147. DOI: 10.1016/S1070-3241(16)30305-4.
- 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.
- Stegbauer, C; Willms, G; Kleine-Budde, K; Bramesfeld, A; Stammann, C; Szecsenyi, J (2017): Development of indicators for a nationwide cross-sectoral quality assurance procedure for mental health care of patients with schizophrenia, schizotypal and delusional disorders in Germany. *ZEFQ – Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 126: 13-22. DOI: 10.1016/j.zefq.2017.07.006.
- Steiger, NJ; Balog, A (2010): Realizing Patient-Centered Care: Putting Patients in the Center, Not the Middle. *Frontiers of Health Services Management* 26(4): 15-25.

- 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.
- Streiner, DL (2003): Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment* 80(1): 99-103. DOI: 10.1207/S15327752JPA8001_18.
- SVR-Gesundheit [Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen] (2002): Bedarfsgerechtigkeit und Wirtschaftlichkeit. Band III: Über-, Unter- und Fehlversorgung. III.1: Grundlagen, Übersichten, Versorgung chronisch Kranker. Gutachten 2000/2001. Baden-Baden: Nomos. ISBN: 3-7890-8099-3.
- SVR-Gesundheit [Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen] (2014): Bedarfsgerechte Versorgung – Perspektiven für ländliche Regionen und ausgewählte Leistungsbereiche. Gutachten 2014. Bern: Hans Huber. ISBN: 978-3-456-85500-4.
- SVR-Gesundheit [Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen] (2018): Bedarfsgerechte Steuerung der Gesundheitsversorgung. Gutachten 2018. Bonn [u. a.]: SVR-Gesundheit. URL: https://www.svr-gesundheit.de/fileadmin/user_upload/Gutachten/2018/SVR-Gutachten_2018_WEBSEITE.pdf (abgerufen am: 23.07.2018).
- 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 (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://effectivehealthcare.ahrq.gov/sites/default/files/pdf/public-reporting-quality-improvement_research.pdf (abgerufen am: 13.03.2019).
- 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.
- Weldring, T; Smith, SM (2013): Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Services Insights* 6: 61-68. DOI: 10.4137/HSI.S11093.

- 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.
- WHO [World Health Organization] (2000): The World Health Report 2000. Health Systems: Improving Performance. Geneva: WHO. ISBN: 92-4-156198-X. URL: http://www.who.int/whr/2000/en/whr00_en.pdf (abgerufen am: 20.06.2018).
- WHO [World Health Organization] (2003): List of Member States by WHO region and mortality stratum. In: WHO; Hrsg.: *The World Health Report 2003: Shaping the Future*. Geneva, CH: WHO, 182-184. ISBN: 92 4 156243 9. URL: http://www.who.int/whr/2003/en/whr03_en.pdf (abgerufen am: 28.02.2019).
- 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.
- Wilbanks, BA; Moss, J (2018): Evidence-Based Guidelines for Interface Design for Data Entry in Electronic Health Records. *CIN – Computers, Informatics, Nursing* 36(1): 35-44. DOI: 10.1097/cin.0000000000000387.
- 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).
- Wünschmann, R (2010): Qualitätsmanagement in der vertragsärztlichen Versorgung. Baden-Baden: Nomos. ISBN: 978-3-8329-5179-5.