



Institut für Qualitätssicherung und  
Transparenz im Gesundheitswesen

# Methodische Grundlagen

Version 2.0

Stand: 27. April 2022

---

# Impressum

**Thema:**

Methodische Grundlagen. Version 2.0

**Ansprechpartner:**

Dr. Dennis Boywitt

**Stand:**

27. April 2022

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

<https://www.iqtig.org>

# Übersicht

Tabellenverzeichnis.....	11
Abbildungsverzeichnis.....	12
Abkürzungsverzeichnis.....	13
Das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen.....	14
Über die „Methodischen Grundlagen“ .....	15
<b>Teil A: Grundlagen der Qualitätssicherung und -förderung.....</b>	<b>18</b>
1 Qualität der Gesundheitsversorgung .....	19
2 Grundlagen der gesetzlichen Qualitätssicherung .....	27
3 Patientenzentrierung in der Qualitätssicherung.....	42
4 Grundlagen der Qualitätsmessung.....	48
<b>Teil B: Entwicklung und Anpassung von Qualitätssicherungsverfahren .....</b>	<b>53</b>
5 Umsetzbarkeitsstudie für Qualitätsmessungen .....	54
6 Entwicklung von dokumentations- und sozialdatenbasierten Qualitätsindikatoren.....	58
7 Entwicklung von Patientenbefragungen im Rahmen von Qualitätssicherungsverfahren ..	73
8 Anpassung indikatorbasierter Instrumente der Qualitätssicherung.....	90
<b>Teil C: Methodische Elemente.....</b>	<b>97</b>
9 Informationsbeschaffung .....	98
10 Beratung durch externe Expertinnen und Experten .....	105
11 Auswahl von Inhalten bei der Entwicklung von Qualitätsmessungen .....	112
12 Beteiligungsverfahren gemäß § 137a Abs. 7 SGB V .....	117
13 Eignung von Qualitätsindikatoren.....	119
14 Eignung von Qualitätsindikatorensets .....	136
15 Veröffentlichung von leistungserbringerbezogenen Indikator- und Kennzahlresultaten.....	140
16 Bewertungskonzepte und Referenzbereiche .....	142
17 Bewertung von Indikatorergebnissen auf Systemebene .....	150
18 Stichprobenziehung und Vollerhebung.....	154
19 Prüfung der Qualität von QS-Dokumentationsdaten.....	158
20 Methoden der Datenanalyse.....	162
Glossar.....	176

<b>Anhang</b> .....	<b>180</b>
Anhang A: Datenbasierte Klassifikation von Leistungserbringerergebnissen .....	181
Anhang B: Illustration der statistisch signifikanten Klassifikation als Bewertungsart .....	192
Literatur.....	194

# Inhaltsverzeichnis

Tabellenverzeichnis.....	11
Abbildungsverzeichnis.....	12
Abkürzungsverzeichnis.....	13
Das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen.....	14
Über die „Methodischen Grundlagen“ .....	15
<b>Teil A: Grundlagen der Qualitätssicherung und -förderung.....</b>	<b>18</b>
1 Qualität der Gesundheitsversorgung .....	19
1.1 Rahmenkonzept für Qualität.....	19
1.2 Qualität auf Systemebene.....	24
2 Grundlagen der gesetzlichen Qualitätssicherung .....	27
2.1 Begriff der Qualitätssicherung .....	27
2.2 Rechtliche Grundlagen der gesetzlichen Qualitätssicherung .....	28
2.3 Theoretische Grundlagen der Qualitätssicherung .....	29
2.4 Wirkmechanismen zur Steigerung der Versorgungsqualität .....	32
2.4.1 Förderung der Verbesserungsbestrebungen der Leistungserbringer .....	34
2.4.2 Anreize für Leistungserbringer .....	35
2.4.3 Auswahlentscheidungen durch Prinzipale .....	37
2.5 Rahmenmodell für die Aufgaben von Qualitätssicherung.....	38
3 Patientenzentrierung in der Qualitätssicherung.....	42
3.1 Verständnis von Patientenzentrierung .....	42
3.2 Adressierung patientenrelevanter Qualitätsaspekte und -merkmale .....	44
3.3 Methoden zur Förderung der Patientenzentrierung .....	45
3.3.1 Analyse wissenschaftlicher Erkenntnisse zur Patientenperspektive .....	46
3.3.2 Beteiligung von Patientinnen und Patienten .....	46
3.3.3 Beteiligung von Patientenvertreterinnen und Patientenvertretern .....	47
4 Grundlagen der Qualitätsmessung.....	48
4.1 Methoden zur Messung von Qualität .....	48
4.2 Datenquellen.....	50
4.3 Qualitätsmessung als formatives Messmodell .....	51

<b>Teil B: Entwicklung und Anpassung von Qualitätssicherungsverfahren .....</b>	<b>53</b>
5	Umsetzbarkeitsstudie für Qualitätsmessungen .....
5.1	Ziel und Fragestellungen .....
5.2	Vorgehen.....
5.2.1	Beschreibung von Patientenpopulation, interessierenden Versorgungsereignissen und adressierten Leistungserbringern.....
5.2.2	Identifizierbarkeit der interessierenden Versorgungsereignisse .....
5.2.3	Für die Qualitätsmessung verfügbare Informationen.....
5.2.4	Neue Datenquellen und Datenflüsse .....
5.3	Ergebnis der Umsetzbarkeitsstudie .....
6	Entwicklung von dokumentations- und sozialdatenbasierten Qualitätsindikatoren.....
6.1	Konkretisierung der Qualitätsaspekte zu Qualitätsmerkmalen .....
6.1.1	Sicherstellung der Bedeutung der Qualitätsmerkmale .....
6.1.2	Berücksichtigung von Verbesserungspotenzial und Zuschreibbarkeit der Verantwortung.....
6.2	Operationalisierung der Qualitätsmerkmale .....
6.2.1	Fall- und einrichtungsbezogene QS-Dokumentation .....
6.2.2	Sozialdaten bei den Krankenkassen .....
6.2.3	Entwicklung von Auswertungsalgorithmen .....
6.3	Zusammenführung zum Qualitätsindikatorenset .....
6.4	Entwicklung von Empfehlungen zur Verfahrensumsetzung .....
6.5	Machbarkeitsprüfung für Dokumentation und Auswertung .....
6.5.1	Gegenstand der Machbarkeitsprüfung .....
6.5.2	Ablauf der Machbarkeitsprüfung .....
6.5.3	Datenauswertung und Optimierung der Qualitätsindikatoren.....
7	Entwicklung von Patientenbefragungen im Rahmen von Qualitätssicherungsverfahren ..
7.1	Faktenorientierte Befragungsansätze.....
7.1.1	Patient-Reported Outcome Measures (PROMs).....
7.1.2	Patient-Reported Experience Measures (PREMs).....
7.2	Ablauf der Fragebogen- und Qualitätsindikatorenentwicklung .....
7.2.1	Literaturrecherche und -analyse .....
7.2.2	Fokusgruppen und Interviews.....
7.2.3	Entwicklung der Qualitätsmerkmale.....

7.2.4	Expertenberatung zu den Qualitätsmerkmalen.....	80
7.2.5	Itementwicklung.....	80
7.2.6	Pretestung .....	81
7.2.7	Stichprobenziehung und Fallzahlenberechnung .....	85
7.2.8	Konstruktion der Qualitätsindikatoren .....	85
7.2.9	Umsetzung der Eignungskriterien .....	86
7.2.10	Risikoadjustierung von Qualitätsindikatoren der Patientenbefragung .....	88
8	Anpassung indikatorbasierter Instrumente der Qualitätssicherung.....	90
8.1	Hinweise auf Anpassungsbedarf .....	92
8.2	Prüfung von Anpassungsbedarf .....	92
8.3	Empfehlungen an den G-BA .....	94
<b>Teil C: Methodische Elemente.....</b>		<b>97</b>
9	Informationsbeschaffung .....	98
9.1	Orientierende Recherche.....	98
9.2	Systematische Recherche .....	99
9.2.1	Recherche nach Leitlinien .....	99
9.2.2	Recherche nach systematischen Übersichtsarbeiten .....	100
9.2.3	Recherche nach Primärliteratur .....	101
9.3	Evidenzaufbereitung und kritische Bewertung relevanter Publikationen.....	102
9.3.1	Evidenzaufbereitung orientiert an der Methodik eines Scoping Reviews .	102
9.3.2	Evidenzaufbereitung mit kritischer Bewertung der eingeschlossenen Publikationen .....	102
10	Beratung durch externe Expertinnen und Experten .....	105
10.1	Begriffsdefinition „Expertin/Experte“ .....	105
10.2	Beratungsgegenstände .....	105
10.3	Methodische Funktion der Beratung durch Expertinnen und Experten.....	107
10.4	Voraussetzungen für die Tätigkeit als Expertin oder Experte.....	108
10.5	Beratung durch Expertengruppen .....	110
11	Auswahl von Inhalten bei der Entwicklung von Qualitätsmessungen .....	112
11.1	Qualitätsaspekte innerhalb eines Themenbereichs.....	112
11.2	Qualitätsmerkmale als Basis der Indikatorenentwicklung.....	115

12	Beteiligungsverfahren gemäß § 137a Abs. 7 SGB V .....	117
13	Eignung von Qualitätsindikatoren .....	119
13.1	Grundsätze zur Eignungsbeurteilung .....	119
13.2	Eignungskriterien des Qualitätsziels .....	122
13.2.1	Bedeutung für die Patientinnen und Patienten .....	122
13.2.2	Zusammenhang mit einem unmittelbar patientenrelevanten Merkmal... ..	123
13.2.3	Potenzial zur Verbesserung.....	124
13.2.4	Brauchbarkeit für den Handlungsanschluss.....	125
13.2.5	Beeinflussbarkeit durch den Leistungserbringer .....	126
13.3	Eignungskriterien des Messverfahrens .....	128
13.3.1	Objektivität der Messung .....	129
13.3.2	Datenqualität .....	130
13.3.3	Reliabilität der Messung.....	130
13.3.4	Validität der Messung .....	131
13.3.5	Praktikabilität der Messung .....	132
13.4	Eignungskriterien des Bewertungskonzepts .....	133
13.4.1	Angemessenheit des Referenzbereichs .....	133
13.4.2	Klassifikationsgüte.....	134
13.4.3	Angemessenheit der Risikoadjustierung.....	135
14	Eignung von Qualitätsindikatorensets .....	136
14.1	Definition „Qualitätsindikatorenset“ .....	136
14.2	Eignungskriterium Inhaltsvalidität .....	136
14.2.1	Herstellung von Inhaltsvalidität .....	138
14.2.2	Retrospektive Einschätzung .....	138
14.3	Aufwände der Messung .....	139
15	Veröffentlichung von leistungserbringerbezogenen Indikator- und Kennzahlergebnissen.....	140
16	Bewertungskonzepte und Referenzbereiche .....	142
16.1	Funktion von Referenzbereichen .....	142
16.2	Arten von Referenzbereichen .....	142
16.3	Entwicklung kriteriumsbezogener Referenzbereiche .....	144
16.3.1	Fachliche Beurteilung des Erreichbaren.....	144
16.3.2	Geforderter Erreichungsgrad .....	147

16.3.3	Festlegung der empfohlenen Referenzbereichsgrenze .....	147
16.3.4	Prüfung der Angemessenheit.....	148
16.4	Sentinel Events.....	148
17	Bewertung von Indikatorergebnissen auf Systemebene .....	150
17.1	Definition des besonderen Handlungsbedarfs .....	150
17.2	Bewertungskriterien .....	150
17.2.1	Leitkriterien .....	150
17.2.2	Weitere Kriterien.....	152
17.3	Beratung durch Expertinnen und Experten .....	153
18	Stichprobenziehung und Vollerhebung.....	154
18.1	Ebenen der Stichprobenziehung und Vollerhebung .....	154
18.2	Kriterien für Stichprobenziehung und Vollerhebungen .....	155
18.3	Stichprobengrößenberechnungen .....	156
19	Prüfung der Qualität von QS-Dokumentationsdaten .....	158
19.1	Prüfung bei Dateneingabe und -annahme.....	159
19.2	Statistische Prüfung nach Datenannahme.....	159
19.3	Datenabgleich mit Referenzquelle.....	161
20	Methoden der Datenanalyse.....	162
20.1	Rahmenkonzept für die statistische Auswertungsmethodik .....	162
20.1.1	Zielsetzung.....	163
20.1.2	Stichprobenart .....	164
20.1.3	Berechnungsart .....	165
20.1.4	Bewertungsart.....	166
20.2	Risikoadjustierung.....	167
20.2.1	Stratifizierung .....	168
20.2.2	Additive Scores .....	169
20.2.3	Statistische Modellierung.....	170
20.3	Auswertung von Follow-up-Indikatoren .....	172
20.3.1	Aktives und passives Follow-up.....	172
20.3.2	Lexis-Diagramme .....	173
20.3.3	Auswertung nach Kohortensichtweise und nach Periodensichtweise .....	174

Glossar .....	176
<b>Anhang .....</b>	<b>180</b>
Anhang A: Datenbasierte Klassifikation von Leistungserbringerergebnissen .....	181
Anhang A.1: Komponenten der Bewertungsart .....	181
Anhang A.2: Herleitung von Entscheidungsregeln anhand der Verlustfunktion .....	182
Anhang A.2.1: Rechnerische Klassifikation .....	184
Anhang A.2.2: Statistisch signifikante Klassifikation .....	185
Anhang A.2.3: Statistisch relevante Klassifikation .....	186
Anhang A.3: Illustration der Konsequenzen unterschiedlicher Entscheidungsregeln zur Klassifikation .....	188
Anhang A.3.1: Schwellenwerte der Entscheidungsregeln .....	188
Anhang A.3.2: Sensitivität und Spezifität der Entscheidungsregeln .....	189
Anhang A.4: Limitationen .....	191
Anhang B: Illustration der statistisch signifikanten Klassifikation als Bewertungsart .....	192
Literatur .....	194

## Tabellenverzeichnis

Tabelle 1: Eignungskriterien für Qualitätsindikatoren und dazugehörige Maßnahmen und Informationsquellen der Patientenbefragung .....	87
Tabelle 2: Fiktive Beispiele für die methodischen Elemente der Qualitätsmessung.....	115
Tabelle 3: Beispiele für unmittelbar und mittelbar patientenrelevante Qualitätsmerkmale...	123
Tabelle 4: Im Krankenhaus verstorbene Patientinnen und Patienten mit isoliertem koronarchirurgischem Eingriff geschichtet nach Dringlichkeit des Eingriffs im Jahr 2015..	169
Tabelle 5: Vier-Felder-Tafel möglicher Aufwandsannahmen des binären Klassifikationsproblems .....	183
Tabelle 6: Vier-Felder-Tafel möglicher Aufwandsannahmen des binären Klassifikationsproblems im Kontext der Qualitätsförderung.....	184
Tabelle 7: Aufwandsannahmen für die statistisch relevante Klassifikationsmethode .....	187
Tabelle 8: Rechenregeln eines Beispielindikators, angelehnt an den Qualitätsindikator „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ .....	192

# Abbildungsverzeichnis

Abbildung 1: Dimensionen des IQTIG-Rahmenkonzepts für Qualität .....	23
Abbildung 2: Beziehungsgefüge im deutschen Gesundheitswesen mit Prinzipal-Agenten- Beziehungen auf verschiedenen Ebenen.....	30
Abbildung 3: Wirkmechanismen zur Steigerung der Versorgungsqualität.....	33
Abbildung 4: Rahmenmodell für die Aufgaben von Qualitätssicherung .....	39
Abbildung 5: Angenommene Kausalrichtung bei formativen und reflektiven Messmodellen...	52
Abbildung 6: Entwicklungsschritte für dokumentations- und sozialdatenbasierte Qualitätsindikatorensets.....	58
Abbildung 7: Ablauf einer Machbarkeitsprüfung.....	68
Abbildung 8: Entwicklungsschritte einer Patientenbefragung .....	76
Abbildung 9: Hierarchische Struktur von Qualitätsaspekten, Qualitätsmerkmalen und Items und hierarchische Aggregation zu Qualitätsindikatoren .....	86
Abbildung 10: Vorgehen bei Anpassungen von Qualitätsmessungen .....	91
Abbildung 11: Strukturierung eines Themenbereichs mit Qualitätsdimensionen und -aspekten am Beispiel Herzkatheteruntersuchungen und PCI .....	113
Abbildung 12: Überblick über die Eignungskriterien für Qualitätsindikatoren .....	120
Abbildung 13: Schematische Darstellung des Ausmaßes der leistungserbringerbedingten und nicht leistungserbringerbedingten Einflüsse auf den Indikatorwert anhand von Beispielen fallbasierter Qualitätsindikatoren .....	126
Abbildung 14: Schritte und zugehörige Eignungskriterien bei der Ermittlung und Aufarbeitung von Informationen für die Berechnung von Indikatorwerten .....	128
Abbildung 15: Strukturierung der Versorgungsqualität im interessierenden Themenbereich mit Qualitätsdimensionen, Qualitätsaspekten und Qualitätsmerkmalen und Operationalisierung durch das Qualitätsindikatorensset.....	137
Abbildung 16: Komponenten eines kriteriumsbezogenen Referenzbereichs; vereinfachte Darstellung ohne Berücksichtigung von Stochastizität.....	148
Abbildung 17: Beispielhafte Darstellung der Daten eines 1-Jahres-Follow-up-Indikators mit 100 Fällen in Form von Lebenslinien in einem Lexis-Diagramm.....	173
Abbildung 18: Lexis-Diagramme für einen 1-Jahres-Follow-up-Indikator in Kohortensichtweise und in Periodensichtweise .....	174
Abbildung 19: Vereinfachtes Einflussdiagramm für einen ratenbasierten Qualitätsindikator. 181	
Abbildung 20: Funnelplot für die Klassifikation der drei Entscheidungsregeln für einen Indikator mit Referenzbereich [0–15 %].....	188
Abbildung 21: Vergleich von Sensitivität und 1-Spezifität für die drei Methoden der Klassifikation .....	190
Abbildung 22: Dichte der A-posteriori-Wahrscheinlichkeit für die zugrunde liegende Rate $\theta$ bei einem Leistungserbringer mit $J = 20$ Fällen im Nenner und $o = 5$ Fällen im Zähler ...	193

## Abkürzungsverzeichnis

Abkürzung	Erläuterung
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V.
DeQS-RL	Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung
G-BA	Gemeinsamer Bundesausschuss
GKV	gesetzliche Krankenversicherung
ICD	Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme ( <i>International Statistical Classification of Diseases and Related Health Problems</i> )
OPS	Operationen- und Prozedurenschlüssel
PCI	perkutane Koronarintervention ( <i>percutaneous coronary intervention</i> )
PICO	Population, Intervention, Comparison, Outcome
plan. QI-RL	Richtlinie zu planungsrelevanten Qualitätsindikatoren
PREM	Patient-Reported Experience Measure
PROM	Patient-Reported Outcome Measure
QS	Qualitätssicherung
SGB	Sozialgesetzbuch
SGB V	Fünftes Buch Sozialgesetzbuch
SMR	standardisiertes Mortalitäts- oder Morbiditätsverhältnis

## Das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen

Das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) ist das zentrale wissenschaftliche Institut für die gesetzlich verankerte Qualitätssicherung im Gesundheitswesen in Deutschland. Nach § 137a Abs. 1 Sozialgesetzbuch (SGB) Fünftes Buch (V) unterstützt das IQTIG als fachlich unabhängiges und wissenschaftliches Institut den Gemeinsamen Bundesausschuss (G-BA) bei der Erfüllung seiner Aufgaben als untergesetzlicher Normgeber für die Qualitätssicherung im System der gesetzlichen Krankenversicherung.

Das IQTIG wird vom G-BA mit der Entwicklung von Maßnahmen der Qualitätssicherung sowie von Maßnahmen zur Verbesserung der Transparenz über die ambulante und stationäre Versorgung beauftragt (§ 137b SGB V). Das IQTIG übermittelt seine Arbeitsergebnisse an den G-BA als Empfehlungen in Form von wissenschaftlichen Berichten. Seine wissenschaftlichen Berichte veröffentlicht das IQTIG auf seiner Website [www.iqtig.org](http://www.iqtig.org).

Darüber hinaus beteiligt sich das IQTIG gemäß § 137a Abs. 3 Nr. 3 SGB V auch an der Umsetzung von Maßnahmen zur Sicherung und Steigerung der Versorgungsqualität. Weitere Informationen über das IQTIG finden sich unter [www.iqtig.org](http://www.iqtig.org).

# Über die „Methodischen Grundlagen“

Die „Methodischen Grundlagen“ erläutern die Methoden und Kriterien, die das IQTIG bei der Erfüllung seines gesetzlichen Auftrags zugrunde legt. Darin beschreibt das IQTIG seine allgemeinen Methoden für die Entwicklung, Weiterentwicklung und Anpassung von Maßnahmen der Qualitätssicherung durch das IQTIG. Damit tragen die „Methodischen Grundlagen“ zur Transparenz über die Arbeit des IQTIG sowie über das Zustandekommen seiner Arbeitsergebnisse bei. Die „Methodischen Grundlagen“ werden fortlaufend weiterentwickelt, um sie den sich kontinuierlich ändernden Erfordernissen der Qualitätssicherung und der Entwicklung der Wissenschaften anzupassen.

Eine fachlich korrekte Beschreibung wissenschaftlicher Methoden erfordert auch Fachbegriffe und -konzepte. Die „Methodischen Grundlagen“ richten sich daher vor allem an Fachwissenschaftlerinnen und Fachwissenschaftler und setzen ein gewisses Maß an Fachwissen voraus. Um jedoch auch weiteren interessierten Personen einen Zugang zu den Arbeitsweisen des IQTIG zu ermöglichen, wurde ein möglichst hohes Maß an Allgemeinverständlichkeit angestrebt.

## Schriftliches Stellungnahmeverfahren

Das IQTIG führt zu jeder Version der „Methodischen Grundlagen“ ein schriftliches Stellungnahmeverfahren durch. Der Entwurf für die Version 2.0 der „Methodischen Grundlagen“ wurde zu diesem Zweck am 16. August 2021 auf der Website des IQTIG veröffentlicht. Bis zum 25. Oktober 2021 hatten alle interessierten Organisationen und Einzelpersonen die Möglichkeit, zu den Methoden des IQTIG schriftlich Stellung zu nehmen.

Auf Grundlage der Stellungnahmen wurden die „Methodischen Grundlagen“ geprüft, überarbeitet und einschließlich entsprechender Würdigung der Stellungnahmen als „Methodische Grundlagen. Version 2.0“ auf der Website des IQTIG veröffentlicht.

## Inhalte und Struktur

Die „Methodischen Grundlagen“ beschreiben diejenigen Methoden, die vom IQTIG zur Erfüllung seines gesetzlichen Auftrags zugrunde gelegt werden, und gliedern sich in drei Teile:

- In Teil A werden rechtliche Grundlagen sowie grundlegende Begriffe, Konzepte und Methoden des IQTIG beschrieben.
- In Teil B werden die Methoden und Kriterien beschrieben, die das IQTIG bei der Entwicklung, Anpassung und Weiterentwicklung von Instrumenten und Maßnahmen der Qualitätssicherung zugrunde legt. Der Schwerpunkt liegt hier auf den Standardprodukten des IQTIG, die regelhaft durch den G-BA beauftragt werden, wie etwa Qualitätsindikatorensets auf Basis der unterschiedlichen Datenquellen.
- In Teil C werden methodische Elemente und Kriterien beschrieben, die an verschiedenen Stellen der Arbeit des IQTIG eine Rolle spielen, wie etwa statistische Methoden oder Eignungskriterien für Qualitätsmessungen.

## Was ist neu in Version 2.0?

Im Vergleich zur Version 1.1 wurden die „Methodischen Grundlagen“ an vielen Stellen mit dem Ziel einer besseren Allgemeinverständlichkeit grundlegend überarbeitet und insgesamt gestrafft. Als größere inhaltliche Änderungen sind zu nennen:

- Teil A wurde um ein Rahmenmodell für die verschiedenen Aufgaben von Qualitätssicherung und die grundlegenden Wirkmechanismen zur Steigerung der Versorgungsqualität ergänzt (Abschnitte 2.4 und 2.5). Zukünftig wird das IQTIG auch Empfehlungen dahingehend aussprechen, welche Qualitätssicherungs- und -weiterentwicklungsmaßnahmen, neben indikatorbasierten Maßnahmen, für einen gegebenen Themenbereich fachwissenschaftlich geeignet erscheinen. Das Rahmenmodell in Teil A stellt dafür eine wichtige Grundlage dar, indem es die vom IQTIG zugrunde gelegten Annahmen zusammenfasst. Sobald die dafür notwendigen Entwicklungen abgeschlossen sind, findet die Beschreibung der relevanten Methoden und Kriterien Eingang in die nächste Version der „Methodischen Grundlagen“.
- Das Produkt „Konzeptstudie“ in der bisherigen Form wird zugunsten einer differenzierteren Vorgehensweise weiterentwickelt. Zukünftig wird das IQTIG bei Bedarf vor der Entwicklung indikatorbasierter Qualitätssicherungsmaßnahmen eine Empfehlung aussprechen, ob ein auf wenige Aspekte fokussiertes Qualitätsindikatorenset entwickelt werden sollte oder ob das zu entwickelnde Qualitätsindikatorenset einen ganzen Themenbereich abdecken sollte. Die Durchführung einer Konzeptstudie in der bisherigen Form ist vor der Entwicklung eines Qualitätsindikatorensets aufgrund von Anpassungen an der Entwicklungsmethodik für Qualitätsindikatorensets (siehe Kapitel 6 und 7) nicht mehr nötig. Die neue Vorgehensweise soll es dem G-BA zukünftig ermöglichen, ein evidenzbasiert bekanntes Qualitätsdefizit frühzeitiger adressieren zu können.
- Das neue Produkt „Umsetzbarkeitsstudie für Qualitätsmessungen“ adressiert die bisher durch das entfallende Produkt „Konzeptstudie“ abgedeckte Fragestellung, ob ein Qualitätsindikatorenset in einem neuen Themenbereich vorab umsetzbar erscheint (Kapitel 5). Anders als die bisherige Konzeptstudie beinhaltet dieses Produkt kein Qualitätsmodell für die Auswahl von Inhaltsbereichen für die Indikatorenentwicklung in dem jeweiligen Themenbereich. Die Entwicklung eines solchen Qualitätsmodells nimmt das IQTIG zukünftig im Rahmen der Indikatorenentwicklung vor (siehe Kapitel 6 und 7).
- Die Machbarkeitsprüfung für Dokumentationsbögen wurde als elementarer Bestandteil der Entwicklung von Qualitätsindikatoren im Sinne eines Pretests und in Analogie zur Methodik für die Entwicklung von Patientenbefragungen in die Entwicklung sozialdaten- und dokumentationsbasierter Qualitätsindikatoren integriert (Abschnitt 6.5).
- Das Kapitel über das Vorgehen bei der Anpassung laufender Qualitätssicherungsverfahren an neue Erkenntnisse wurde mit dem Ziel einer besseren Allgemeinverständlichkeit neu strukturiert (Kapitel 8).
- Die „Methodischen Grundlagen“ wurden um eine Beschreibung des Vorgehens für das gesetzlich vorgeschriebene Beteiligungsverfahren für Entwicklungen durch das IQTIG ergänzt (Kapitel 12).

- Eine Beschreibung der Methoden für die Durchführung von Fokusgruppen und Leitfadeninterviews entfällt, da diese Methoden der einschlägigen Literatur entnommen werden können.
- Die Systematik der Eignungskriterien für Qualitätsindikatoren wurde weiterentwickelt und konsequenter auf die drei Komponenten eines Qualitätsindikators ausgerichtet (Kapitel 13).
- Teil C wurde auch um ein Konzept zur Beurteilung der Güte von Qualitätsindikatorensets ergänzt (Kapitel 14).
- Die Methodik für die Setzung von festen Referenzbereichen für Qualitätsindikatoren wurde weiterentwickelt (Kapitel 16).
- Die Beschreibung der Methodik für die Feststellung besonderen Handlungsbedarfs wurde in Teil C verortet und gestrafft (jetzt Kapitel 17). Eine Weiterentwicklung dieser Methodik ist für Folgeversionen der „Methodischen Grundlagen“ vorgesehen.
- Die statistischen Methoden zur Auswertung von Qualitätsindikatorergebnissen wurden zu einem grundlegenden statistischen Rahmenkonzept weiterentwickelt (Abschnitt 20.1).

# **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 Medicine,<sup>1</sup> 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

---

<sup>1</sup> 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 Dimensionen<sup>2</sup> 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, Institute of Medicine/Committee on Quality of Health Care in America 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 anlehnt (Institute of Medicine/Committee on Quality of Health Care in America 2001: 39 f.), 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).

---

<sup>2</sup> 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. Donabedian (2003: 47) 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 [...]“

Um diese Ergebnisse zu erreichen, müssen die Strukturen und Prozesse der Versorgung dem aktuellen Wissensstand entsprechen und die Wahrscheinlichkeit angestrebter Ergebnisse erhöhen (DIN EN 15224 2017: 78, JCAHO 1990: 9, Institute of Medicine/Committee on Quality of Health Care in America 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 (biopsychosoziale Perspektive, Scholl et al. 2014).

- **Patientensicherheit** (*safety*)

Patientensicherheit ist die Abwesenheit unerwünschter, schädlicher Ereignisse im Rahmen der Versorgung (Kohn et al. 2000: 4, APS [2016]). 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 nimmt diese als Individuen wahr, berücksichtigt deren individuelle Bedürfnisse und Bedarfe, Präferenzen und Werte und stellt sicher, dass alle Versorgungsentscheidungen durch diese geleitet werden (Institute of Medicine/Committee on Quality of Health Care in America 2001: 40) (siehe Kapitel 3). Dazu gehören insbesondere der respekt- und würdevolle Umgang von Gesundheitsprofessionen mit den Patientinnen und Patienten, ausgehend vom Angebot einer partnerschaftlichen Beziehungsgestaltung, die bedarfsgerechte Information, Kommunikation und Unterstützung (Empowerment), das Angebot zur aktiven Beteiligung an Versorgungsprozessen und Entscheidungen (Shared Decision Making), die Berücksichtigung der Diversität von Patientinnen und Patienten und ihrer Lebenslagen und Lebensphasen sowie die Einbeziehung des sozialen Umfelds bei der Versorgungsgestaltung. Diese Aspekte werden in der Literatur mit dem Begriff *responsiveness* (Arah et al. 2006, WHO 2000: 31, Pluut 2016) oder ähnlich auch *acceptability* (Maxwell 1992) beschrieben.<sup>3</sup>

- **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“ (Institute of Medicine/Committee on Quality of Health Care in America 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*)

---

<sup>3</sup> 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.

(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 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 niedrigschwelligen 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, Institute of Medicine/Committee on Quality of Health Care in America 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 (Institute of Medicine/Committee on Quality of Health Care in America 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.

Abbildung 1 fasst die Qualitätsdimensionen für die leistungserbringerbezogene Qualitätssicherung des IQTIG-Rahmenkonzepts zusammen.

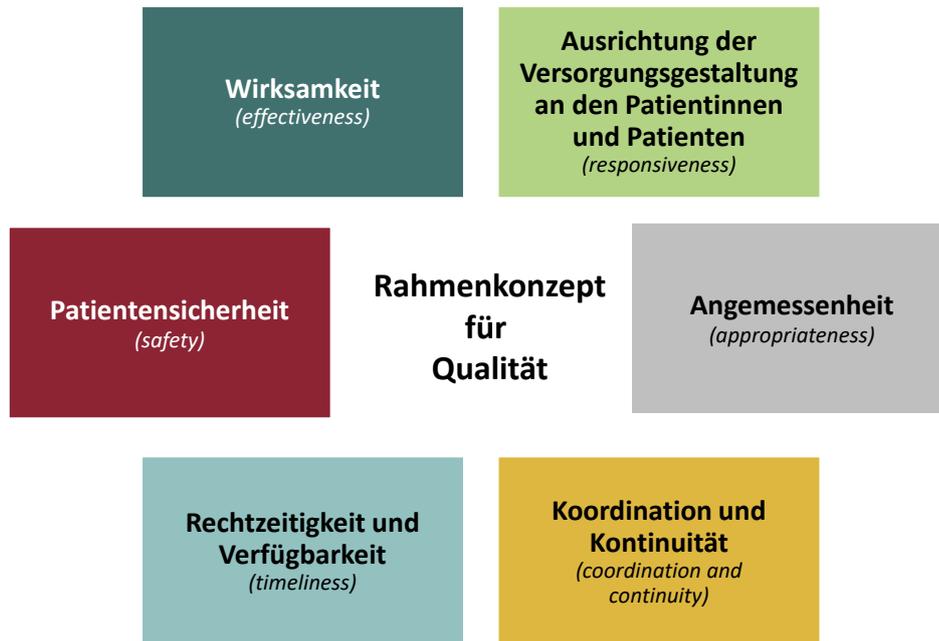


Abbildung 1: Dimensionen des IQTIG-Rahmenkonzepts für Qualität

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 „Wirksamkeit“ und „Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten“. Qualität der Gesundheitsversorgung wird demnach als mehrdimensionales Konstrukt verstanden.

Die Qualitätsdimensionen des Rahmenkonzepts dienen als Raster, um zu prüfen, ob ein Qualitätsmodell oder ein Qualitätsindikatorensatz (siehe Kapitel 11) alle Dimensionen adressiert oder ob unerwünschte „blinde Flecke“ bestehen. Der Abgleich der Inhalte eines Qualitätsmodells bzw. eines Qualitätsindikatorensatzes mit den Qualitätsdimensionen des Rahmenkonzepts erfolgt vor allem bei der Entwicklung, aber auch bei der Prüfung von Anpassungsbedarf laufender QS-Verfahren.

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 2.2).
- 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 jedoch keine patientenzentrierten Versorgungsziele, sondern Methoden, um diese zu erreichen.

## 1.2 Qualität auf Systemebene

Qualität im Gesundheitswesen kann auf verschiedenen Ebenen betrachtet werden. Die gesetzliche, 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 Mikroebene genannt). Auf übergeordneter Ebene kann die Qualität der Gesundheitsversorgung auch für die Gesamtpopulation betrachtet werden, die sogenannte Systemebene (Makroebene). Qualität auf Systemebene wird hier definiert als die Qualität der Gesundheitsversorgung für die Gesamtpopulation. 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 Anforder-

derungen, wie beispielsweise an 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 zu Lasten 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, Institute of Medicine/Committee on Quality of Health Care in America 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 Gesundheitswesen 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“ (sogenannter „new universalism“, siehe 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. Das IQTIG formuliert im Rahmen der leistungserbringerbezogenen Qualitätssicherung daher keine Anforderungen an die Leistungserbringer bezüglich der Kosten bzw. der Effizienz der Gesundheitsversorgung und betrachtet diese wie die Autoren des Modells der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (Organisation for Economic Co-operation and Development, OECD) (Arah et al. 2006) getrennt von 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 (Institute of Medicine/Committee on Quality of Health Care in America 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, das finanzielle Anreize zur Bevorzugung

bestimmter Personengruppen vermeiden soll. Versorgungsgerechtigkeit wird daher in erster Linie als Eigenschaft des Systems und nicht als Eigenschaft einzelner Leistungserbringer betrachtet, also als eine Dimension von Systemqualität (Campbell et al. 2000). Sofern sich bei der Entwicklung von Qualitätsmessungen durch das IQTIG dennoch leistungserbringerbezogene Qualitätsdefizite der Versorgungsgerechtigkeit zeigen, beispielsweise eine Benachteiligung bestimmter Patientengruppen bei der Bereitstellung von Versorgungsmaßnahmen, werden diese Aspekte der jeweiligen Qualitätsdimension (z. B. „Rechtzeitigkeit und Verfügbarkeit“) zugeordnet. Die zusätzlichen Anforderungen „Kosten“ bzw. „Effizienz“ und „Versorgungsgerechtigkeit“ liegen größtenteils außerhalb des Einflussbereichs der Leistungserbringer und richten sich daher an Adressaten auf Ebene des Gesundheitssystems (Makroebene oder Systemebene), beispielsweise an Einrichtungen der Gesundheitspolitik, Institutionen der Selbstverwaltung im Gesundheitswesen oder Berufsverbände.<sup>4</sup> Systemqualität umfasst also nicht nur die Versorgungsqualität der Gesamtheit der Leistungserbringer, sondern auch den Grad der Erfüllung dieser zusätzlichen Anforderungen.

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.

---

<sup>4</sup> 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 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.

## 2 Grundlagen der gesetzlichen Qualitätssicherung

Die Entwicklung und transparente Herleitung von Empfehlungen zur Qualitätssicherung und -verbesserung setzt ein Rahmenmodell voraus, das die Aufgaben der Qualitätssicherung, mögliche Wirkmechanismen von Qualitätssicherungsmaßnahmen sowie deren Voraussetzungen und Grundannahmen zusammenfasst. Nur auf Grundlage eines solchen Rahmenmodells ist es möglich, nachvollziehbare und widerspruchsfreie Empfehlungen für ein komplexes Thema wie die Qualitätssicherung der Gesundheitsversorgung zu geben. Ein solches Rahmenmodell wird im Folgenden anhand der Einordnung des Begriffs „Qualitätssicherung“ (Abschnitt 2.1), des rechtlichen Rahmens der gesetzlichen Qualitätssicherung sowie der Aufgaben des IQTIG (Abschnitt 2.2), der Beziehungen zwischen den verschiedenen Akteuren im Gesundheitswesen (Abschnitt 2.3) und der grundlegenden Wirkmechanismen von Qualitätssicherungsmaßnahmen (Abschnitt 2.4) beschrieben.

### 2.1 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 EN ISO 9000 ff. wird Qualitätssicherung als Teil des Qualitätsmanagements begriffen. Qualitätsmanagement umfasst danach mehrere Komponenten, wobei der Begriff „Qualitätssicherung“ den „Teil des Qualitätsmanagements [bezeichnet], der auf das Erzeugen von Vertrauen darauf gerichtet ist, dass Qualitätsanforderungen erfüllt werden“, 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 Darlegung der Qualität (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 im Sinne der Behebung von Qualitätsdefiziten. 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 gesetzlichen 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 Qualitätssicherung ist, sondern auch die Weiterentwicklung von Versorgungsqualität (Wünschmann 2010: 23 f.). So verpflichtet das SGB V die Leistungserbringer zur Weiterentwicklung der Versorgungsqualität entsprechend dem aktuellen Stand der wissenschaftlichen Erkenntnisse (§ 135a SGB V). Dieses Verständnis von Qualitätssicherung findet sich beispielsweise auch im Gesetz zur Reform der Strukturen der Krankenhausversorgung (Krankenhausstrukturgesetz, KHSG), das u. a. selektivvertragliche Anreize für höherwertige Qualitätsanforderungen in bestimmten Leistungsbereichen (Qualitätsverträge nach § 110a SGB V) vorsieht.

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 Gesamtheit der Aufgaben und Maßnahmen, durch die die Qualität der Gesundheitsversorgung transparent gemessen, dargestellt und gesteigert werden soll.

## 2.2 Rechtliche Grundlagen der gesetzlichen Qualitätssicherung

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 in den §§ 135 ff. die „Sicherung der Qualität der Leistungserbringung“. So hat der Gesetzgeber den G-BA mit der Festlegung und Durchsetzung einheitlicher (Mindest-)Anforderungen an die Struktur-, Prozess- und Ergebnisqualität für alle zugelassenen Krankenhäuser und für die vertragsärztliche Versorgung beauftragt (§§ 136 bis 136b und 137d SGB V), um sicherzustellen, dass Leistungen für die Patientinnen und Patienten dem jeweiligen Stand der wissenschaftlichen Erkenntnisse entsprechen und in der fachlich gebotenen Qualität erbracht werden (§§ 70 Abs. 1 und 135a Abs. 1 SGB V). Die verpflichtende Beteiligung der Leistungserbringer an einrichtungsübergreifenden Maßnahmen der Qualitätssicherung hat insbesondere zum Ziel, die Ergebnisqualität für die Patientinnen und Patienten zu verbessern (§ 135a Abs. 2 Nr. 1 SGB V).

Zur Erarbeitung der fachlichen Grundlagen für seine Regelungen beauftragt der G-BA nach § 137a Abs. 1 SGB V als fachlich unabhängiges und wissenschaftliches Institut das IQTIG mit der Entwicklung und Durchführung von einrichtungsübergreifenden QS-Verfahren. Das IQTIG arbeitet im Auftrag des G-BA an Maßnahmen zur Qualitätssicherung und zur Darstellung der Versorgungsqualität im Gesundheitswesen. Gemäß § 137a SGB V soll das IQTIG 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 G-BA 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 IQTIG unmittelbar mit Untersuchungen und Handlungsempfehlungen zu den oben genannten Aufgaben für den G-BA beauftragen. Das IQTIG leitet seine Arbeitsergebnisse gemäß § 137b SGB V dem G-BA als Empfehlungen zu. Dieser berücksichtigt sie im Rahmen seiner Aufgaben 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 Theoretische Grundlagen der Qualitätssicherung**

Die Gesundheitsversorgung entspricht nicht immer den in Kapitel 1 dargestellten Anforderungen. Dies äußert sich in Unterschieden in der Versorgung, die nicht auf Innovationen, Unterschiede in Patientenpräferenzen oder zufällige Einflüsse zurückzuführen sind, sondern unerwünschte Variation darstellen (Appleby et al. 2011, Corallo et al. 2014). Maßnahmen des Qualitätsmanagements und der Qualitätssicherung dienen dazu, diese unerwünschte Variation zu verringern und dadurch die Versorgungsqualität zu verbessern (Berwick 1991).

Für die Erklärung solcher Qualitätsunterschiede lassen sich verschiedene Modelle heranziehen und Gründe für die Notwendigkeit einer nicht nur einrichtungsinternen, sondern externen und verpflichtenden Qualitätssicherung ableiten. Als ein etablierter Erklärungsansatz beschreibt die Prinzipal-Agenten-Theorie ganz allgemein die Beziehung von Auftraggeber und Auftragnehmer und rückt die Bedeutung von Qualitätstransparenz für beide bei potenziell divergierenden Interessen in den Vordergrund (Hurley 2000, Mooney und Ryan 1993). Der Auftraggeber (Prinzipal) beauftragt einen Auftragnehmer seiner Wahl (Agent) mit der Erbringung von Leistungen, für die dem Auftraggeber meist das hierfür notwendige Wissen oder spezifische Ressourcen fehlen

(Hurley 2000, Holmstrom und Milgrom 1991). Der Prinzipal kann aufgrund dieser Informationsasymmetrie die Qualität der Leistungserbringung jedoch nicht überprüfen.

Diese mangelnde Qualitätstransparenz prägt auch das komplexe Gefüge von Beziehungen im deutschen Gesundheitswesen (BT-Drucksache 17/10323 vom 10.07.2021), die sich in Teilaspekten auch als Prinzipal-Agenten-Beziehungen auf mehreren Ebenen auffassen lassen. Dieses Gefüge von Beziehungen wird stark vereinfacht in Abbildung 2 dargestellt. Auf der Mikroebene geht es um die Beziehung zwischen Patientinnen/Patienten und Versorgenden (z. B. Ärztinnen und Ärzten, Pflegepersonal). Hier agieren Patientinnen und Patienten als Prinzipale, die versorgungsbezogene Aufgaben wie beispielsweise die Diagnose, Akutbehandlung oder Dauerversorgung von Erkrankungen an die Versorgenden übertragen. Im Vergleich mit anderen Anwendungsgebieten der Prinzipal-Agenten-Theorie liegen jedoch die Ziele von Prinzipal und Agent auf der Mikroebene im Gesundheitswesen deutlich näher zusammen, da davon ausgegangen wird, dass sich die Versorgenden den Interessen der Patientinnen und Patienten verpflichtet fühlen (Mooney und Ryan 1993). Dieser Aspekt wird beispielsweise durch die Stewardship-Theorie aufgegriffen (Van Slyke 2007), die annimmt, dass die Versorgenden grundsätzlich im Interesse der Patientinnen und Patienten handeln.

Auf der Mikroebene kann jedoch mangelnde Transparenz über Qualität z. B. die Qualifikation der Versorgenden, die medizinische Indikation für eine Behandlung und die Kenntnis von (alternativen) Behandlungsmöglichkeiten betreffen (Hurley 2000, Mooney und Ryan 1993). Daraus resultiert ein besonderes Schutzbedürfnis der Patientinnen und Patienten, für die Qualitätsmängel schwer zu identifizieren und gegenüber Leistungserbringern auch schwer geltend zu machen sind.

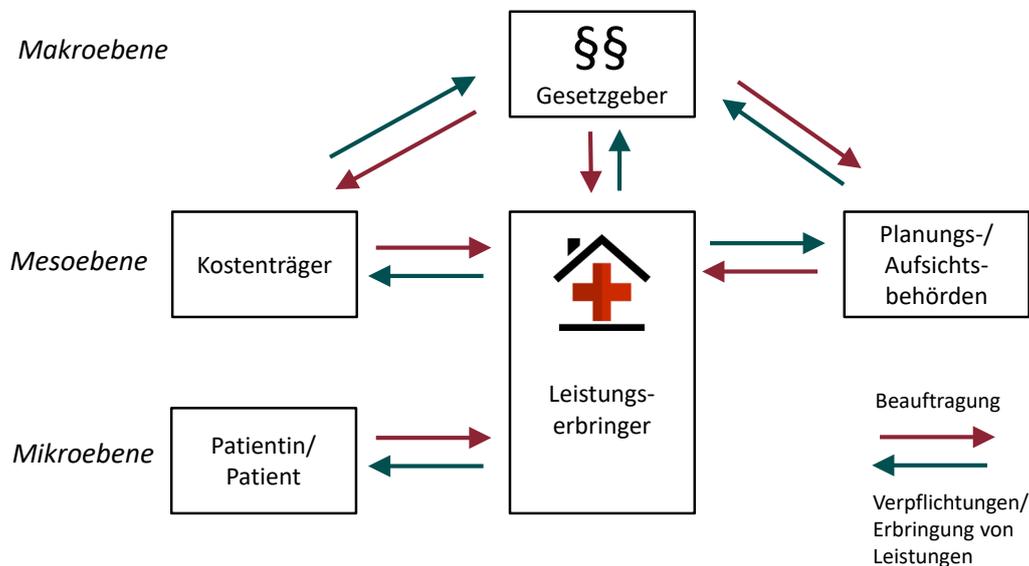


Abbildung 2: Beziehungsgefüge im deutschen Gesundheitswesen mit Prinzipal-Agenten-Beziehungen auf verschiedenen Ebenen

Auf der Mesoebene können Prinzipal-Agenten-Beziehungen aus der Sicht der Organisationen im Gesundheitswesen betrachtet werden. Dies betrifft insbesondere die Beziehungen zwischen

den jeweils für die Zulassung, Aufsicht und Vergütung zuständigen Stellen als Prinzipalen sowie den Leistungserbringern als Agenten. Auch auf der Mesoebene ist für die jeweiligen Prinzipale ohne Qualitätsmessung kaum ersichtlich, inwieweit Leistungserbringer die an sie gestellten Qualitätsanforderungen erfüllen (Blomqvist 1991, Sanderson et al. 2019).

Auf der Makroebene geht es mit dem SGB V und weiteren Bestimmungen um den rechtlichen Rahmen des Gesundheitssystems, in dem die Organisationen des Gesundheitswesens – im Auftrag des Gesetzgebers als Prinzipal – agieren. Mangelnde Qualitätstransparenz auf der Makroebene betrifft die Qualität auf Systemebene, das heißt den Grad, in dem Anforderungen an die Gesundheitsversorgung für die Gesamtpopulation erfüllt werden (vgl. Abschnitt 1.2).

Transparenz über die Qualität der Gesundheitsversorgung ist aus mehreren Gründen wichtig. Zum einen kann Transparenz als Wert an sich in einer demokratischen Gesellschaftsordnung verstanden werden (vgl. z. B. Fox 2007, Van Belle und Mayhew 2016, Paschke et al. 2018). Aus diesem Verständnis leitet sich beispielsweise die Empfehlung des IQTIG ab, Ergebnisse der Qualitätssicherung grundsätzlich zu veröffentlichen (siehe Kapitel 15). Zum anderen stellt Qualitätstransparenz eine notwendige<sup>5</sup> Voraussetzung für die Steigerung der Versorgungsqualität dar, da erst sie die Beteiligten in die Lage versetzt, Defizite zu erkennen und zu beheben.<sup>6</sup> Mangelnde Qualitätstransparenz kann dagegen bestehende Qualitätsprobleme verschleiern und damit einer Behebung im Weg stehen. Qualitätstransparenz reduziert also die Informationsasymmetrie zwischen Prinzipalen und Agenten und stellt darüber hinaus den Leistungserbringern Informationen zur Verbesserung ihrer Versorgungsqualität zur Verfügung (Hurley 2000, Mooney und Ryan 1993, Holmstrom und Milgrom 1991).

Die Versorgung im Rahmen der gesetzlichen Krankenversicherung (GKV) wird durch eine Vielzahl divergierender Interessen geprägt. So steht die Verbesserung des Gesundheitszustands der Versicherten als zentrales Ziel der gesetzlichen Krankenversicherung (vgl. § 1 SGB V) zwar im Einklang mit dem selbstgesetzten berufsethischen Ziel der Versorgenden, für die Gesundheit und das Wohlergehen ihrer Patientinnen und Patienten zu handeln (z. B. Weltärztebund 2017). Um dieses Ziel jedoch im Kontext begrenzter Ressourcen zu erreichen, werden im Rahmen der gesetzlichen Krankenversicherung weitere, intermediäre Anforderungen an die Versorgung gestellt. So soll die Versorgung gemäß § 70 Abs. 1 SGB V bedarfsgerecht und gleichmäßig sein, dem allgemein anerkannten Stand der medizinischen Erkenntnisse entsprechen und in der fachlich gebotenen Qualität erbracht werden, und zugleich soll sie aber auch wirtschaftlich, ausreichend, notwendig und zweckmäßig sein, wobei der Gesetzgeber diese Anforderungen nicht näher definiert. Leistungserbringer stehen hier aufgrund ihrer Einbindung in mehrere Prinzipal-Agenten-Beziehungen zum Teil unterschiedlichen Interessen vonseiten unterschiedlicher Prinzipale gegenüber – darunter z. B. dem Wunsch von Patientinnen und Patienten nach optimaler Versorgung auf der Mikroebene und dem Ziel der Kostenträger auf der Mesoebene, sowohl Qualität

---

<sup>5</sup> Jedoch nicht hinreichende, siehe Abschnitt 2.4 für die Darstellung weiterer Voraussetzungen.

<sup>6</sup> Transparenz über Qualität kann unterschiedliche Adressaten haben (Kosack und Fung 2014). Im Gesundheitswesen in Deutschland können dies die Agenten selbst, die Prinzipale auf der Mikroebene (d. h. die Patientinnen und Patienten), auf der Mesoebene (z. B. Kostenträger, Aufsichts- und Planungsbehörden) und auf der Makroebene (d. h. der Gesetzgeber) sein (siehe hierzu auch die unterschiedlichen Adressaten von Qualitätstransparenz für die in Abschnitt 2.4 dargestellten Wirkmechanismen).

als auch Wirtschaftlichkeit der Versorgung im Auftrag der GKV-Solidargemeinschaft sicherzustellen – sowie der organisationsinternen Berücksichtigung von betriebswirtschaftlichen Interessen (vgl. hierzu auch Blomqvist 1991).

Inwieweit diese mitunter unterschiedlichen Interessen tatsächlich zu Zielkonflikten führen, die die Versorgungsqualität insgesamt und die verschiedenen Qualitätsdimensionen (siehe Abschnitt 1.1) beeinträchtigen, hängt von zahlreichen Einflussfaktoren auf der Makro-, Meso- und Mikroebene ab. Beispielsweise setzen die Organisation des Gesundheitssystems und die Gestaltung der Vergütungssysteme auf der Makroebene Anreize für die Leistungserbringer auf der Mesoebene, die dem Ziel der Qualitätssteigerung zum Teil entgegenstehen können. So steht dem Ziel des Prinzipals einer angemessenen Indikationsstellung der Anreiz für eine Mengenausweitung aufseiten des Agenten gegenüber, der aus einer fallbezogenen Vergütung resultiert (Fürstenberg et al. 2013: 28, Schreyögg et al. 2014: 85 f.). Daraus ergibt sich beispielsweise ein Bedarf an der Sicherstellung der Indikationsqualität. Das Ziel des Prinzipals einer kontinuierlichen Versorgung wiederum konfliktiert potenziell mit den Interessen der Agenten, die aus der Aufteilung der Leistungserbringung zwischen den Sektoren sowie innerhalb des stationären und des ambulanten Sektors entstehen (SVR Gesundheit 2018). Daraus ergibt sich beispielsweise der Bedarf an der Sicherstellung einer koordinierten Versorgung, wenn Patientinnen und Patienten von unterschiedlichen Leistungserbringern versorgt werden. Auf der Mikroebene wiederum können beispielsweise fehlendes Wissen oder fehlende Ressourcen dem Ziel, die Gesundheit der Patientinnen und Patienten zu verbessern und hierfür eine Versorgung in der fachlich gebotenen Qualität sicherzustellen, entgegenstehen (z. B. Nilsen 2015, Mc Cluskey et al. 2013, Mazza et al. 2013, Michie et al. 2011).

Die mangelnde Qualitätstransparenz auf der Mikro-, Meso- und Makroebene des Gesundheitssystems, insbesondere in Kombination mit den möglichen Zielkonflikten zwischen Prinzipalen und Agenten, stellt einen wichtigen Grund für eine gesetzlich verankerte externe Qualitätssicherung dar. Aufgrund mangelnder Transparenz sind bestehende Qualitätsdefizite, die den Grundsätzen und Zielen der Versorgung nach Maßgabe des SGB V widersprechen, für die Prinzipale, und unter Umständen auch für die Agenten selbst, häufig nicht als solche erkennbar. Daraus folgt die Notwendigkeit, die Qualität der Gesundheitsversorgung zu messen und transparent darzustellen. Daher haben Fördermaßnahmen, in deren Rahmen den Leistungserbringern Informationen über ihre eigene Versorgungsqualität zur Verfügung gestellt werden, eine hohe Bedeutung für die kontinuierliche Weiterentwicklung der Versorgungsqualität (siehe auch Abschnitt 2.4.1). Aus potenziell divergierenden Interessen im Rahmen der unterschiedlichen Prinzipal-Agenten-Beziehungen wiederum folgt die Notwendigkeit, entsprechende Maßnahmen zur Steigerung der Versorgungsqualität umzusetzen.

## **2.4 Wirkmechanismen zur Steigerung der Versorgungsqualität**

Ausgehend von dem Ziel der Verbesserung von Gesundheit und der Steigerung der Versorgungsqualität als intermediärem Ziel unterscheidet das IQTIG drei grundlegende Wirkmechanismen,

die unterschiedliche Annahmen darüber treffen, wie eine Steigerung der Versorgungsqualität erreicht werden kann<sup>7</sup>:

- Förderung der eigenständigen Verbesserungsbestrebungen der Leistungserbringer
- Anreize für Leistungserbringer
- Auswahlentscheidungen durch Prinzipale

Um die eigenständigen Verbesserungsbestrebungen der Leistungserbringer zu fördern, erhalten diese durch die Qualitätsmessung Informationen über ihre Versorgungsprozesse und -ergebnisse. Diese Informationen können die Leistungserbringer zur Veränderung und Verbesserung der Versorgungsstrukturen und -prozesse nutzen (siehe Abbildung 3, Pfeil am rechten Bildrand). Im Rahmen der gesetzlichen Qualitätssicherung erfolgt die **Förderung** dieser eigenständigen Verbesserungsbestrebungen der Leistungserbringer insbesondere durch die Rückmeldung von Informationen zur Versorgungsqualität an die Leistungserbringer sowie ggf. durch Maßnahmen zur Beratung und Unterstützung, damit Leistungserbringer Ursachen für Qualitätsdefizite identifizieren und beheben können.

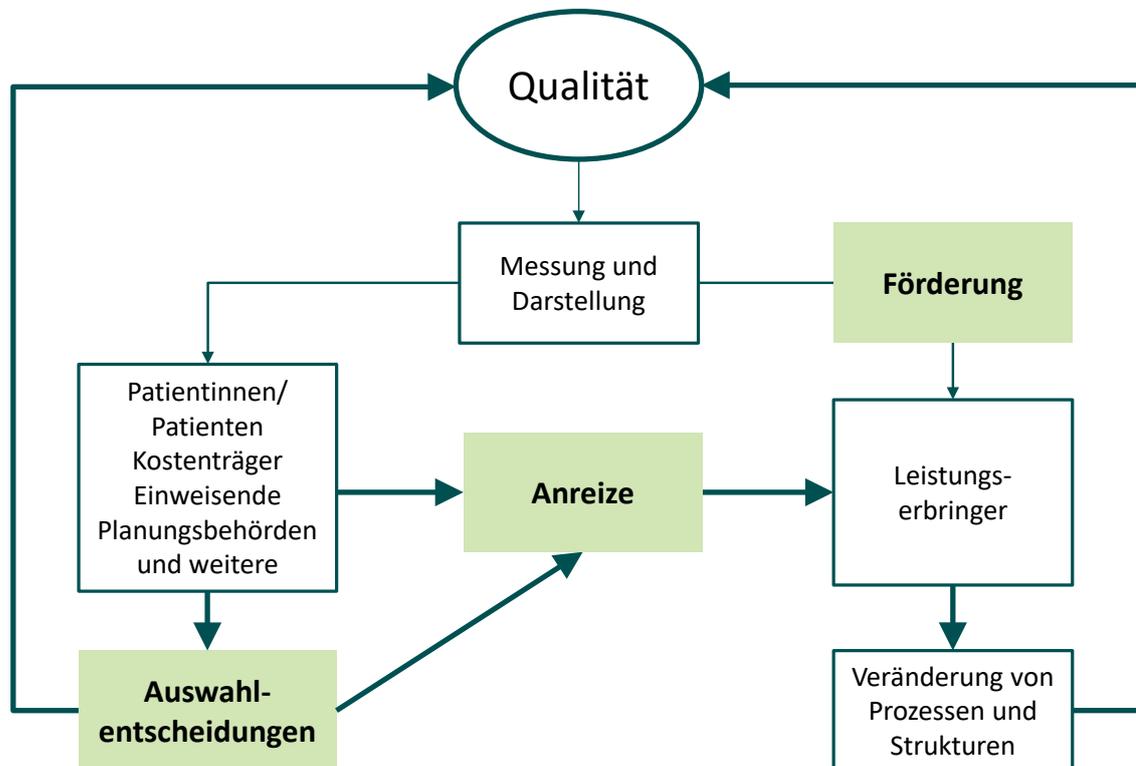


Abbildung 3: Wirkmechanismen zur Steigerung der Versorgungsqualität (in Anlehnung an Berwick et al. (2003))

Informationen über die Versorgungsqualität können auch zusätzlich den jeweiligen Prinzipalen (z. B. Kostenträgern, Patientinnen und Patienten, Planungsbehörden, einweisenden Ärztinnen

<sup>7</sup> In der Literatur wird häufig vereinfacht zwischen *improvement* und *accountability* unterschieden, wobei unter *accountability* die beiden Wirkmechanismen Anreize und Auswahlentscheidungen zusammengefasst werden (Berwick et al. 2003, Freeman 2002, Gardner et al. 2018).

und Ärzten) zur Verfügung gestellt werden, um Qualitätstransparenz herzustellen. Die Prinzipale auf der Mikro- und Mesoebene können mit diesen Informationen qualitätsorientierte **Auswahlentscheidungen** treffen, durch die mehr Patientinnen und Patienten von Leistungserbringern mit höherer Qualität versorgt werden (siehe Abbildung 3, Pfeil am linken Bildrand). Die Prinzipale auf der Meso- und Makroebene können die qualitätsbezogenen Informationen darüber hinaus für die Gestaltung von **Anreizen** einsetzen, die die Leistungserbringer motivieren, ihre Versorgungsqualität zu steigern. Diese Anreize können direkt aus den qualitätsbezogenen Informationen resultieren, die den Prinzipalen vorliegen (siehe Abbildung 3, mittiger Verbindungspfeil von links nach rechts) oder indirekt wirken, weil die Leistungserbringer Auswahlentscheidungen zu ihren Gunsten erzielen wollen (siehe Abbildung 3, diagonaler Verbindungspfeil von links nach rechts).

Während mit jedem dieser grundsätzlichen Wirkmechanismen eine Steigerung der Versorgungsqualität erreicht werden soll, unterscheiden sich die jeweils benötigten Voraussetzungen (Berwick et al. 2003, Greenhalgh et al. 2018, Contandriopoulos et al. 2014). Der Frage, welcher Wirkmechanismus bzw. welche Wirkmechanismen sich für die gesetzliche Qualitätssicherung in Bezug auf ein konkretes Versorgungsthema eignen, kommt daher eine entscheidende Bedeutung für die Empfehlung geeigneter Steuerungsinstrumente zu.

#### 2.4.1 Förderung der Verbesserungsbestrebungen der Leistungserbringer

Der Wirkmechanismus „Förderung“ zielt darauf ab, die eigenständigen Verbesserungsbestrebungen der Leistungserbringer zu unterstützen. Dieser Wirkmechanismus basiert auf der Annahme, dass Leistungserbringer primär der Gesundheit ihrer Patientinnen und Patienten verpflichtet sind und in diesem Sinne als ideale Agenten der Patientinnen und Patienten die Qualität der Versorgung sicherstellen (siehe auch Stewardship-Ansatz, Abschnitt 2.3). Eine zentrale Voraussetzung für die Wirksamkeit der Qualitätsförderung ist daher die Motivation der Versorgenden, Patientinnen und Patienten in der fachlich gebotenen Qualität zu versorgen.

Der Wirkmechanismus „Förderung“ setzt auch voraus, dass die Versorgenden fachlich qualifiziert sind und z. B. durch regelmäßige Fort- und Weiterbildungen sicherstellen, dass sie eine Versorgung entsprechend dem aktuellen Stand der medizinischen Erkenntnis gewährleisten (Vrangbæk und Byrkjeflot 2016). Denn nur wenn die intrinsische Motivation der Versorgenden zur Qualitätssteigerung durch konsenterte, evidenzbasierte wissenschaftliche Erkenntnisse gestützt wird, kann sie zum Wohle der Patientinnen und Patienten eingesetzt werden.

Damit Qualitätsförderung wirken kann, müssen qualitätsbezogene Informationen darüber hinaus zielgerichtet innerhalb von Organisationen der Leistungserbringung genutzt werden. So müssen Individuen und Teams von den entsprechenden Informationen Kenntnis erhalten (z. B. in Form von Rückmeldeberichten), diese als aussagekräftig akzeptieren sowie als anwendbar auf ihre konkrete Situation bewerten (Glasziou und Haynes 2005). Darüber hinaus müssen Leistungserbringer über die notwendigen Ressourcen verfügen, um z. B. im Rahmen ihres internen Qualitätsmanagements die Ursachen für festgestellte Defizite analysieren, entsprechende Maßnahmen umsetzen und deren Erfolg evaluieren zu können (Furnival et al. 2018, Karlun et al.

2020). Dies setzt auch ein Umfeld voraus, in dem es üblich und selbstverständlich ist, Qualitätsergebnisse zu reflektieren und für die Verbesserung von Strukturen und Prozessen der Versorgung zu nutzen (Freeman 2002, Braithwaite 2018, Gardner et al. 2018).

### **Zusammenfassung der grundlegenden Voraussetzungen für erfolgreiche Qualitätsförderung**

- Motivation der Versorgenden zur Qualitätssteigerung
- fachliche Qualifikation der Versorgenden entsprechend dem aktuellen Stand der Wissenschaft
- aussagekräftige und anwendbare Qualitätsinformationen
- Ressourcen, um entsprechende Veränderungen in Strukturen und Prozessen umzusetzen
- ein unterstützendes Umfeld

#### **2.4.2 Anreize für Leistungserbringer**

Leistungserbringer stehen auch betriebswirtschaftlichen Interessen gegenüber, die sich beispielsweise aus dem Wettbewerb mit anderen Leistungserbringern und der Notwendigkeit einer betriebswirtschaftlich einträglichen Organisationsführung ergeben (Mooney und Ryan 1993). Wenn betriebswirtschaftliche Interessen einen erheblichen Einfluss auf Art und Umfang der Leistungserbringung ausüben und die Qualität der Versorgung dadurch beeinträchtigen, reichen Maßnahmen der Qualitätsförderung nicht aus, um eine angemessene Versorgungsqualität sicherzustellen.

Das Ziel von Anreizsystemen ist es, die Interessen des Agenten so zu lenken, dass sie sich den Interessen des Prinzipals annähern (Vrangbæk und Byrkjeflot 2016, Fritsch 2018). Anreizsysteme sind daher besonders dann relevant, wenn qualitätsbezogene Anforderungen an die Versorgung durch Zielkonflikte des Agenten beeinträchtigt werden. Dies ist beispielsweise bei Fehlanreizen auf der Makro- und Mesoebene des Gesundheitssystems (z. B. im Rahmen des Vergütungssystems) der Fall oder auch bei fehlender Motivation zur Qualitätssteigerung oder Interessenkonflikten auf der Mikroebene bei einzelnen Versorgenden.

Anreize können sowohl positive Konsequenzen (z. B. Zuschläge, Reputationssteigerung) als auch negative Konsequenzen (z. B. Abschläge, Reputationsverluste) für die Agenten umfassen. Damit Anreizsysteme zur Steigerung der Versorgungsqualität beitragen können, sollte das mit den Anreizen verbundene Qualitätsziel eindeutig und für die Versorgenden erkennbar auf eine Steigerung der Versorgungsqualität ausgerichtet sein: Denn nur, wenn Anreize im Einklang mit der berufsethisch begründeten Motivation der Versorgenden, die Versorgungsqualität zu verbessern, wirken, können sie zu tatsächlichen Verbesserungsbestrebungen führen (Kronick et al. 2015, Roland und Dudley 2015, Van Herck et al. 2010, Eckhardt et al. 2019).

Das Spektrum möglicher Anreize ergibt sich aus den im jeweiligen Kontext als maßgeblich eingeschätzten Einflussfaktoren auf das Verhalten (Hamblin 2008). Beispielsweise können Anreize am betriebswirtschaftlichen Interesse von Leistungserbringern ansetzen, um das gewünschte Verhalten zu erzielen. Bei direkten finanziellen Anreizen zieht das Ergebnis der Qualitätsmessung eine direkt darauf bezogene Belohnung bzw. Sanktionierung nach sich (Hamblin 2008, Cattell et al. 2020). Finanzielle Anreize können jedoch auch indirekt auf betriebswirtschaftliche

Interessen wirken (Hamblin 2008). Beispielsweise können qualitätsorientierte Auswahlentscheidungen durch Prinzipale (vgl. Abschnitt 2.4.3) zu Veränderungen in den Fallzahlen eines Leistungserbringers führen. Solche Anreize setzen die Überzeugung aufseiten der Leistungserbringer voraus, dass qualitätsorientierte Auswahlentscheidungen durch Prinzipale getroffen werden und somit Fallzahlen und damit auch betriebswirtschaftliche Ziele tatsächlich beeinflussen.

Anreize können auch an nicht materiellen Einflussfaktoren auf Verhalten ansetzen, beispielsweise an der Reputation der Leistungserbringer (Contandriopoulos et al. 2014, Bevan und Fasolo 2013). Solche Anreize setzen nicht voraus, dass Prinzipale (z. B. Kostenträger, Patientinnen und Patienten) qualitätsorientierte Auswahlentscheidungen tatsächlich vornehmen – vielmehr ist es die leicht zugängliche Veröffentlichung, Verbreitung und allgemeinverständliche Darstellung der einrichtungsbezogenen Versorgungsqualität selbst, welche die Leistungserbringer aus Reputationsgründen zu Verbesserungsbestrebungen motiviert (Hibbard 2008, Bevan und Fasolo 2013, Bevan et al. 2019).

Der effektive Einsatz von Anreizsystemen setzt zudem voraus, dass unbeabsichtigte Folgen bereits in der Gestaltung der Anreizsysteme antizipiert oder in der Umsetzungsphase frühzeitig erkannt werden (Milstein und Schreyoegg 2016, Kondo et al. 2016, Scott et al. 2011). Mögliche unbeabsichtigte, negative Auswirkungen von Anreizsystemen umfassen u. a. Verschlechterungen der Versorgung in nicht incentivierten Bereichen und die Fokussierung auf leichtere Fälle, mit denen das Qualitätsziel und daran hängende Anreize durch weniger Aufwand erreicht werden können (Risikoselektion), und reichen bis hin zur Manipulation von Daten, um Vorteile bei Anreizsystemen zu erzielen (Eckhardt et al. 2019, Bevan und Hood 2006, Smith 1995). Negative Auswirkungen der Anreize sollten durch entsprechende Gegenmaßnahmen möglichst behoben werden. Wenn dies nicht für alle negativen Folgen möglich ist, sollte abgewogen werden, ob der mit den Anreizen verbundene Nutzen diese negativen Folgen überwiegt.

Für die Entwicklung von Anreizsystemen besteht eine entscheidende Herausforderung darin, die maßgeblichen Einflussfaktoren auf das Verhalten zu identifizieren, das Anreizsystem unter Berücksichtigung weiterer Einflussfaktoren auf Verhalten, wie z. B. gesetzlicher Regelungen oder sozialer Normen, im jeweiligen Kontext zu entwickeln und mögliche unerwünschte Effekte zu adressieren (Kondo et al. 2016, Scott et al. 2011, Cattel et al. 2020).

### **Zusammenfassung der grundlegenden Voraussetzungen für Qualitätssteigerung durch Anreize**

- eindeutig erkennbare Verknüpfung des Anreizes mit einem Qualitätsziel
- Anreize setzen an maßgeblichen Einflussfaktoren auf das Verhalten an (z. B. betriebswirtschaftliche Interessen, Reputation)
- bei indirekten Anreizen: Überzeugung der Leistungserbringer, dass Auswahlentscheidungen qualitätsorientiert getroffen werden
- Erkennung unbeabsichtigter Folgen der Anreize und Ergreifen von Gegenmaßnahmen

### 2.4.3 Auswahlentscheidungen durch Prinzipale

Während die Wirkmechanismen „Förderung“ und „Anreize“ bei den Leistungserbringern ansetzen, können qualitätsorientierte Auswahlentscheidungen der Prinzipale die Verbesserungsbestrebungen der Leistungserbringer in unterschiedlicher Hinsicht ergänzen oder ersetzen (Berwick et al. 2003). Auswahlentscheidungen können sich grundsätzlich auf unterschiedliche Bezugsebenen beziehen, wie beispielsweise die Auswahl von Leistungserbringern, von Fachabteilungen, von Leistungsgruppen und -bereichen, oder von Einzelleistungen.

Eine wichtige Funktion von Auswahlentscheidungen besteht in der Sicherstellung von (Mindest-) Anforderungen für die Teilnahme an der Versorgung im Rahmen der gesetzlichen Krankenversicherung. Formal stellen Anforderungen, gemäß denen Leistungen erbracht und im Rahmen der gesetzlichen Krankenversicherung abgerechnet werden dürfen, Zutrittsbarrieren für Leistungserbringer dar (Fritsch 2018). Die Durchsetzung solcher Zutrittsbarrieren schafft damit einen einheitlichen Rahmen für alle Leistungserbringer, innerhalb dessen sich die Versorgung der Patientinnen und Patienten bewegt. Auswahlentscheidungen zur Teilnahme an der Versorgung umfassen sowohl die Zulassung (z. B. Abrechnungsgenehmigungen in der ambulanten Versorgung) als auch den Ausschluss von Einheiten der jeweiligen Bezugsebene (z. B. Vergütungswegfall).

Innerhalb der Gruppe der Leistungserbringer, die an der Versorgung der gesetzlich Versicherten teilnehmen, können Auswahlentscheidungen weitere Funktionen haben. Auf der Mikroebene der Versorgung können einzelne Patientinnen und Patienten bzw. deren Vertreter (z. B. Angehörige oder einweisende Ärztinnen und Ärzte) geeignete Leistungserbringer auswählen, die die Versorgung der Patientin bzw. des Patienten mit einem bestimmten Versorgungsanlass übernehmen. Auf der Mesoebene können Kostenträger beispielsweise Informationen über die einrichtungsbezogene Qualität der Versorgung für die Auswahl von Leistungserbringern z. B. für die Gestaltung von Versorgungsverträgen für bestimmte Bevölkerungs- oder Patientengruppen nutzen (Sanderson et al. 2019).

Damit durch Auswahlentscheidungen der Prinzipale mehr Patientinnen und Patienten besser versorgt werden, müssen die Prinzipale in die Lage versetzt werden, qualitätsorientierte Auswahlentscheidungen zu treffen. Dies setzt voraus, dass entsprechende Informationsasymmetrien (vgl. Abschnitt 2.3) aufgelöst werden. Die jeweiligen Prinzipale müssen über die einrichtungsbezogene Qualität der Versorgung informiert werden, diese Informationen verstehen und als aussagekräftig und relevant für ihr konkretes Entscheidungsproblem bewerten (Hibbard et al. 2003, Hibbard et al. 2005). Außerdem müssen die Prinzipale bereit sein, den qualitätsbezogenen Informationen eine bedeutende Rolle im Auswahlprozess zu geben (Contandriopoulos et al. 2014).

Schließlich muss eine Auswahl möglich sein. Bezogen auf die Auswahl von Leistungserbringern bedeutet dies, dass ausreichend alternative Anbieter mit entsprechenden Kapazitäten verfügbar sind, unter denen eine Auswahl stattfinden kann (Cacace et al. 2019). Die jeweiligen Entscheidungsträger müssen über die notwendigen Ressourcen verfügen, um Auswahlentscheidungen praktisch umzusetzen. Aufseiten von Planungsbehörden und Kostenträgern umfasst dies z. B.

entsprechendes Personal, aufseiten von Patientinnen und Patienten setzt dies u. a. eine entsprechende Mobilität der Patientinnen und Patienten bzw. Erreichbarkeit der Leistungserbringer voraus.

### **Zusammenfassung der grundlegenden Voraussetzungen für Qualitätssteigerung durch Auswahlentscheidungen**

- Information der Prinzipale über Qualitätsunterschiede zwischen Leistungserbringern
- aussagekräftige und handlungsrelevante Qualitätsinformationen
- Bereitschaft der Prinzipale zu qualitätsorientierten Entscheidungen
- Verfügbarkeit mehrerer alternativer Leistungserbringer
- Ressourcen für die Umsetzung der Auswahlentscheidung bei den Prinzipalen

## **2.5 Rahmenmodell für die Aufgaben von Qualitätssicherung**

Die Qualitätssicherung der Gesundheitsversorgung lässt sich in grundsätzliche Aufgaben gliedern, die in verschiedenen Modellen beschrieben wurden. Der Plan-Do-Check-Act-Zyklus (PDCA-Zyklus)<sup>8</sup> beispielsweise umfasst vier Schritte – die Planung, Umsetzung, Überprüfung sowie die erneute Anpassung von Maßnahmen –, die in ihrer Gesamtheit einen kontinuierlichen Verbesserungsprozess sicherstellen sollen (Taylor et al. 2014, Reed und Card 2016). Das „Model for Improvement“ wiederum beinhaltet eine Definition der Anforderungen, die Messung des Grads der Erfüllung dieser Anforderungen und mögliche Verbesserungsmaßnahmen (Provost und Langley 1998). Die Juran-Trilogie differenziert in ähnlicher Weise zwischen den Managementaufgaben Planung, Kontrolle und Verbesserung von Qualität (De Feo 2017: 40 ff.). Auch wenn sich die Begrifflichkeiten zwischen den verschiedenen Modellen und Ansätzen teilweise unterscheiden (vgl. Tello et al. 2020, Panteli et al. 2019, Furnival et al. 2017), weisen sie große konzeptuelle Ähnlichkeiten auf. Daraus leitet das IQTIG ein Rahmenmodell für die Qualitätssicherung mit den folgenden, für eine effektive Qualitätssicherung grundsätzlich erforderlichen Aufgaben ab (siehe Abbildung 4):

- Die **Definition von Anforderungen an die Versorgung** dient dazu, Klarheit zu schaffen, was von den Leistungserbringern erwartet wird.
- Die **Messung und Bewertung der Versorgungsqualität** ermöglicht einen Soll-Ist-Abgleich zwischen den Anforderungen an die Versorgung und dem Ist-Zustand der Versorgung.
- Die **Entwicklung und Umsetzung von Maßnahmen** zur Steigerung der Versorgungsqualität ist notwendig, um die Anforderungen an die Versorgungsqualität durchzusetzen.
- Die **Evaluation** der Maßnahmen dient dazu, deren Auswirkungen auf die Versorgungsqualität zu überprüfen.

Die Unterscheidung zwischen diesen zusammenhängenden Aufgaben macht einerseits deutlich, welche unterschiedlichen Instrumente im Rahmen der Qualitätssicherung benötigt werden und ggf. entwickelt werden müssen. Da die Aufgaben andererseits aufeinander aufbauen und gleichermaßen für die Verbesserung der Versorgungsqualität wichtig sind, müssen die Instrumente

---

<sup>8</sup> Auch bekannt als Plan-Do-Study-Act-Zyklus (PDSA-Zyklus).

zur Erfüllung dieser Aufgaben aufeinander abgestimmt sein, um die Qualität der Gesundheitsversorgung möglichst wirksam zu steigern.

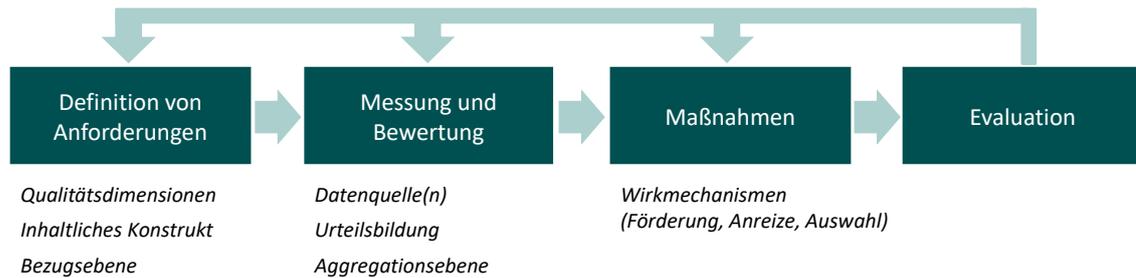


Abbildung 4: Rahmenmodell für die Aufgaben von Qualitätssicherung

### Definition von Anforderungen

Eine Bewertung der Versorgungsqualität ist nur möglich, wenn zuvor Anforderungen an die Versorgung festgelegt wurden. Durch die Definition von Anforderungen wird der Soll-Zustand der Versorgung bestimmt. Eine fachlich begründete Definition von Anforderungen erfordert Festlegungen zur adressierten Qualitätsdimension, zum interessierenden inhaltlichen Konstrukt und zur Bezugsebene. Anforderungen an die Versorgung lassen sich gemäß dem Rahmenkonzept für Qualität in sechs grundlegende *Qualitätsdimensionen* gruppieren: Wirksamkeit, Patientensicherheit, Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten, Rechtzeitigkeit und Verfügbarkeit, Angemessenheit sowie Koordination und Kontinuität der Versorgung (siehe Abschnitt 1.1).

Im Rahmen der Qualitätsmessung müssen sich Anforderungen dabei auf ein bestimmtes *inhaltliches Konstrukt* beziehen, wie z. B. die Qualität einer bestimmten Leistung oder eines Leistungsbereichs (siehe Kapitel 11). Anforderungen müssen sich zudem auf eine *Bezugsebene* beziehen. So können Anforderungen beispielsweise an eine Fachabteilung, einen Standort, eine Einrichtung insgesamt oder an mehrere Einrichtungen gestellt werden, die einen Versorgungsprozess gemeinsam verantworten (z. B. accountable care organizations, siehe z. B. McClellan et al. 2010, McClellan et al. 2017, McClellan et al. 2014, Kaufman et al. 2017, Pimperl et al. 2017).

### Qualitätsmessung und -bewertung

Aussagekräftige Informationen über die Qualität der Versorgung geben Auskunft über Art und Ausmaß bestehender Qualitätsdefizite und sind daher notwendig, um zielgerichtete Maßnahmen umsetzen zu können. Die Messung und Bewertung von Qualität ermöglicht einen Abgleich zwischen dem Soll-Zustand und dem Ist-Zustand der Versorgungsqualität. Instrumente der Qualitätsmessung lassen sich nach der *Datenquelle* (siehe Abschnitt 4.2) und dem *Prozess der Urteilsbildung* unterscheiden.

Der *Prozess der Urteilsbildung* beschreibt, wie die gesammelten Informationen schließlich zu einem Urteil über die Versorgungsqualität integriert werden. Beim mechanischen (auch: statistischen) Vorgehen erfolgt dies nach expliziten, konsistent anzuwendenden Regeln (Meehl 1954, Grove et al. 2000); z. B. bei der quantitativen Auffälligkeitseinstufung von Qualitätsindikatoren

(siehe Kapitel 16). Beim sogenannten klinisch-intuitiven Vorgehen wird dagegen ein individuelles Urteil nach impliziten Regeln getroffen (Meehl 1954, Grove et al. 2000). Mischformen greifen sowohl auf die mechanische als auch auf die intuitive Urteilsbildung zurück.

Bei der Darstellung und Bewertung der Ergebnisse von Qualitätsmessungen muss darüber hinaus festgelegt werden, ob für die jeweilige Maßnahme zur Qualitätssteuerung eine Detailbetrachtung der Ergebnisse je Qualitätsmerkmal oder aber eine höhere *Aggregationsebene* der Informationen, beispielsweise durch Indexbildung, zielführend ist.

### **Maßnahmen**

Um die im ersten Schritt definierten Anforderungen an die Versorgung durchzusetzen, stehen die grundsätzlichen *Wirkmechanismen* „Förderung“, „Anreize“ und „Auswahl“ zur Verfügung, deren Wirksamkeit von unterschiedlichen Voraussetzungen abhängig ist (siehe Abschnitt 2.4). Maßnahmen zur Durchsetzung der Anforderungen an die Versorgung bedienen sich immer mindestens eines dieser Wirkmechanismen. Manche Maßnahmen wirken auch über mehrere Wirkmechanismen. Beispielsweise kann Public Reporting sowohl durch Auswahlentscheidungen der Patientinnen und Patienten als auch durch Anreize auf die Leistungserbringer zu Steigerungen der Versorgungsqualität führen. Die Unterscheidung zwischen den verschiedenen Wirkmechanismen macht deutlich, dass bei der Entwicklung und Ausgestaltung von Maßnahmen zur Qualitätssteuerung eine Festlegung getroffen werden muss, über welche Wirkmechanismen die Maßnahme wirken soll, damit die jeweiligen Voraussetzungen für ihre Wirksamkeit geschaffen werden können.

### **Evaluation**

Die Evaluation der Maßnahmen zeigt deren Nutzen, Kosten sowie mögliche Implementierungsprobleme auf und schafft damit die Grundlage für deren Weiterentwicklung. Gemäß § 136d SGB V ist der G-BA zur Evaluation der Wirksamkeit der eingeführten Qualitätssicherungsmaßnahmen, zur Feststellung des Stands der Qualitätssicherung im Gesundheitswesen und des sich daraus ergebenden Weiterentwicklungsbedarfs verpflichtet. Die Evaluation von Maßnahmen ist eine wichtige Aufgabe der gesetzlichen Qualitätssicherung, um eine kontinuierliche Überprüfung der Umsetzung und der Wirksamkeit der Maßnahmen zu gewährleisten. Erkenntnisse aus der Evaluation können dabei zu Anpassungen an der Definition der Anforderungen, an den Instrumenten zur Messung und Bewertung der Versorgungsqualität sowie an den eingesetzten Maßnahmen zur Qualitätssteuerung führen.

### **Aufgabenbereiche des IQTIG im Rahmenmodell**

Das Rahmenmodell für die Aufgaben von Qualitätssicherung strukturiert gleichzeitig die zentralen Aufgaben des IQTIG (siehe Abschnitt 2.2): die Entwicklung und Ableitung von Anforderungen, die Entwicklung von Instrumenten zur Messung und Bewertung der Versorgungsqualität („Messinstrumente“), die Entwicklung und Umsetzung von Maßnahmen zur Durchsetzung der Anforderungen („Qualitätssicherungsmaßnahmen“) sowie deren Evaluation. Es fasst die maßgeblichen Grundannahmen des IQTIG zu den qualitätsbezogenen Messinstrumenten und Quali-

tätssicherungsmaßnahmen des SGB V zusammen, die für die fachlich unabhängige, wissenschaftliche Beratung des G-BA gemäß dem gesetzlichen Auftrag des IQTIG (vgl. § 137a Abs. 3 SGB V, siehe Abschnitt 2.2) relevant sind. So wird beispielsweise deutlich, dass die angestrebten Wirkmechanismen bei der Entwicklung von Instrumenten zur Qualitätsmessung und -bewertung berücksichtigt werden sollten, da sie unterschiedliche Voraussetzungen aufweisen. Das Rahmenmodell dient dazu, sichtbar zu machen, welche Voraussetzungen der Wirkmechanismen berücksichtigt und welche Festlegungen zu den einzelnen Aufgaben der Qualitätssicherung getroffen werden müssen, um konsistente Strategien und Verfahren der Qualitätssicherung zu entwickeln.

## 3 Patientenzentrierung in der Qualitätssicherung

Eine an den Erwartungen der Bevölkerung ausgerichtete Gesundheitsversorgung, die zu guter Gesundheit führt (vgl. Kapitel 1), erfordert, dass Patientinnen und Patienten in das Zentrum der Bemühungen im Gesundheitssystem gestellt werden (Battles 2006, Arah et al. 2006, Gerlinger 2009, Berwick 2002, Deutscher Ethikrat 2016). Der Begriff „Patientenzentrierung“ ist national und international allerdings nicht einheitlich definiert (Scholl et al. 2014). 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 auf diese unterschiedlichen Ansätze zurück, um nachfolgend sein eigenes Verständnis von Patientenzentrierung für den Kontext der externen Qualitätssicherung darzulegen.

### 3.1 Verständnis von Patientenzentrierung

„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), 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 verbrieftete Rechte und Pflichten verfügt“ (Klemperer 2000: 15).<sup>9</sup> Diese Forderung gilt zum einen für die direkte Interaktion zwischen Gesundheitsprofessionen und Patientinnen bzw. Patienten, die sogenannte Mikroebene der Gesundheitsversorgung (Institute of Medicine/Committee on Quality of Health Care in America 2001, Gerlinger 2009). Sie zielt zum anderen aber auch auf der sogenannten Meso- und Makroebene auf die Institutionen und Einrichtungen im Gesundheitswesen. Diese können beispielsweise durch organisatorische, gesetzgeberische und andere regulatorische Maßnahmen geeignete Bedingungen für eine patientenzentrierte Versorgung schaffen, Anforderungen an eine patientenzentrierte Versorgung definieren oder – bezogen auf die gesetzliche Qualitätssicherung – deren Erfüllung durch Qualitätsmessungen prüfen (Gerlinger 2009, Delaney 2018, Greene et al. 2012, Steiger und Balog 2010, Epstein et al. 2010).

Um Über-, Unter- oder Fehlversorgung entgegenzusteuern, sollten dabei – neben dem sogenannten objektiven Bedarf – auch der sogenannte subjektive sowie der latente Bedarf<sup>10</sup> von Patientinnen und Patienten Berücksichtigung finden, unter vordringlicher Beachtung von deren Selbstbestimmungsrecht (vgl. SVR Gesundheit 2018, SVR-Gesundheit 2014, SVR-Gesundheit 2002). Anforderungen der Patientenzentrierung, wie die Ausrichtung der Versor-

---

<sup>9</sup> Dieses Verständnis steht im Einklang mit international einflussreichen Betrachtungsweisen wie der des Institute of Medicine (Institute of Medicine/Committee on Quality of Health Care in America 2001).

<sup>10</sup> 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 einschätzt oder [...] die Nutzung der entsprechenden Gesundheitsleistung aus anderen Gründen, beispielsweise wegen mangelnder Angebotskapazitäten, nicht erfolgt“ (SVR Gesundheit 2018: 50).

gestaltung an den Patientinnen und Patienten, setzen ein verändertes Verständnis der Beziehungsgestaltung zwischen Patientinnen und Patienten und Gesundheitsprofessionen voraus – weg von paternalistischen Ansätzen hin zu einem partnerschaftlichen Verhältnis und partizipativer Entscheidungsfindung (Shared Decision Making; Klemperer 2000, Delaney 2018). Die damit verbundene Anerkennung der Patientenbedürfnisse, die Informiertheit über die Behandlung und die Einbindung in Entscheidungsprozesse haben mit dem Gesetz zur Verbesserung der Rechte von Patientinnen und Patienten (Patientenrechtegesetz) bereits Eingang in die Gesetzgebung gefunden. Patientenzentrierung begründet sich also auch durch gesellschaftliche Werte. Insgesamt sollten die Strukturen, Prozesse und Ergebnisse der Gesundheitsversorgung entsprechend den Anforderungen, die Patientinnen und Patienten bei der Nutzung des Gesundheitssystems stellen, kontinuierlich weiterentwickelt werden (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, Ansmann und Hower 2020).

Für das IQTIG ergeben sich aus dieser Betrachtung von Patientenzentrierung zwei Konsequenzen: Erstens verfolgt das IQTIG mit seinen Arbeitsprodukten den Anspruch, eine patientenzentrierte Ausrichtung der Gesundheitsversorgung zu fördern. Zweitens setzt sich das IQTIG als Institution des Gesundheitswesens das Ziel, sich selbst in seinen Tätigkeiten und Verantwortlichkeiten patientenzentriert auszurichten und dahingehend beständig weiterzuentwickeln.

Das IQTIG versteht Patientenzentrierung in diesem Zusammenhang als übergeordnete Anforderung an die Gesundheitsversorgung (Scholl et al. 2014, Zeh et al. 2019). Eine patientenzentrierte Versorgung spiegelt sich dabei in der Erfüllung der in Abschnitt 1.1 definierten Qualitätsdimensionen:

- Wirksamkeit (*effectiveness*)
- Patientensicherheit (*safety*)
- Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten (*responsiveness*)
- Rechtzeitigkeit und Verfügbarkeit (*timeliness*)
- Angemessenheit (*appropriateness*)
- Koordination und Kontinuität (*coordination, continuity*)

Im Verständnis des IQTIG ist eine qualitativ hochwertige Versorgung demnach immer auch eine patientenzentrierte Versorgung.

Um die genannten Anforderungen an die Gesundheitsversorgung bei der Entwicklung von Empfehlungen zur Qualitätssicherung und -verbesserung umzusetzen, strebt das IQTIG an, im Rahmen von QS-Verfahren solche Inhalte und Themen zu adressieren, die für Patientinnen und Patienten von Belang sind. Abgebildet werden diese über Qualitätsaspekte und -merkmale, die über Qualitätsindikatoren erfasst werden (siehe Abschnitt 3.2).

Um sich als Institution des Gesundheitswesens selbst patientenzentriert auszurichten, setzt sich das IQTIG zusätzlich folgende Ziele:

- **Transparenz und Information für Patientinnen und Patienten**
  - die Bereitstellung von allgemeinverständlichen, relevanten und aktuellen Informationen über Prozesse und Ergebnisse des IQTIG
  - die Verfügbarkeit von Ansprechpartnerinnen bzw. -partnern im IQTIG
- **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
- **kontinuierliche Organisationsentwicklung des IQTIG 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

Zu einer Ausrichtung als patientenzentrierte Institution gehört, wie oben dargestellt, die Bereitstellung von Informationen für Patientinnen und Patienten in einer allgemeinverständlichen Art und Weise. Das IQTIG berät keine Einzelpersonen, sondern strebt an, Patientinnen und Patienten entsprechend seinem Auftrag nach § 137a SGB V transparent und verständlich über die Ergebnisse der Qualitätssicherung sowie seine Arbeitsprozesse zu informieren. Dies umfasst auch den kontinuierlichen Ausbau allgemeinverständlicher Berichtsformate.

### **3.2 Adressierung patientenrelevanter Qualitätsaspekte und -merkmale**

Das IQTIG beschreibt und strukturiert Themenbereiche der Versorgung in Form von Qualitätsaspekten und -merkmalen (vgl. Kapitel 11). Um sicherzustellen, dass die darauf basierenden Entwicklungsergebnisse und Empfehlungen des IQTIG die Anforderungen einer patientenzentrierten Qualitätssicherung erfüllen, sollen die Qualitätsaspekte und -merkmale Themen beschreiben, die patientenrelevant sind. Zu den unmittelbar patientenrelevanten Themen gehören zum einen gesundheitsbezogene Aspekte wie Mortalität, Morbidität, Krankheitssymptome, körperliches Wohlbefinden und Lebensqualität (vgl. Abschnitt 13.2.1). Zum anderen handelt es sich um Themen, die eine hohe Ausrichtung der interpersonellen, kontextuellen und organisatorischen Aspekte der Versorgungsgestaltung an Patientinnen und Patienten kennzeichnen, z. B. Patienteninformation und -edukation, Berücksichtigung der Patientenpräferenzen und Mitbestimmung, Versorgungskontinuität, Zugang zur Versorgung, psychische und emotionale Unterstützung, Einbeziehung der Familie. Mittelbar patientenrelevante Themen sind nur dann für eine patientenzentrierte Qualitätssicherung geeignet, wenn ein ausreichend belegter Zusammenhang mit einem unmittelbar patientenrelevanten Ziel bzw. Thema vorliegt (z. B. „Antibiotikapro-

phylaxe vor einer Operation“ als mittelbar patientenrelevantes Thema für das unmittelbar patientenrelevante Ziel „Vermeidung von Wundinfektionen bei der Operation“; vgl. Abschnitt 13.2.2).

Welche konkreten Qualitätsaspekte und -merkmale bedeutsam für Patientinnen und Patienten sind, ist projektbezogen zu ermitteln (Mühlhauser und Müller 2009: 36 f.). Das IQTIG stellt dies durch eine konsequente und systematische Einbindung der Patientenperspektive sicher: zum einen, indem es bestehende wissenschaftliche Erkenntnisse zur Patientenperspektive nutzt, zum anderen, indem es betroffene Patientinnen und Patienten sowie Patientenvertreterinnen und -vertreter direkt an verschiedenen Stellen des Entwicklungsprozesses einbezieht (vgl. Abschnitt 3.3).

Wie beispielsweise bei der Entwicklung klinischer Leitlinien deutlich wird, werden durch die direkte Einbindung von betroffenen Patientinnen und Patienten Themen erfasst, die Vertreterinnen und Vertreter der Gesundheitsprofessionen allein nicht miteinbringen (Tong et al. 2012, Armstrong et al. 2018, den Breejen et al. 2016). Auch werden Themen in ihrer Bedeutsamkeit von Patientinnen und Patienten teilweise anders gewertet als von Vertreterinnen und Vertretern der Gesundheitsprofessionen (z. B. Kersting et al. 2020, IQWiG 2013). Trotz dieser empirischen Hinweise erfolgt eine direkte Einbindung der Patientenperspektive bei der Leitlinienentwicklung bisher nicht im erforderlichen Ausmaß (Armstrong und Bloom 2017, Ollenschläger et al. 2018, Blackwood et al. 2020). Vor diesem Hintergrund ist es für das IQTIG von zentraler Bedeutung, betroffene Patientinnen und Patienten bei der Erstellung des Qualitätsmodells direkt einzubeziehen (vgl. Kapitel 11).

Ergänzend nutzt das IQTIG bei der Entwicklung von Qualitätsmessungen die Qualitätsdimensionen des Rahmenkonzepts für Qualität als Raster, um zu prüfen, ob das Qualitätsmodell oder das Qualitätsindikatorenset alle Dimensionen adressiert oder ob unerwünschte „blinde Flecke“ bestehen (vgl. Kapitel 11 und Abschnitt 14.2). Werden nicht alle Dimensionen durch das Qualitätsmodell bzw. das Qualitätsindikatorenset abgedeckt, weist das IQTIG darauf hin, wenn potenziell patientenrelevante Themen nicht adressiert werden.<sup>11</sup>

### 3.3 Methoden zur Förderung der Patientenzentrierung

Der Ansatz des IQTIG sieht vor, sowohl Patientinnen und Patienten direkt einzubeziehen als auch im Sinne von Patientinnen und Patienten, also für sie, zu agieren (Gerlinger 2009). Die systematische Einbindung der Patientenperspektive ist das Mittel der Wahl, um Versorgungsprozesse und -ergebnisse patientenzentriert nachzuvollziehen. Das IQTIG erachtet sie bei der Entwicklung von Maßnahmen der Qualitätssicherung insbesondere als nutzbringend, um zu erfahren, welche spezifischen Bedürfnisse, Erfahrungen, Sichtweisen oder Erleben die Betroffenen einer Krankheit oder Behandlung selbst haben. Dabei ist die Einbindung von Patientinnen und Patienten bzw. deren Vertreterinnen und Vertretern kein Selbstzweck oder ein eigenes Ergebnis, sondern

---

<sup>11</sup> Bestehende Indikatorensets, denen bislang kein Qualitätsmodell zugrunde liegt, fokussieren meist nur auf einige wenige Qualitätsdimensionen (Beaussier et al. 2020, Döbler und Geraedts 2018). Das IQTIG strebt an, dies in den Empfehlungen zur Weiterentwicklung von Indikatorensets durch retrospektiven Abgleich der Qualitätsindikatoren mit den Qualitätsdimensionen des Rahmenkonzepts perspektivisch stärker zu adressieren (vgl. Abschnitt 14.2).

erfolgt immer mit dem Ziel der Verbesserung der Gesundheitsversorgung (Carman und Workman 2017). Die Einbindung der Patientenperspektive kann sowohl über eine Analyse bereits vorliegender wissenschaftlicher Erkenntnisse als auch über eigene empirische Erhebung und strukturierte Beteiligung von Patientinnen und Patienten geschehen, jedoch nicht durch Berichte von Gesundheitsprofessionen ersetzt werden. Darüber hinaus werden bei der Bearbeitung patientenrelevanter Themen in übergeordneten Kontexten Patientenvertreterinnen und Patientenvertreter beteiligt.

### **3.3.1 Analyse wissenschaftlicher Erkenntnisse zur Patientenperspektive**

Bei der Analyse bereits vorliegender wissenschaftlicher Erkenntnisse werden Informationen zur Patientenperspektive erfasst, 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 zu erfassen sowie 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 Hinweise auf Perspektiven besonders vulnerabler Subgruppen (z. B. pflegebedürftige Menschen, Menschen mit terminalem Krankheitsverlauf) generiert werden. Da diese an eigenen empirischen Erhebungen des IQTIG bzw. strukturierten Beteiligungen in der Regel nicht selbst teilnehmen können, besteht die Gefahr, dass ihre Perspektiven, Bedürfnisse und Bedarfe vernachlässigt werden. Dem wird durch die Nutzung bereits vorliegender Erkenntnisse entgegengewirkt.

### **3.3.2 Beteiligung von Patientinnen und Patienten**

Bei der empirischen Erhebung kommen Methoden zum Einsatz, die betroffene Patientinnen und Patienten selbst in den Arbeitsprozess einbinden. Dies erfolgt in der Regel in Form einer strukturierten Beteiligung mittels Fokusgruppen oder Interviews (siehe Abschnitt 7.2.2). 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 und Geschlechtern. Das IQTIG strebt an, primär Patientinnen und Patienten einzubeziehen, die selbst von einer Erkrankung bzw. Behandlung betroffen sind. Weitere Kriterien wie Alter, Geschlecht, Bildungshintergrund, Krankheitsdauer und -schwere werden projektspezifisch nach methodischen Überlegungen wie der epidemiologischen Verteilung und dem angestrebten Erkenntnisinteresse bestimmt. Dabei werden, so weit wie möglich, die Diversität und unterschiedlichen Lebenslagen von Patientinnen und Patienten berücksichtigt.

Projektspezifisch bezieht das IQTIG ergänzend zu betroffenen Patientinnen und Patienten ggf. auch ihre Angehörigen ein oder nicht betroffene Patientinnen und Patienten, die eine spezifische Funktion einnehmen (z. B. Patientenfürsprecher im Krankenhaus). Gründe dafür wären beispielsweise, dass diese eine besondere Rolle im Versorgungsprozess der betroffenen Patientinnen und Patienten spielen (z. B. pflegende Angehörige, Eltern erkrankter Kinder) oder dass die

im Projekt primär adressierte Patientengruppe infolge von Gebrechlichkeit oder krankheitsspezifischer Einschränkungen an Fokusgruppen oder Interviews selbst nur eingeschränkt teilnehmen kann.

### **3.3.3 Beteiligung von Patientenvertreterinnen und Patientenvertretern**

Vom direkten Einbezug von betroffenen Patientinnen und Patienten, die in keinerlei Verbindung zur Qualitätssicherung oder zu Patientenverbänden stehen, ist der Einbezug der Patientenvertretung der maßgeblichen Organisationen nach § 140f SGB V (benannt über die Patientenvertretung des G-BA bzw. deren Mitgliedsorganisationen) zu unterscheiden. Der Einbezug und die Auswahl von Patientenvertreterinnen und -vertretern erfolgt projektspezifisch nach methodischen Überlegungen und dem angestrebten Erkenntnisinteresse. Bedarf es einer Einordnung und Vertretung von patientenrelevanten Themen in übergeordneten Kontexten, z. B. in Expertengremien (vgl. Kapitel 10), bindet das IQTIG primär Patientenvertreterinnen und -vertreter ein. Diese werden von der Patientenvertretung des G-BA bzw. deren Mitgliedsorganisationen (Vertretung der maßgeblichen Organisationen nach § 140f SGB V) entsandt. Darüber hinaus können Patientenvertreterinnen und -vertreter eingeladen werden, um spezifische Themen zu beraten, die für Patientinnen und Patienten von Belang sind, z. B. zum Thema Barrierefreiheit. Bei Bedarf können die Patientenvertretung des G-BA bzw. deren Mitgliedsorganisationen zudem unterstützend in die Rekrutierung von Patientinnen und Patienten für Fokusgruppen und Interviews involviert werden.

## 4 Grundlagen der Qualitätsmessung

Um Versorgungsqualität zu verbessern und die Ziele der Gesundheitsversorgung zu erreichen, ist die Messung und Bewertung der Qualität der zentrale Schritt (Lohr 1990: 45 f., Donabedian 2003: xxvi, Berwick et al. 2003), der die Grundlage für alle weiteren Maßnahmen bildet (vgl. Kapitel 2).

### 4.1 Methoden zur Messung von Qualität

Für die Messung und Darstellung der Versorgungsqualität auf Ebene einzelner Leistungserbringer sowie auf Ebene des Gesundheitssystems als Ganzes betrachtet das IQTIG quantitative Methoden als besonders geeignet, da sie im Vergleich zu qualitativen Methoden weitgehend standardisierbar sind und eine hohe Objektivität und eine hohe Reliabilität der Messung ermöglichen (siehe Eignungskriterien für Qualitätsmessungen, Abschnitt 13.3). Quantitative Methoden eignen sich daher besonders für die Prüfung von Annahmen (Döring und Bortz 2016: 184 f., Maats 2007: 293 f.) wie etwa, ob die Versorgung bestimmte Qualitätsanforderungen erfüllt. Die weitgehend automatisierbare Auswertung großer Datenmengen hat zudem im Vergleich zu qualitativen Methoden den Vorteil, dass ihr Aufwand nahezu unabhängig von der Anzahl der Behandlungsfälle ist und sie eine Qualitätsmessung anhand einheitlicher Regeln bei einer Vielzahl von Leistungserbringern und über viele Behandlungsfälle erlaubt.

#### Qualitätsindikatoren

Für die quantitative Analyse von Qualität werden Merkmale der Versorgung, wie etwa Komplikationen nach einem bestimmten Eingriff, messbar gemacht (Operationalisierung; Meyer 2004: 28 und 51, Döring und Bortz 2016: 223 und 228, Schnell et al. 2013: 117 und 121), über quantitative Größen dargestellt und mit den an sie gestellten Anforderungen verglichen. Diese qualitätsbezogenen Merkmale der Versorgung bezeichnet das IQTIG als Qualitätsmerkmale und die quantitativen Größen zu ihrer Messung werden als Qualitätsindikatoren bezeichnet. Da Qualität multidimensional ist und verschiedene Aspekte des Behandlungsprozesses von einem Leistungserbringer mit unterschiedlicher Qualität durchgeführt werden können (siehe Abschnitt 4.3), bezieht sich die Aussage eines Qualitätsindikators dabei ausschließlich auf das von ihm adressierte Qualitätsmerkmal. Beispielsweise soll ein Qualitätsindikator zur Sterblichkeit bei akutem Herzinfarkt nicht die Qualität der Infarktbehandlung insgesamt beschreiben, sondern nur darstellen, ob die Sterblichkeit von Patientinnen und Patienten mit diesem Krankheitsbild bei einem Leistungserbringer zu hoch ist.

Qualitätsindikatoren beschreiben den Grad der Erfüllung konkreter Qualitätsanforderungen. Bei Prozess- oder Ergebnisindikatoren 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 16) erlaubt dabei eine Aussage über die Erfüllung der Qualitätsanforderungen (Soll-Ist-Abgleich) durch Vergleich des Indikatorwerts mit

einem 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. eine Komplikationsrate von höchstens 5 %. Ein Qualitätsindikator umfasst somit nach Definition des IQTIG die folgenden drei Komponenten:

- 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 objektive und reliable Beurteilung erlauben, ob eine Qualitätsanforderung erfüllt ist.

Für die Messung 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 Erhebungsmethoden, wie die fallbezogene und die einrichtungsbezogene QS-Dokumentation und die Patientenbefragung, sowie sekundäre Datenanalysen auf Basis standardisierter Daten, wie die Analyse von Sozialdaten, die die Krankenkassen erheben.

### **Qualitative Analysemethoden**

Neben standardisierten quantitativen Methoden zur Qualitätsmessung stehen auch qualitative Analysemethoden zur Verfügung. 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). In einem zweiten Schritt werden die gewonnenen Erkenntnisse fallübergreifend interpretiert (vgl. Döring und Bortz 2016: 603 f.). Qualitative Methoden sind meist dann angezeigt, wenn Sachverhalte erklärt oder Hypothesen generiert werden sollen (Döring und Bortz 2016: 184 f., Maats 2007: 293 f.). 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 Ergebnisse 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., Grove et al. 2000, Kuncel et al. 2013, Dawes et al. 1989). Das IQTIG empfiehlt daher qualitative Methoden meist dann, wenn die Analyse nicht mit dem Ziel möglichst objektiver, standardisierter Bewertungen und Vergleiche erfolgt (beispielsweise, weil

die Bewertungsergebnisse wegen einer geringen Zahl von Untersuchungseinheiten oder einer hohen Komplexität des Sachverhalts mit einer hohen Unsicherheit einhergehen würden), sondern wenn z. B. die Ursachen für ein bestimmtes Indikatorergebnis oder wenn mögliche Maßnahmen zur Qualitätsverbesserung ermittelt werden sollen.

## 4.2 Datenquellen

Für die Messung der Versorgungsqualität stehen bisher drei verschiedene Datenquellen zur Verfügung:

### ▪ **Dokumentation der Leistungserbringer**

Dabei handelt es sich um Daten, die durch die Leistungserbringer zum Zweck der Qualitätssicherung dokumentiert werden. Dabei kann es sich um die Dokumentation von Behandlungsfällen (fallbezogene QS-Dokumentation) sowie um Angaben auf Einrichtungsebene (einrichtungsbezogene QS-Dokumentation), z. B. Strukturangaben, handeln.

### ▪ **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).

### ▪ **Befragung von Patientinnen und Patienten**

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

Behandlungsfallbezogene QS-Dokumentationen machen die Qualität anhand von Versorgungsmaßnahmen und erreichten Ergebnissen abbildbar. Einrichtungsbezogene QS-Dokumentationen dienen in erster Linie der Erhebung von Strukturparametern, beispielsweise bzgl. des Vorliegens von Hygieneregeln. Die Sozialdaten bei den Krankenkassen erlauben es darüber hinaus, den Behandlungsverlauf auch über einzelne Behandlungsepisoden in einer Einrichtung hinaus abzubilden. Des Weiteren stellen die Patientinnen und Patienten selbst eine zentrale Informationsquelle für die Beurteilung der Versorgungsqualität dar. Sie sind die Adressaten der Gesundheitsversorgung und können beispielsweise Auskunft über Behandlungsabläufe, Behandlungsergebnisse und Rahmenbedingungen der Behandlung geben. Patientenbefragungen sind daher eine wichtige Methode zur gezielten Erfassung patientenrelevanter Qualitätsaspekte und -merkmale (siehe Kapitel 7).

Die Datenquellen unterscheiden sich auch im Hinblick auf den Erhebungsaufwand und an welcher Stelle dieser anfällt. So ist der Dokumentationsaufwand für die Leistungserbringer unter Umständen hoch, wenn viele fallbezogene Daten nur für die Qualitätssicherung erfasst werden sollen. Geringer fällt der Aufwand in der Regel bei einer einrichtungsbezogenen QS-Dokumentation aus. Für die Nutzung der Sozialdaten bei den Krankenkassen entsteht bei den Leistungserbringern dagegen kein zusätzlicher Aufwand, 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.

Darüber hinaus unterscheiden sich die Datenquellen in ihrer Anfälligkeit für Fehlangaben (siehe Kapitel 19). Bei den Dokumentationsdaten der Leistungserbringer handelt es sich um selbstberichtete Daten, während durch Patientenbefragung erhobene Daten fremdberichtete Daten darstellen. Werden an Qualitätsmessungen mit selbstberichteten Daten beispielsweise Konsequenzen wie Anreize oder Auswahlentscheidungen geknüpft (siehe Abschnitt 2.4), entsteht ein Anreiz, diese Konsequenzen durch gezielte Dokumentation zu beeinflussen (vgl. Eckhardt et al. 2019). Die Sozialdaten bei den Krankenkassen werden von den Leistungserbringern primär für Abrechnungszwecke dokumentiert. Damit unterliegen sie als selbstberichtete Daten einerseits möglichen Fehlanreizen, andererseits werden sie im Rahmen der Abrechnung zusätzlichen Prüfungen unterzogen.

### 4.3 Qualitätsmessung als formatives Messmodell

Ein theoretisches Konstrukt – wie die Qualität der Gesundheitsversorgung – ist nicht direkt beobachtbar und muss über Indikatoren der Messung zugänglich gemacht werden. Die Beziehung zwischen einem theoretischen Konstrukt und den Indikatoren wird im Messmodell näher spezifiziert (Weiber und Mühlhaus 2014: 41). Dabei hat die Art des Messmodells wichtige Konsequenzen für die Auswahl von Indikatoren und für die Auswertung der Indikatorergebnisse (Bollen und Lennox 1991, Weiber und Mühlhaus 2014: 41 ff.).

Grundlegend werden formative und reflektive Messmodelle unterschieden. Wichtigstes Unterscheidungsmerkmal ist die angenommene Kausalrichtung zwischen Konstrukt und Indikatoren (Bollen und Lennox 1991). Bei formativen Messmodellen verläuft die Kausalrichtung von den Indikatoren zum Konstrukt, die Indikatoren *formen* das Konstrukt. Beispielsweise werden verschiedene Vitalfunktionen eines Neugeborenen (Indikatoren) im Apgar-Score kombiniert, um den Vitalstatus (Konstrukt) abzubilden (Apgar 1953)<sup>12</sup>. Dies wird in Abbildung 5 durch die Pfeilrichtung von den Indikatoren ( $x_1$  bis  $x_3$ ) hin zum Konstrukt ( $\eta$ ) symbolisiert. Bei reflektiven Messmodellen verläuft die Kausalrichtung vom Konstrukt zum Indikator. Dabei gehen die gemessenen Indikatorwerte auf eine einzige, gemeinsame Ursache – das Konstrukt – zurück und *reflektieren* diese. Beispielsweise gehen die Antwortausprägungen auf Befragungssitems zur Geselligkeit und Herzlichkeit eines Menschen (Indikatoren) auf die zugrunde liegende Persönlichkeitseigenschaft Extraversion (Konstrukt) zurück. Dies wird in Abbildung 5 durch die Pfeilrichtung vom Konstrukt ( $\eta$ ) hin zu den gemessenen Variablen ( $x_1$  bis  $x_3$ ) symbolisiert. Dementsprechend setzen reflektive Messmodelle ein unidimensionales Konstrukt voraus, dies ist bei formativen Messmodellen nicht der Fall (Bollen und Lennox 1991). Ob ein Konstrukt mit einem

---

<sup>12</sup> Eine ähnliche Unterscheidung wurde von Feinstein (1987) vorgeschlagen, der zwischen „clinimetrics“ und „psychometrics“ differenziert und den Apgar-Score als Beispiel für „clinimetrics“ anführt. „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.

formativen oder reflektiven Messmodell operationalisiert werden sollte, kann nicht formal getestet, sondern nur nach inhaltlichen Überlegungen entschieden werden (Jarvis et al. 2003, Weiber und Mühlhaus 2014: 261).<sup>13</sup>

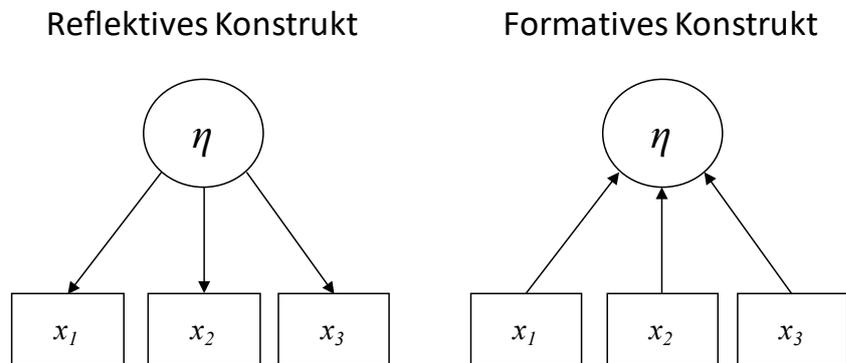


Abbildung 5: Angenommene Kausalrichtung bei formativen und reflektiven Messmodellen: Im Fall reflektiver Messmodelle verläuft die Kausalrichtung vom Konstrukt ( $\eta$ ) hin zu den gemessenen Indikatoren ( $x_1$  bis  $x_3$ ), im Fall formativer Messmodelle von den gemessenen Indikatoren hin zum Konstrukt.

Das IQTIG legt bei der Messung des Konstrukts „Qualität der Gesundheitsversorgung“ ein formatives Messmodell zugrunde, da die Qualität der Gesundheitsversorgung nicht auf eine einzige, zugrunde liegende Eigenschaft der Leistungserbringer zurückgeht, sondern durch verschiedene Qualitätsmerkmale bestimmt wird. Gemäß dem Rahmenkonzept für Qualität mit seinen sechs Dimensionen wird Qualität der Gesundheitsversorgung als mehrdimensionales Konstrukt verstanden (vgl. Abschnitt 1.1). Die sechs Dimensionen stellen verschiedene grundsätzliche Anforderungen an die Versorgung dar und können von den Leistungserbringern in jeweils unterschiedlichen Graden erfüllt werden (Shwartz et al. 2011); es ist beispielsweise denkbar, dass zwar hervorragende Behandlungserfolge (Qualitätsdimension: Wirksamkeit) erzielt werden, die Patientinnen und Patienten aber nicht in wünschenswertem Maße respekt- und würdevoll behandelt werden (Qualitätsdimension: Ausrichtung der Versorgungsgestaltung an Patientinnen und Patienten). Auch die Qualitätsmerkmale selbst stehen für unterschiedliche Anforderungen an die Versorgung, die unabhängig voneinander erfüllt sein können. Daher werden Qualitätsindikatoren grundsätzlich so abgeleitet, dass sie jeweils ein eigenständiges qualitätsrelevantes Ereignis oder eine spezifische qualitätsrelevante Situation erfassen. Aus dem Verständnis von Qualität im Sinne eines formativen Messmodells, bei dem es nicht „die“ Qualität z. B. eines Leistungserbringers oder einer Fachabteilung gibt, folgt darüber hinaus, dass es wichtig ist, für die Qualitätsmessung geeignete Qualitätsmerkmale zielgerichtet auszuwählen (siehe Kapitel 11).

<sup>13</sup> Je nach gewähltem Messmodell wird auch das Konstrukt selbst als formatives Konstrukt oder als reflektives Konstrukt bezeichnet.

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

## 5 Umsetzbarkeitsstudie für Qualitätsmessungen

Bei der Entwicklung indikatorbasierter Qualitätssicherungsverfahren (QS-Verfahren) ist die Umsetzbarkeit der Qualitätsmessungen von zentraler Bedeutung. Wenn Unsicherheit über die Umsetzbarkeit besteht, kann daher vor Entwicklung eines indikatorbasierten QS-Verfahrens für einen Themenbereich die Durchführung einer Umsetzbarkeitsstudie durch das IQTIG angezeigt sein. Dies ist beispielsweise unter folgenden Umständen der Fall:

- Indikatorbasierte Messungen für die gesetzliche Qualitätssicherung sind im betreffenden Versorgungsbereich noch nicht etabliert.
- Aus Vorarbeiten ergeben sich Hinweise auf Umsetzungshürden (z. B. auf Schwierigkeiten der Identifikation interessierender Behandlungsfälle in bestehenden Datenquellen).
- Neue Datenquellen sollen für die gesetzliche Qualitätssicherung nutzbar gemacht werden.

Durch eine Umsetzbarkeitsstudie soll sichergestellt werden, dass mögliche Umsetzungshürden frühzeitig erkannt werden und damit eine Entwicklung von unangemessen aufwendigen oder nicht praktikabel umsetzbaren QS-Verfahren vermieden wird.

### 5.1 Ziel und Fragestellungen

Das Ziel einer Umsetzbarkeitsstudie ist die vorläufige Einschätzung der Umsetzbarkeit der Qualitätsmessung in einem vom G-BA geplanten indikatorbasierten QS-Verfahren für einen Themenbereich. Mittels einer Umsetzbarkeitsstudie werden dazu vor einer möglichen Entwicklung von Qualitätsindikatoren Einschätzungen und Empfehlungen zu folgenden Fragen gegeben:

- Mittels welcher Daten und Erhebungsinstrumente kann eine Qualitätsmessung im interessierenden Themenbereich voraussichtlich erfolgen?
- Müssen über die bestehenden Instrumente der gesetzlichen Qualitätssicherung hinaus weitere technische (z. B. Datenflüsse) oder rechtliche Rahmenbedingungen (z. B. Zugriff auf neue Datenquellen) geschaffen werden?
- Welche Schritte werden für die Einführung und Umsetzung einer Qualitätsmessung für den interessierenden Themenbereich voraussichtlich benötigt?

### 5.2 Vorgehen

Für die Durchführung einer Umsetzbarkeitsstudie durch das IQTIG werden als Ausgangspunkt folgende Informationen vom G-BA benötigt:

- der Themenbereich, für den ein indikatorbasiertes QS-Verfahren entwickelt werden soll
- die Gründe, warum ein Bedarf an externer Qualitätssicherung für diesen Themenbereich gesehen wird; dies schließt die Information ein, für welche Patientengruppe und für welche Leistungserbringer Verbesserungspotenziale gesehen werden

Dabei kann beispielsweise auf Informationen aufgebaut werden, die im G-BA-Verfahren zur Themenfindung und Priorisierung ermittelt wurden (Kriterienkatalog). Soweit bereits Anhaltspunkte für mögliche Umsetzungshürden eines QS-Verfahrens vorliegen, werden diese bei der Umsetzbarkeitsstudie berücksichtigt.

### **5.2.1 Beschreibung von Patientenpopulation, interessierenden Versorgungsereignissen und adressierten Leistungserbringern**

Die Einschätzung der Umsetzbarkeit einer Qualitätsmessung mit Indikatoren erfordert eine Abgrenzung, welche Patientengruppen, Leistungserbringer und Versorgungsereignisse in das QS-Verfahren eingeschlossen werden können und sollen. Dazu nimmt das IQTIG eine Beschreibung des Versorgungsgeschehens auf Basis einer orientierenden Recherche und einer Beratung durch Expertinnen und Experten vor und erstellt einen Versorgungspfad. Gegebenenfalls greift das IQTIG zusätzlich auf allgemein verfügbare anonymisierte Datensätze zur Versorgungsforschung zurück, die auftragsspezifisch zum Projektstart durch anonymisierte Abrechnungsdaten der Kassenärztlichen Bundesvereinigung (KBV) und des Spitzenverbands Bund der Krankenkassen (GKV-SV) ergänzt werden können. Anhand dieser Informationen werden die Häufigkeit eines Versorgungsereignisses, die Gruppen und die Zahl der an der Versorgung beteiligten Leistungserbringer und ihre Verteilung auf die einzelnen Sektoren, die Fallzahlen pro Leistungserbringer, Behandlungsverläufe und Ergebnisse der Versorgung eingeschätzt. Darüber hinaus wird untersucht, welche Auswirkungen die normativen Rahmenbedingungen im Themenbereich auf eine indikatorbasierte Qualitätsmessung für die gesetzliche Qualitätssicherung haben.

### **5.2.2 Identifizierbarkeit der interessierenden Versorgungsereignisse**

Es wird untersucht, ob die interessierenden Patientengruppen und Versorgungsereignisse in den bestehenden Datenquellen (siehe Abschnitt 4.2) identifiziert werden können (z. B. anhand von Codes für die Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (*International Statistical Classification of Diseases and Related Health Problems*, ICD) oder Operationen- und Prozedurenschlüssel (OPS)) und ob deren Abgrenzung zwecks Einschluss in ein QS-Verfahren automatisiert erfolgen kann (sogenannte Auslösung, z. B. über den QS-Filter für QS-Dokumentationen oder den Patientenfilter für Sozialdaten bei den Krankenkassen). Außerdem wird ermittelt, ob sich dabei für verschiedene Sektoren, Versorgungsformen oder Leistungserbringergruppen Unterschiede ergeben.

### **5.2.3 Für die Qualitätsmessung verfügbare Informationen**

Ausgehend von der interessierenden Grundgesamtheit der Patientinnen und Patienten sowie den zu adressierenden Leistungserbringern und unter Berücksichtigung der Zielsetzung des potenziellen QS-Verfahrens (siehe Abschnitt 5.2.1) prüft das IQTIG, welche Informationen für die Qualitätsmessung im Themenbereich in bestehenden Daten grundsätzlich verfügbar sind. Diese können bereits vorliegen (z. B. in Abrechnungsdaten) oder durch Erhebungsinstrumente der gesetzlichen Qualitätssicherung verfügbar gemacht werden (z. B. mittels QS-Dokumentation oder Patientenbefragung). Es wird eine vorläufige Empfehlung erarbeitet, welche Datenquellen und Erhebungsinstrumente der gesetzlichen Qualitätssicherung voraussichtlich benötigt werden.

Diese Empfehlung berücksichtigt qualitätsrelevante Themen, die bereits aus der Beauftragung bekannt sind oder die im Rahmen der Recherche zum Versorgungsgeschehen ermittelt wurden. Eine detaillierte Prüfung erfolgt bei der späteren Entwicklung von Qualitätsindikatoren auf Grundlage eines Qualitätsmodells (siehe Kapitel 6 und 7).

Des Weiteren wird untersucht, mit welchem zeitlichen Abstand zur Versorgungsleistung die benötigten Daten für die Qualitätsmessung zur Verfügung stehen und inwieweit sich Informationen aus verschiedenen Datenquellen oder von verschiedenen Erfassungszeitpunkten grundsätzlich verknüpfen lassen.

#### 5.2.4 Neue Datenquellen und Datenflüsse

Werden für eine angemessene Qualitätsdarstellung im beauftragten Themenbereich voraussichtlich Daten benötigt, die nicht mit den etablierten Erhebungsinstrumenten der gesetzlichen Qualitätssicherung nach §§ 136 ff. SGB V erfassbar sind (siehe Abschnitt 4.2), so prüft das IQTIG, ob alternative oder ergänzende Datenquellen in Betracht kommen. Für sie wird untersucht,

- welche Informationen aus diesen Datenquellen grundsätzlich ermittelt werden können,
- ob Informationen vorliegen, wie die Datenqualität einzuschätzen ist, und welche Möglichkeiten zu deren Prüfung bestehen,
- welche Verwendungsmöglichkeiten sich aus rechtlicher Sicht ergeben,
- welche Maßnahmen zur Gewährleistung der Datensicherheit und des Datenschutzes (z. B. Verschlüsselungsverfahren, Pseudonymisierung und Anonymisierung der Daten) voraussichtlich benötigt werden und
- welche Schritte für die Einrichtung der Datenflüsse und die Erhebung der Daten erforderlich sind.

Auf dieser Grundlage gibt das IQTIG eine vorläufige Einschätzung ab, ob diese Datenquellen bei der Entwicklung von Qualitätsmessungen im beauftragten Themenbereich herangezogen werden sollten. Wenn diese Einschätzung positiv ausfällt, wird ein Datenflusskonzept skizziert, das die beteiligten Akteure sowie technische Aspekte der Umsetzung darstellt.

### 5.3 Ergebnis der Umsetzbarkeitsstudie

Auf Grundlage der o. g. Analysen gibt das IQTIG Einschätzungen und Empfehlungen zu folgenden Aspekten der Qualitätsmessung in einem möglichen indikatorbasierten QS-Verfahren ab:

- Identifikation und Abgrenzung der Grundgesamtheit der Patientinnen und Patienten, für die Versorgungsqualität gemessen werden soll
- grundsätzliche Möglichkeiten zur Abbildung der Qualität des Versorgungsgeschehens innerhalb des beauftragten Themenbereichs
- Möglichkeiten einer automatisierten Identifikation der interessierenden Behandlungsfälle in den Datenquellen (sogenannte Auslösung für die Qualitätssicherung)
- für die Entwicklung der Qualitätsmessung empfohlene Datenquellen und Erhebungsinstrumente der gesetzlichen Qualitätssicherung
- ggf. Unterschiede zwischen verschiedenen Gruppen von Leistungserbringern und zwischen verschiedenen Versorgungsformen hinsichtlich der Verfügbarkeit von Daten

- ggf. Skizze eines Konzepts zur Nutzung ergänzender oder alternativer Datenquellen und Datenflüsse
- ggf. technische oder rechtliche Rahmenbedingungen für die Etablierung neuer Datenflüsse
- erforderliche Schritte für die Etablierung und für die Durchführung einer Qualitätsmessung im jeweiligen Themenbereich

Auf dieser Basis spricht das IQTIG eine Empfehlung aus, ob im jeweiligen Themenbereich grundsätzlich eine angemessene und praktikable Qualitätsmessung mit Indikatoren voraussichtlich möglich ist und welche Voraussetzungen ggf. vor Einführung und Umsetzung eines QS-Verfahrens geschaffen werden müssen.

## 6 Entwicklung von dokumentations- und sozialdatenbasierten Qualitätsindikatoren

Die Entwicklung von Qualitätsindikatoren auf Basis von QS-Dokumentationsdaten und Sozialdaten bei den Krankenkassen beginnt mit der Entwicklung eines Qualitätsmodells (siehe Kapitel 11). Dieses stellt eine Strukturierung der Inhalte eines Themenbereichs dar und beschreibt sie anhand von sogenannten Qualitätsaspekten. Damit bestimmen die Qualitätsaspekte des Qualitätsmodells die inhaltlichen Schwerpunkte der Indikatorenentwicklung. Die Indikatorenentwicklung erfolgt in mehreren Schritten (Abbildung 6): Zunächst werden die Qualitätsaspekte in Form von Qualitätsmerkmalen konkretisiert, diese werden anschließend zu Qualitätsindikatoren operationalisiert und zu einem Indikatorenset zusammengeführt. Die Qualitätsindikatoren werden einem Beteiligungsverfahren unterzogen (siehe Kapitel 12) und – sofern sie auf QS-Dokumentationsdaten basieren – in einer Machbarkeitsprüfung erprobt. Anschließend werden die optimierten Indikatoren zum final empfohlenen Indikatorenset einschließlich seines Erhebungsinstruments zusammengestellt.

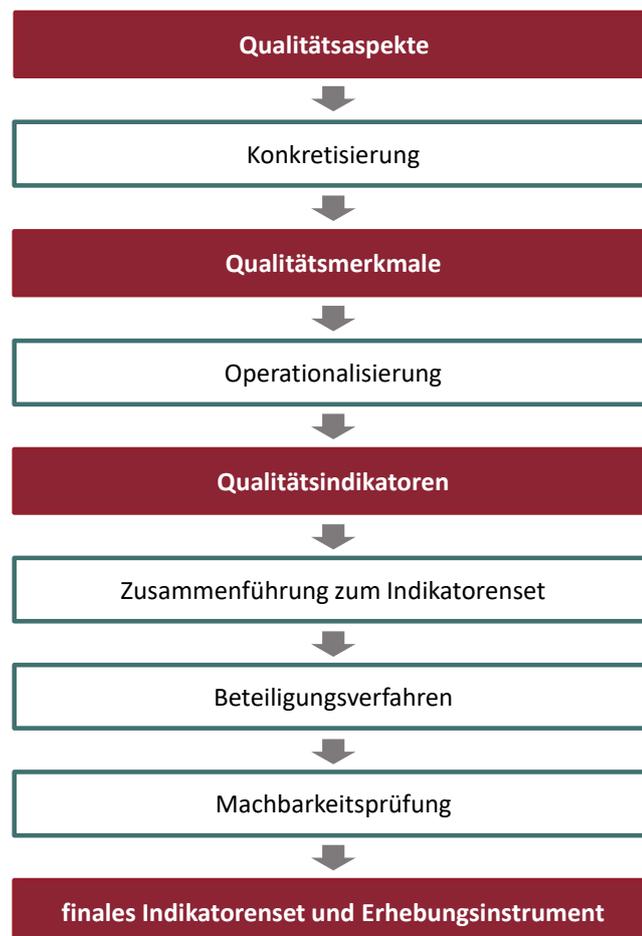


Abbildung 6: Entwicklungsschritte für dokumentations- und sozialdatenbasierte Qualitätsindikatorensets

Im gesamten Verlauf der Entwicklung von Qualitätsindikatoren auf Basis von fallbezogenen bzw. einrichtungsbezogenen QS-Dokumentationen und Sozialdaten bei den Krankenkassen wird ein Expertengremium hinzugezogen. Das Expertengremium hat bei allen Entwicklungsschritten eine beratende Funktion für das IQTIG (siehe Kapitel 10), wobei die Entscheidungen für die Empfehlungen an den G-BA in der Verantwortung des IQTIG liegen.

Bei der Zusammensetzung des Expertengremiums wird insbesondere darauf geachtet, Personen aus unterschiedlichen Gesundheitsprofessionen einzubeziehen und dass in diesem beratenden Gremium nicht nur Personen mit Leitungsposition, sondern auch die mittlere Führungsebene, unter Berücksichtigung ausreichender Berufserfahrung und wissenschaftlicher Expertise, angemessen vertreten ist. Ebenso wird eine angemessene Anzahl an Expertinnen und Experten aus den beteiligten Sektoren angestrebt. Neben klinischer und wissenschaftlicher Expertise können auch Kodierexpertinnen und -experten Teil des Expertengremiums sein. Darüber hinaus werden Patientenvertreterinnen und Patientenvertreter regelhaft einbezogen (siehe Kapitel 3).

## **6.1      Konkretisierung der Qualitätsaspekte zu Qualitätsmerkmalen**

Für jeden Qualitätsaspekt des Qualitätsmodells werden patientenrelevante Qualitätsmerkmale identifiziert, die wichtige Teile des jeweiligen Aspekts abbilden. Diese Qualitätsmerkmale sind mit konkreten, begründeten Qualitätsanforderungen an die Versorgungspraxis (Struktur, Prozess, Ergebnis) verbunden, deren Erfüllung im Verfahren der gesetzlichen Qualitätssicherung gemessen werden soll. Dabei orientieren sich die Qualitätsmerkmale am Rahmenkonzept für Qualität (siehe Abschnitt 1.1) und sie müssen insbesondere die Eignungskriterien des Qualitätsziels (siehe Abschnitt 13.2) erfüllen. Die Bewertung der Eignungskriterien erfolgt qualitativ, wobei die Gründe für diese Bewertung im Abschlussbericht dargelegt werden. Dabei erfolgt die Prüfung der Eignungskriterien unter eingehender Beratung mit dem Expertengremium. 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 weiteren Entwicklungsschritte übernommen.

### **6.1.1    Sicherstellung der Bedeutung der Qualitätsmerkmale**

Die Qualitätsmerkmale werden entsprechend den Eignungskriterien „Bedeutung für die Patientinnen und Patienten“ und „Zusammenhang mit einem unmittelbar patientenrelevanten Merkmal“ entwickelt (siehe Abschnitt 13.2). Um relevante Abläufe oder Ereignisse der Patientenversorgung zu identifizieren, die für einen Qualitätsaspekt des Qualitätsmodells als Qualitätsmerkmal genutzt werden könnten, wird auf die Wissensbestände zurückgegriffen, die im Rahmen der Entwicklung des Qualitätsmodells recherchiert bzw. erhoben wurden (siehe Kapitel 11). Die Ableitung der Qualitätsmerkmale erfolgt entlang des Versorgungspfads auf Grundlage der besten verfügbaren Evidenz dieser Wissensbestände. 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 Qualitätsmerkmale nur aus jenen Qualitätsanforderungen abgeleitet, die einen Empfehlungsgrad A oder B der deutschen Leitlinien der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V. (AWMF) (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.

In der weiteren Entwicklung wird festgelegt, bei welchen Patienten- und Leistungserbringergruppen innerhalb der Grundpopulation des QS-Verfahrens das jeweilige Qualitätsmerkmal erhoben werden soll und welches Erhebungsinstrument 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.

### **6.1.2 Berücksichtigung von Verbesserungspotenzial und Zuschreibbarkeit der Verantwortung**

Im nächsten Schritt werden die abgeleiteten Qualitätsmerkmale daraufhin geprüft, ob ihre Messung der Sicherstellung und Verbesserung der Versorgungsqualität dienen kann (Eignungskriterium „Potenzial zur Verbesserung“) und ob sie von den Leistungserbringern soweit beeinflussbar sind, dass diesen die Verantwortung dafür zugeschrieben werden kann (Eignungskriterium „Beeinflussbarkeit durch den Leistungserbringer“; siehe Kapitel 13). Die Prüfung der Qualitätsmerkmale anhand dieser Eignungskriterien durch das IQTIG erfolgt auf Basis der folgenden Wissensquellen:

- Ergebnisse der Literaturrecherche
- Ergebnisse der Sozialdatenanalyse
- Einschätzung des Expertengremiums

Hierfür werden die Literaturrecherchen und Sozialdatenanalysen aus der Entwicklung des Qualitätsmodells ggf. 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. Dabei wird auch der Zusammenhang zwischen Struktur- und Prozessmerkmalen und patientenrelevanten Outcomes untersucht.

Am Ende der Konkretisierung der Qualitätsmerkmale werden diese durch das beratende Expertengremium hinsichtlich der beiden oben genannten Eignungskriterien bewertet. Die Erfassung der Expertenmeinung erfolgt dabei in Anlehnung an die RAND/UCLA-Appropriateness-Method (RAM) (Fitch et al. 2001, Anwendungen in der Qualitätsindikatorenentwicklung 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 schriftlichen Vorabbewertungen als auch in 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 dienen dem IQTIG als Empfehlung und werden im Ergebnisbericht dokumentiert.

Die Entscheidung, welche Qualitätsmerkmale Grundlage der weiteren Indikatorenentwicklung sind, wird durch das IQTIG nach einer umfassenden Aufbereitung und Gegenüberstellung der Ergebnisse aus den oben genannten Wissensquellen getroffen. Damit liegen die Qualitätsmerkmale vor, die die oben genannten Eignungskriterien des Qualitätsziels erfüllen. Diese Qualitätsmerkmale werden im nächsten Entwicklungsschritt operationalisiert.

## **6.2    Operationalisierung der Qualitätsmerkmale**

Um Qualitätsmerkmale messen und deren Ausprägung leistungserbringerbezogen bewerten zu können, werden sie mittels Qualitätsindikatoren operationalisiert. Im Folgenden wird die Operationalisierung für Qualitätsindikatoren beschrieben, die auf Basis von fall- bzw. einrichtungsbezogener QS-Dokumentation oder von Sozialdaten bei den Krankenkassen berechnet werden sollen.

### **6.2.1    Fall- und einrichtungsbezogene QS-Dokumentation**

Zur Erfassung fall- und einrichtungsbezogener Daten für die Qualitätssicherung werden vom IQTIG Dokumentationsbögen entwickelt. Diese werden später durch die Hersteller der Softwarelösungen für die Dokumentation elektronisch umgesetzt. Durch das IQTIG werden dabei alle in den Dokumentationsbögen enthaltenen Texte (wie z. B. Abschnittsüberschriften, Datenfeldbezeichnungen oder Bezeichnungen der Antwortoptionen) vorgegeben; die visuelle Gestaltung der Dokumentationsbögen (z. B. Anordnung der Datenfelder auf der Benutzeroberfläche, Umsetzung der Antwortoptionen als Dropdown-Liste oder über Checkboxes) liegt in der Verantwortung der Hersteller von QS-Dokumentationssoftware, dafür macht das IQTIG keine Vorgaben.

Die Entwicklung der Dokumentationsbögen durch das IQTIG erfolgt mit dem Ziel, eine Datenerfassung entsprechend den Eignungskriterien des Messverfahrens (siehe Abschnitt 13.3) zu ermöglichen, insbesondere die Objektivität, Reliabilität und Praktikabilität der Messung des Qualitätsindikators sowie eine hohe Datenqualität zu gewährleisten.

### **Auswahl der für die Qualitätssicherung zu dokumentierenden Daten**

Die Aufnahme von Datenfeldern in den Dokumentationsbogen folgt dem Gebot der Datensparsamkeit: Für die Qualitätssicherung werden nur Daten erfasst, die auch für die Durchführung der Qualitätssicherung benötigt werden. Dokumentationsbögen enthalten daher nur Datenfelder, die

- für die Fallidentifikation,
- zur Berechnung eines Qualitätsindikators oder einer Kennzahl,
- für die Risikoadjustierung (siehe Abschnitt 20.2),
- für wichtige Zusatzinformationen für das Stellungsverfahren nach § 17 der Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung (DeQS-RL)<sup>14</sup>
- aus technischen Gründen
- oder in Ausnahmefällen auch zur Erhöhung der Dokumentationsqualität eines anderen Datenfelds

benötigt werden. Dokumentationsbögen enthalten keine Datenfelder, die von den Leistungserbringern optional auszufüllen sind.

Weiterhin werden im Dokumentationsbogen nur Daten erfasst, die in der Versorgung, z. B. an den Patientinnen und Patienten, ausreichend objektiv und reliabel erhoben werden. Dies wird dadurch erzielt, dass keine Meinungen, Einstellungen oder Überzeugungen der Dokumentierenden erfasst werden, sondern Daten und Informationen aus den bei den Leistungserbringern vorliegenden Originalunterlagen wie insbesondere der Patientenakte. Dies ermöglicht auch aufgrund der Überprüfbarkeit der Daten eine hohe Datenqualität. Eine hohe Praktikabilität wird dadurch gewährleistet, dass bei den Leistungserbringern nach Möglichkeit leicht verfügbare oder automatisch in die QS-Dokumentation zu übertragende Daten erfasst werden.

### **Entwicklung von Dokumentationsbögen**

Beim Zusammenfügen der einzelnen Datenfelder zu einem Bogen steht die Praktikabilität für die Dokumentierenden im Vordergrund: So entspricht die Reihenfolge der Datenfelder auf dem Dokumentationsbogen so weit wie möglich dem Behandlungsablauf; die Übersichtlichkeit wird durch das Anlegen von thematisch gegliederten Abschnitten erhöht; und Filterfelder sowie abhängige Datenfelder werden gezielt so eingesetzt, dass in möglichst vielen Fällen möglichst wenige Datenfelder auszufüllen sind.

---

<sup>14</sup> Richtlinie des Gemeinsamen Bundesausschusses zur datengestützten einrichtungsübergreifenden Qualitätssicherung. In der Fassung vom 19. Juli 2018, zuletzt geändert am 16. September 2021, in Kraft getreten am 1. Januar 2022. URL: <https://www.g-ba.de/informationen/richtlinien/105/> (abgerufen am 21.03.2022).

## Entwicklung von Datenfeldern

Das IQTIG orientiert sich bei der Entwicklung von Datenfeldern an den Grundsätzen, die in der sozialwissenschaftlichen Literatur für die Entwicklung von Fragebögen formuliert werden (Porst 2014, Faulbaum et al 2009), da die kognitiven Prozesse beim Ausfüllen eines Fragebogens und bei der Dokumentation von Daten für die Qualitätssicherung vergleichbar sind: In beiden Fällen muss zuerst die Frage / das Datenfeld verstanden werden und es müssen relevante Informationen identifiziert werden. Auf Basis dieser Informationen muss beurteilt werden, welcher Sachverhalt zu dokumentieren ist, und dieser in ein vorgegebenes Antwortformat eingepasst werden. Daher legt das IQTIG folgende Anforderungen an die Entwicklung von Datenfeldern an:

- Alle Texte in Datenfeldern, einschließlich der Antwortoptionen, sollen möglichst präzise und eindeutig, gleichzeitig aber auch so kurz und einfach wie möglich formuliert sein.
- Antwortoptionen müssen vollständig und überschneidungsfrei sein und den Stufen/Kategorien entsprechen, die in der Versorgung für den betreffenden Sachverhalt verbreitet sind.
- Datenfelder dürfen nicht Sachverhalte voraussetzen, die nicht in allen Fällen zutreffen.
- Datenfelder dürfen nicht mehrere Sachverhalte enthalten, die nicht gemeinsam zu beantworten sind („double-barrelled“). Andernfalls besteht die Gefahr, dass das Datenfeld nicht beantwortbar ist oder dass manche Teilnehmerinnen und Teilnehmer die Frage mit Bezug auf den einen Sachverhalt beantworten und andere mit Bezug auf den anderen Sachverhalt.
- Das Datenfeld muss alle für eine korrekte Dokumentation notwendigen Informationen enthalten. Informationen, die nur für die Dokumentation bestimmter Sonder- bzw. Ausnahmefälle relevant sind, werden dabei in einem im Bedarfsfall abzurufenden Ausfüllhinweis platziert.
- Um Kategorisierungsfehler zu vermeiden, werden in der Versorgung vorliegende kontinuierliche Daten als kontinuierliche Werte und nicht in Kategorien erfasst; ggf. erfolgt eine Kategorisierung bei der Auswertung.
- Datenfelder, die eine Berechnung durch die Dokumentierenden erfordern, werden vermieden.
- Um die Wahrscheinlichkeit „sozial erwünschter“ Antworten gering zu halten, wird die Erfüllung von in der Versorgung gut bekannten und als verbindlich angesehenen Normen (z. B. Leitlinien) möglichst nicht in direkter Form abgefragt.

Die Verständlichkeit der Texte für Datenfelder richtet das IQTIG auf die Berufsgruppe aus, die für die Gesundheitsversorgung oder Behandlung im betreffenden Bereich verantwortlich ist (z. B. Ärztinnen und Ärzte, Pflege, Therapeutinnen und Therapeuten).

Um der Unterdokumentation von unerwünschten Ereignissen wie z. B. Komplikationen (Maass et al. 2011, Wessling et al. 2016, Jakob et al. 2010, Döbler et al. 2016) entgegenzuwirken, erfolgt die Erfassung von eindeutig als negativ im Hinblick auf das Qualitätsziel des Indikators zu wertenden Sachverhalten (z. B. von Komplikationen) durch eine verpflichtende Ja-Nein-Abfrage („forced-choice“). Dieses Vorgehen gründet sich auf Befunde aus der Literatur, dass per Forced-choice-Abfrage durchgängig mehr Antwortoptionen als zutreffend ausgewählt werden als bei Erfassung durch Mehrfachauswahl (Review in Callegaro et al. 2015, s. auch Delnevo et al. 2017).

## QS-Filter

Die Zielpopulation des QS-Verfahrens, die zu Beginn der Entwicklung festgelegt wird (siehe Abschnitt 6.1.1), muss für die Auslösung der Datenerhebung operationalisiert werden. Für die fall- bzw. einrichtungsbezogene QS-Dokumentation wird daher zusätzlich zu den Datenfeldern und Dokumentationsbögen der QS-Filter entwickelt. Dieser umfasst mindestens alle Behandlungsfälle, die durch das Indikatorenset innerhalb der Zielpopulation des QS-Verfahrens abgebildet werden sollen.

### 6.2.2    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 inhaltsvalide 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
- welche 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 Prozedurenkodes oder anderen zugänglichen kodierten Informationen, welche eine möglichst gut zutreffende Beschreibung des mit dem Qualitätsmerkmal thematisierten Versorgungsereignisses erlaubt
2. die Formulierung von Rechenregeln, die festlegen, in welcher Kombination und (bei Follow-up-Indikatoren) wann die Codes dokumentiert worden sein sollen; die verbale Beschreibung einer Rechenregel 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“

Für die Sozialdaten bei den Krankenkassen werden die einzuschließenden Behandlungsfälle über den Patientenfilter beschrieben. Darüber hinaus werden Leistungs- und Medikationsfilter entwickelt, die die zu übermittelnden Daten beschreiben.

### 6.2.3    Entwicklung von Auswertungsalgorithmen

Neben der beschriebenen Festlegung der für die Qualitätssicherung zu übermittelnden Daten stellen das Konzept zur Datenaufbereitung und die Entwicklung von Rechenregeln wichtige Bestandteile der Auswertungsalgorithmen und damit der Operationalisierung eines Qualitätsmerkmals dar, die teilweise erst in späteren Entwicklungsschritten bzw. bei Umsetzung eines QS-Verfahrens in den Regelbetrieb abgeschlossen werden können.

Das Konzept zur Datenaufbereitung umfasst die Aufteilung der Daten in verschiedene Auswertungsebenen wie Patienten- und Prozedurebene sowie die Verknüpfung von mehreren Behandlungsfällen einer Patientin oder eines Patienten (z. B. im Rahmen von Follow-up-Indikatoren) und die Verknüpfung von fallbasierten QS- und Sozialdaten. Des Weiteren werden Kriterien festgelegt, mit denen die Plausibilität der Daten überprüft werden soll.

Mittels der Rechenregeln werden die relevanten Populationen und die interessierenden Ereignisse für einen Qualitätsindikator operationalisiert. Dafür werden üblicherweise auf der untersten Auswertungsebene, beispielsweise der Prozedur- oder Patientenebene, Ein- und Ausschlusskriterien definiert, die beschreiben, welche Fälle zur Population eines Qualitätsindikators gehören. Alle relevanten Fälle auf der festgelegten Auswertungsebene bilden dann die Menge der eingeschlossenen Fälle eines Qualitätsindikators (siehe Abschnitt 18.1). Dabei wird angestrebt, dass die über die Rechenregeln definierte Menge an Fällen möglichst vollständig mit der eigentlich interessierenden Zielpopulation übereinstimmt, um eine hohe Validität der Messung zu erreichen (siehe Abschnitt 13.3.4). Zunächst wird die konkrete Operationalisierung der Indikatoren in Form einer präzisen textlichen Beschreibung vorgenommen. Sobald Echtdaten verfügbar sind, werden diese Festlegungen in ausführbaren Programmcode umgesetzt.

#### Referenzbereiche und Risikofaktoren

Zur Operationalisierung eines Qualitätsmerkmals mittels Qualitätsindikator gehört auch die Empfehlung eines Referenzbereichs. Es kann sich um einen verteilungsunabhängigen (festen) Referenzbereich oder um einen verteilungsabhängigen (perzentilbasierten) Referenzbereich, z. B. festgelegt über das 5. oder 95. Perzentil, handeln (siehe Kapitel 16). Das IQTIG empfiehlt in der Regel verteilungsunabhängige Referenzbereiche, da sie ein klar definiertes, prospektives Ziel für den Leistungserbringer vorgeben.

Sofern in den Literaturrecherchen zu einem Qualitätsmerkmal patientenbezogene Risikofaktoren, die das Ergebnis eines entsprechenden Indikators beeinflussen könnten, identifiziert wurden, werden diese in die Beschreibung des Qualitätsindikators aufgenommen. Darüber hinaus werden die bei der Risikoadjustierung zu berücksichtigenden Faktoren auch im Rahmen der Expertenberatungen thematisiert und die Angemessenheit der vorgeschlagenen Referenzbereiche für die Abbildung der vorgesehenen Bewertungskategorien wird mit den Expertinnen und Experten diskutiert. Die Risikofaktoren werden zu einem späteren Zeitpunkt bei der Entwicklung eines Risikoadjustierungsmodells berücksichtigt.

### 6.3      Zusammenführung zum Qualitätsindikatorenset

Gegenstand dieses Entwicklungsschritts ist die Betrachtung der Qualitätsindikatoren insgesamt. Ziel dieser Gesamtbetrachtung ist u. a. die Sicherstellung eines inhaltsvaliden Sets an Qualitätsindikatoren (siehe Kapitel 14). Zur Übersicht werden die Qualitätsindikatoren den sechs Qualitätsdimensionen des Rahmenkonzepts des IQTIG für Qualität (siehe Abschnitt 1.1) und den Dimensionen der Struktur-/Prozess-/Ergebnisqualität zugeordnet. Aus dieser Betrachtung können Einschätzungen zu Stärken und Schwächen des Qualitätsindikatorensets für das im geplanten QS-Verfahren vorgesehene Steuerungsinstrument getroffen werden.

Konkret werden folgende Fragen adressiert:

- Liegen Überschneidungen zwischen den Indikatoren vor?
- Gibt es Widersprüche in den Zielen und Anreizen der Indikatoren?
- Handelt es sich um ein in der Erhebung datensparsames, im Aufwand für die Leistungserbringer angemessenes Indikatorenset?
- Ist das Indikatorenset auf die wichtigsten Verbesserungsbedarfe fokussiert?

Auf dieser Grundlage werden die Qualitätsindikatoren zu einem Indikatorenset zusammengeführt. Darüber hinaus wird durch das IQTIG eine Würdigung vorgenommen, ob das Indikatorenset dem intendierten Zweck des Verfahrens entspricht und das Qualitätsmodell für den Themenbereich angemessen abbildet.

### 6.4      Entwicklung von Empfehlungen zur Verfahrensumsetzung

Zusätzlich zum Qualitätsindikatorenset werden Empfehlungen und Informationen zu folgenden weiteren Elementen der Umsetzung des vorgesehenen QS-Verfahrens ausgearbeitet und im Hinblick auf Auftrag und Richtlinienanforderungen geprüft und ggf. angepasst:

- Details zur Datenerhebung:
  - Datenflüsse
  - Erfassungszeiträume (Follow-up)
  - Verfügbarkeit der Daten (Lieferfristen, Datenexporte)
  - Überprüfung der Dokumentationspflicht (Sollstatistik)
  - Anforderungen an eine Datenvalidierung
  - Vollerhebung oder Stichprobenverfahren
- Berichtskonzept:
  - Berichtszeitpunkte
  - Inhalte für die Rückmeldeberichte an die beteiligten Leistungserbringer (Qualitätsindikatoren, Kennzahlen)
- möglicher Zeitplan bis zur Umsetzung des QS-Verfahrens in den Regelbetrieb
- Empfehlungen zu weiteren Schritten bis zum Regelbetrieb einschließlich Hinweisen zum Umgang mit ggf. erwarteten Umsetzungshürden für das QS-Verfahren

Die Ergebnisse der Indikatorenentwicklung und die Empfehlungen zur Umsetzung des QS-Verfahrens werden einem Beteiligungsverfahren unterzogen (siehe Kapitel 12), in dem die Möglichkeit zur Kommentierung der Entwicklungen und des vorgeschlagenen Indikatorensets besteht. Die Rückmeldungen und Anregungen der beteiligten Organisationen und Institutionen werden vom IQTIG bei den weiteren Entwicklungen angemessen berücksichtigt.

## **6.5    Machbarkeitsprüfung für Dokumentation und Auswertung**

Im Rahmen der Entwicklung von Qualitätsindikatoren auf Basis von QS-Dokumentationen der Leistungserbringer erfolgt eine Pretestung der Qualitätsindikatoren. 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, Unschärfen der Operationalisierung von Qualitätsmerkmalen oder der Definition des QS-Filters 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 13.3) 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).

### **6.5.1    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 (Prüfung der Datenfeld- und Schlüsselbeschriftungen, Ausfüllhinweise und Plausibilitätsregeln)
2. der QS-Filter (Prüfung der Ein- und Ausschlusskriterien für die zu dokumentierenden Fälle)
3. die Qualitätsindikatoren, soweit sie auf der Basis von QS-Dokumentationsdaten berechnet werden können (Prüfung des Anteils auswertbarer Behandlungsfälle bzw. fehlender Werte)

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. eine Erprobung 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. Auf Grundlage der Machbarkeitsprüfung werden ggf. die Empfehlungen zu Datenfeldern oder zur Operationalisierung der Indikatoren überarbeitet. 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 unter Richtlinienbedingungen erfolgen.

Sozialdatenbasierte Qualitätsindikatoren werden in einer Machbarkeitsprüfung nicht berücksichtigt. Die Dokumentationsqualität und Rechenregeln für rein sozialdatenbasierte Indikatoren werden bereits im Rahmen der Indikatorenentwicklung durch Probeauswertungen auf der Basis anonymisierter Sozialdaten einer großen Krankenkasse geprüft und optimiert (vgl. Abschnitt 6.2.2). Qualitätsindikatoren, die eine patientenbezogene Verknüpfung von QS-Dokumentationsdaten und Sozialdaten bei den Krankenkassen benötigen, können dagegen erst in der initialen Phase im Regelbetrieb geprüft werden, da diese Verknüpfung datenschutzrechtlich erst im Rahmen des verpflichtenden Regelbetriebs erlaubt ist.

### 6.5.2 Ablauf der Machbarkeitsprüfung

Die zentralen Elemente einer Machbarkeitsprüfung sind eine Probedokumentation durch Leistungserbringer sowie eine Nachbefragung der Leistungserbringer. Ausgangspunkt einer Machbarkeitsprüfung ist das Qualitätsindikatorenset sowie der Entwurf eines Dokumentationsbogens zur Erfassung der erforderlichen Daten. Abbildung 7 zeigt den Ablauf einer Machbarkeitsprüfung im Überblick. Die einzelnen Arbeitsschritte werden im Folgenden beschrieben.

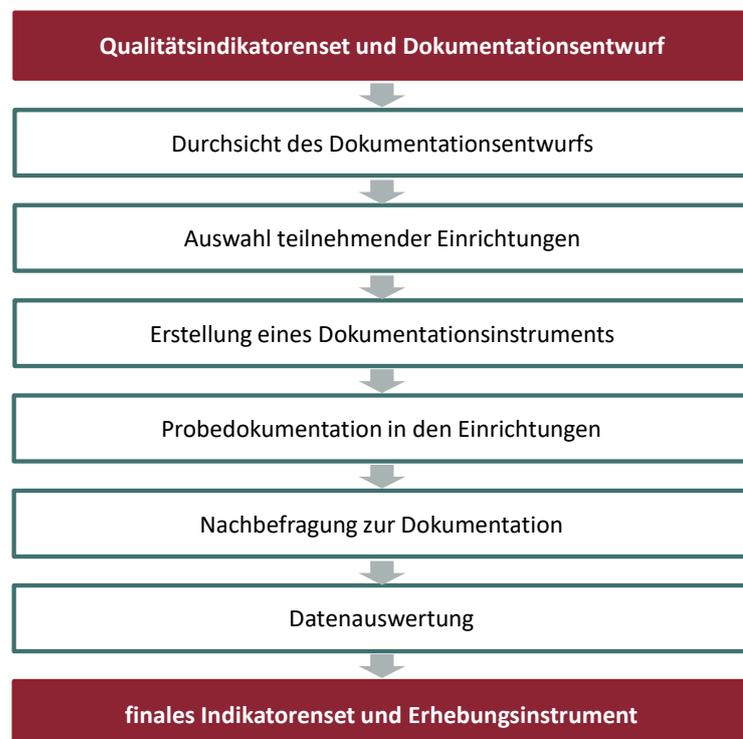


Abbildung 7: Ablauf einer Machbarkeitsprüfung

### **Durchsicht des Dokumentationsentwurfs**

Für jedes Datenfeld wird festgehalten, welcher Klärungsbedarf im Rahmen der Nachbefragung (s. u.) beim Anwender der Dokumentation, d. h. bei den Leistungserbringern, besteht. Der entsprechende Klärungsbedarf kann geringer sein, 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.

### **Auswahl teilnehmender Einrichtungen**

Für die Durchführung einer Machbarkeitsprüfung ist das IQTIG auf die Unterstützung freiwillig teilnehmender ambulanter und stationärer Leistungserbringer angewiesen. In der Regel wird eine Teilnehmerzahl von 20 bis 30 Einrichtungen angestrebt. Im Rahmen des öffentlichen Rekrutierungsaufrufs auf der Website des IQTIG werden auch die Trägerorganisationen des G-BA sowie 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. Um dennoch eine möglichst hohe Aussagekraft der Ergebnisse der Machbarkeitsprüfung für den späteren Regelbetrieb zu erreichen, wird eine gezielte Auswahl von Leistungserbringern und Behandlungsfällen im Sinne eines *purposive sampling* (Creswell und Plano Clark 2018, Palinkas et al. 2015, Shadish et al. 2002) vorgenommen.

Für die Auswahl von Einrichtungen, die an der Machbarkeitsprüfung teilnehmen, 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. In jeder teilnehmenden Einrichtung sollen ca. 15 bis 25 Falldokumentationen vorgenommen werden. Zusätzlich bittet das IQTIG 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.

### **Erstellung des Dokumentationsinstruments und Probedokumentation in den Einrichtungen**

Die Datenerfassung erfolgt mithilfe von durch das IQTIG bereitgestellten elektronischen Dokumenten oder einer Webapplikation. Da eine Verknüpfung des vom IQTIG für Machbarkeitsprüfungen bereitgestellten Erfassungsinstruments mit Klinik- oder Praxisinformationssystemen und damit eine automatische Auswahl der Behandlungsfälle derzeit nicht möglich ist, erhalten die teilnehmenden Einrichtungen die ICD- und OPS-Listen des QS-Filters in Form von (elektronischen) Dokumenten und nehmen ihre Fallauswahl manuell vor.

Wird ein Behandlungsfall dokumentiert, der gemäß den Kodelisten des QS-Filters nicht zur Grundgesamtheit des neuen QS-Verfahrens gehört, gibt das vom IQTIG bereitgestellte Erfassungsinstrument einen Hinweis aus; 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 der Leistungserbringer werden nach Abschluss der Probedokumentation in den Einrichtungen mit möglichst mehreren Personen durchgeführt, die an der Behandlung der Patientinnen und Patienten und der Dokumentation ihrer Behandlungsdaten beteiligt sind, wie Ärztinnen und Ärzten, Study Nurses, Dokumentarinnen und Dokumentaren, Qualitätsbeauftragten und Medizincontrollerinnen und -controllern. Die Befragten werden gebeten zu berichten, was ihnen bei der Dokumentation positiv und was ihnen negativ aufgefallen ist.

Der überwiegende Teil des Interviews bezieht sich auf Fragen zu den Datenfeldern der Dokumentation. Dabei werden – abhängig vom Klärungsbedarf – nicht zu jedem Datenfeld alle Fragen gestellt. Zu jedem Datenfeld kommen die folgenden Fragen in Betracht:

- 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?

Für Datenfelder mit Antwortalternativen können zusätzlich die folgenden Nachfragen gestellt werden:

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

Damit der zeitliche Aufwand für die Teilnehmer an der Nachbefragung zumutbar bleibt, werden in der Nachbefragung vorrangig diejenigen Datenfelder und Themen behandelt, für die erkennbarer Klärungsbedarf besteht oder die bereits bei der Durchsicht der Dokumentation vorab als klärungsbedürftig eingestuft wurden. Dagegen können sich Nachfragen insbesondere für Datenfelder erübrigen, die eindeutig definierte Inhalte erfassen oder die aus bereits existierenden Dokumentationen bekannt sind. Beispiele für Datenfelder dieser Art sind Geschlecht, Alter oder Körpergröße der Patientinnen und Patienten.

Zusätzlich zur 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 anhand des QS-Filters 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 erwarten Sie aus dem QS-Verfahren für Ihre Einrichtung?

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.

### **6.5.3    Datenauswertung und Optimierung der Qualitätsindikatoren**

Die Daten aus der Probedokumentation 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. Es wird geprüft, ob die Qualitätsindikatoren entsprechend den bei der vorläufigen Operationalisierung entwickelten Rechenregeln und dem Dokumentationsentwurf über alle Fälle ihrer jeweiligen Grundgesamtheit berechnet werden können. Für den Fall, dass Berechnungen für einen Indikator nicht möglich sind, wird analysiert, welche Ursachen eine

erfolgreiche Berechnung verhindern. Es kann z. B. vorkommen, dass wegen eines hohen Anteils fehlender Werte in erforderlichen Datenfeldern ein erheblicher Teil der dokumentierten Behandlungsfälle aus der Berechnung ausgeschlossen werden muss. Auf Basis dieser Auswertungen und der Ergebnisse der Nachbefragung werden der QS-Filter, der Dokumentationsbogen und die Rechenregeln der Qualitätsindikatoren optimiert.

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

- Berechnung und grafische Darstellung der Kennwerte der Verteilung der einrichtungsbezogenen Indikatorergebnisse
- probeweise Anwendung der Referenzbereiche

### **Rückmeldeberichte**

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.

## 7 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. Die folgenden Ausführungen fokussieren 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, 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.

Das IQTIG legt in seiner Definition von Qualität (vgl. Kapitel 1) und bei der Bewertung der Versorgungsqualität den Schwerpunkt auf eine patientenzentrierte Ausrichtung (vgl. Kapitel 3). Dies schlägt sich unmittelbar in den Entwicklungsmethoden für Patientenbefragungen nieder. Befragungsergebnisse von Patientinnen und Patienten sollen nicht bzgl. bestimmter Situationen und Zustände Auskunft geben, die ausschließlich durch Dritte als qualitätsrelevant definiert wurden. Vielmehr sollen sich die Inhalte von Patientenbefragungen explizit nach den Bedürfnissen der Patientinnen und Patienten und nach der Bedeutung, die sie für die Patientinnen und Patienten haben, ausrichten. Die entsprechenden Methoden, die dies in der Entwicklung sicherstellen sollen, werden im Folgenden dargestellt.

## 7.1 Faktenorientierte Befragungsansätze

Bei Patientenbefragungen, die zur Bewertung medizinischer Behandlungs- und Versorgungsaspekte eingesetzt werden, können zwei grundlegende Ansätze unterschieden werden: Der Reporting-Ansatz, zu dem in der Regel *Patient-Reported Experience Measures* (PREMs) und *Patient-Reported Outcome Measures* (PROMs) gezählt werden, sowie der Rating-Ansatz, dem Instrumente zur Erfassung der Patientenzufriedenheit üblicherweise folgen. Beim Rating-Ansatz werden die Befragten gebeten, Aussagen zu bewerten, etwa indem sie die Stärke ihrer Zustimmung oder Ablehnung zu einer Aussage mittels vorgegebener Antwortmöglichkeiten ausdrücken. Beim Reporting-Ansatz werden die Befragten hingegen aufgefordert, die Häufigkeit, die Intensität bzw. das Eintreten/Nichteintreten bestimmter Ereignisse zu berichten (z. B. Lloyd et al. 2014, Fullam und VanGeest 2014, Cleary 1999, Gerteis et al. 1993).

Zwar wird Patientenzufriedenheit oft als Indikator für Behandlungsqualität herangezogen, wobei das Ausmaß der Zufriedenheit häufig als Maß für die Übereinstimmung von vorab getroffenen Erwartungen und den tatsächlichen Erfahrungen verstanden wird (Pascoe 1983, Lingenfelder und Schneider 1990, Jacob und Bengel 2000). Die Annahme, dass Zufriedenheit ein Maß für Qualität ist, wird jedoch vielfach infrage gestellt (Williams 1994, Linder-Pelz 1982). Denn trotz anderweitig erfasster Qualitätsunterschiede, z. B. in den Behandlungsergebnissen oder Patienteninterviews, zeigen Zufriedenheitsbefragungen oft keine wesentlichen 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, welche Einflussfaktoren eine Rolle spielen und ob es tatsächlich zu einer Gegenüberstellung von Erwartungen und Erfahrungen kommt, wurde bisher nicht vollständig geklärt. Die Ergebnisse von Zufriedenheitsbefragungen sind von einer hohen Subjektivität geprägt, die einen großen Interpretationsspielraum lassen, aber wenig konkrete Handlungsansätze für Verbesserungsmaßnahmen bieten.

Im Rahmen der externen Qualitätssicherung müssen Qualitätsindikatoren auf Basis von Patientenbefragung qualitätsrelevante Merkmale adressieren, die konkrete Hinweise auf Verbesserungspotenzial geben und von Patientinnen und Patienten weitgehend objektiv beantwortet werden können. Das heißt, der individuelle Interpretationsspielraum der Patientinnen und Patienten wird so gering wie möglich gehalten und es müssen wesentliche Merkmale der Versorgungsqualität im Fokus stehen. Daher sind Patientenzufriedenheitsbefragungen für die externe Qualitätssicherung ungeeignet, da zu starke Limitationen hinsichtlich der Objektivität der Messung und der Interpretation der Ergebnisse bestehen.

Deshalb legt das IQTIG bei der Entwicklung von Patientenbefragungen faktenorientierte Ansätze zugrunde, wie sie im Rahmen der PROMs und PREMs verfolgt werden. „Faktenorientiert“ bedeutet in diesem Zusammenhang, dass individuelle Werturteile von Patientinnen und Patienten durch die Verwendung des Reporting-Ansatzes möglichst vermieden werden.

### 7.1.1 Patient-Reported Outcome Measures (PROMs)

Patient-Reported Outcome Measures (PROMs) können als Überbegriff für Instrumente zur Erfassung verschiedener Aspekte des patientenberichteten Gesundheitszustandes verstanden

werden, bei denen keine Interpretation der berichteten Inhalte durch andere Personen (z. B. durch Behandlerinnen/Behandler) erfolgt (CDER et al. 2006, Lützner et al. 2017). PROMs zeigen einen starken Zusammenhang mit klinischen Parametern der Ergebnisqualität (Tang et al. 2018) und werden explizit zur Erfassung von solchen Endpunkten empfohlen, die am besten von Patientinnen und Patienten selbst beurteilt werden können (CDER et al. 2006).

Grundsätzlich können vier übergeordnete, sich gegenseitig beeinflussende bzw. inhaltlich überschneidende Dimensionen des Gesundheitszustands unterschieden werden, die typischerweise anhand von PROMs erfasst werden, nämlich *Symptomstatus* (somatisch, psychisch oder kognitiv), *funktionaler Status* (z. B. körperliche Funktionsfähigkeit hinsichtlich der Ausübung von Alltagsaktivitäten), *allgemeine Gesundheitswahrnehmung* und *gesundheitsbezogene Lebensqualität* (Valderas und Alonso 2008). PROMs selbst können wiederum hinsichtlich ihrer Zielgruppenspezifität unterschieden ausgerichtet sein (Lützner et al. 2017, Obbarius et al. 2018). Für die Patientenbefragung des IQTIG wird entsprechend der Ausrichtung des QS-Verfahrens nach verfahrensspezifischen Indikatoren in der Regel ein zielgruppenspezifischer Ansatz verfolgt, sodass relevante Outcomes für eine spezifische Subpopulation (z. B. nach Krankheit, Behandlungsform, Alter) erfasst werden. Welche Dimensionen des Gesundheitszustands adressiert werden, richtet sich nach der thematischen Ausrichtung des jeweiligen QS-Verfahrens bzw. der Patientenbefragung, die im Rahmen der Entwicklung mit der Definition der Qualitätsaspekte und Qualitätsmerkmale festgelegt wird.

### 7.1.2 Patient-Reported Experience Measures (PREMs)

Als Patient-Reported Experience Measures (PREMs) werden Instrumente bezeichnet, die Erfahrungen der Patientinnen und Patienten erfassen, die sie während ihrer medizinischen Versorgung und Behandlung gemacht haben bzw. regelmäßig machen (Kingsley und Patel 2017, Lützner et al. 2017).

Eine Form von PREMs, die einen faktenorientierten Ansatz verfolgt, ist die ereignisorientierte Patientenbefragung (Cleary 1999, Klein 2004, Gerteis et al. 1993). Hier stehen konkrete Erfahrungen oder Beobachtungen bzw. Ereignisse im Vordergrund, über deren Eintreten oder Nicht-eintreten die Patientinnen und Patienten berichten. Die Ereignisse sollen Schlüsselsituationen in der Versorgung aus Sicht der Patientinnen und Patienten darstellen, die 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.

Behandlungserfahrungen sind wichtige Grundlagen zur Bewertung qualitätsrelevanter Prozesse und Strukturen, die jenseits und unabhängig der Behandlungsergebnisse patientenrelevante Aspekte der Versorgungsqualität adressieren (Thompson et al. 2016, Rhee et al. 2018).

Typische Beispiele für Themen, die anhand von PREMs im Rahmen der Qualitätssicherung adressiert werden können, sind Patientenaufklärung, die Interaktion zwischen Ärztin/Arzt und Patientin/Patient, Hygienebedingungen, Wartezeiten oder die Einbindung bei Behandlungsentscheidungen (Breckenridge et al. 2015, Lützner et al. 2017).

Zusammenfassend zielen PROMs und PREMs 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 sowie Erfahrungen und deren individuelle Bewertungen voneinander abgekoppelt. Damit wird die Interpretation der Ergebnisse eindeutiger, wodurch konkrete Hinweise für mögliche Handlungsansätze gegeben werden können. Durch den kombinierten Einsatz von PREMs und PROMs können die unterschiedlichen Qualitätsdimensionen, die sich auf unabhängige Sachverhalte beziehen, adäquat abgebildet werden (Black et al. 2014).

## 7.2 Ablauf der Fragebogen- und Qualitätsindikatorenentwicklung

Die Fragebogenentwicklung beginnt mit der Entwicklung eines Qualitätsmodells (siehe Kapitel 11) für den jeweiligen Themenbereich. Dieses stellt eine Strukturierung der Inhalte eines Themenbereichs dar und beschreibt sie anhand von sogenannten Qualitätsaspekten. Damit bestimmen die Qualitätsaspekte des Qualitätsmodells die inhaltlichen Schwerpunkte der Befragung.

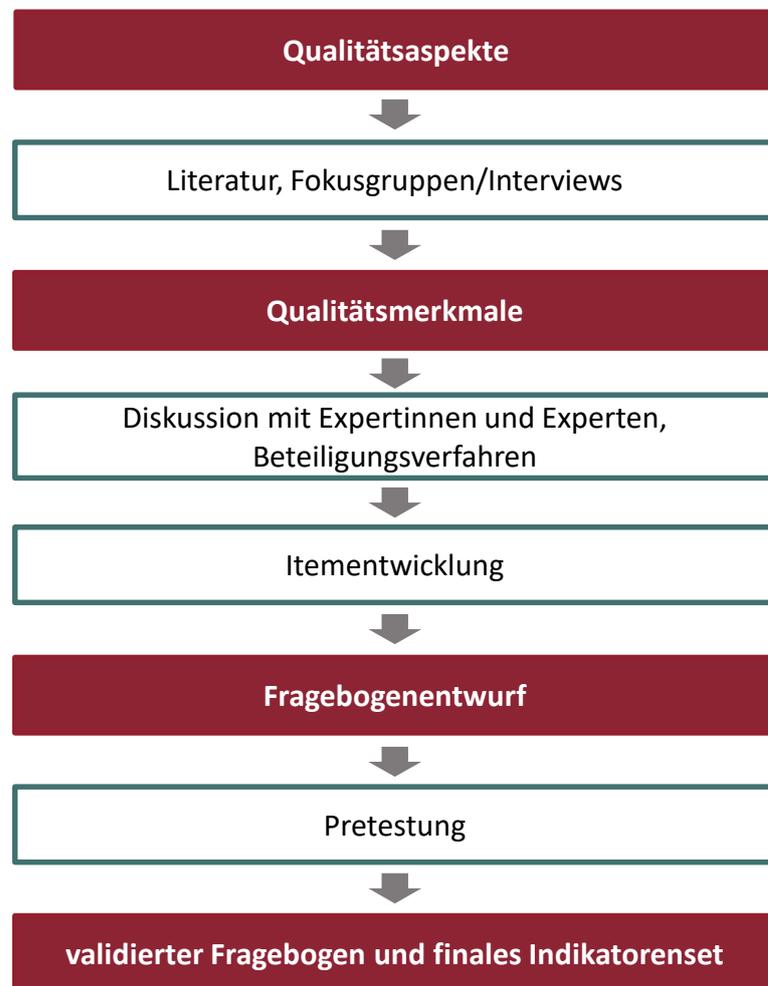


Abbildung 8: Entwicklungsschritte einer Patientenbefragung

Bei der in Abbildung 8 schematisch dargestellten Fragebogenentwicklung werden dazu in einer ersten, explorativen Phase die abzubildenden Inhalte ermittelt und anhand von Qualitätsmerkmalen konkretisiert. Dazu führt das IQTIG eine systematische Literaturrecherche und Fokusgruppen bzw. leitfadengestützte Interviews durch. Die auf dieser Basis entwickelten Qualitätsmerkmale werden mit einem Expertengremium diskutiert und ggf. überarbeitet. Da die Qualitätsmerkmale die Basis für die Fragebogenentwicklung bilden, markiert dieser Entwicklungsschritt einen wichtigen Meilenstein. Deshalb ist mit Abgabe des Zwischenberichts, in dem alle bis zu diesem Zeitpunkt abgeschlossenen Entwicklungen dokumentiert sind, ein Beteiligungsverfahren vorgesehen (siehe Kapitel 12). Zu diesem Zeitpunkt können Anregungen und Hinweise aus dem Beteiligungsverfahren in die weiteren Entwicklungen einfließen, d. h., bevor die Fragebogenentwicklung und -testung beginnt. 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 Beteiligungsverfahren beginnt die Entwicklung der Items zur Erfassung der Qualitätsmerkmale, die in einem ersten Fragebogenentwurf mündet. Dieser Fragebogenentwurf wird in der folgenden Phase der Fragebogenoptimierung zur Sicherstellung der Verständlichkeit und Interpretation durch die Patientinnen und Patienten zunächst einem kognitiven Pretest unterzogen und auf der Basis der Ergebnisse überarbeitet. In einem anschließenden Standard-Pretest wird der Fragebogen hinsichtlich seiner messtheoretischen Eigenschaften getestet. Wenn die Ergebnisse des Standard-Pretests erneuten Anpassungsbedarf von Fragebogenitems aufzeigen, der über rein redaktionelle Veränderungen hinausgeht, wird der Fragebogen einem nochmaligen kognitiven Pretest unterzogen (Retest), der gezielt auf die Modifikationen ausgelegt ist, um die Verständlichkeit und Interpretierbarkeit der veränderten Fragebogenitems sicherzustellen.

Abschließend werden auf Grundlage des so validierten Fragebogens ein Aggregationsmodell für die Fragebogenitems und Qualitätsindikatoren definiert. Diese Entwicklungsschritte werden in den folgenden Abschnitten jeweils genauer erläutert. Darüber hinaus werden u. a. auch Fragebogenlogistik und Datenflüsse entwickelt, da sich vor allem Befragungszeiträume und Auslösekriterien unterscheiden können.

### **7.2.1 Literaturrecherche und -analyse**

Die Literaturrecherche und -analyse (siehe Kapitel 9) ist eine Wissensquelle zur Identifizierung, Analyse und Beschreibung der für die Patientenbefragung relevanten Qualitätsmerkmale. Außerdem dienen die Analyseergebnisse der Generierung von Themen für die Diskussionsleitfäden der Fokusgruppen bzw. die Interviewleitfäden und geben auch Anregungen und Hinweise für die spätere Entwicklung der Fragebogenitems. Anhand der Literaturanalyse wird sichtbar, mit welcher Deutlichkeit die recherchierten Quellen Hinweise auf konkrete patientenrelevante Qualitätsmerkmale geben.

Die recherchierte Literatur wird sowohl hinsichtlich der eingangs formulierten Fragestellungen analysiert als auch hinsichtlich der Eignungskriterien für Qualitätsmessungen (siehe Kapitel 13), die im gesamten Entwicklungsprozess zum Tragen kommen (siehe Abschnitt 7.2.9). 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. Vielmehr ist die Ausrichtung der Behandlungsprozesse an den Bedürfnissen der Patientinnen und Patienten als gesellschaftlicher Wert an sich anerkannt (siehe Kapitel 3). Die Auswirkungen der Patientenzentrierung – beispielweise auf die Therapietreue, Lebensqualität oder Ängste – sind Gegenstand verschiedener Untersuchungsbereiche (z. B. Härter et al. 2005, Scheibler et al. 2003, Haynes et al. 1996).

Folglich leitet bei der Entwicklung einer Patientenbefragung vor allem das Eignungskriterium „Bedeutung für die Patientinnen und Patienten“ die Literaturrecherche (siehe Kapitel 13). 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 potenziell beurteilbar sind, also erlebt wurden, und inwiefern über sie im Rahmen einer Befragung berichtet werden kann.

### 7.2.2 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 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, Tausch und Menold 2015). Das Ziel der Fokusgruppen liegt dementsprechend in der Vertiefung und Konkretisierung der Qualitätsaspekte, die über eine Patientenbefragung adressiert werden sollen, so dass 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 qualitätsrelevanten Themen 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.

Ziel der Fokusgruppen mit Vertreterinnen und Vertretern der Gesundheitsprofessionen ist, diese Merkmale aus ihrem Erfahrungshorizont heraus zu ergänzen.

In der Regel bestehen die Fokusgruppen jeweils aus fünf bis zehn Teilnehmerinnen und Teilnehmern. Die Zusammensetzung der Fokusgruppen erfolgt über eine gezielte Auswahl charakteristischer Repräsentanten der jeweiligen Zielgruppe, um deren wesentliche Meinungsbilder zur jeweiligen Fragestellung zu gewinnen (Nyumba et al. 2018, Scheuren 2004: 34, Krueger und Casey 2015). Die Moderation der Fokusgruppen erfolgt anhand eines sogenannten Topic Guides, in dem konkrete Themen aufgeführt sind, die den Diskussionsprozess leiten. Von allen Fokusgruppensitzungen werden Transkripte erstellt, die gruppenübergreifend und in Anlehnung an die inhaltlich-strukturierende Inhaltsanalyse ausgewertet werden (Kuckartz 2014, Mayring 2015).

Die Anzahl der durchzuführenden Fokusgruppen wird vom IQTIG themenspezifisch festgelegt. Dabei werden einerseits die für den benötigten Erkenntnisgewinn voraussichtlich erforderliche Gruppenszahl (Krueger und Casey 2015, Kühn und Koschel 2011) und andererseits Praktikabilität und Aufwand berücksichtigt. Im Falle von Rekrutierungsschwierigkeiten oder wenn die Personen physisch oder psychisch nicht in der Lage sind, an Gruppendiskussionen teilzunehmen, werden alternativ leitfadengestützte Interviews mit den gleichen Zielgruppen durchgeführt.

### 7.2.3 Entwicklung der Qualitätsmerkmale

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

Bevor die Qualitätsmerkmale anhand von Fragebogenitems operationalisiert werden, wird überprüft, ob sie die folgenden Eignungskriterien (siehe Kapitel 13) erfüllen:

- Bedeutung für die Patientinnen und Patienten
- Zusammenhang mit einem unmittelbar patientenrelevanten Merkmal
- Potenzial zur Verbesserung
- Beeinflussbarkeit durch den Leistungserbringer

Zudem ist es im Kontext der Patientenbefragung von Bedeutung, ob die Patientinnen und Patienten die Inhalte der Qualitätsmerkmale beurteilen und die darauf aufbauenden Fragebogenitems auch beantworten können, d. h., ob die Sachverhalte und Situationen für die Patientinnen und Patienten erlebbar und beobachtbar sind. Insofern ist die „Beurteilbarkeit durch Patientinnen und Patienten“ ein weiteres Kriterium, das für die Entwicklung von Patientenbefragungen herangezogen wird. Weiterhin geben Spezifika der Beauftragung Anforderungen vor, die ebenfalls zu diesem Zeitpunkt geprüft werden (z. B. Diagnoseunabhängigkeit der Qualitätsmerkmale).

#### 7.2.4 Expertenberatung zu den Qualitätsmerkmalen

Die entwickelten Qualitätsmerkmale werden hinsichtlich der Erfüllung der o. g. Eignungskriterien mit externen Fachexpertinnen und -experten diskutiert. Zu diesem Zweck wird ein Expertengremium einberufen. Um ein möglichst breites Spektrum an Fachexpertise zu erhalten und die Vielfalt der Einschätzungen und Erkenntnisse abzubilden (Niederberger und Wassermann 2015, Krick 2013), 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 – in Abhängigkeit von der Zielgruppe und des Versorgungskontextes des jeweils avisierten QS-Verfahrens – nicht mehr als 25 Expertinnen und Experten vertreten sein. Dabei wird eine Ausgewogenheit vor allem der gesundheitsprofessionsspezifischen Fachexpertise und der Fachexpertise der Patientinnen und Patienten sichergestellt (siehe Abschnitt 10.5).

Die Expertenberatung besteht aus zwei Phasen. In der ersten Phase werden die erarbeiteten Qualitätsmerkmale mit Zuordnung zum Qualitätsaspekt und Darstellung ihrer Ableitung (Fokusgruppen, Literatur) den Expertinnen und Experten im Rahmen einer webbasierten Befragung vorgestellt. Es erfolgt eine Einschätzung eines jeden Qualitätsmerkmals anhand der o. g. Eignungskriterien. Die Ergebnisse dieser 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. Auf Basis dieser Einschätzungen und Diskussion wird in der Zusammenschau aller Erkenntnisse seitens des IQTIG entschieden, ob und wie die Qualitätsmerkmale in Form von Fragebogenitems in den Fragebogen übernommen werden sollen.

#### 7.2.5 Itementwicklung

Grundlage der Itementwicklung bilden die Qualitätsmerkmale. Bei den Qualitätsmerkmalen handelt es sich in der Regel um latente Konstrukte, die nicht direkt erfasst werden können. Vielmehr wird die Ausprägung des Konstrukts auf Basis gemessener (manifeste) Variablen abgeleitet. Damit die Ausprägung der Qualitätsmerkmale von den Patientinnen und Patienten beurteilt werden kann, werden sie daher mithilfe von Fragen und Antwortkategorien operationalisiert, die als Items in einem Fragebogen zusammengefasst werden. Mit dem Begriff „Item“ ist die Fragestellung gemeint, zu der auch immer eine Antwortskala gehört.<sup>15</sup>

---

<sup>15</sup> Eine Ausnahme stellen offene Fragen dar, die im vorliegenden Kontext allerdings eine untergeordnete Rolle spielen.

Die Patientenbefragung verfolgt im Schwerpunkt einen faktenorientierten Ansatz. Die Qualitätsmerkmale werden mithilfe von Items so operationalisiert, dass sie die qualitätsrelevanten Ereignisse oder Situationen des jeweiligen Qualitätsmerkmals abbilden. Dabei adressiert jedes Qualitätsmerkmal für sich genommen einen qualitätsrelevanten Sachverhalt, der durch Fragebogenitems messbar gemacht wird. Je nachdem, wie komplex ein Qualitätsmerkmal ist, werden ein oder auch mehrere Items benötigt, um den dahinterliegenden Sachverhalt verständlich und beantwortbar in die Befragung zu „übersetzen“.

Neben grundsätzlichen Prinzipien der Itemformulierung (z. B. Überschneidungsfreiheit der Antwortkategorien, keine doppelten Stimuli, eindeutige Begrifflichkeiten)<sup>16</sup> 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 Abschnitt 7.2.6).

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).

Neben den Items zu den Qualitätsmerkmalen wird ein Fragebogen immer auswertungsrelevante Zusatzfragen beinhalten, wie z. B. für Zwecke der Risikoadjustierung oder Stratifizierung, wie soziodemografische Angaben oder Angaben zum Gesundheitszustand oder dem Gesundheitsverhalten, Items zur Stichprobenbeschreibung.

### 7.2.6 Pretestung

Um Hinweise zur Funktionsfähigkeit oder zu eventuellen Mängeln von Items oder des Fragebogens insgesamt zu erhalten, durchläuft jeder Fragebogenentwurf eine Testung. Pretests gelten als unabdingbarer Bestandteil jeder Fragebogenentwicklung und als eine Voraussetzung für

---

<sup>16</sup> Eine gute Übersicht von allgemeinen Regeln der Itemformulierung findet sich u. a. bei Porst (2014).

hohe Datenqualität (Lenzner et al. 2015, Prüfer und Rexroth 1996, Bradburn et al. 2004, Porst 2014). Das IQTIG verfolgt mit der Pretestung folgende Ziele (Porst 2014):

- 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

Dazu bedient sich das IQTIG einem zweistufigen Testverfahren bestehend aus einem kognitiven Pretest und einem Standard-Pretest<sup>17</sup>.

### **Kognitiver Pretest**

Kognitive Pretests geben Einblicke in die kognitiven Prozesse, die bei den Befragten während der Beantwortung der Fragen ablaufen. So werden die Verständlichkeit und Interpretation der Fragen, aber auch die Erinnerbarkeit der erfragten Erlebnisse 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).

Das IQTIG setzt in den Entwicklungen zur Patientenbefragung u. a. die folgenden, häufig verwendeten Techniken ein:

- 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
- 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

---

<sup>17</sup> In der Literatur werden für „Standard-Pretest“ häufig auch die Bezeichnungen „Standardbeobachtungspretest“ oder „klassischer Pretest“ verwendet.

Befragten sehr konkret Aussagen zum Verständnis einer Frage oder einzelner Begriffe machen. 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 Untersuchungsgegenstand pro Fragebogen zwischen 10 und 30 Patientinnen und Patienten. Das IQTIG verfolgt bei der kognitiven Pretestung in der Regel zwei verschiedene Testansätze, die in Kombination bei jeder Fragebogenentwicklung eingesetzt werden.

Im ersten Testansatz werden die Testpersonen gebeten, den Fragebogen zunächst vollständig und ungestört auszufüllen. Erst nach Beantwortung des gesamten Fragebogens wird um eine Gesamteinschätzung des Fragebogens gebeten sowie vertiefende Nachfragen zu einzelnen Items gestellt. Ziel dieses Testansatzes ist, einen Überblick über generelle Schwierigkeiten bei der Beantwortung des betreffenden Fragebogens zu erhalten. Die Nachfragen beziehen sich folglich auf Probleme, die bei der Fragebogenbearbeitung beobachtet oder nach dem Ausfüllen von den Testpersonen angesprochen wurden. Zugleich wird die Zeit erfasst, die die einzelnen Testpersonen zum Ausfüllen des Fragebogens benötigt haben.

Beim zweiten Testansatz werden einzelne Items gezielt geprüft (Lenzner et al. 2015). Hierfür wird vor Durchführung des Pretests 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. Nach der Beantwortung der jeweiligen Items werden der Testperson unter Anwendung der Pretest-Techniken Fragen zu den betreffenden Items gestellt.

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.

Alle Interviews werden mit Einverständnis der Testpersonen aufgezeichnet und im Anschluss transkribiert und systematisch 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 nach dem kognitiven Pretest ein Standard-Pretest durchgeführt. Im Rahmen des Standard-Pretests wird der Fragebogen unter Bedingungen eingesetzt, die dem späteren Regelbetrieb des QS-Verfahrens so ähnlich wie möglich sind. 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 durch Leistungserbringer, die freiwillig am Standard-Pretest teilnehmen.

Auf Basis der im Standard-Pretest gewonnenen Daten findet die Validierung des Fragebogens statt. Hierzu werden Verfahren zur Evaluation der Fragenqualität, wie die Analyse und Darstellung der Antwortverteilungen, und Verfahren zur Überprüfung der Gütekriterien der Messung angewandt. Sollten die Ergebnisse des Standard-Pretests auf Probleme von Itemformulierungen hinweisen, werden diese nochmals angepasst. Diese Items werden einem erneuten kognitiven Pretest unterzogen, bei dem der zweite Testansatz gewählt wird (kognitiver Retest). Zu diesem Entwicklungszeitpunkt können ausschließlich Itemanpassungen erfolgen, die nicht zu inhaltlichen Veränderungen führen, da andernfalls ein erneuter Standard-Pretest durchgeführt werden müsste, um die messtheoretischen Eigenschaften für die Indikatorkonzeption zu prüfen.

Da meist mehrere Items durch Aggregation in einem Qualitätsindikator zusammengefasst werden (siehe Abschnitt 7.2.8), wird mittels der im Standard-Pretest gewonnenen Daten untersucht, inwieweit die zugrunde liegenden Items miteinander in Beziehung stehen und mit dem Wert des jeweiligen Qualitätsindikators zusammenhängen. Positive Interkorrelationen der Items eines Merkmals sowie positive Korrelationen der Items mit dem Gesamtwert des Indikators erleichtern die Interpretation eines Qualitätsindikators. Je stärker die einzelnen Items untereinander und mit der Gesamtskala assoziiert sind, desto eindeutiger ist das Indikatorergebnis zu interpretieren. Sehr niedrige oder negative Interkorrelationen der Items können dazu führen, dass hohe Werte auf einem Item niedrige Werte auf einem anderen Item ausgleichen können, und reduzieren damit die Variabilität des Indikatorwerts. Insbesondere negative Korrelationen zwischen Items, die in einem Qualitätsindikator zusammengefasst werden, würden eine sinnvolle Interpretation der Indikatorwerte erschweren. Um entsprechende Hinweise hinsichtlich Richtung und Höhe des Zusammenhangs zwischen den Items eines Qualitätsindikators zu generieren, werden Kennzahlen der Homogenität und der internen Konsistenz der Items herangezogen (Goeman und De Jong 2018).

Deuten die verwendeten Kennzahlen auf sehr niedrige bzw. negative Interkorrelationen der Items eines Indikators hin, wird in einem folgenden Schritt überprüft, inwieweit eine Modifikation des Indikators (z. B. Aufteilung der zusammengefassten Qualitätsmerkmale auf mehrere Qualitätsindikatoren) auf Basis inhaltlicher Überlegungen notwendig bzw. sinnvoll ist.

### 7.2.7 Stichprobenziehung und Fallzahlenberechnung

Gemäß § 299 Abs. 1 SGB V ist vorgesehen, dass die Datenerhebung im Rahmen von QS-Verfahren in der Regel auf einer Stichprobe der betroffenen Patientinnen und Patienten basiert. Im Rahmen von Patientenbefragungen wird projektspezifisch anhand einer klassischen Poweranalyse ermittelt, wie viele behandelte Patientinnen und Patienten befragt werden müssen, um statistisch aussagekräftige Indikatorergebnisse für die Leistungserbringer zu erhalten. Dazu wird auf Basis einer Binomialannahme berechnet, ab welcher Fallzahl bei vorgegebenem Referenzwert und zugrunde liegendem Indikatorwert eine bestimmte Teststärke in Abhängigkeit von der Anzahl der Merkmale, Items und Antwortkategorien des Qualitätsindikators erreicht werden kann. Anhand dieser Berechnungen werden Empfehlungen ausgesprochen, ab welchen Fallzahlen eine Stichprobenziehung erfolgen kann und wie groß diese Stichprobe sein sollte und wann Vollerhebungen durchgeführt werden sollten. Die Art der Stichprobenziehung (z. B. einfache Zufallsstichprobe oder geschichtete Zufallsstichprobe) wird projektspezifisch methodisch abgewogen.

Bei der Bewertung von Qualitätsindikatoren auf Basis von Patientenbefragungen wird anhand von Vertrauensbereichen um den gemessenen Indikatorwert beurteilt, inwiefern der zugrunde liegende Indikatorwert den Referenzbereich einhält. Sind nur wenige patientenseitige Angaben zur Qualität verfügbar, bildet sich das in entsprechend breiten Vertrauensbereichen ab. Mit steigender Fallzahl wächst die Wahrscheinlichkeit, dass ein Leistungserbringer mit einem zugrunde liegenden Indikatorwert, der außerhalb des Referenzbereichs liegt, auch anhand der vorliegenden Daten als auffällig eingestuft wird. Prinzipiell gilt jedoch, dass auch bei kleinen Fallzahlen Aussagen über das Erreichen des jeweiligen Referenzwertes anhand des Vertrauensintervalls eindeutig interpretiert werden können. Folglich werden keine Mindestfallzahlen für die Durchführung einer Patientenbefragung definiert, da sich die Unsicherheit durch geringe Fallzahlen in den Indikatorergebnissen und den Auffälligkeitseinstufungen niederschlägt.

### 7.2.8 Konstruktion der Qualitätsindikatoren

Die Konstruktion der Qualitätsindikatoren basiert auf der Aggregation mehrerer Items zu einem Qualitätsindikator. Hierfür wird ein zweistufiges Vorgehen gewählt, das die geschachtelte Struktur von Qualitätsaspekten (QA) und Qualitätsmerkmalen (QM) widerspiegelt (siehe Abbildung 9). Sofern ein Qualitätsmerkmal anhand mehrerer Items gemessen wird, soll dabei im ersten Schritt eine Aggregation der Items (I) auf Merkmalsebene erfolgen. Im zweiten Schritt – sofern sich ein Qualitätsindikator aus mehreren Merkmalen zusammensetzt – erfolgt eine Aggregation der Werte aus dem ersten Schritt zu einem Indexwert. Da alle Merkmale als gleichwertig angesehen werden und um eine intuitive Interpretation der Ergebnisse zu erleichtern, werden sie bei der Aggregation gleichgewichtet (Shwartz et al. 2015). Auf Ebene der Qualitätsmerkmale erfolgt die Aggregation der Items ebenfalls gleichgewichtet. Dieses Vorgehen führt dazu, dass Qualitätsmerkmale mit mehr Items das gleiche Gewicht bei der Aggregation in einen Index erhalten, wie Qualitätsmerkmale mit wenigen Items oder nur einem Item. Weiterhin müssen alle Qualitätsmerkmale in den Qualitätsindikatoren vertreten sein.

Abgeleitet aus den Qualitätsaspekten und Qualitätsmerkmalen wird für jeden Qualitätsindikator ein möglichst konkretes Qualitätsziel und ein Referenzbereich definiert (siehe Kapitel 16).

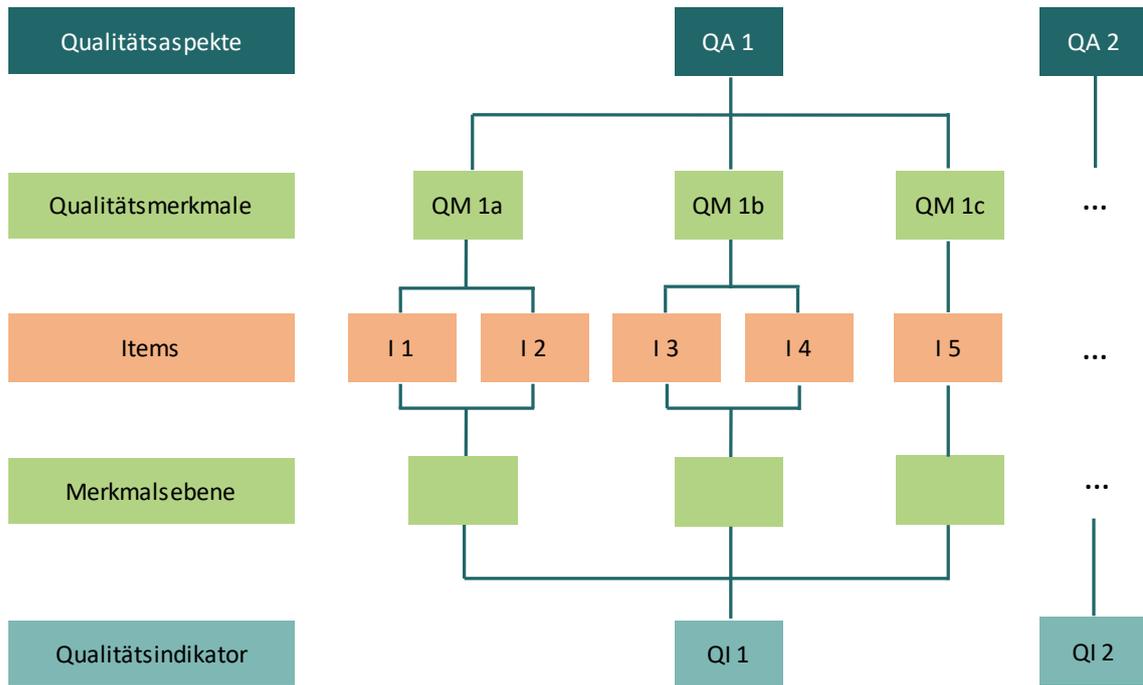


Abbildung 9: Hierarchische Struktur von Qualitätsaspekten, Qualitätsmerkmalen und Items und hierarchische Aggregation zu Qualitätsindikatoren

Die statistische Beziehung zwischen der Merkmalsebene und der hierarchisch übergeordneten Ebene des Qualitätsindikators entspricht einem formativen Messmodell (siehe Abschnitt 4.3), wobei für die konstituierenden Qualitätsmerkmale stets eine konzeptuelle Einheit angenommen wird. Konzeptuelle Einheit liegt bei formativen Messmodellen vor, wenn sich die konstituierenden Variablen auf jeweils ein gemeinsames Konzept beziehen, das anhand thematisch zusammenhängender Inhalte klar definiert ist (Bollen und Diamantopoulos 2017).

### 7.2.9 Umsetzung der Eignungskriterien

Die Entwicklung von Qualitätsindikatoren auf Basis von Patientenbefragungen orientiert sich an den Eignungskriterien für Qualitätsmessungen (siehe Kapitel 13). Diese werden als Leitfaden über den gesamten Entwicklungsprozess hinweg herangezogen und deren Erfüllung wird auf Basis unterschiedlicher Maßnahmen und Wissensbestände sichergestellt. In Tabelle 1 werden die Eignungskriterien für Qualitätsmessungen und die jeweiligen Maßnahmen und Wissensbestände, die für deren Sicherstellung herangezogen werden, gegenübergestellt.

Tabelle 1: Eignungskriterien für Qualitätsindikatoren und dazugehörige Maßnahmen und Informationsquellen der Patientenbefragung

Eignungskriterium	Maßnahmen und Wissensbestände
<b>Qualitätsziel des Indikators</b>	
Bedeutung für die Patientinnen und Patienten	Literaturrecherche insbesondere zur Patientenperspektive, Fokusgruppen/Interviews insbesondere mit Patientinnen und Patienten, Expertengremium
Zusammenhang mit einem unmittelbar patientenrelevanten Merkmal	Literaturrecherche insbesondere zur Patientenperspektive, Fokusgruppen/Interviews insbesondere mit Patientinnen und Patienten, Expertengremium
Potenzial zur Verbesserung	Literaturrecherche, Fokusgruppen/Interviews, Expertengremium
Brauchbarkeit für den Handlungsanschluss	Zusammenschau der entwickelten Indikatoren vor dem Hintergrund der Zielsetzung des betreffenden QS-Verfahrens
Beeinflussbarkeit durch den Leistungserbringer	Literaturrecherche, Expertengremium
<b>Messverfahren</b>	
Objektivität der Messung	faktenorientierter Befragungsansatz, kognitiver Pretest, Standard-Pretest
Datenqualität	kognitiver Pretest, Standard-Pretest
Reliabilität der Messung	Standard-Pretest
Validität der Messung	Literaturrecherche, Fokusgruppen/Interviews, Expertengremium, kognitiver Pretest, Standard-Pretest
Praktikabilität der Messung	kognitiver Pretest, Standard-Pretest
<b>Bewertungskonzept</b>	
Angemessenheit des Referenzbereichs	Literaturrecherche, Indikatorkonstruktion
Klassifikationsgüte	Indikatorkonstruktion und Auswertungsmethodik
Angemessenheit der Risikoadjustierung	Literaturrecherche, Daten des Standard-Pretests (Weiterentwicklung im Regelbetrieb)

Im Folgenden werden einige Kriterien erläutert, die bei der Messung mit Fragebogeninstrumenten besonders zu berücksichtigen sind.

### **Objektivität der Messung**

Als Objektivität wird das Maß der Unabhängigkeit eines Messergebnisses von den Untersuchungsumständen bezeichnet (Bühner 2011, Moosbrugger und Kelava 2012b). 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 (Versendestelle) 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.

### **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 7.2.6).

#### **7.2.10 Risikoadjustierung von Qualitätsindikatoren der Patientenbefragung**

Das grundsätzliche Vorgehen bei der Entwicklung von Risikoadjustierungsmodellen in der Qualitätssicherung wird in Abschnitt 20.2 dargestellt. Im Folgenden wird 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 innerhalb einer orientierenden Literaturrecherche grundsätzliche, patientenseitige Risikofaktoren im jeweiligen Themenfeld 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 jeweiligen Regelbetriebs.

## 8 Anpassung indikatorbasierter Instrumente der Qualitätssicherung

Das IQTIG übernimmt für die vom G-BA beschlossenen indikatorbasierten Verfahren der Qualitätssicherung<sup>18</sup> zwei grundsätzlich verschiedene Arten von Aufgaben: Zum einen ist das IQTIG an der richtlinienkonformen Durchführung der Qualitätssicherung beteiligt (§ 137a Abs. 3 Satz 2 Nr. 3 SGB V), zum anderen analysiert es die Ergebnisse der Qualitätsmessungen und entwickelt Empfehlungen, um die eingesetzten Instrumente der Qualitätssicherung dem Stand der Wissenschaften anzupassen.

Unter „Durchführung von QS-Verfahren“ werden die folgenden Aufgaben verstanden, die das IQTIG bei der Umsetzung der vom G-BA beschlossenen QS-Verfahren in dessen Auftrag übernimmt und deren Ausgestaltung sich nach den Richtlinien des G-BA richtet:

- Schaffung der Datengrundlage für die Qualitätsmessung (z. B. die Veröffentlichung von Spezifikationen für die Hersteller von QS-Software, deren Anpassung an technische Rahmenbedingungen oder an Änderungen in den zugrunde liegenden medizinischen Klassifikationen, die Annahme von Daten, Prüfungen der Datenqualität),
- Auswertung der Daten (z. B. Berechnung der Indikatorergebnisse anhand der vom G-BA beschlossenen Rechenregeln und Erstellung von Berichten),
- Bewertung der Indikatorergebnisse (z. B. anhand der vom G-BA beschlossenen Referenzbereiche; in einigen QS-Verfahren<sup>19</sup> darüber hinaus durch Stellungnahmeverfahren mit Leistungserbringern, welche die durch die Referenzbereiche definierten Anforderungen nicht erfüllen) sowie
- Ergebnisdarstellung und Rückmeldung an die Leistungserbringer (z. B. durch Rückmeldebereiche) sowie Durchführung von Maßnahmen zur Qualitätsverbesserung (z. B. Zielvereinbarungen mit Leistungserbringern<sup>20</sup> und Veröffentlichung leistungserbringerbezogener Ergebnisse)

Auch die Anpassung der QS-Verfahren an geänderte Vorgaben in den Richtlinien des G-BA gehört in der Regel zu den Durchführungsaufgaben des IQTIG.

Instrumente der Qualitätssicherung müssen allerdings auch wiederkehrend auf ihre Aktualität geprüft und an neue Erkenntnisse oder geänderte Rahmenbedingungen angepasst werden (CHP/PCOR/Battelle Memorial Institute 2011). Anpassungsbedarf entsteht beispielsweise durch technischen Fortschritt in der Medizin oder neue wissenschaftliche Erkenntnisse, z. B. durch die Überarbeitung von Leitlinien oder durch neue Forschungsergebnisse. Darüber hinaus können

---

<sup>18</sup> Zu den Verfahren der Qualitätssicherung gehören zum einen die Messinstrumente zur Darstellung der Versorgungsqualität und zum anderen die Steuerungsinstrumente (Handlungsanschlüsse), mit denen eine Verbesserung der Versorgungsqualität erreicht werden soll (siehe Abschnitt 2.4).

<sup>19</sup> Derzeit betrifft dies die sogenannten bundesbezogenen Verfahren nach DeQS-RL, für die der Unterausschuss Qualitätssicherung des G-BA das verantwortliche Lenkungsgremium ist, sowie das Verfahren nach plan. QI-RL.

<sup>20</sup> In den Verfahren nach DeQS-RL, für die der Unterausschuss Qualitätssicherung des G-BA das verantwortliche Lenkungsgremium ist.

sich im Rahmen der Durchführung der QS-Verfahren Hinweise ergeben, dass Anpassungen der eingesetzten Instrumente aus fachlich-medizinischer, methodischer oder statistischer Sicht notwendig sind. Dieses Kapitel stellt dar, wie das IQTIG den Anpassungsbedarf indikatorbasierter Instrumente der Qualitätssicherung prüft und Empfehlungen für solche Anpassungen für den G-BA entwickelt.

Hinweise auf Anpassungsbedarf beziehen sich hauptsächlich auf die im Rahmen der QS-Verfahren eingesetzten Instrumente zur Qualitätsmessung. Der Anpassungsbedarf wird vom IQTIG daher mit Blick auf die Sicherstellung der Eignung der Qualitätsindikatoren ermittelt (siehe Kapitel 13). Abbildung 10 zeigt das Vorgehen bei der Anpassung von Qualitätsmessungen mit Indikatoren. Ausgangspunkt dafür sind Hinweise auf Anpassungsbedarf, die das IQTIG von extern erhält, insbesondere von Beteiligten an den QS-Verfahren, oder die es selbst im Rahmen der Verfahrensdurchführung ermittelt (siehe Abschnitt 8.1). Bei der anschließenden Prüfung des Anpassungsbedarfs (siehe Abschnitt 8.2) untersucht das IQTIG, in Bezug auf welche Eignungskriterien von Qualitätsmessungen Hinweise auf Anpassungsbedarf bestehen. Ergibt sich nach vorläufiger Beurteilung eine Notwendigkeit zur Anpassung, nimmt das IQTIG nachfolgend eine detaillierte Prüfung des Anpassungsbedarfs unter Hinzuziehung weiterer Informationen vor. Auf dieser Grundlage entwickelt das IQTIG Empfehlungen an den G-BA, welche Anpassungen vorgenommen werden sollten, oder empfiehlt dem G-BA die Beauftragung einer Weiterentwicklung, falls die Anpassungen umfangreiche Analysen oder Entwicklungsarbeiten erfordern (siehe Abschnitt 8.3).

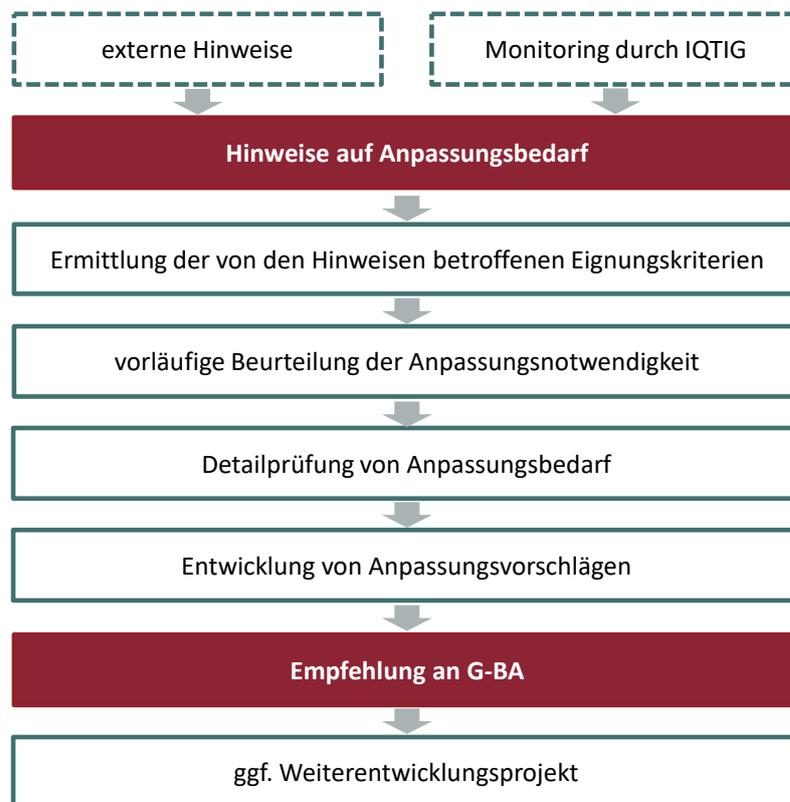


Abbildung 10: Vorgehen bei Anpassungen von Qualitätsmessungen

## 8.1 Hinweise auf Anpassungsbedarf

Hinweise auf mögliche Anpassungen an den Instrumenten der Qualitätsmessung werden auf mehreren Wegen ermittelt. Eine wichtige Quelle für solche Hinweise sind die durchgeführten Verfahren selbst, beispielsweise

- Rückmeldungen von Leistungserbringern, von Patientinnen und Patienten, von den für die Qualitätssicherung beauftragten Stellen auf der Landesebene, von Fachgesellschaften oder Verbänden,
- Hinweise der Trägerorganisationen in den Arbeitsgruppen und Gremien des G-BA ,
- Auswertungsergebnisse im Rahmen der regulären Berichterstellung,
- Hinweise der beratenden Expertengruppen z. B. auf neue Entwicklungen in der Versorgungspraxis,
- Ergebnisse der Stellungnahmeverfahren mit Leistungserbringern oder
- Ergebnisse der Prüfung der Datengrundlage im Rahmen der Auswertung sowie der Verfahren zur Datenvalidierung.

Weitere Hinweise können sich aus ergänzenden Untersuchungen ergeben, die das IQTIG zum Monitoring des Anpassungsbedarfs vornimmt, wie

- der Überwachung von Änderungen der medizinischen Klassifikationscodes oder Kodierrichtlinien,
- dem Monitoring neuer wissenschaftlicher Publikationen, z. B. von Leitlinienaktualisierungen,
- der fortlaufenden Beurteilung, wie sich vorangegangene Änderungen an den QS-Verfahren, z. B. an den Erfassungsinstrumenten, Indikatoren oder Rahmenbedingungen, ausgewirkt haben,
- explorativen Datenanalysen über die reguläre Berichterstattung hinaus,
- der Einstufung der bundesweiten Ergebnisse bzgl. des Potenzials zur Verbesserung und bzgl. eines besonderen Handlungsbedarfs (siehe Kapitel 17) und
- der wiederkehrenden Konsultation der beratenden Expertinnen und Experten hinsichtlich neuer Entwicklungen in Versorgungsalltag und Forschung.

Die Grundlage für die Prüfung, ob die Hinweise tatsächlichen Anpassungsbedarf an den Messinstrumenten bedeuten, sind die Eignungskriterien für Qualitätsmessungen (siehe Kapitel 13). Die oben genannten Informationen werden darauf untersucht, ob sie für einen oder mehrere Qualitätsindikatoren eine Einschränkung in der Erfüllung eines oder mehrerer Eignungskriterien bedeuten könnten.

## 8.2 Prüfung von Anpassungsbedarf

Sofern die oben genannten Hinweise Anpassungsbedarf an den Instrumenten der Qualitätssicherung signalisieren, wird zunächst eine vorläufige Beurteilung vorgenommen, welche Art von Anpassung (siehe Abschnitt 8.3) erforderlich sein könnte und welche zusätzlichen Informationen für die weitere Prüfung des Anpassungsbedarfs und ggf. für die Entwicklung von Empfehlungen an den G-BA benötigt werden.

Beispielsweise kann der Hinweis, dass ein wichtiger Einflussfaktor auf die Indikatorergebnisse in der Rechenregel eines Indikators bisher nicht berücksichtigt ist, auf Anpassungsbedarf an der Risikoadjustierung hindeuten und Recherchen nach Risikofaktoren notwendig machen. Ein anderes Beispiel sind Hinweise auf Qualitätsdefizite bei einem neu eingeführten Behandlungsverfahren, die ggf. durch einen neuen Qualitätsindikator abgebildet werden sollten und Informationen zu diesem Behandlungsverfahren erfordern. Die weitergehenden Informationen werden über eine Literaturrecherche sowie durch die Einbindung von Expertinnen und Experten ermittelt. Sofern entsprechende Daten verfügbar sind, werden ggf. zusätzlich Sonderauswertungen vorgenommen.

### **Literaturrecherche**

Informationen aus der wissenschaftlichen Literatur dienen bei der Entwicklung von Empfehlungen zu den Messinstrumenten insbesondere der Überprüfung folgender Eignungskriterien für die betreffenden Qualitätsindikatoren:

- Bedeutung für die Patientinnen und Patienten
- Potenzial zur Verbesserung
- Beeinflussbarkeit durch den Leistungserbringer
- Angemessenheit des Referenzbereichs
- Angemessenheit der Risikoadjustierung

Zu diesem Zweck wird mittels Literaturrecherche untersucht, auf welche Patientengruppen sich das Qualitätsmerkmal eines Indikators beziehen sollte, welche Unterschiede in der Erfüllung des Qualitätsmerkmals zwischen den Leistungserbringern bestehen und welche vom Leistungserbringer beeinflussbaren und nicht beeinflussbaren Faktoren sich auf die Ausprägung des Qualitätsmerkmal auswirken. Besteht der voraussichtliche Anpassungsbedarf in der Anpassung eines bestehenden Qualitätsindikators, wird geprüft, welche Informationen zu den Fragestellungen bereits vorliegen (z. B. aus dem Entwicklungsprozess des Indikators) und welche Informationen aktualisiert oder ergänzt werden müssen.

Je nach Fragestellung werden orientierende oder systematische Recherchen eingesetzt (siehe Kapitel 9). So kann bei eher explorativen Fragestellungen, die einen Überblick über einen Sachverhalt (z. B. übliche Behandlungsverfahren für eine Erkrankung oder Risikofaktoren für ein Behandlungsergebnis) geben sollen, eine orientierende Recherche ausreichend sein.

### **Einbindung von Expertengruppen**

Expertinnen und Experten werden vom IQTIG bei der Prüfung von Anpassungsbedarf beratend hinzugezogen, um die Informationen aus Literatur und Datenauswertungen aus fachlich-wissenschaftlicher Sicht und aus ihrer praktischen Erfahrung zu ergänzen. Dazu wird in der Regel jedes QS-Verfahren von einer Expertengruppe beim IQTIG begleitet<sup>21</sup> und es werden bei Bedarf weitere Fachexpertinnen und Fachexperten einbezogen. Die Zusammensetzung dieser Expertengruppen wird mit dem Ziel vorgenommen, durch Einbindung eines breiten Wissens- und

---

<sup>21</sup> In den Verfahren nach DeQS-RL wird diese Funktion von den Expertengremien auf Bundesebene nach §26 Teil 1 DeQS-RL wahrgenommen, in der plan. QI-RL durch die Gremien zur Systempflege nach § 14 plan. QI-RL

Erfahrungsspektrums eine fundierte fachliche Beratung sicherzustellen, und folgt den entsprechenden Grundsätzen des IQTIG (siehe Kapitel 10).

Das medizinisch-fachliche Wissen der Expertinnen und Experten wird bei der Prüfung von Anpassungsbedarf an den Messinstrumenten vom IQTIG schwerpunktmäßig zur Einschätzung folgender Eignungskriterien genutzt:

- Bedeutung für die Patientinnen und Patienten
- Potenzial zur Verbesserung
- Beeinflussbarkeit durch den Leistungserbringer
- Angemessenheit des Referenzbereichs
- Angemessenheit der Risikoadjustierung

Beispielsweise werden die Expertinnen und Experten zu ihrer Einschätzung befragt, inwieweit und mit welchen Maßnahmen das Qualitätsziel eines Indikators durch die Leistungserbringer erreicht werden kann, welche Indikatorergebnisse bei optimaler Versorgungsqualität erreicht werden könnten und welche besonderen und häufigen fachlichen Konstellationen bei der Qualitätsmessung und -bewertung zu berücksichtigen sind. Diese Beratungsergebnisse fließen u. a. in die Beurteilung von besonderem Handlungsbedarf (siehe Kapitel 17) und in die Festlegung oder Anpassung von Referenzbereichen ein. Darüber hinaus beraten die Expertinnen und Experten das IQTIG dazu, mittels welcher Anpassungen die Eignung der jeweiligen Qualitätsindikatoren verbessert werden könnte.

### **8.3 Empfehlungen an den G-BA**

Auf Grundlage der genannten Informationen nimmt das IQTIG eine Gesamtbeurteilung vor, ob und ggf. welche Anpassungen an den Instrumenten der Qualitätssicherung empfohlen werden. Die Empfehlungen können grundsätzlich jeden Bestandteil der Instrumente betreffen, beispielsweise die Datenerfassung, Datenflüsse, Maßnahmen zur Datenvalidierung, Rechenregeln der Qualitätsindikatoren und Bewertungsmethodik inkl. Referenzbereiche. Die Entwicklung der Empfehlungen zu Anpassungen an den Messinstrumenten sowie deren Darstellung erfolgt in allen Fällen anhand der in Kapitel 13 dargestellten Eignungskriterien.

#### **Anpassung von Referenzbereichen**

Anpassungen kriteriumsbezogener Referenzbereiche (siehe Abschnitt 16.3) werden empfohlen, wenn die Erreichbarkeit des Qualitätsziels durch die Leistungserbringer anders als zuvor beurteilt wird. Dies ist beispielsweise der Fall, wenn Einflussfaktoren auf die Indikatorergebnisse bekannt werden, die nicht von den Leistungserbringern zu verantworten sind und die die Rechenregel des Indikators bisher nicht berücksichtigt, oder wenn medizinischer Fortschritt seit der letzten Überprüfung des Indikators und seines Referenzbereichs höhere Versorgungsqualität ermöglicht.

Die Prüfung auf Anpassungsbedarf kann darüber hinaus zu einer geänderten Einschätzung darüber führen, wie sicher die Beurteilung der Erreichbarkeit des Qualitätsziels ist. Ermöglichen eine verbesserte Datenbasis aus dem Regelbetrieb oder das Vorliegen neuer evidenzbasierter

Anforderungen (z. B. in Leitlinien), zu benennen, welches Maß an Qualität erreichbar ist, so empfiehlt das IQTIG bei Indikatoren mit verteilungsbezogenem Referenzbereich die Einführung eines kriteriumsbezogenen Referenzbereichs.

### **Anpassung von Qualitätsindikatoren**

Zeigt die Prüfung, dass andere Eignungskriterien als die Angemessenheit des Referenzbereichs nicht hinreichend erfüllt sind, so wird eine Anpassung des betreffenden Qualitätsindikators empfohlen. Das kann beispielsweise der Fall sein, wenn besondere, häufige Behandlungskonstellationen in der Rechenregel nicht berücksichtigt sind oder wenn durch neue Abrechnungsmodalitäten ein Datenfeld anders als zuvor ausgefüllt wird. In solchen Fällen werden Vorschläge entwickelt, wie die Rechenregel ggf. einschließlich Risikoadjustierung geändert werden sollte oder welche Anpassungen an den Instrumenten zur Datenerfassung (z. B. Auslösekriterien, Datenfeldern, Ausfüllhinweisen, Plausibilitätsregeln, Datenflüssen) notwendig sind, um die Messeigenschaften zu optimieren.

### **Aussetzen von Qualitätsindikatoren**

Sofern die Prüfung eines Qualitätsindikators hinsichtlich der Eignungskriterien nach Kapitel 13 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 Aussetzung empfohlen. Falls das Qualitätsmerkmal weiterhin abgebildet werden soll, wird durch das IQTIG geprüft, ob es bereits hinreichend durch einen anderen Qualitätsindikator abgedeckt wird oder ob die Einführung eines neu entwickelten Qualitätsindikators empfohlen wird.

### **Entwicklung neuer Qualitätsindikatoren**

Ein Ergebnis der Prüfung kann auch sein, dass das verwendete Indikatorenset um neue Indikatoren ergänzt werden sollte, beispielsweise, wenn ein bereits adressierter Qualitätsaspekt nicht ausreichend durch die bestehenden Indikatoren abgedeckt ist. Auch neu aufgetretene Qualitätsdefizite in der Versorgung, z. B. bei neu eingeführten Behandlungsverfahren, können Anlass für die Entwicklung neuer Indikatoren sein. Für neue Qualitätsindikatoren wird grundsätzlich eine Einschätzung hinsichtlich aller Eignungskriterien vorgenommen, auf deren Basis eine Gesamtempfehlung ausgesprochen wird, ob der Indikator im QS-Verfahren eingesetzt werden sollte.

### **Weitergehende Anpassungen von QS-Verfahren**

Bei umfangreichem Änderungsbedarf erfolgt die Anpassung der QS-Verfahren in gesonderten Weiterentwicklungsprojekten. Beispiele für solche weitergehenden Anpassungen sind

- die Zusammenlegung von Datenerfassung und Auswertung bestehender QS-Verfahren oder die Aufteilung eines QS-Verfahrens in mehrere Teilverfahren,
- die Hinzunahme einer neuen Datenquelle zu einem QS-Verfahren, z. B. durch die Entwicklung von Follow-up-Indikatoren auf Grundlage von Sozialdaten bei den Krankenkassen, oder die Entwicklung von Indikatoren auf Grundlage von Patientenbefragungen,

- die größere Änderung des Qualitätsmodells eines bestehenden QS-Verfahrens, z. B. durch die zusätzliche Betrachtung einer weiteren Eingriffsart über mehrere Indikatoren eines QS-Verfahrens hinweg,
- die Weiterentwicklung der grundlegenden statistischen Auswertungsmethodik in einem QS-Verfahren.

Weiterentwicklungsprojekte für QS-Verfahren werden dem G-BA vom IQTIG vorgeschlagen und von ihm im Benehmen mit dem IQTIG beauftragt.

# Teil C: Methodische Elemente

## 9 Informationsbeschaffung

Die Informationsbeschaffung gehört zu den grundlegenden Bausteinen für die Erfüllung der Aufgaben des IQTIG. In diesem Kapitel werden die verschiedenen Arten und deren jeweilige Kernelemente der Informationsgewinnung aus wissenschaftlicher Literatur und ergänzenden Quellen dargestellt, die in der Arbeit des IQTIG angewandt werden:

- orientierende Recherche
- systematische Literaturrecherche
- Evidenzaufbereitung und kritische Bewertung relevanter Publikationen

### 9.1 Orientierende Recherche

Im Rahmen von orientierenden Recherchen wird gezielt nach Informationen gesucht, die zur Beantwortung der jeweiligen Fragestellung benötigt werden. Die Quellen werden daher so ausgewählt, dass möglichst vertrauenswürdige Informationen identifiziert werden können. Beispiele hierfür sind Hintergrundrecherchen zu Beginn eines Projekts, die zur Einarbeitung in das Thema dienen, oder die gezielte Suche nach epidemiologischen Daten. Als Informationsquellen werden hierfür u. a. Point-of-Care-Tools wie DynaMed, die Cochrane Library, der Social Sciences Citation Index (SSCI) und Leitliniendatenbanken z. B. der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V. (AWMF) herangezogen. Auch eine einfache Suche ohne Entwicklung einer Suchstrategie und meist nur in einer bibliografischen Datenbank, beispielsweise in MEDLINE via PubMed mit voreingestellten Suchfiltern für die Suche nach systematischen Reviews oder nach Primärstudien zu Fragen der Therapie, Diagnostik, Ätiologie und Prognose (PubMed Clinical Queries), wird genutzt.

Themenspezifisch, z. B. bei Fragen zum Versorgungsprozess, können Recherchen beispielsweise in Morbiditätsregistern, Arzneimittelregistern, Berichten des Statistischen Bundesamts oder der Gesundheitsberichterstattung des Bundes erfolgen. Bei Bedarf werden weitere Informationsquellen (z. B. spezifische Websites) für die Recherche hinzugezogen.

In Abhängigkeit von der Fragestellung kann es zur Bearbeitung der Projekte notwendig sein, z. B. nach bereits vorhandenen Messinstrumenten zu recherchieren. 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) sowie Itembanken wie PROMIS (Patient-Reported Outcomes Measurement Information System) genutzt.

Des Weiteren werden ergänzende Recherchetechniken angewandt: Als Ausgangspunkt dienen hierbei ein oder mehrere relevante Schlüsselartikel, die genutzt werden, um damit weitere relevante Artikel zu identifizieren. Angewandt wird hierbei zum einen die Snowball-Recherche: Bei dieser Methode der Informationsbeschaffung erfolgt sowohl eine Recherche nach Artikeln, die den Ausgangsartikel zitieren (*forward citation* über Google Scholar), als auch eine Sichtung der im Artikel selbst zitierten Publikationen (*backward citation* = Sichten der Referenzlisten) (EUnetHTA 2017). Des Weiteren wird die „Similar articles“-Funktion von PubMed mit Sichtung

der ersten 20 Publikationen, die dem Ausgangsartikel ähnlich sind, genutzt (Waffenschmidt et al. 2013).

Für orientierende Recherchen werden keine komplexen Suchstrategien entworfen, es erfolgt auch keine Dokumentation der Suche. Die Suche und Auswahl der Literatur findet durch eine Person statt.

## 9.2 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 aus der Literatur, 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 Projekts, die themenspezifische Auswahl der Informationsquellen, die Entwicklung der Suchstrategie inklusive der Durchführung der Suche, die Auswahl relevanter Publikationen sowie die Dokumentation der gesamten Recherche. Die Umwandlung der Fragestellung in eine präzise, recherchierbare Frage ist dabei der erste zentrale Schritt. Hieraus werden alle weiteren Schritte abgeleitet.

### Entwicklung einer präzisen, recherchierbaren Fragestellung und Definition von Ein- und Ausschlusskriterien

Zur Vorbereitung einer systematischen Recherche wird zunächst eine präzise, recherchierbare Fragestellung erarbeitet. Verschiedene etablierte Schemata werden in Abhängigkeit von der Zielsetzung der Literaturrecherche angewendet wie z. B. das PICO-Schema (Population, 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) (Kloda und Bartlett 2013). Hiermit wird die Voraussetzung dafür geschaffen, dass die für die Fragestellung relevanten Publikationen auch in den ausgewählten Informationsquellen identifiziert und Ein- und Ausschlusskriterien für die Auswahl der Evidenz definiert werden können. Je nach Fragestellung werden auch Kriterien für die Übertragbarkeit der Inhalte auf den deutschen Versorgungskontext definiert (Beispiel: Nur Publikationen aus Ländern, die wie Deutschland zum WHO Stratum A der Staateneinteilung des Weltgesundheitsberichts 2003 der World Health Organization (WHO) gehören, werden eingeschlossen; WHO 2003).

Ausgehend von der Fragestellung wird auch definiert, welche Publikationsarten (Leitlinien, systematische Reviews, Primärstudien) von Interesse sind.

#### 9.2.1 Recherche nach Leitlinien

Für Fragestellungen, die z. B. Standards der Versorgung betreffen, wird systematisch nach evidenzbasierten Leitlinien gesucht (entsprechend S2e- und S3-Leitlinien nach AWMF-Methodik). Für die Leitliniensuche werden Quellen wie z. B. das „Portal der wissenschaftlichen Medizin“ der

AWMF<sup>22</sup>, die „International Guideline Library“ des Guideline International Network (G-I-N)<sup>23</sup> und die Metasuchmaschine Trip<sup>24</sup> genutzt. Abhängig von der Thematik findet eine Suche in weiteren Datenbanken von nationalen und/oder internationalen Leitlinienerstellern sowie auf Websites der Fachgesellschaften statt.

Die Suche wird für jede Datenbank bzw. Informationsquelle dokumentiert, inklusive des Suchzeitraums sowie sprachlicher und anderer Einschränkungen bzw. Filter.

Die Titel aller recherchierten Leitlinien werden von zwei Personen unabhängig voneinander hinsichtlich ihrer inhaltlichen Relevanz bezogen auf die a priori festgelegte Frage überprüft. Inhaltlich für die Fragestellung 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 eine Einigung herbeigeführt. Die Ausschlussgründe aller im Volltext ausgeschlossenen Leitlinien werden dokumentiert.

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

### 9.2.2 Recherche nach systematischen Übersichtsarbeiten

Systemische Reviews dienen u. a. im Rahmen der Entwicklung eines QS-Verfahrens zur Identifizierung und Beschreibung von Qualitätsaspekten und Qualitätsmerkmalen (siehe Kapitel 6 und 7) oder auch im Rahmen der Verfahrenspflege zur Identifizierung von Weiterentwicklungsbedarfen (siehe Kapitel 8).

Bei der Suche nach systematischen Übersichtsarbeiten wird in der Cochrane Library 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).

Ausgehend von der Fragestellung z. B. gemäß PICO-Schema und von einem vorab vorliegenden Testset aus bereits bekannten, relevanten Publikationen werden geeignete Suchbegriffe abgeleitet. Für die Recherche wird zunächst eine Strategie für die Literaturliteraturdatenbank MEDLINE (via Ovid) entwickelt und dann entsprechend an die anderen Datenbanken angepasst. Hierbei wird auch ein Suchfilter für systematische Reviews eingesetzt (z. B. modifiziert nach Scottish Intercollegiate Guidelines Network; SIGN [kein Datum]). Die Syntax wird regelhaft von einer zweiten Person überprüft.

Die Suche wird für jede Datenbank bzw. Informationsquelle dokumentiert, inklusive des Suchzeitraums sowie sprachlicher und anderer Einschränkungen bzw. Filter.

Die Titel und Abstracts der recherchierten Publikationen werden von zwei Personen unabhängig voneinander hinsichtlich ihrer inhaltlichen Relevanz bezogen auf die a priori festgelegte Frage

---

<sup>22</sup> <https://www.awmf.org> (abgerufen am 24.09.2020)

<sup>23</sup> <https://g-i-n.net> (abgerufen am 24.09.2020)

<sup>24</sup> <https://www.tripdatabase.com/> (abgerufen am 24.09.2020)

geprüft und ausgewählt. Uneinheitliche Bewertungen werden diskutiert und eine Einigung herbeigeführt. Die Volltexte der ausgewählten Publikationen werden ebenfalls von zwei Personen unabhängig voneinander dahingehend überprüft, ob die a priori festgelegten Einschlusskriterien zutreffen. Uneinheitliche Bewertungen werden diskutiert und eine Einigung herbeigeführt. Es erfolgt eine Dokumentation des Volltext-Screenings und der Ausschlussgründe für alle ausgeschlossenen Publikationen. Der gesamte Recherche- und Auswahlprozess nach systematischen Reviews in bibliographischen Datenbanken wird in einem Flussdiagramm dokumentiert (Liberati et al. 2009).

### 9.2.3 Recherche nach Primärliteratur

Zur Beschreibung der Patientenperspektive hinsichtlich der Behandlung und Versorgung wird bei der Entwicklung von Qualitätssicherungsmaßnahmen regelhaft auch nach Primärliteratur recherchiert (siehe Kapitel 6 und 7). Hierbei sind neben Forschungsergebnissen aus qualitativen Studien gerade auch quantitative (Befragungs-)Studien ein wesentlicher Teil der Informationsbeschaffung. Sie werden zur Ermittlung und Abbildung der Patientenperspektive hinsichtlich Erfahrungen, Einstellungen, Haltungen und Akzeptanz einbezogen. Weitere Fragestellungen der Recherche nach Primärstudien betreffen u. a. die Versorgungssituation in Deutschland sowie die Identifikation von Anhaltspunkten auf eine mögliche Über-, Unter- und Fehlversorgung bzw. von Qualitätsdefiziten / Qualitätspotenzialen in der Versorgung von Patientinnen und Patienten.

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, erfolgen ergänzend eine Suche nach unpublizierten Daten (z. B. Studienregister, Autorenanfragen) und eine Handsuche, z. B. in den Referenzlisten ausgewählter Schlüsselartikel bzw. in ausgewählten Fachzeitschriften.

Ausgehend von der Fragestellung z. B. gemäß PICO-Schema und von einem vorab vorliegenden Testset aus bereits bekannten, relevanten Publikationen werden geeignete Suchbegriffe abgeleitet. Für die Recherche wird zunächst eine Strategie für die Literaturdatenbank MEDLINE (via Ovid) entwickelt und dann entsprechend an die anderen Datenbanken angepasst. Abhängig von der Fragestellung kann die Recherche auf Publikationen aus Deutschland (bzw. auf eine deutsche Studienpopulation) eingeschränkt werden. Hierzu wird meist ein Suchfilter basierend auf dem „Deutschland-Filter“ von Pieper et al. (2015) eingesetzt und für die jeweilige Suchoberfläche angepasst. Die Syntax wird regelhaft von einer zweiten Person überprüft.

Der Auswahl- und Dokumentationsprozess erfolgt analog zum Vorgehen bei systematischen Reviews (siehe Abschnitt 9.2.2).

### **9.3 Evidenzaufbereitung und kritische Bewertung relevanter Publikationen**

Unter Evidenzaufbereitung versteht das IQTIG den Prozess der systematischen Identifizierung, Auswahl und Ergebnisdarstellung vorhandener Evidenz zu einer Fragestellung. In Bezug auf die Zielsetzung kommen in der Arbeit des IQTIG vorrangig zwei Formen der Evidenzaufbereitung zum Einsatz.

#### **9.3.1 Evidenzaufbereitung orientiert an der Methodik eines Scoping Reviews**

Nach Colquhoun et al. (2014) und von Elm et al. (2019) 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 zu erstellen. Zu einer (oft eher breiten) Fragestellung erfolgt eine systematische Recherche (entsprechend Abschnitt 9.2), die auch nicht vorab festgelegte, sondern sich schrittweise annähernde Prozessschritte umfassen kann, z. B. bei der Entwicklung oder Anpassung von Ein- und Ausschlusskriterien für die Literatursuche. Eine kritische Bewertung der Literatur erfolgt hier nicht regelhaft (Schmucker et al. 2013, von Elm et al. 2019). Die Aufbereitung der Ergebnisse erfolgt deskriptiv. Sie dienen als Hinweise bei der Bearbeitung der Projektfragestellungen. Dabei werden sie in der Regel nicht als gesonderter Evidenzbericht herausgegeben, sondern direkt in die Projektberichte integriert.

Beispiele hierfür sind Fragestellungen zu Qualitätsaspekten bei der Entwicklung des Qualitätsmodells sowie Fragen nach der Patientenperspektive und Versorgungssituation oder im Rahmen der Themenerschließung von Entwicklungsprojekten.

#### **9.3.2 Evidenzaufbereitung mit kritischer Bewertung der eingeschlossenen Publikationen**

Ziel dieser Form der Evidenzaufbereitung ist eine systematische Identifizierung und Qualitätsbewertung aller verfügbaren und bekannten Publikationen, die zur Beantwortung der Fragestellung von Bedeutung sind. Unter Berücksichtigung der methodischen Qualität der einzelnen Publikationen werden die Ergebnisse deskriptiv aufbereitet und zusammengefasst. Dabei werden die Ergebnisse in der Regel nicht als gesonderter Evidenzbericht herausgegeben, sondern direkt in die Projektberichte integriert. Beispiele hierfür sind die regelhaft durchgeführte systematische Aufbereitung und Bewertung von evidenzbasierten Leitlinien für Fragestellungen, die die Standards der Versorgung betreffen, oder auch die themenspezifische Aufbereitung von systematischen Reviews bei der Entwicklung und Anpassung von Qualitätssicherungsmaßnahmen (siehe Abschnitt 9.2).

Im Unterschied zum Scoping-Review-Ansatz wird regelhaft die Qualität für jede im Volltext eingeschlossene Publikation bewertet. Qualitätsbewertung heißt hier, dass für jede Publikation eingeschätzt wird, welchem möglichen Risiko von Verzerrungen (systematischer Fehler / Bias) sie

unterliegt (Cochrane Deutschland/AWMF-IMWi 2016). Dies ermöglicht, eine Aussage über die Vertrauenswürdigkeit der Ergebnisse der Publikation zu treffen bzw. darzustellen. Die Verzerrungsmöglichkeiten und deren kritische Bewertung sind zwischen Leitlinien, systematischen Übersichtsarbeiten und Primärstudien unterschiedlich. Hierfür stehen verschiedene Instrumente/Hilfsmittel zur Verfügung.

### **Kritische Bewertung von Leitlinien und Darstellung der evidenzbasierten Empfehlungen**

Um die methodische Genauigkeit und Transparenz bei der Entwicklung von Leitlinien kritisch zu beurteilen, verwendet das IQTIG das AGREE-II-Instrument (Appraisal of Guidelines for Research & Evaluation; AGREE Next Steps Consortium 2017). Es besteht aus 23 Items, die sechs Domänen zugeordnet sind, welche die Qualität einer Leitlinie bestimmen:

- Domäne 1: Scope and Purpose
- Domäne 2: Stakeholder Involvement
- Domäne 3: Rigour of Development
- Domäne 4: Clarity of Presentation
- Domäne 5: Applicability
- Domäne 6: Editorial Independence

Daran angelehnt werden die Leitlinien von zwei Personen unabhängig voneinander bewertet, begrenzt auf die Domänen 2, 3 und 6. Dies erfolgt im Hinblick auf die Zielsetzung, vor allem die methodische Qualität der eingeschlossenen Leitlinien zu vergleichen und hinsichtlich ihrer Berücksichtigung diskutieren zu können. Die anderen Domänen beschreiben die Qualität einer Leitlinie aus der Sicht der Anwender in der praktischen Versorgung.

Die entsprechenden Items werden auf einer Punkteskala von 1 (trifft überhaupt nicht zu) bis 7 (trifft vollständig zu) bewertet. Bei Unstimmigkeiten von mehr als zwei Punkten wird das Item von den bewertenden Personen diskutiert und ein finaler Konsens herbeigeführt. Daraus wird ein standardisierter Wert je Domäne errechnet, der den prozentualen Anteil an der maximal erreichbaren Punktzahl je Domäne darstellt.

Hohe standardisierte Domänenwerte deuten auf eine hohe Qualität der Leitlinie hin und niedrige standardisierte Domänenwerte weisen auf eine geringe Qualität der Leitlinie hin. Die standardisierten Domänenwerte ermöglichen es, die verschiedenen Leitlinien hinsichtlich ihrer methodischen Qualität untereinander zu vergleichen.

Für jede eingeschlossene Leitlinie werden Charakteristika wie z. B. federführende/beteiligte Fachgesellschaft(en)/Institution(en), Herkunftsland/Geltungsbereich, Thema und Ziel der Leitlinie sowie Zielpopulation extrahiert. Zudem erfolgt eine Zusammenstellung der in den Leitlinien verwendeten Schemata für die Evidenz- und Empfehlungsgraduierung.

Die einzelnen zur Beantwortung der Fragestellung relevanten Empfehlungen werden von einer Person extrahiert, tabellarisch dargestellt und mit der jeweiligen Evidenz- und Empfehlungsstärke im Projektbericht abgebildet. Eine weitere Person prüft die Extraktion. Die weitere Synthese erfolgt projektspezifisch.

### **Kritische Bewertung und Darstellung 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 (AMeasurement Tool to Assess systematic Reviews; Shea et al. 2017) oder dem ROBIS-Instrument (Risk Of Bias In Systematic reviews; Whiting et al. 2016) bewerten zwei Personen unabhängig voneinander die methodische Qualität systematischer Übersichtsarbeiten. Uneinheitliche Bewertungen werden diskutiert und eine Einigung herbeigeführt.

Die Datenextraktion der eingeschlossenen systematischen Reviews erfolgt von einer Person und wird von einer weiteren geprüft. Je nach zu beantwortender Fragestellung werden relevante Informationen wie z. B. Publikationsjahr, Intervention, Outcomes, Ergebnisse extrahiert und tabellarisch dargestellt. Die weitere Synthese erfolgt projektspezifisch unter Berücksichtigung der kritischen Bewertung.

### **Kritische Bewertung und Darstellung von Primärliteratur**

Mögliche Verzerrungen in der Primärliteratur lassen sich im Wesentlichen folgendermaßen einteilen (Fletcher und Fletcher 2007):

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

Mithilfe standardisierter Bewertungsinstrumente wie z. B. RoB 2 (Cochrane risk-of-bias tool for randomized trials; Sterne et al. 2019) oder ROBINS-I (Risk Of Bias In Non-randomised Studies - of Interventions; Sterne et al. 2016) bewerten zwei Personen unabhängig voneinander das Verzerrungspotenzial von Einzelstudien. Uneinheitliche Bewertungen werden diskutiert und eine Einigung herbeigeführt.

Eingeschlossene qualitative Studien werden bzgl. ihrer methodischen Qualität mithilfe eines Bewertungsinstruments, z. B. mit der Checkliste des Critical Appraisal Skills Programme (CASP 2013), bewertet.

Die Datenextraktion der eingeschlossenen Publikationen erfolgt von einer Person und wird von einer weiteren geprüft. Je nach zu beantwortender Fragestellung werden relevante Informationen wie z. B. Publikationsjahr, Studiendesign, Studienpopulation, Fragestellung, Ergebnisse extrahiert und tabellarisch dargestellt. Die weitere Synthese erfolgt projektspezifisch unter Berücksichtigung der kritischen Bewertung.

## 10 Beratung durch externe Expertinnen und Experten

Das IQTIG bezieht neben publiziertem Wissen z. B. aus der wissenschaftlichen Literatur (siehe Kapitel 9) auch das Wissen, die Erfahrungen und die Perspektiven Dritter in seine Arbeit ein, insbesondere von betroffenen Patientinnen und Patienten, von an der Versorgung Beteiligten sowie von Expertinnen und Experten. Dadurch soll sowohl für die Entwicklungsarbeiten des IQTIG als auch bei der Durchführung von QS-Verfahren eine möglichst hochwertige Informationsgrundlage unter Berücksichtigung aller wichtigen Wissensbestände erreicht werden. Für Fragestellungen, bei denen Expertenwissen im Vordergrund des Informationsbedarfs steht, erfolgt eine Beratung durch externe Expertinnen und Experten.

### 10.1 Begriffsdefinition „Expertin/Experte“

Für den Begriff „Expertin/Experte“ gibt es keine allgemeine, von der Fragestellung unabhängige Definition (siehe z. B. Baker et al. 2006, Niederberger und Wassermann 2015, Bogner et al. 2014: 11 ff.). 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 und dieses als solches anerkannt ist (Meuser und Nagel 1991: 443 f., Bogner et al. 2014: 13 f., Hitzler 1994: 25 ff.). Das IQTIG versteht unter „Expertin/Experte“ Personen, die in besonderem Maße Wissen, Erfahrung oder Deutungskompetenz (Bogner et al. 2014: 17 ff.) für Fragestellungen im Aufgabenbereich des IQTIG aufweisen und auf deren Wissen zurückgegriffen werden soll. Bei Expertinnen und Experten kann es sich im Kontext der Qualitätssicherung demnach nicht nur um Angehörige der Gesundheitsberufe handeln, sondern auch um andere Berufsgruppen (z. B. Medizincontrollerinnen und -controller, Kodierexpertinnen und -experten, Datenschutzbeauftragte, Juristinnen und Juristen, Sozialarbeiterinnen und -arbeiter, Informatikerinnen und Informatiker, Qualitätsmanagerinnen und -manager) sowie um sachkundige Personen der Interessenvertretung der Patientinnen und Patienten.

### 10.2 Beratungsgegenstände

Das IQTIG zieht externe Expertinnen und Experten zur Beratung unterschiedlicher Sachverhalte heran. Die Beratung kann im Rahmen von Entwicklungsaufgaben des IQTIG erfolgen oder im Rahmen der Beteiligung des IQTIG an der Durchführung von QS-Verfahren des G-BA. Darüber hinaus zieht das IQTIG auch bei weiteren wissenschaftlichen Fragestellungen sowie bei der Entwicklung von Methoden nach Bedarf externe Expertise hinzu.

## Entwicklung und Anpassung von Instrumenten der Qualitätssicherung

Bei der Entwicklung und Anpassung von Instrumenten der Qualitätssicherung für einen Themenbereich beraten die Expertinnen und Experten das IQTIG insbesondere zu folgenden Themen:

- Versorgungspraxis in einem bestimmten Themenbereich, aus Perspektive der Versorgenden sowie der Patientinnen und Patienten
- Bedeutung für die Patientinnen und Patienten, Beeinflussbarkeit und Verbesserungspotenzial von Qualitätsmerkmalen der Versorgung
- medizinische Fragestellungen
- Praxis der Behandlungsdokumentation
- Abrechnungsfragen und Einflüsse des Vergütungssystems auf die Versorgung
- rechtliche Fragen zur Versorgungspraxis
- erwartete Auswirkungen von Qualitätsmessungen
- Anpassungsbedarf bei Qualitätsindikatoren (z. B. aufgrund neuer medizinischer Entwicklungen)
- Beurteilung von Indikatorergebnissen auf Bundesebene (z. B. hinsichtlich besonderen Handlungsbedarfs)

## Durchführung von QS-Verfahren im Auftrag des G-BA

Bei einigen vom G-BA beschlossenen QS-Verfahren, mit deren Durchführung das IQTIG betraut ist, macht der G-BA teilweise Vorgaben für die Beratung des IQTIG durch externe Expertinnen und Experten im Rahmen seiner Durchführungsaufgaben. Diese Vorgaben können sich u. a. auf die Auswahl der Expertinnen und Experten, auf die Zusammensetzung und Größe von Expertengruppen, auf formale Regelungen wie Benennungsverfahren, auf Beratungsgegenstände und auf Abstimmungsverfahren beziehen.<sup>25</sup> In diesen Fällen werden die Richtlinienvorgaben des G-BA vom IQTIG bei der Experteneinbindung berücksichtigt. In allen anderen Fällen geht das IQTIG nach den in diesem Kapitel beschriebenen Grundsätzen vor.

Die Beratungsgegenstände durch externe Expertinnen und Experten bei der Durchführung von QS-Verfahren umfassen über die oben genannten Beratungsgegenstände hinaus vor allem folgende Themen:

- Bewertung von Einrichtungsergebnissen (z. B. bei den sogenannten bundesbezogenen Verfahren nach DeQS-RL und beim Verfahren nach der Richtlinie zu planungsrelevanten Qualitätsindikatoren, plan. QI-RL<sup>26</sup>)
- Durchführung von Maßnahmen zur Qualitätsverbesserung mit den Leistungserbringern (z. B. Fachgespräche, Zielvereinbarungen)

---

<sup>25</sup> Beispielsweise enthält die Geschäftsordnung des G-BA für die Fachkommissionen gemäß § 12 plan. QI-RL die Regelung, dass die Fachkommission Empfehlungen mit Mehrheit beschließen soll.

<sup>26</sup> Richtlinie zu planungsrelevanten Qualitätsindikatoren. In der Fassung vom 15. Dezember 2016, zuletzt geändert am 17. Juni 2021, in Kraft getreten am 22. Juli 2021. URL: <https://www.g-ba.de/informationen/richtlinien/91/> (abgerufen am 21.03.2022).

### 10.3 Methodische Funktion der Beratung durch Expertinnen und Experten

Je nach interessierendem Sachverhalt kann die Beratung durch externe Expertinnen und Experten verschiedene methodische Funktionen erfüllen, aus denen sich die Ausgestaltung der Beratung (z. B. in Form von Experteninterviews oder in Form von Gruppendiskussionen) ableitet. Dabei werden in Anlehnung an Niederberger und Wassermann (2015: 37 ff.) folgende methodische Funktionen von Expertenberatung unterschieden:

- **Ermittlung oder Prüfung von Fakten**

In der Entwicklung und Umsetzung von Instrumenten der Qualitätssicherung ergibt sich häufig Beratungsbedarf zu Sachverhalten, die als Fakten, d. h. als unstrittig verstanden werden, beispielsweise, wenn der übliche Ablauf eines Operationsverfahrens ermittelt werden soll. In diesen Fällen besteht die Aufgabe der hinzugezogenen Expertinnen und Experten in der Bereitstellung von technischem Wissen oder Prozesswissen (Bogner et al. 2014: 17 f.). Die Beratung kann entweder durch gezielte Hinzuziehung einzelner Personen mit entsprechender Kompetenz erfolgen, z. B. eines Spezialisten oder einer Spezialistin für ein bestimmtes Operationsverfahren, oder durch eine größere Zahl von Expertinnen und Experten (z. B. eine Expertengruppe), wenn angenommen wird, dass die Breite oder Komplexität eines Sachverhalts dies erfordert.

- **Ermittlung unterschiedlicher Perspektiven**

Ist davon auszugehen, dass zu einer Fragestellung unterschiedliche Perspektiven von den Expertinnen und Experten eingenommen werden, wird eine möglichst umfassende Ermittlung und Abbildung dieser Perspektiven angestrebt. Dies ist besonders dann sinnvoll, wenn Deutungswissen im Vordergrund des Beratungsinteresses steht (Bogner et al. 2014: 18 f.), beispielsweise, wenn Einschätzungen zu Ursachen und Hintergründen von Qualitätsdefiziten erfragt werden sollen. In diesem Fall werden mehrere Personen zur Beratung hinzugezogen. Die Expertise kann dabei entweder von mehreren Einzelpersonen separat erhoben werden (z. B. in Form von Experteninterviews) oder, vor allem wenn ein Dialog und Austausch zwischen den Personen für die Erschließung der interessierenden Fragestellung als wichtig angesehen wird (Xyländer et al. 2020), über eine Expertengruppe.

- **Bewertung von Sachverhalten durch eine Expertengruppe**

Sofern unterschiedliche Perspektiven auf einen Sachverhalt zu berücksichtigen sind, kann auch ein Ziel sein, die Ansichten der einbezogenen Expertinnen und Experten zu einem gemeinsamen Ergebnis der Gruppe zusammenzuführen sowie eine ggf. quantitative Aussage über die Bewertung des Sachverhalts zu treffen. Beispielsweise kann das Urteil einer Expertengruppe eingeholt werden, welche Indikatorergebnisse bei optimaler Versorgungsqualität erreicht werden könnten. In diesem Fall erfolgt die Beratung grundsätzlich auf Basis einer Gruppendiskussion.

Die Beratungen einer Expertengruppe können auch mehrere der genannten methodischen Funktionen umfassen, wenn dieselbe Expertengruppe zu verschiedenen Beratungsgegenständen einbezogen wird, beispielsweise, wenn sowohl Fakten zu einem Versorgungsprozess zu klären als auch Einschätzungen zur Ursachen für Versorgungsprobleme vorzunehmen sind. In diesem Fall wird das methodische Vorgehen bei der Beratung an die jeweiligen Beratungsgegenstände angepasst.

Um kompetente und unvoreingenommene Beratungen durch Expertinnen und Experten sicherzustellen, verfolgt das IQTIG mehrere Ansätze:

- die Formulierung konkreter, themenspezifischer Voraussetzungen für die Tätigkeit als Expertin oder Experte
- der Ausschluss bzw. die Berücksichtigung relevanter Interessenkonflikte
- eine Zusammensetzung von Expertengruppen, die möglichst viele verschiedene Perspektiven berücksichtigt
- die Nutzung vertraulicher Befragungen bei Bewertungen durch Expertengruppen

## **10.4 Voraussetzungen für die Tätigkeit als Expertin oder Experte**

### **Fachliche Voraussetzungen**

Die Anforderungen an die persönlichen und beruflichen Kenntnisse externer Expertinnen und Experten für eine Beratungstätigkeit hängen von der jeweiligen Fragestellung ab. Sie werden daher hier nur in allgemeiner Form genannt und für die jeweilige Beratungstätigkeit themenspezifisch festgelegt. In der Regel beziehen sie sich auf formale Qualifikationen, die Dauer der Tätigkeit in einem Bereich der Gesundheitsversorgung, wissenschaftliche Tätigkeiten, Erfahrungen in der Qualitätssicherung oder auf die Aktualität der Kenntnisse (vgl. Baker et al. 2006, Bogner und Menz 2009: 68 f.).

### **Formale Regelungen**

Neben fachlichen Kriterien legt das IQTIG auch formale, vom Beratungsthema unabhängige Regelungen für die Tätigkeit als Expertin oder Experte fest. Diese regeln

- das Verfahren zur Bewerbung, Benennung von Mitgliedern in Expertengruppen,
- das Verfahren zur Prüfung und Berücksichtigung von Interessenkonflikten (s. u.),
- die Teilnahme an und Organisation von Besprechungen und Sitzungen,
- den Umgang mit Dokumenten,
- die Vertraulichkeit der Beratungen sowie
- die Beendigung der Mitarbeit in Expertengruppen.

### **Umgang mit Interessenkonflikten**

Interessenkonflikte sind definiert als Situationen, die ein Risiko dafür schaffen, dass professionelles Urteilen oder Handeln, welches sich auf ein primäres Interesse bezieht, durch ein sekundäres Interesse unangemessen beeinflusst wird (IOM 2009: 6, Thompson 1993). 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 können die Expertinnen oder Experten beeinflussen, ohne dass sie es bemerken (Thompson 2009).

Um eine ausschließlich an fachlichen Erwägungen orientierte Beratung zu erzielen und die fachliche Unabhängigkeit des IQTIG zu gewährleisten müssen Interessenkonflikte von beratenden Expertinnen und Experten, die zu einer unsachgemäßen Verzerrung der Beratungsleistung führen könnten, transparent gemacht und möglichst ausgeschlossen werden.<sup>27</sup> Die Beurteilung von Interessenkonflikten erfolgt vor einer möglichen Beratungstätigkeit für das IQTIG auf Grundlage einer Selbstauskunft der Expertinnen und Experten zu folgenden Arten von Beziehungen:

- Hauptberufliche, nebenberufliche und ehrenamtliche Tätigkeiten (z. B. Anstellungsverhältnisse, Tätigkeiten auf Honorarbasis, Beratungstätigkeiten)
- finanzielle oder geldwerte Zuwendungen (z. B. Unterstützung bei Forschungsaktivitäten, bei der Teilnahme an oder Ausrichtung von Veranstaltungen)
- Geschäftsanteile (z. B. Aktienbesitz, Fonds)

Die inhaltliche Bewertung der Angaben nimmt eine dafür eingerichtete Interessenkonflikt-Kommission als neutrale Stelle im IQTIG auf Grundlage einer Geschäftsordnung und festgelegter Verfahrensregeln vor. Diese prüft, ob die angegebenen Interessen zu relevanten Interessenkonflikten für die vorgesehenen Beratungsthemen führen und als wie ausgeprägt diese zu bewerten sind. Da Interessenkonflikte alltäglich und unvermeidlich sind (Klemperer 2011: 16, Thompson 1993), erfolgt nur dann ein Ausschluss von Beratungstätigkeiten, wenn sich aufgrund der vorgelegten Informationen erhebliche Bedenken hinsichtlich einer Verzerrung des Beratungsergebnisses ergeben.

### **Transparenz über die Beratung durch Expertinnen und Experten**

Die Namen der beratenden Expertinnen und Experten werden vom IQTIG in geeigneter Weise veröffentlicht (z. B. in seinen Berichten oder auf der Website des IQTIG). Angaben zu den Interessenkonflikten werden vom IQTIG in allgemeiner Form veröffentlicht, d. h., es wird dargelegt, ob Interessenkonflikte bestehen, jedoch keine Einzelheiten z. B. über die Höhe von Zuwendungen oder über Partner.

---

<sup>27</sup> Das IQTIG ist auch per Gesetz dazu verpflichtet, Interessenkonflikte von Beschäftigten und von allen anderen an der Aufgabenerfüllung des IQTIG beteiligten Personen und Institutionen zu vermeiden (§ 137a Abs. 9 SGB V).

## 10.5 Beratung durch Expertengruppen

### Kriterien für die Zusammensetzung von Expertengruppen

Erfolgt die Einbindung von Expertinnen und Experten in Form von Expertengruppen, so werden neben der Festlegung der persönlichen Voraussetzungen für die Tätigkeit als Expertin oder als Experte auch Auswahlkriterien benötigt, die eine für die Fragestellung sachgerechte Zusammensetzung der Gruppe sicherstellen sollen (Pohontsch et al. 2018). Da die Einbindung von Expertengruppen die Funktion hat, bei der Beratung eines Themas verschiedene Perspektiven und Ideen zusammenzutragen und zu diskutieren (siehe Abschnitt 10.3), strebt das IQTIG bei der Zusammensetzung von Expertengruppen keine Repräsentativität für eine bestimmte Zielgruppe an, sondern in der Regel eine möglichst heterogene Gruppenzusammensetzung (Häder 2000: 4 f., 7 f., Mullen 2003: 42). Bei Beratungen zu Fragen der Gesundheitsversorgung und Qualitätssicherung werden die Expertengruppen daher soweit möglich und in Abhängigkeit vom Thema insbesondere hinsichtlich folgender Aspekte strukturiert:

- Versorgungssektor
- Einrichtungsart und -größe sowie Versorgungsangebot
- räumliches Versorgungsumfeld (z. B. ländlich, urban; ggf. Bundesland)
- Rolle hinsichtlich des Verfahrens (z. B. Leistungserbringer, Patientin/Patient, Qualitätsmanagerin/Qualitätsmanager, Wissenschaftlerin/Wissenschaftler)
- Beruf und Spezialisierung (z. B. Fachgebiet)
- Geschlecht

Durch die gezielte Strukturierung der Gruppenzusammensetzung soll darüber hinaus vermieden werden, dass die Beratungsergebnisse einseitig sind (Häder 2000: 8 f.). Expertengruppen, die in Verantwortung des IQTIG zur Beratung hinzugezogen werden, besetzt das IQTIG nach den o. g. inhaltlichen Kriterien in einem Bewerbungsverfahren.

Allgemeingültige Mindest- und Höchstzahlen an Mitgliedern für Expertengruppen legt das IQTIG nicht fest. Die Gruppengröße wird abhängig vom Aufgabenbereich so gewählt, dass einerseits möglichst alle wichtigen Perspektiven in die Beratung eingebracht werden können und andererseits noch ein ausreichender Austausch zwischen den Teilnehmern möglich ist (Pohontsch et al. 2018). Darüber hinaus werden praktische Aspekte wie der zeitliche, personelle und finanzielle Aufwand bei den externen Expertinnen und Experten und aufseiten des IQTIG berücksichtigt.

### Bewertungen durch Expertengruppen

Für manche Fragestellungen kann neben der Ermittlung unterschiedlicher Perspektiven auf einen Sachverhalt auch die Bewertung eines Sachverhalts durch die Expertengruppe von Interesse sein. Die Funktion der Diskussion in der Gruppe besteht dabei nicht in der Herstellung eines Konsenses (Pohontsch et al. 2018), da ein solcher das Risiko birgt, nicht durch inhaltliche Übereinstimmung, sondern durch soziale Gruppenprozesse zustande zu kommen (Asch 1955, Fitzpatrick 1989, Sackman 1974: 58, Mullen 2003: 43). Zudem kann eine Expertengruppe mit einer handhabbaren Anzahl von Teilnehmern nicht als repräsentative Stichprobe konstruiert werden, sodass die Bewertung nur die Ansicht der Gruppe wiedergibt (Häder 2000: 7) und stark

von der Gruppenzusammensetzung beeinflusst wird (Campbell et al. 2004, Coulter et al. 1995).<sup>28</sup> Bewertungen durch Expertengruppen dienen vielmehr dem Ziel, die Ansicht der Expertengruppe zu ermitteln und dabei auch fehlende Übereinstimmung zwischen den Gruppenmitgliedern transparent zu machen.

Um gruppendynamisch oder psychologisch bedingte Einflüsse auf das Bewertungsverhalten zu verringern und möglichst nur inhaltlich bedingte Meinungsänderungen der Gruppenmitglieder zu erreichen, bedient sich das IQTIG bei Bedarf vertraulicher Befragungen in den Expertengruppen (Fitzpatrick 1989), beispielsweise in Form eines Delphi-Verfahrens (Übersicht bei Mullen 2003). Da die Informationsgrundlage und der Ablauf des Bewertungsprozesses Einfluss auf das Ergebnis von Bewertungen durch Gruppen haben (Campbell et al. 2003, Murphy et al. 1998, Sackman 1974), werden diese Aspekte bei Bewertungen durch Expertengruppen vom IQTIG transparent gemacht.

---

<sup>28</sup> Repräsentative (Zufalls-)Stichproben können zwar grundsätzlich in Delphi-Befragungen zur Konsensierung zum Einsatz kommen, allerdings nehmen die Teilnehmer in diesem Fall keine Expertenrolle ein und das Vorgehen erlaubt keine Gruppendiskussion (Häder 2000: 10 f.).

# 11 Auswahl von Inhalten bei der Entwicklung von Qualitätsmessungen

Keine Erhebung kann sämtliche Facetten der Qualität der Gesundheitsversorgung in einem Themenbereich gleichzeitig abbilden. Wird das IQTIG mit der Darstellung der Qualität in einem bestimmten Themenbereich (z. B. Herzkatheteruntersuchungen und perkutane Koronarintervention (*percutaneous coronary intervention*, PCI)) beauftragt, muss sich die Entwicklung von Mess- und Steuerungsinstrumenten daher immer auf eine begrenzte Anzahl von Anforderungen an die Versorgung in diesem Bereich beschränken (z. B. Evans et al. 2009). Es muss eine Teilmenge von Anforderungen zielgerichtet und begründet ausgewählt werden, anhand derer die Qualität in dem interessierenden Themenbereich abgebildet und gesteigert werden soll. Bei der Entwicklung von Qualitätssicherungsmaßnahmen legt das IQTIG daher vor dem Hintergrund des Rahmenkonzepts für Qualität (siehe Abschnitt 1.1) im ersten Schritt fest, welche konkreten Inhalte innerhalb eines Themenbereichs für die Qualitätssicherung ausgewählt werden sollen (McGlynn und Asch 1998, Evans et al. 2009, Donabedian 2003: 29 ff.). Die ausgewählten Inhalte bezeichnet das IQTIG in ihrer Summe als Qualitätsmodell für diesen Themenbereich und beschreibt sie anhand von sogenannten Qualitätsaspekten. Diese strukturieren die darauf aufbauende Entwicklung von Qualitätsindikatoren und erlauben im Anschluss einen Abgleich zwischen dem eigentlich interessierenden Konstrukt und dem Qualitätsindikatorensatz (siehe Kapitel 14).

## 11.1 Qualitätsaspekte innerhalb eines Themenbereichs

Unter Qualitätsaspekten versteht das IQTIG spezifische Inhalte eines bestimmten Themenbereichs der Gesundheitsversorgung, anhand derer sich die Qualität der Versorgung beschreiben lässt.<sup>29</sup> Beispielsweise könnten die Indikationsstellung zur Herzkatheteruntersuchung, Information und Aufklärung vor elektiver Prozedur oder die Veränderung der Symptombelastung nach PCI Qualitätsaspekte der Versorgung mit Herzkatheteruntersuchungen und PCI sein. Anders als das Rahmenkonzept für Qualität in der Gesundheitsversorgung, das verschiedene grundlegende Dimensionen der Versorgungsqualität unterscheidet (siehe Abschnitt 1.1), benennen die Qualitätsaspekte konkrete Inhalte in einem bestimmten Themenbereich. Die Qualitätsaspekte für einen Themenbereich entwickelt das IQTIG auf Basis folgender Wissensbestände:

### ▪ Patientenperspektive

Um eine patientenzentrierte Qualitätssicherung zu gewährleisten (siehe Kapitel 3), berücksichtigt das IQTIG grundsätzlich die Patientenperspektive bei der Entwicklung von Qualitätsaspekten. Dazu führt das IQTIG eine gezielte Literaturrecherche durch, bindet Patientinnen und Patienten direkt ein und führt Fokusgruppen mit Patientinnen und Patienten durch (siehe z. B. Tausch und Menold 2015, Kühn und Koschel 2011, Krueger und Casey 2015 für Erläuterungen zu der Methode von Fokusgruppen). So können qualitätsrelevante Themen

---

<sup>29</sup> 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.

oder wichtige Merkmale und Ereignisse einer Behandlung aus Sicht der Patientinnen und Patienten ermittelt werden.

#### ▪ **Perspektive der Versorgenden**

Um das Erfahrungswissen aus der Versorgungspraxis zu berücksichtigen, werden Fokusgruppen mit Vertreterinnen und Vertretern von Gesundheitsprofessionen durchgeführt. Dazu werden Praktikerinnen und Praktiker aus allen beteiligten Sektoren der Versorgung und aus allen relevanten Berufsgruppen rekrutiert, die am alltäglichen Behandlungsgeschehen direkt beteiligt sind.

#### ▪ **wissenschaftliche Literatur**

Neben eigenen Erhebungen nutzt das IQTIG wissenschaftliche Literatur für Erkenntnisse zu Qualitätsthemen aus Perspektive der Patientinnen und Patienten, der Gesundheitsprofessionen oder aus wissenschaftlicher Perspektive. Diese Informationen werden durch orientierende und systematische Literaturrecherchen gewonnen (siehe Kapitel 9).

### Synthese der ermittelten Informationen

Die Informationen aus diesen Wissensbeständen werden vom IQTIG mittels einer qualitativen Synthese zu voneinander abgrenzbaren Teilbereichen, den Qualitätsaspekten, zusammengefasst (siehe Abbildung 11). Dabei werden folgende Punkte beachtet:

- Ein Qualitätsaspekt muss patientenrelevante Themen adressieren (siehe Kapitel 3 und Abschnitt 13.2.1) erfüllen.
- Ein Qualitätsaspekt muss mindestens einer Dimension des Rahmenkonzepts für Qualität inhaltlich zuzuordnen sein.
- Ein Qualitätsaspekt soll so wenig Dimensionen des Rahmenkonzepts inhaltlich zuzuordnen sein wie möglich (angestrebte Überschneidungsfreiheit).
- Ein Qualitätsaspekt muss konkreter sein als die Dimensionen des Rahmenkonzepts.

Durch die Berücksichtigung dieser Punkte soll erreicht werden, dass die Qualitätsaspekte patientenrelevante, hinreichend konkrete und abgrenzbare Inhalte im jeweiligen Themenbereich beschreiben.

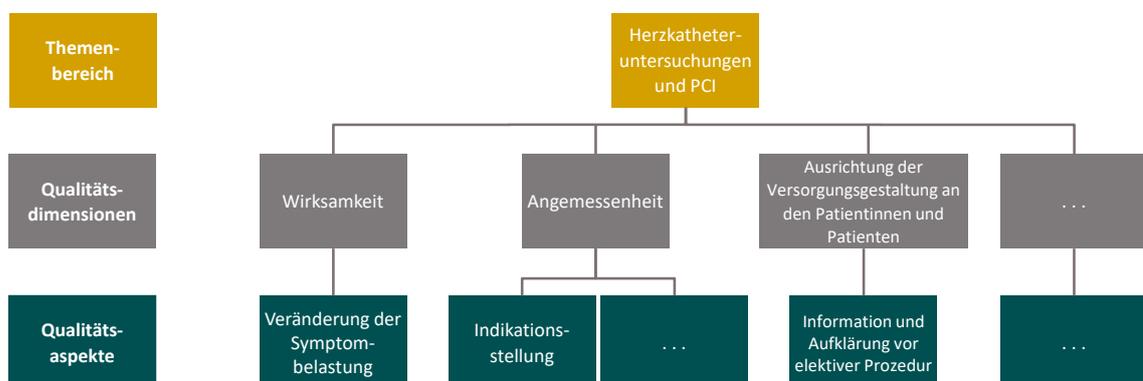


Abbildung 11: Strukturierung eines Themenbereichs mit Qualitätsdimensionen und -aspekten am Beispiel Herzkatheteruntersuchungen und PCI

## Qualitätsmodell

Von den für den Themenbereich identifizierten Qualitätsaspekten empfiehlt das IQTIG diejenigen für die Entwicklung von Qualitätsindikatoren, die folgende Kriterien erfüllen:

- Es gibt Hinweise auf ein Verbesserungspotenzial oder es besteht die Gefahr der Verschlechterung der Versorgungsqualität (Eignungskriterium „Potenzial zur Verbesserung“, siehe Abschnitt 13.2.3).
- Es ist plausibel, dass sich die Versorgungsqualität durch Anstrengungen der Leistungserbringer positiv beeinflussen lässt (Eignungskriterium „Beeinflussbarkeit durch den Leistungserbringer“, siehe Abschnitt 13.2.5).
- Der Qualitätsaspekt liegt im Regelungsbereich des G-BA: Dies bedeutet die Begrenzung auf das SGB V, d. h. beispielsweise den Ausschluss einer Qualitätsmessung von Leistungen der Rehabilitation (SGB IX) und der Pflegeversicherung (SGB XI).

Für die Analyse des Verbesserungspotenzials werden insbesondere folgende Quellen herangezogen:

- 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 Registern und größeren einrichtungsübergreifenden Qualitätsinitiativen und -projekten sowie Versorgungsmodellen können anhand der Verteilung oder des Zeitverlaufs der Ergebnisse Verbesserungspotenziale abgeleitet werden.
- explorative Sozialdatenanalysen
- Fokusgruppen mit Vertreterinnen und Vertretern von Gesundheitsprofessionen und mit Patientinnen und Patienten: Zu vielen Dimensionen der Versorgungsqualität liegen im deutschen Gesundheitswesen meist keine Informationen vor, z. B. Angemessenheit oder Ausrichtung der Versorgungsgestaltung an den Patientinnen und Patienten (vgl. Abschnitt 1.1). Daher wertet das IQTIG dementsprechende Ergebnisse aus Fokusgruppen auch als Hinweise auf Verbesserungspotenziale.

Darüber hinaus prüft das IQTIG, ob alle Qualitätsdimensionen des Rahmenkonzepts durch die Qualitätsaspekte adressiert werden. Falls eine oder mehrere Qualitätsdimensionen nicht adressiert werden, prüft das IQTIG, ob dies nachvollziehbar begründet ist, beispielsweise, weil für deren Erfassung keine Datenquelle zur Verfügung steht oder weil die Beauftragung thematisch enger eingegrenzt ist.

Im Qualitätsmodell werden die für einen Themenbereich relevanten Inhalte definiert und die Grundlage für die zielgerichtete Entwicklung von Qualitätsindikatoren gelegt. Anhand des Abgleichs des Indikatorensets mit dem Qualitätsmodell wird sichtbar, welche Aspekte durch die Qualitätsindikatoren abgebildet werden und bei welchen Aspekten eine Darstellung der Versorgungsqualität derzeit nicht oder nur partiell möglich ist, beispielsweise, weil sich die Inhalte nicht anhand von Indikatoren abbilden lassen. Mit dem Qualitätsmodell werden nicht nur die Inhalte

eines QS-Verfahrens definiert, zugleich wird eine transparente Darstellung der Aussagekraft und Limitationen eines Indikatorensets für einen Themenbereich möglich.

## 11.2 Qualitätsmerkmale als Basis der Indikatorenentwicklung

Bei der Entwicklung von Qualitätsindikatoren werden für jeden in das Qualitätsmodell aufgenommenen Qualitätsaspekt konkrete Merkmale identifiziert, die wichtige Facetten des jeweiligen Aspekts darstellen und die mit konkreten Anforderungen an die Versorgungspraxis (Strukturen, Prozesse, Ergebnisse) verknüpft sind. Diese Merkmale bezeichnet das IQTIG als Qualitätsmerkmale.<sup>30</sup> So ließe sich im oben genannten Beispiel der Qualitätsaspekt „Information und Aufklärung vor elektiver Prozedur“ weiter inhaltlich differenzieren u. a. in die Merkmale „Information über alternative Behandlungsmöglichkeiten“ und „Information über Folgen der Behandlung“. Die Qualitätsmerkmale stellen also weitere inhaltliche Ausdifferenzierungen der Qualitätsaspekte und damit konkrete Eigenschaften der Gesundheitsversorgung dar, anhand derer die Versorgungsqualität bewertet werden kann. Die Anforderungen an die Qualitätsmerkmale werden als Qualitätsziele bezeichnet. Beispielsweise wird mit dem Merkmal „Information über alternative Behandlungsmöglichkeiten“ das Qualitätsziel verbunden, dass alle betreffenden Patientinnen und Patienten des betrachteten Leistungserbringers vor einer PCI über alternative Behandlungsmöglichkeiten informiert werden sollten.

Auf Basis der Qualitätsmerkmale entwickelt das IQTIG Qualitätsindikatoren, 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  $\geq 95\%$ <sup>31</sup>) als konkret zu erfüllende Anforderung bewertet werden. Tabelle 2 verdeutlicht die Unterschiede zwischen den verschiedenen methodischen Elementen der Qualitätsmessung anhand von Beispielen.

Tabelle 2: Fiktive Beispiele für die methodischen Elemente der Qualitätsmessung

Element	Beispiel 1	Beispiel 2
<b>Qualitätsaspekt</b>	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)
<b>Qualitätsmerkmal</b>	Information über alternative Behandlungsmöglichkeiten	Sterblichkeit im Krankenhaus bei Pneumonie

<sup>30</sup> Dies ist konsistent 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).

<sup>31</sup> Diese Prozentangabe ist ein fiktiver Wert und dient nur der Erläuterung des Beispiels.

Element	Beispiel 1	Beispiel 2
<b>Qualitätsanforderung (Qualitätsziel)</b>	Information aller Patientinnen und Patienten über alternative Behandlungsmöglichkeiten	möglichst geringe Sterblichkeit von Patientinnen und Patienten mit Pneumonie
<b>Qualitätsindikator</b>	Anteil von Patientinnen und Patienten, die über alternative Behandlungsmöglichkeiten informiert wurden	Anteil von Patientinnen und Patienten mit Pneumonie, die im Krankenhaus verstorben sind
<b>Referenzbereich</b>	Anteil von Patientinnen und Patienten, die über alternative Behandlungsmöglichkeiten informiert wurden, ist $\geq 95\%$	Anteil von Patientinnen und Patienten mit Pneumonie, die im Krankenhaus verstorben sind, ist $\leq 10\%$ .

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

## 12 Beteiligungsverfahren gemäß § 137a Abs. 7 SGB V

Gemäß § 137a Abs. 3 SGB V soll das IQTIG insbesondere mit der Entwicklung und Durchführung von Maßnahmen der Qualitätssicherung einschließlich der Entwicklung von Instrumenten zur Messung und Darstellung der Versorgungsqualität vom G-BA beauftragt werden (siehe Abschnitt 2.2). Bei der Entwicklung solcher Inhalte sind gemäß § 137a Abs. 7 SGB V vom IQTIG folgende Organisationen zu beteiligen:

- die Kassenärztliche Bundesvereinigung
- die Kessenzahnärztliche Bundesvereinigung
- die Deutsche Krankenhausgesellschaft
- der Spitzenverband Bund der Krankenkassen
- der Verband der Privaten Krankenversicherung
- die Bundesärztekammer
- die Bundeszahnärztekammer
- 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

Für die Berufsorganisationen der Krankenpflegeberufe, die wissenschaftlichen medizinischen Fachgesellschaften sowie für die Bundesoberbehörden im Geschäftsbereich des Bundesministeriums für Gesundheit ist nicht festgelegt, welche konkreten Organisationen zu beteiligen sind. Daher legt das IQTIG dies nach eigenem Ermessen fest.

Laut der Gesetzesbegründung zu dem Gesetz zur Weiterentwicklung der Finanzstruktur und der Qualität in der gesetzlichen Krankenversicherung (GKV-Finanzstruktur- und Qualitätsweiterentwicklungsgesetz, GKV-FQWG, BR-Drucksache 151/14 vom 11.04.2014 2014) soll mit der Pflicht zur Beteiligung gewährleistet werden, dass das IQTIG bei der Bearbeitung seiner Aufträge den Sachverstand der genannten, fachlich betroffenen Organisationen und Institutionen einbezieht. Als Beteiligung ist hierbei die Einbeziehung in wesentliche Entwicklungsphasen der Auftragsbearbeitung zu verstehen (Roters 2020). Das konkrete Verfahren zur Durchführung der Beteiligung ist allerdings nicht im § 137a Abs. 7 SGB V beschrieben. Aus der Gesetzesbegründung geht hervor, dass das IQTIG bei der Ausgestaltung des Beteiligungsverfahrens frei ist (BT-

Drucksache 18/1307 vom 05.05.2014: 36). Dabei ist die Mitwirkung der zu Beteiligten an Arbeits- und Expertengruppen während der Entwicklungsphase, aber auch die Durchführung schriftlicher Stellungnahmeverfahren während oder am Ende der Entwicklungsphase möglich (Roters 2020).

### **Umsetzung bei der Entwicklung indikatorbasierter Qualitätssicherungsmaßnahmen**

Das IQTIG setzt seine Pflicht zur Beteiligung der in § 137a Abs. 7 SGB V erwähnten Organisationen und Institutionen bei der Entwicklung indikatorbasierter Qualitätssicherungsmaßnahmen in der Regel als schriftliches Stellungnahmeverfahren um. Ziel dieses Vorgehens ist, eine möglichst breite Beurteilung der zu diesem Zeitpunkt vorliegenden Entwicklungsergebnisse einzuholen. Die Wahl des Zeitpunkts für das Beteiligungsverfahren steht dem IQTIG frei und wird so gewählt, dass die Sachkenntnis der zu Beteiligten zielführend in die fachliche Entwicklungsarbeit eingebunden werden kann (siehe Kapitel 6 und 7 für die konkreten Zeitpunkte für die Beteiligungsverfahren).

Den zu Beteiligten wird nach vorheriger Ankündigung ein Bericht übermittelt, zu dem in der Regel innerhalb von 6 Wochen eine schriftliche Stellungnahme eingereicht werden kann. Die eingegangenen Stellungnahmen werden durch das IQTIG aufbereitet und die darin vorgebrachten Argumente fachlich geprüft. Ergibt sich daraus begründeter Anpassungsbedarf an den Entwicklungsergebnissen, wird dieser vorgenommen. Mit Veröffentlichung des jeweiligen Berichts werden auch die eingegangenen Stellungnahmen sowie deren schriftliche Würdigung durch das IQTIG veröffentlicht.

### **Umsetzung bei anderen Entwicklungen**

Das IQTIG kann neben der Entwicklung indikatorbasierter Qualitätssicherungsmaßnahmen auch mit der Entwicklung anderer Qualitätssicherungsmaßnahmen und Inhalte beauftragt werden. Für solche Inhalte prüft das IQTIG, welche Form der Beteiligung am besten geeignet ist, die Sachkenntnis der Organisationen nach § 137a Abs. 7 SGB V zielführend in den Entwicklungsprozess einzubinden, wie etwa schriftliche Stellungnahmeverfahren, Workshops oder Anhörungsverfahren.

### **Anpassung bestehender Qualitätssicherungsmaßnahmen**

Qualitätssicherungsmaßnahmen, mit deren Durchführung das IQTIG gemäß § 137a Abs. 3 SGB V beauftragt ist, werden regelmäßig durch das IQTIG auf Anpassungsbedarf geprüft (siehe Kapitel 8). Die für die Durchführung der Maßnahmen nötigen technischen Anpassungen, wie etwa Überarbeitungen der Spezifikation für die QS-Software oder Anpassungen von OPS- und ICD-Kodes stellen dabei keine inhaltlichen Entwicklungen im Sinne von § 137a Abs. 3 SGB V dar. Daher führt das IQTIG für solche Anpassungen keine Beteiligungsverfahren durch. Für über technische Anpassungen hinausgehende inhaltliche Anpassungen an Qualitätssicherungsmaßnahmen, wie die Entwicklung neuer Qualitätsindikatoren, wird dagegen in der Regel ein Beteiligungsverfahren durchgeführt.

## 13 Eignung von Qualitätsindikatoren

In diesem Kapitel werden die grundsätzlichen Anforderungen beschrieben, die Messinstrumente in der Gesundheitsversorgung erfüllen müssen, damit sie sich als Grundlage für qualitätsorientierte Entscheidungen eignen.

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 bzgl. 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). 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).

Da kein allgemein anerkannter Standard für die Beurteilung der Eignung von Qualitätsindikatoren existiert (Rode et al. 2016, Geraedts et al. 2017: e103), wurde unter Berücksichtigung der wissenschaftlichen Fachliteratur eine eigene Systematik von Anforderungen an Qualitätsmessungen entwickelt.

### 13.1 Grundsätze zur Eignungsbeurteilung

Messungen der Versorgungsqualität werden in verschiedenen Kontexten und als Grundlage für unterschiedliche Maßnahmen zur Qualitätssteuerung eingesetzt (siehe Kapitel 2). Die Eignung einer Qualitätsmessung für eine bestimmte Verwendung bestimmt sich danach, inwieweit die aus ihr gezogenen Schlussfolgerungen und Entscheidungen (z. B. über den Einsatz von Steuerungsinstrumenten) angemessen sind. Dies wird anhand der Eignungskriterien beschrieben.

Das IQTIG beurteilt alle Messinstrumente einheitlich mit denselben Eignungskriterien unabhängig vom auf die Qualitätsmessung folgenden Handlungsanschluss. Inwieweit die Kriterien durch ein Messinstrument tatsächlich erfüllt sind, hängt dabei nicht nur von der Konstruktion des Messinstruments (Qualitätsmerkmal, Rechenregeln des Indikators etc.), sondern auch von der konkreten Datengrundlage, der betrachteten Population, den Rahmenbedingungen der Qualitätsmessung und -bewertung sowie dem Handlungsanschluss ab (siehe Abschnitt 2.4). Die Eig-

nung eines Instruments zur Qualitätsmessung und -bewertung ist folglich keine inhärente Eigenschaft des Instruments, sondern muss im Kontext ihrer Verwendung beurteilt werden (Meyer 2004: 24, Kane 2016, McGlynn 2003). Qualitätsmessungen sollen diese Eignungskriterien grundsätzlich in möglichst hohem Maße erfüllen, damit die aus ihnen abgeleiteten Schlussfolgerungen und Entscheidungen gerechtfertigt sind.

Das IQTIG gliedert die Eignungskriterien für Qualitätsmessungen anhand der drei Komponenten eines Qualitätsindikators (Qualitätsziel, Messverfahren, Bewertungskonzept; siehe Abschnitt 4.1; siehe Abbildung 12).

Eignung		
des Qualitätsziels	des Messverfahrens	des Bewertungskonzepts
Bedeutung für die Patientinnen und Patienten	Objektivität der Messung	Angemessenheit des Referenzbereichs
Zusammenhang mit patientenrelevantem Merkmal	Datenqualität	Klassifikationsgüte
Potenzial zur Verbesserung	Reliabilität der Messung	Angemessenheit der Risikoadjustierung
Brauchbarkeit für den Handlungsanschluss	Validität der Messung	
Beeinflussbarkeit	Praktikabilität der Messung	

Abbildung 12: Überblick über die Eignungskriterien für Qualitätsindikatoren

Die Anwendung der Eignungskriterien erfolgt auf verschiedene Weisen: Zum einen leiten die Eignungskriterien die Entwicklung neuer Qualitätsmessungen (prospektive Anwendung), zum anderen werden sie zur Beurteilung bestehender Qualitätsmessungen herangezogen (retrospektive Anwendung).

### Prospektive Anwendung

Die Entwicklung von Qualitätsmessungen erfolgt in mehreren Schritten von den Qualitätsaspekten über die Qualitätsmerkmale bis hin zu den Qualitätsindikatoren. Die Eignungskriterien leiten dabei insbesondere die Auswahl und den Zuschnitt der Qualitätsmerkmale sowie die Operationalisierung der Indikatoren (siehe Kapitel 6 und 7). Beispielsweise lässt sich aus dem Kriterium „Brauchbarkeit für den Handlungsanschluss“ bei Entwicklung von Indikatoren für Auswahlentscheidungen von Patientinnen und Patienten ableiten, dass die abzubildenden Qualitätsmerk-

male auf elektive Versorgungsanlässe zugeschnitten werden. Ein anderes Beispiel ist das Kriterium „Objektivität der Messung“, aus dem abgeleitet wird, wie Datenfelder für QS-Dokumentationen sachgerecht konstruiert werden.

### **Retrospektive Anwendung**

Retrospektiv werden die Eignungskriterien für die Beurteilung bestehender Indikatoren eingesetzt. Diese erfolgt z. B. bei der Überprüfung von Indikatoren am Ende der Erprobungsphase eines neuen QS-Verfahrens, bei der Anpassung von sich im Gebrauch befindlichen Indikatoren (siehe Kapitel 8) oder bei der Prüfung, ob sich bestehende Indikatoren für zusätzliche Handlungsanschlüsse einsetzen lassen. Beispielsweise macht das Kriterium „Brauchbarkeit für den Handlungsanschluss“ bei Anwendung auf einen Qualitätsindikator zur „Door-to-balloon-Zeit bei akutem Herzinfarkt“ sichtbar, dass dieser Indikator zwar für Maßnahmen der Qualitätsförderung, aber kaum für Auswahlentscheidungen von Patientinnen und Patienten geeignet ist, da bei Herzinfarkt meist keine Auswahlentscheidungen durch die Patientinnen und Patienten möglich sind. Zur Beurteilung der Eignung werden dabei nicht nur einzelne Qualitätsindikatoren, sondern auch deren Auswahl, Zusammenspiel und Verwendung im Set betrachtet (siehe Kapitel 14).

### **Einschätzung der Kriterienerfüllung**

Die Einschätzung hinsichtlich der Erfüllung jedes Eignungskriteriums durch einen Qualitätsindikator erfolgt durch das IQTIG mittels Einordnung, Beurteilung und Zusammenschau der verfügbaren Informationen wie Literatur, eigene Erhebungen und externe Expertise. Sie erfordert für die meisten Eignungskriterien eine Beurteilung komplexer Sachverhalte. Beispielsweise kann die Angemessenheit der Risikoadjustierung eines Qualitätsindikators nicht allein anhand eines statistischen Maßes betrachtet werden, sondern es muss auch beurteilt werden, ob alle wichtigen Einflussfaktoren auf die Indikatorergebnisse berücksichtigt wurden, ob deren Einfluss fachlich plausibel ist und ob ein Einflussfaktor der Verantwortung des Leistungserbringers zuzurechnen ist oder nicht.<sup>32</sup> Die Einschätzung, inwieweit ein Kriterium erfüllt ist, erfolgt daher in der Regel durch eine qualitative Beurteilung und wird unter Abwägung der positiven Belege und Hinweise gegenüber möglichen gegenteiligen Hinweisen inhaltlich begründet und dargestellt.

### **Gesamtbeurteilung der Eignung**

Die Gesamtbeurteilung der Indikatoreignung – auch im Vergleich zu anderen möglichen Indikatoren – nimmt das IQTIG in Form einer Abwägung der Erfüllung der Einzelkriterien vor, die gemeinsam ein Indikatorprofil darstellen. Die Entscheidung für einen Indikator wird auf Basis der Gesamtschau der Vor- und Nachteile, die für oder gegen den Indikator sprechen, getroffen und lässt sich nicht sinnvoll durch einen festen Algorithmus vornehmen (Meyer 2004: 40). Transparenz über die Eignung von Indikatoren entsteht somit durch die nachvollziehbare Darlegung der Gründe, weshalb das IQTIG einen Indikator für geeignet hält.

---

<sup>32</sup> Beispielsweise kann es auch Faktoren geben, die zwar nicht direkt vom Leistungserbringer zu beeinflussen sind (z. B. kognitive Beeinträchtigung oder sprachliche Verständnisschwierigkeiten von Patientinnen und Patienten), die aber nicht in eine Risikoadjustierung für Qualitätsmessungen eingehen, weil es in der Verantwortung des Leistungserbringers liegt, die Auswirkung dieser Faktoren zu kompensieren (z. B. bei einem Qualitätsmerkmal zur Information und Aufklärung zur Behandlung).

Im Folgenden werden die Eignungskriterien als grundlegende Anforderungen an Qualitätsmessungen beschrieben. Das konkrete Vorgehen bei der Anwendung dieser Kriterien erfolgt je nach Zeitpunkt (bei der Entwicklung von Qualitätsmessungen vs. für bestehende Qualitätsindikatoren) unterschiedlich und ist nicht Gegenstand dieses Kapitels.

## 13.2 Eignungskriterien des Qualitätsziels

Qualitätsmessungen eignen sich nur dann als Grundlage für Qualitätssicherungsmaßnahmen und Entscheidungen, wenn die betrachteten Merkmale der Versorgung tatsächlich auch Qualität widerspiegeln, für die Patientinnen und Patienten relevant sind und die Messergebnisse für die vorgesehenen Qualitätssicherungsmaßnahmen und Entscheidungen genutzt werden können. Vereinfacht ausgedrückt muss beurteilt werden, ob die Messung für die richtigen Qualitätsziele erfolgt. Diese Beurteilung wird in fünf Eignungskriterien untergliedert.

### 13.2.1 Bedeutung für die Patientinnen und Patienten

Damit sich Qualitätsmessungen für eine patientenzentrierte Qualitätssicherung eignen, müssen sie Qualitätsaspekte und Qualitätsmerkmale abbilden, die im Sinne des Rahmenkonzepts für Qualität relevant für die Patientinnen und Patienten sind (siehe Kapitel 1 und 3), sich also einer Qualitätsdimension des Rahmenkonzepts zuordnen lassen. Qualitätsaspekte und -merkmale sowie die mit ihnen verbundenen Qualitätsziele können unmittelbar oder mittelbar für Patientinnen und Patienten relevant sein. Qualitätsmerkmale werden als unmittelbar patientenrelevant bezeichnet, wenn sie *per se* einen Wert für die Patientinnen und Patienten darstellen und nicht nur Mittel zum Zweck sind. Sie basieren also auf einer sozialen Werteübereinkunft und umfassen gesundheitsbezogene Versorgungsergebnisse sowie Merkmale der Versorgungsgestaltung (Kersting et al. 2020, Scholl et al. 2014). Beispiele hierfür sind:

- Überleben bzw. verringerte Sterblichkeit
- verminderte Krankheitsbeschwerden und -folgen
- keine negativen gesundheitlichen Folgen einer Behandlung
- gesundheitsbezogene Lebensqualität
- Behandlung der Patientinnen und Patienten mit Respekt
- emotionale Unterstützung
- Information der Patientinnen und Patienten
- Partizipation der Patientinnen und Patienten

Demgegenüber werden mittelbar patientenrelevante Qualitätsmerkmale erst dadurch für Patientinnen und Patienten relevant, dass sie in einem Zusammenhang mit einem unmittelbar patientenrelevanten Qualitätsmerkmal stehen. So macht beispielsweise eine Antibiotikaphylaxe bei Operationen Infektionen weniger wahrscheinlich. Die Antibiotikaphylaxe selbst hat jedoch keine unmittelbare Bedeutung für die Patientinnen und Patienten, sie ist Mittel zum Zweck der Vermeidung von Infektionen. Dieser Zusammenhang zwischen einem mittelbar relevanten Qualitätsmerkmal und einem unmittelbar relevanten Qualitätsmerkmal wird daher für solche Merkmale als *zusätzliches* Eignungskriterium durch das IQTIG geprüft (siehe Abschnitt 13.2.2).

Merkmale aller drei Messdimensionen (Strukturen, Prozesse, Ergebnisse) können unmittelbar oder mittelbar patientenrelevante Qualitätsmerkmale sein (siehe Beispiele in Tabelle 3).

Tabelle 3: Beispiele für unmittelbar und mittelbar patientenrelevante Qualitätsmerkmale

	<b>unmittelbar patientenrelevantes Qualitätsmerkmal</b>	<b>mittelbar patientenrelevantes Qualitätsmerkmal</b>
<b>Strukturen</b>	barrierefreier Zugang	Desinfektionsmittelpender in Patientenzimmer
<b>Prozesse</b>	partizipative Entscheidungsfindung	Antibiotikaphylaxe bei Operation
<b>Ergebnisse</b>	niedrige Sterblichkeit	CRP-Verringerung bei Rheumabehandlung

Welche konkreten Aspekte Bedeutung für Patientinnen und Patienten haben, muss in jedem Themenbereich gezielt ermittelt werden (Mühlhauser und Müller 2009: 36 f.). Das IQTIG nutzt hierfür als Informationsgrundlage insbesondere Literatur, eigene qualitative Erhebungen wie Fokusgruppen mit Patientinnen und Patienten sowie mit Vertreterinnen und Vertretern von Gesundheitsprofessionen und ergänzend Hinweise von Expertinnen und Experten einschließlich der Patientenvertretung (siehe Abschnitte 6.1 und 7.2).

Das Ausmaß der Bedeutung, die ein unmittelbar patientenrelevantes Qualitätsmerkmal für die Patientinnen und Patienten hat, hängt beispielsweise von Ausprägung und Dauer einer Erkrankung ab, von der Schwere einer Komplikation oder vom Ausmaß der Verkürzung oder Verlängerung von Lebensdauer oder Krankheitssymptomen. Das Ausmaß der Bedeutung eines mittelbar relevanten Qualitätsmerkmals wird sowohl von der Bedeutung des zugrunde liegenden unmittelbar relevanten Qualitätsmerkmals als auch von der Stärke des Zusammenhangs mit diesem bestimmt (vgl. Porzsolt und Gaus 1993). Beispielsweise hängt die Bedeutung einer Antibiotikaphylaxe bei Operation davon ab, wie schwer die möglichen Infektionskomplikationen der Operation sind und mit welcher Wahrscheinlichkeit ihr Auftreten durch die Antibiotikaphylaxe vermieden wird.

Die Beurteilung der Bedeutung für die Patientinnen und Patienten erfolgt bei verschiedenen Entwicklungsschritten. Beispielsweise wird bei der Entwicklung eines Qualitätsmodells (siehe Kapitel 11) untersucht, welche übergeordneten Themen im Versorgungsgeschehen Bedeutung für die Patientinnen und Patienten haben und als Qualitätsaspekte benannt werden können. Bei der Entwicklung eines Qualitätsindikators muss die Bedeutung für Patientinnen und Patienten für das vom Indikator abgebildete Qualitätsmerkmal (z. B. „Anteil von Patientinnen und Patienten mit Wundinfektionen“) gegeben sein.

### 13.2.2 Zusammenhang mit einem unmittelbar patientenrelevanten Merkmal

Mittelbar patientenrelevante Qualitätsmerkmale sind Ergebnisse, Prozesse oder Strukturen, denen zwar kein Wert an sich zugeschrieben wird, deren Vorliegen bzw. deren Durchführung aber

mit einer höheren bzw. niedrigeren Wahrscheinlichkeit des Auftretens unmittelbar patientenrelevanter Ergebnisse einhergeht (siehe Abschnitt 13.2.1). Im Sinn einer patientenzentrierten Qualitätssicherung sieht das IQTIG mittelbare Qualitätsmerkmale nur dann als geeignet für die Qualitätsmessung an, wenn der Zusammenhang mit einem unmittelbar patientenrelevanten Merkmal ausreichend sicher ist.<sup>33</sup> So kann beispielsweise das Qualitätsmerkmal, bestimmte Werte in einem Laborparameter zu erreichen (mittelbar patientenrelevantes Merkmal), mit einer verminderten Sterblichkeit einhergehen (unmittelbar patientenrelevantes Merkmal). Um diesen Zusammenhang zwischen mittelbar und unmittelbar patientenrelevanten Qualitätsmerkmalen zu belegen, erfolgt bei der Entwicklung von Qualitätsmessungen eine systematische Literaturrecherche und -bewertung (siehe Abschnitt 9.2). Falls keine ausreichende externe Evidenz für diesen Zusammenhang vorliegt, kann auf eine konsentrierte Einschätzung von Expertinnen und Experten zurückgegriffen werden. Ein fester Grenzwert, ab welcher Qualität und Konsistenz der recherchierten Evidenz der Zusammenhang mit einem unmittelbar patientenrelevanten Qualitätsmerkmal als ausreichend belegt anzusehen ist, wird nicht definiert (siehe Grundsätze zur Eignungsbeurteilung in Abschnitt 13.1).

### 13.2.3 Potenzial zur Verbesserung

Die Wichtigkeit einer Qualitätsmessung bemisst sich neben der Bedeutung für die Patientinnen und Patienten auch nach ihrem möglichen Nutzen. Dazu muss die Messung eines bestimmten Qualitätsmerkmals entscheidungsrelevante Informationen liefern, die der Sicherstellung und Verbesserung der Versorgungsqualität dienen (McGlynn 2003, Turpin et al. 1996) (siehe Abschnitt 2.5).

Eine Voraussetzung für qualitätsorientierte Entscheidungen ist das Vorliegen von Qualitätsunterschieden oder -defiziten in der Versorgung (McGlynn 2003, Mainz et al. 2003) oder ein nachvollziehbares Risiko für Qualitätsunterschiede, dem durch eine Überwachung im Sinne eines Qualitätsmonitorings begegnet werden kann (Berwick 1991, NQF 2015: 45). So kann es Qualitätsmerkmale geben, die zwar eine hohe Bedeutung für die Patientinnen und Patienten haben, aber deren Messung keine entscheidungsrelevante Information bietet, da hohe Versorgungsqualität über alle Leistungserbringer hinweg vorliegt.<sup>34</sup> Aus Aufwand-Nutzen-Gründen vermeidet das IQTIG die Einführung oder Beibehaltung von Indikatoren, deren Nutzen begrenzt ist oder deren Ergebnisse das Versorgungsziel erreicht haben (Meyer 2004: 39, NQF 2015: 44 ff., Reeves et al. 2010). Es sollen also Qualitätsmerkmale gemessen werden, bei denen ein Potenzial für positive Effekte im Sinne einer Qualitätsverbesserung oder einer Vermeidung von Qualitätsdefiziten besteht.

Die Beurteilung des Verbesserungspotenzials erfolgt anhand einer Analyse der Versorgungssituation, beispielsweise auf Grundlage von Literatur, Ergebnissen aus Fokusgruppen, einer Beratung durch Expertinnen und Experten oder eigener Datenanalysen des IQTIG (z. B. von QS-

---

<sup>33</sup> Für Struktur- und Prozessindikatoren entspricht dieses Eignungskriterium dem im QUALIFY-Instrument als Indikatorevidenz bezeichneten Kriterium (Reiter et al. 2007).

<sup>34</sup> 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.

Dokumentationsdaten oder Sozialdaten bei den Krankenkassen). Anhaltspunkte für Qualitätsunterschiede und -defizite und damit einhergehendes Verbesserungspotenzial ergeben sich z. B. bei großen Unterschieden in der Versorgung oder durch Berichte über Fehlentwicklungen oder Berichte zu Über-, Unter- oder Fehlversorgung. Für die Feststellung eines Verbesserungspotenzials muss nicht die gesamte betrachtete Population von Patientinnen und Patienten oder von Leistungserbringern gleichmäßig betroffen sein. Verbesserungspotenzial kann auch dann vorliegen, wenn einzelne Subgruppen besonders betroffen sind, beispielsweise bei regionalen Unterschieden oder bei Defiziten in der Versorgung bestimmter Patientengruppen. Ist ein Verbesserungspotenzial gegeben, so kann dessen Ausmaß beispielsweise anhand der Zahl betroffener Patientinnen und Patienten (McGlynn und Asch 1998, NQF 2015: 45), anhand der finanziellen Mittel für nicht indizierte Behandlungen, die dann für andere Maßnahmen nicht zur Verfügung stehen, oder anhand der Zahl der Leistungserbringer mit Verbesserungsbedarf (NQF 2015: 45) näher eingeschätzt werden.

#### **13.2.4 Brauchbarkeit für den Handlungsanschluss**

Aus den Ergebnissen von Qualitätsmessungen müssen die Adressaten der Messergebnisse entsprechende Entscheidungen ableiten können. Mit diesem Eignungskriterium soll sichergestellt werden, dass sich Formulierung und Auswahl der Qualitätsmerkmale an den Informationsbedarfen für die jeweilige anschließende Maßnahme zur Qualitätsverbesserung orientieren (siehe Abschnitt 2.4). So sollen für Steuerungsinstrumente, denen der Wirkmechanismus „Qualitätsförderung“ zugrunde liegt, möglichst solche Qualitätsmerkmale gemessen werden, die den Leistungserbringern Ansatzpunkte für Maßnahmen zur Verbesserung der Versorgungsqualität liefern (McGlynn 2003, Turpin et al. 1996). Bei Qualitätsmessungen für Steuerungsinstrumente, die über Auswahlentscheidungen von Prinzipalen wirken sollen, sind die Informationsbedarfe und die Handlungsmöglichkeiten der jeweiligen Prinzipale zu berücksichtigen (siehe Abschnitt 2.4.3) (Birkmeyer et al. 2006: 194 f.). Beispielsweise sind Auswahlentscheidungen von Patientinnen und Patienten in erster Linie bei elektiven Behandlungen, weniger aber bei Notfallbehandlungen möglich. Das Qualitätsmerkmal „Door-to-balloon-Zeit bei akutem Herzinfarkt“ ist für Auswahlentscheidungen durch Patientinnen und Patienten mit akutem Herzinfarkt kaum hilfreich, da wegen der Dringlichkeit der Behandlung hier meist kein sinnvoller Entscheidungsspielraum besteht.

Die Brauchbarkeit von Qualitätsmessungen hängt auch davon ab, ob die Messergebnisse früh genug für den vorgesehenen Handlungsanschluss zur Verfügung stehen. Beispielsweise muss für einen Qualitätsindikator zur Überlebensrate von Patientinnen und Patienten festgelegt werden, auf welchen Zeitpunkt er sich beziehen soll (z. B. In-Hospital-Sterblichkeit oder 1-Jahres-Sterblichkeit) und aus welcher Datenquelle die Informationen ermittelt werden. Der Indikator würde dementsprechend je nach Wahl des Zeitpunkts und der benötigten Datenquelle bei der Formulierung des Qualitätsmerkmals eine unterschiedliche Brauchbarkeit aufweisen.

Maßnahmen der Qualitätssicherung zielen in der Regel nicht auf einzelne, isolierte Qualitätsmerkmale, sondern auf Qualitätsverbesserungen in einem Themenbereich. Die Beurteilung der Brauchbarkeit für einen Handlungsanschluss erfolgt daher nicht nur für einzelne Indikatoren, sondern auch in Zusammenschau der Indikatoren eines Indikatorensets.

### 13.2.5 Beeinflussbarkeit durch den Leistungserbringer

In der externen Qualitätssicherung werden Qualitätsmerkmale und Anforderungen an die Leistungserbringer formuliert und ihre Erfüllung mittels Messinstrumenten wie Qualitätsindikatoren dargestellt (siehe Kapitel 11). Einem oder mehreren Leistungserbringern wird also die Verantwortung für die Ergebnisse der Qualitätsmessung zugeschrieben. Damit diese Zuschreibung angemessen ist, muss die Erfüllung der Anforderungen auch innerhalb der Handlungsmöglichkeiten des Leistungserbringers liegen, also die Ausprägung des gemessenen Merkmals maßgeblich von ihm beeinflussbar sein (NQF 2016: 25, McGlynn und Asch 1998).

Bei der Abbildung von Qualitätsmerkmalen mittels Prozess- und Ergebnisindikatoren werden die Indikatorwerte nicht nur durch die Leistungserbringer selbst beeinflusst, sondern potenziell auch von patientenseitigen und sonstigen (z. B. systembedingten oder strukturellen) Faktoren sowie von Wechselwirkungen zwischen diesen. Das Ausmaß, in dem die Leistungserbringer die Indikorausprägung beeinflussen können, ist daher 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 13). Andere Prozess- und Ergebnisindikatoren (z. B. Eingehen auf das Informationsbedürfnis der Patientinnen und Patienten, Tod bei elektiver Operation ohne Vorerkrankungen) können dagegen so spezifisch formuliert werden, dass patientenseitige und sonstige Einflüsse weitgehend 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.

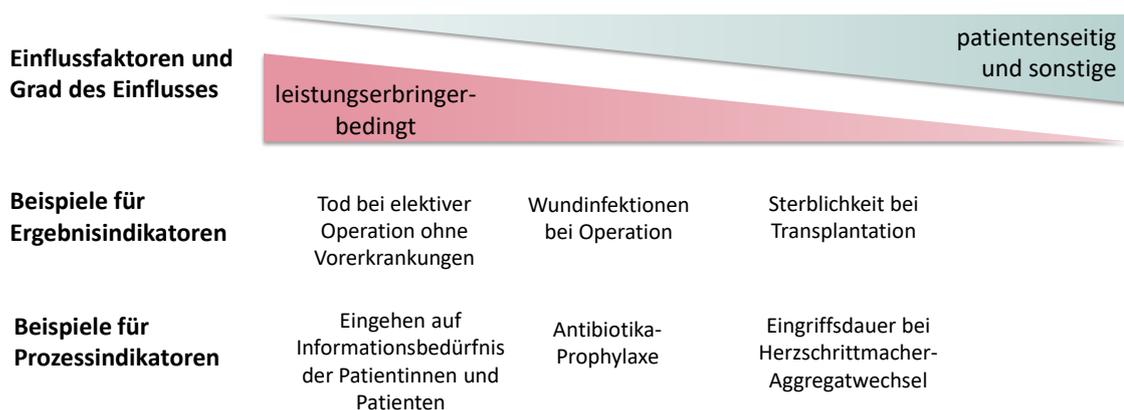


Abbildung 13: Schematische Darstellung des Ausmaßes der leistungserbringerbedingten und nicht leistungserbringerbedingten Einflüsse auf den Indikatorwert anhand von Beispielen fallbasierter Qualitätsindikatoren

Das Eignungskriterium beschreibt also, in welchem Maße das Indikatorergebnis durch denjenigen Leistungserbringer, dem das Indikatorergebnis zugeschrieben wird, in Richtung des Qualitätsziels beeinflusst werden kann.

Am Zustandekommen eines Messergebnisses können mehrere Leistungserbringer beteiligt sein, insbesondere, wenn an der Patientenversorgung mehrere Fachdisziplinen oder Berufsgruppen beteiligt sind oder wenn lange Beobachtungszeiträume vorliegen (Langzeit-Follow-up). Beispielsweise kann ein Ergebnisindikator „Häufigkeit von Abstoßungsreaktionen einer Transplantatnieren“ gemessen werden, dessen Indikatorwert sowohl von der Qualität des Krankenhauses (Nierentransplantationszentrum) als auch der Qualität der ambulanten Betreuung (z. B. niedergelassene Nephrologin oder niedergelassener Nephrologe) abhängt.

Für jede Qualitätsmessung wird untersucht, welchen Anteil jeder der beteiligten Leistungserbringer am Zustandekommen der Messergebnisse hat. Dabei kann der jeweilige Anteil der Beeinflussbarkeit nicht quantifiziert werden, sondern erfordert eine qualitative Abschätzung auf Grundlage der Informationen zum Versorgungsgeschehen, insbesondere aus Literatur, Fokusgruppen sowie aus der Beratung durch Expertinnen und Experten. Sollen die auf die Qualitätsmessung folgenden Qualitätssicherungsmaßnahmen nur auf einzelne, konkret benennbare Leistungserbringer angewendet werden<sup>35</sup>, so muss die Beeinflussung des Messergebnisses im Wesentlichen diesem Leistungserbringer zugeschrieben werden können. Es können grundsätzlich (bei Schaffung entsprechender rechtlicher Grundlagen) aber auch Qualitätsmessungen mit auf mehrere Leistungserbringer verteilter Verantwortung für die Versorgung eingesetzt werden, wenn eine Gruppe von beteiligten Leistungserbringern gemeinschaftlicher Adressat von Qualitätssicherungsmaßnahmen sein soll (z. B. Accountable Care Organizations, McClellan et al. 2014) oder wenn Versorgungsqualität auf regionaler Ebene dargestellt und gefördert werden soll.

Neben der Einschätzung, ob ein Qualitätsaspekt vom Leistungserbringer grundsätzlich beeinflussbar ist (z. B. bei der Erstellung eines Qualitätsmodells), muss die Beeinflussbarkeit auch anhand des konkreten Messinstruments geprüft werden, da sie von der Operationalisierung abhängt, beispielsweise von der konkreten Berechnungsvorschrift eines Indikators.

---

<sup>35</sup> In den Verfahren der externen Qualitätssicherung werden bisher die Messergebnisse jeweils einem einzelnen Leistungserbringer zugeschrieben.

### 13.3 Eignungskriterien des Messverfahrens

Unter den Eignungskriterien des Messverfahrens werden Anforderungen an Qualitätsmessungen mit Indikatoren zusammengefasst, die die messtheoretischen Eigenschaften und die Praktikabilität des Indikators beschreiben. Nur wenn die Operationalisierung des Qualitätsmerkmals adäquat ist und eine belastbare und nachvollziehbare Informationsgrundlage zur Verfügung stellt, können anhand der Indikatorergebnisse sinnvoll Entscheidungen getroffen werden.

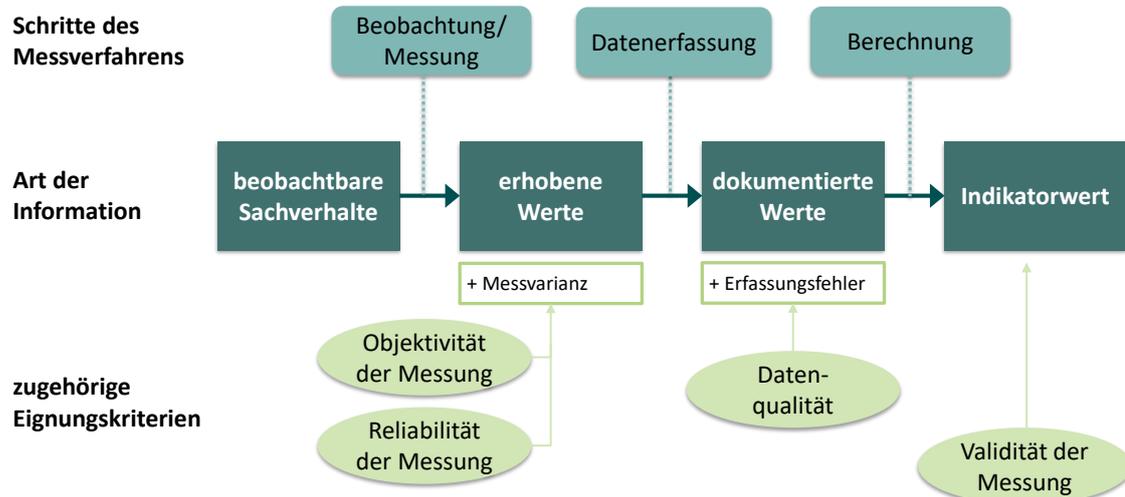


Abbildung 14: Schritte und zugehörige Eignungskriterien bei der Ermittlung und Aufarbeitung von Informationen für die Berechnung von Indikatorwerten

Bei der Qualitätsmessung mittels Indikatoren stellen die Gewinnung der Informationen und Berechnung von Indikatorwerten eine Kette von drei aufeinanderfolgenden Schritten dar (Abbildung 14). Der erste Schritt besteht in der Beobachtung eines Sachverhalts und Erhebung eines Wertes für diesen Sachverhalt auf Ebene der Patientenversorgung, 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 der Qualitätsmessung mittels Prozess- oder Ergebnisindikatoren erfolgt dieser Schritt auf Patienten- bzw. Fallebene, bei Strukturindikatoren auf Ebene des Leistungserbringers. In einem zweiten Schritt werden die primär für die Patientenversorgung erhobenen Daten für den Zweck der Qualitätssicherung dokumentiert und übermittelt (Datenerfassung).<sup>36</sup> Bei Patientenbefragungen für die Qualitätssicherung fallen Messung auf Versorgungsebene und Datenerfassung zusammen, d. h., beides erfolgt gleichzeitig beim Ausfüllen des Fragebogens. Der dritte Schritt besteht schließlich in der Berechnung des Indikatorwerts auf Ebene des Leistungserbringers gemäß den Festlegungen zu den verwendeten Datenfeldern, Dateninhalten, Rechenregeln und Risikoadjustierungsmethoden.

<sup>36</sup> Bei der Nutzung von Abrechnungsdaten für die Qualitätsmessung besteht die Datenerfassung ihrerseits aus zwei Schritten: der Dokumentation und Übermittlung zu Abrechnungszwecken und der Übermittlung von den Krankenkassen an die Auswertungsstellen für die Qualitätssicherung.

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.

### **13.3.1 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, deren Durchführungsobjektivität z. B. durch Interpretationsspielraum bei Datenfeldern beeinträchtigt sein kann.

Um die Objektivität der Daten für die Qualitätsmessung weitestgehend sicherzustellen, werden solche Sachverhalte zur Erfassung ausgewählt, für die eine weitgehend objektive Messung angenommen werden kann (siehe auch Abschnitt 7.1). Darüber hinaus wird die Objektivität der Messung durch bestimmte Maßnahmen der Erfassungs- bzw. Fragebogengestaltung gesteigert. Für Qualitätsindikatoren, die auf der Dokumentation der Leistungserbringer basieren, geschieht dies durch eindeutige Spezifikationen zum Erhebungsprozess einschließlich präziser und verständlicher Datenfelder und ggf. Ausfüllhinweise (siehe Abschnitt 6.2.1). Die Durchführungsobjektivität von Patientenbefragungen (vgl. Moosbrugger und Kelava 2012a) wird durch standardisierte Befragung mittels Fragebogen – in Abgrenzung zu beispielsweise mündlichen Befragungen – und durch Beachtung der wissenschaftlichen Prinzipien der Itemformulierung sichergestellt (siehe Abschnitt 7.2.5).

### 13.3.2 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) 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 maßgeblich in der Verantwortung der Dokumentierenden liegt und erst in laufenden QS-Verfahren adäquat empirisch beurteilt werden kann.

### 13.3.3 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). Dem 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 dieser 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 zurückgehen müssen, sondern auch durch Messfehler bedingt sein können.<sup>37</sup> 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). Die Reliabilität ist also ein Maß für die Präzision der Messung bei sonst gleichen Untersuchungsumständen.

---

<sup>37</sup> Dies ist nicht zu verwechseln mit einer fehlenden Übereinstimmung von dem für die Qualitätssicherung dokumentierten Wert mit dem Wert aus einer Patientenakte. Dies ist kein Reliabilitätsproblem im Sinne eines zufälligen Messfehlers, sondern ein Dokumentationsproblem, das durch das Eignungskriterium „Datenqualität“ adressiert wird.

Aufgrund der Mehrebenenstruktur der Qualitätsmessung (siehe Einleitung zu Abschnitt 13.3) kann Reliabilität zum einen auf Fallebene und zum anderen auf Leistungserbringerebene betrachtet werden. Auf Fallebene sind Einschränkungen der Reliabilität von den zu messenden Sachverhalten abhängig und könnten beispielsweise durch unterschiedliche Interpretationen von Befragungssitems durch die Patientinnen und Patienten und durch die Leistungserbringer zustande kommen. Auch die Messung direkt beobachtbarer Merkmale einzelner Patientinnen und Patienten kann mit Messfehlern einhergehen, beispielsweise, wenn bei der sonografischen Bestimmung der Organgröße dieselbe untersuchende Person bei wiederholten Messungen an derselben Patientin oder demselben Patienten unterschiedliche Messwerte bestimmt (Test-Retest-Reliabilität).

Darüber hinaus kann die Reliabilität der Messung auf Ebene der Leistungserbringer adressiert werden. Bei Prozess- und Ergebnisindikatoren, für deren Berechnung Daten über die Behandlungsfälle eines Leistungserbringers aggregiert werden, hängt die Variabilität der Daten eines Leistungserbringers dabei von der Reliabilität der Messung auf Fallebene ab: Je geringer diese ist, desto größer wird die Variabilität der Daten eines Leistungserbringers und desto geringer ist die Präzision der Schätzung des zugrunde liegenden Indikatorwerts des Leistungserbringers. Dagegen steigt die Präzision der Schätzung des zugrunde liegenden Indikatorwerts eines Leistungserbringers mit zunehmender Stichprobengröße, d. h. mit steigender Fallzahl des Leistungserbringers. Aufgrund dieses Aggregationsvorteils sind, im Gegensatz zum individualdiagnostischen Einsatz, sinnvolle Aussagen auf Ebene der Leistungserbringer auch schon bei niedrigerer Reliabilität auf Fallebene möglich (s. a. Frost et al. 2007).

Die Reliabilität der Messung auf Ebene der Leistungserbringer kann beispielsweise durch die Breite des Vertrauensbereichs um die Indikatorwerte der Leistungserbringer quantifiziert werden: Der Vertrauensbereich wird umso breiter, je größer die Variabilität der Daten innerhalb eines Leistungserbringers ist, und umso schmaler, je größer die Fallzahl des Leistungserbringers ist. Bei der Entwicklung von Qualitätsmessungen wird eine möglichst hohe Reliabilität dadurch erzielt, dass möglichst eindeutige und objektive Sachverhalte (siehe Abschnitt 13.3.1) herangezogen werden.

#### **13.3.4 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 transparent zu ma-

chen. Validität der Messung 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 Qualitätsmerkmal abbildet, das durch die Indikatorbezeichnung beschrieben wird.

Die Validität der Messung direkt beobachtbarer Sachverhalte wird dadurch bestimmt, dass die gewählte Indikatorbezeichnung, die ausgewählten Datenfelder bzw. Fragebogenitems und die Rechenregeln im Sinne des interessierenden Qualitätsmerkmals spezifiziert sind. 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. Dasselbe gilt für faktenorientierte Items von Patientenbefragungen (siehe Abschnitt 7.1), beispielsweise „Wurde Ihnen erklärt, weshalb Sie die Ihnen verschriebenen Medikamente einnehmen sollen?“ oder „Wie viele Treppenstufen können Sie nach der Behandlung wieder problemlos steigen?“.

Dieser Sachverhalt wird auch als Inhaltsvalidität bezeichnet und in erster Linie mittels sachlogischer Überlegungen eingeschätzt (Hartig et al. 2012: 148 ff.). 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 bzw. am besten geeigneten Befragungsisems für die Berechnung zu ermitteln und Besonderheiten bestimmter Daten (z. B. wegen Kodierung zu Abrechnungszwecken) zu berücksichtigen.

### 13.3.5 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 der Messung, sondern auch Aufwände, die bei der erstmaligen oder wiederholten Einrichtung der Messung für einen Indikator anfallen, z. B. für die Spezifikation neuer Datenfelder, für die Einrichtung von Datenflüssen bei Nutzung neuer Datenquellen, für die Aktualisierung der Software für die Erfassung oder für die Information der an der Datenerfassung Beteiligten. In die Beurteilung der Praktikabilität fließen auch die Verfügbarkeit der Informationen für einen Indikator und der benötigte zeitliche Aufwand für deren Bereitstellung ein. Die Praktikabilität der Messung ist damit keine Messeigenschaft eines Qualitätsindikators im eigentlichen Sinn, ist aber ein wichtiger Faktor bei der Beurteilung, wie gut das Messverfahren dieses Indikators in der Anwendung geeignet ist und ob für die vorgesehenen Maßnahmen zur Qualitätssteuerung (siehe Abschnitt 2.5) ein angemessenes Verhältnis von Aufwand und Nutzen besteht.

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 Einleitung zu Abschnitt 13.3 und Abbildung 14). 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 z. B. 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 in der Fragebogenlogistik (z. B. Versand der Fragebögen durch die durchführende Stelle) und bei den Patientinnen und Patienten. Die Zumutbarkeit der Fragebögen für die Patientinnen und Patienten in zeitlicher, psychischer und physischer Hinsicht ist ein wichtiger Teilaspekt der Praktikabilität und wird dadurch befördert, dass die Fragen nur die für den konkreten Qualitätsaspekt relevanten Ereignisse berücksichtigen und keine darüber hinaus.

### **13.4 Eignungskriterien des Bewertungskonzepts**

Um auf Grundlage von Qualitätsmessungen Entscheidungen zu treffen, muss nicht nur das Messverfahren objektiv, reliabel und valide sein, sondern auch das Bewertungskonzept, mit dem Messergebnisse interpretiert und in Beziehung zu einer Entscheidung gesetzt werden (Kane 2016). Auswertungs- und Interpretationsobjektivität (Moosbrugger und Kelava 2012a: 9 f.) werden bei Messungen mit Qualitätsindikatoren dadurch gewährleistet, dass explizite, standardisierte Bewertungskriterien in Form von Referenzbereich und zugehöriger Klassifikationsregel verwendet werden, die unabhängig von den beteiligten Personen auf die Messergebnisse angewendet werden (siehe Kapitel 16). Sie sind daher hier nicht als separate Eignungskriterien aufgeführt. Die Eignung des Bewertungskonzepts eines Indikators im Rahmen eines QS-Verfahrens lässt sich zusammenfassend anhand der drei nachfolgend dargestellten Kriterien beschreiben.

#### **13.4.1 Angemessenheit des Referenzbereichs**

Der Referenzbereich dient bei Qualitätsbewertungen mit Indikatoren als Kriterium, ob eine bestimmte Entscheidung (z. B. hinsichtlich der Anwendung einer Qualitätssicherungsmaßnahme) getroffen wird.<sup>38</sup> Die angemessene Interpretation von Indikatorergebnissen hängt somit maßgeblich von der Wahl des Referenzbereichs ab, mit dem der Indikatorwert verglichen wird (Goodwin 1996). Da der Einsatz unterschiedlicher Instrumente zur Qualitätssicherung (siehe Ab-

---

<sup>38</sup> Werden qualitätsorientierte Entscheidungen auf Basis eines Index aus mehreren Qualitätsindikatoren getroffen, sind ggf. zusätzliche methodische Anforderungen zu beachten.

schnitt 2.5) an unterschiedlich hohe Anforderungen geknüpft werden kann, können auch unterschiedliche Referenzbereiche für einen Indikator angemessen sein. Beispielsweise ist es sinnvoll, für Entscheidungen über Maßnahmen der Qualitätsförderung den Indikatorwert eines Leistungserbringers anders zu bewerten als hinsichtlich der Frage, ob eine Mindestanforderung erfüllt ist, die Voraussetzung zur Teilnahme des Leistungserbringers an der Versorgung ist. Die Angemessenheit eines Referenzbereichs beschreibt also, ob der Referenzbereich ein für die Entscheidung passendes Anforderungsniveau operationalisiert. Dies kann als ein Aspekt der Validität des Indikators hinsichtlich seiner Konsequenzen verstanden werden (Kane 2016, Goodwin 1996).

Zur Einschätzung der Angemessenheit gehört eine Beurteilung, inwieweit die Anforderung, die durch einen Referenzbereich an den Leistungserbringer gestellt wird, erreichbar ist. Dazu kann beispielsweise untersucht werden, wie viele Leistungserbringer die Anforderung voraussichtlich erreichen (Norcini und Shea 1997: 44, Berk 1996: 230) und ob es für den jeweiligen Handlungsanschluss plausibel ist, wenn die Anforderung von sehr vielen oder sehr wenigen Leistungserbringern erreicht wird (Doran 2014). Stünde schon vor der Messung fest, dass keines der Ergebnisse oder alle Ergebnisse im Referenzbereich liegen würden, würde sich die Klassifikation mittels des Indikators erübrigen. Beispielsweise ist ein Bewertungskonzept für einen Indikator „In-Hospital-Sterblichkeit bei Lebertransplantation“ mit einer oberen Referenzbereichsgrenze von 1 % nicht für die Entscheidung über Maßnahmen der Qualitätsförderung geeignet, da dieser Wert in der Regel auch bei optimaler Qualität nicht erreicht wird. Somit könnte mit diesem Referenzbereich nicht entschieden werden, welche Leistungserbringer eine Qualitätsförderung erhalten sollen und welche nicht. Details zu den Methoden zur Festlegung von Bewertungskonzepten und Referenzbereichen sind in Kapitel 16 dargestellt.

#### 13.4.2 Klassifikationsgüte

Bei Prozess- und Ergebnisindikatoren, welche durch Aggregation über die Fälle eines Leistungserbringers berechnet und mit einer analytischen Zielsetzung ausgewertet werden, werden anhand der beobachteten Daten Rückschlüsse auf den zugrunde liegenden Indikatorwert eines Leistungserbringers gezogen. Da diese Schätzung mit Unsicherheit behaftet ist (z. B. wegen der Fallzahlabhängigkeit der Schätzgenauigkeit), wird die Eignung des Bewertungskonzepts auch von der Klassifikationsgüte bestimmt (siehe Abschnitt 20.1.4). Die Klassifikationsgüte beschreibt, 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, und kann beispielsweise anhand der Maße Sensitivität und Spezifität beschrieben werden. Dies entspricht dem Konzept der Teststärkeanalyse (vgl. Döring und Bortz 2016: 807 ff.).<sup>39</sup>

---

<sup>39</sup> Dies ist von anderen Verständnissen von Klassifikationsgüte, Sensitivität und Spezifität abzugrenzen, die diese Begriffe für den Vergleich von Indikatorergebnissen mit einem Außenkriterium nutzen (z. B. bei QUALIFY, Reiter et al. 2007) und damit nicht Teststärke, sondern Kriteriumsvalidität beschreiben. Ein solches Außenkriterium ist aber nicht verfügbar, da ein Goldstandard zur Messung 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

Je nach Verwendung der Qualitätsmessung können für den gemessenen Indikatorwert unterschiedliche Vergleichswerte von Interesse sein. Beispielsweise kann ein Vergleich mit dem Referenzbereich des Indikators oder mit den Indikatorwerten anderer Leistungserbringer vorgenommen werden. In der externen Qualitätssicherung soll in der Regel entschieden werden, ob der zugrunde liegende Indikatorwert des Leistungserbringers außerhalb des Referenzbereichs liegt oder nicht. Die Festlegung eines bestimmten Signifikanzniveaus, ab dem eine statistisch signifikante Abweichung vom Referenzbereich festgestellt wird, geht dabei mit einer Abwägung von Sensitivität und Spezifität einher. Die Klassifikationsgüte hängt in diesem Fall zum einen von den beobachteten Daten, insbesondere den Fallzahlen der Leistungserbringer, und zum anderen von der Wahl von Referenzbereich, Klassifikationsregel und Signifikanzniveau ab (siehe Abschnitt 20.1.4).

### **13.4.3 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 mit unterschiedlich zusammengesetzten Patientenkollektiven fair ist, müssen Einflussfaktoren auf den Indikatorwert, die die Leistungserbringer nicht durch eine gute Versorgung kompensieren können, bei der Bewertung der Messergebnisse berücksichtigt werden. Dies geschieht mittels Methoden der Risikoadjustierung (siehe Abschnitt 20.2). Das Kriterium beschreibt somit das Ausmaß, in dem patientenseitige Einflussfaktoren auf die Indikatorwerte durch Risikoadjustierung kontrolliert sind.

---

stattdessen zwischen den verschiedenen Gesichtspunkten, die Einfluss auf die Klassifikation nehmen, anhand der Eignungskriterien.

## 14 Eignung von Qualitätsindikatorensets

Ein einzelner Qualitätsindikator bildet einen kleinen Ausschnitt der Versorgung ab. Oft ist aber nicht nur ein kleiner Ausschnitt (z. B. die Vermeidung postoperativer Wundinfektionen), sondern die Qualität der Versorgung in einem größeren Themenbereich (z. B. elektive Hüftendoprothesenversorgung) von Interesse. Um Aussagen über die Versorgungsqualität in einem Themenbereich treffen zu können, sind mehrere Indikatoren nötig, die zu einem Qualitätsindikatorenset zusammengeführt werden. Die Eignung eines Qualitätsindikatorensets hängt nicht allein von der Güte der einzelnen Indikatoren (siehe Kapitel 13 „Eignung von Qualitätsindikatoren“), sondern auch von der Zusammensetzung des Sets ab (Döbler und Geraedts 2018, Döbler et al. 2019, NHS Institute for Innovation and Improvement/APHO [2007]). Bislang existieren keine konsentierten Kriterien für die Eignung von Qualitätsindikatorensets (Döbler und Geraedts 2018). In der wissenschaftlichen Literatur werden vor allem zwei Aspekte diskutiert (Übersicht bei Schang et al. 2021): zum einen das fachlich-inhaltliche Kriterium der Güte von Schlussfolgerungen auf Basis des Sets; dieses wird typischerweise als Inhaltsvalidität bezeichnet. Zum anderen werden die Aufwände thematisiert, die mit der Entwicklung, Erhebung und Auswertung des Indikatorensets verbunden sind. In diesem Kapitel wird erläutert, wie das IQTIG diese beiden Aspekte bei der Entwicklung neuer Indikatorensets und bei der Beurteilung bestehender Indikatorensets berücksichtigt.

### 14.1 Definition „Qualitätsindikatorenset“

Eine einheitliche Definition des Begriffs „Qualitätsindikatorenset“ steht in der wissenschaftlichen Literatur noch aus. Das IQTIG definiert ein Qualitätsindikatorenset als Summe aller Qualitätsindikatoren, die Versorgungsqualität in einem gemeinsamen Themenbereich erfassen sollen. Der interessierende Themenbereich kann breit (z. B. Palliativversorgung (Leemans et al. 2013)) oder enger (z. B. Indikationsstellung bei Kniegelenk-Arthroskopie) gefasst sein. Die Ergebnisse der Qualitätsindikatoren eines Sets können nebeneinander dargestellt oder für bestimmte Verwendungszwecke (z. B. qualitätsorientierte Auswahlentscheidungen) zu einem Index aggregiert werden.

### 14.2 Eignungskriterium Inhaltsvalidität

Sollen Schlussfolgerungen über die Versorgungsqualität in einem Themenbereich gezogen werden, muss das Indikatorenset diesen Themenbereich angemessen abbilden. Die Inhaltsvalidität eines Sets beschreibt dies als das Ausmaß, in dem die Auswahl der Indikatoren relevant und repräsentativ für die Versorgungsqualität im Themenbereich und die externe Qualitätssicherung ist (vgl. Haynes et al. 1995).

Sie lässt sich anhand dreier Subkriterien qualitativ bewerten (American Educational Research Association et al. 2014, Haynes et al. 1995, Sireci und Sukin 2013, Schang et al. 2021):

- **Abdeckung:** Ausmaß, in dem die relevanten Inhalte des Themenbereichs durch das Qualitätsindikatorenset abgebildet werden.
- **Proportionalität:** Ausmaß, in dem die Qualitätsindikatoren des Sets in einem – entsprechend ihrer Bedeutung für die Versorgungsqualität im Themenbereich – ausgewogenen Verhältnis zueinander stehen.
- **Vermeidung von Kontamination:** Ausmaß, in dem keine für den Themenbereich irrelevanten Inhalte durch das Qualitätsindikatorenset abgebildet werden.

Zur Illustration ist in Abbildung 15 ein Qualitätsindikatorenset mit hoher Inhaltsvalidität dargestellt. In diesem Beispiel werden alle Inhalte, die zuvor als relevant identifiziert wurden, durch Indikatoren abgebildet; die Indikatoren stehen in einem ausgewogenen Verhältnis zueinander und es sind keine irrelevanten Indikatoren enthalten.

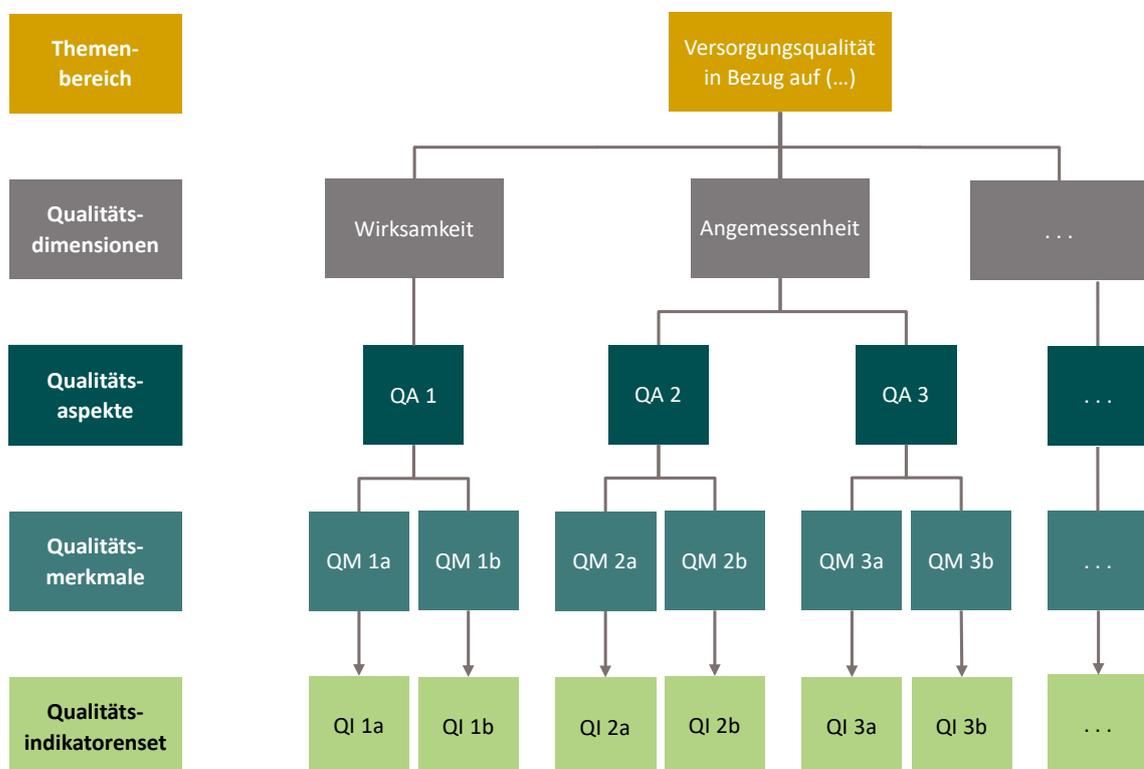


Abbildung 15: Strukturierung der Versorgungsqualität im interessierenden Themenbereich mit Qualitätsdimensionen, Qualitätsaspekten und Qualitätsmerkmalen und Operationalisierung durch das Qualitätsindikatorenset. Die Qualitätsindikatoren spiegeln die definierten relevanten Inhalte im Themenbereich (Qualitätsdimensionen des Rahmenkonzepts, Qualitätsaspekte und Qualitätsmerkmale) wider. Das Qualitätsindikatorenset erlaubt somit aussagekräftige Schlussfolgerungen über die Versorgungsqualität im Themenbereich.

Die Anwendung des Eignungskriteriums kann zu unterschiedlichen Zeitpunkten erfolgen. Prospektiv leitet es die Entwicklung des Qualitätsindikatorensets – von den Qualitätsaspekten über

die Qualitätsmerkmale bis hin zur Operationalisierung und Zusammenführung der Qualitätsindikatoren. Die Beurteilung der Inhaltsvalidität kann aber auch retrospektiv für bereits bestehende Qualitätsindikatorensets erfolgen.

#### 14.2.1 Herstellung von Inhaltsvalidität

Das IQTIG setzt folgende Schritte gemäß der wissenschaftlichen Literatur bei der Entwicklung von Qualitätsindikatorensets um, um eine möglichst hohe Inhaltsvalidität zu erzielen (z. B. American Educational Research Association et al. 2014, Cronbach 1971, Haynes et al. 1995, Schang et al. 2021):

- **Definition relevanter Inhalte und theoretisches Modell:** Um ein inhaltsvalides Indikatorenset entwickeln zu können, müssen zunächst die qualitätsrelevanten Inhalte im Themenbereich, die abgebildet werden sollen, definiert und abgegrenzt werden. Die begründete Auswahl der relevanten Inhalte im Themenbereich erfolgt auf Basis der wissenschaftlichen Literatur, der Perspektive der Patientinnen und Patienten, der klinischen Perspektive der Versorgenden, weiterer Expertinnen und Experten sowie von Sozialdatenanalysen und orientiert sich am Rahmenkonzept für Qualität (siehe Abschnitt 1.1). Das resultierende theoretische Modell für die Versorgungsqualität im Themenbereich bezeichnet das IQTIG als Qualitätsmodell (siehe Kapitel 11 für die Beschreibung der zugrunde liegenden Methodik).
- **Entwicklung entlang des theoretischen Modells:** Um eine hohe Inhaltsvalidität in Bezug auf das Qualitätsmodell zu erzielen, entwickelt das IQTIG Qualitätsindikatoren entlang der Qualitätsaspekte des Qualitätsmodells (siehe Kapitel 6 und 7).
- **Abgleich mit theoretischem Modell:** Während der Entwicklung der Qualitätsindikatoren erfolgt wiederholt der Abgleich mit dem Qualitätsmodell im Hinblick auf die drei Subkriterien der Inhaltsvalidität (s. o.). Gegebenenfalls werden Qualitätsindikatoren modifiziert, gestrichen oder ergänzt.

Durch dieses Vorgehen soll eine hohe Inhaltsvalidität des Sets und Transparenz darüber hergestellt werden, inwieweit sich die qualitätsrelevanten Inhalte im Themenbereich durch die Instrumente der gesetzlichen Qualitätssicherung abbilden lassen.

#### 14.2.2 Retrospektive Einschätzung

Für die retrospektive Einschätzung der Inhaltsvalidität eines Qualitätsindikatorensets sind insbesondere folgende Schritte bedeutsam:

- **Definition relevanter Inhalte und theoretisches Modell:** Um die Inhaltsvalidität eines Qualitätsindikatorensets retrospektiv prüfen zu können, muss ebenfalls definiert sein, welche Inhalte relevant für die Versorgungsqualität im Themenbereich sind und abgebildet sein sollten. Liegt für den Themenbereich kein Qualitätsmodell vor, greift das IQTIG orientierend auf sein Rahmenkonzept für Qualität zurück (siehe Abschnitt 1.1).
- **Zuordnung zum theoretischen Modell und Abgleich:** Zur retrospektiven Prüfung der Inhaltsvalidität eines bestehenden Qualitätsindikatorensets erfolgt eine Zuordnung und ein Abgleich mit den Qualitätsaspekten des Qualitätsmodells bzw. orientierend mit den Qualitätsdimensi-

onen des Rahmenkonzepts. Anhand der drei Subkriterien der Inhaltsvalidität (s. o.) wird beurteilt, inwieweit das Qualitätsindikatorenset die Qualitätsaspekte des Qualitätsmodells bzw. die Qualitätsdimensionen des Rahmenkonzepts angemessen abbildet. Gegebenenfalls wird eine Weiterentwicklung des Indikatorensets empfohlen.

### **14.3    Aufwände der Messung**

Das zentrale inhaltliche Kriterium für die Eignung eines Qualitätsindikatorensets ist die Güte der Schlussfolgerungen, die auf der Basis des Indikatorensets gezogen werden. Auf der anderen Seite stehen die Aufwände, die mit der Entwicklung, Erhebung und Auswertung eines Indikatorensets verbunden sind (Schang et al. 2021). Für diese beiden Aspekte kann ein Trade-off erforderlich sein, da eine höhere Inhaltsvalidität mit höheren Aufwänden einhergehen kann. Die Aufwände bei der Erhebung hängen für die Leistungserbringer jedoch auch von der gewählten Datenquelle ab; werden Sozialdaten oder Befragungen von Patientinnen und Patienten als Datenquelle verwendet, führt ein größeres Set nicht oder nur in geringem Ausmaß zu einem Mehraufwand für die Leistungserbringer. Diese Aspekte werden bei der Entwicklung von Indikatorensets und bei der Beurteilung der Eignung durch das IQTIG berücksichtigt.

## 15 Veröffentlichung von leistungserbringerbezogenen Indikator- und Kennzahlergebnissen

Eine gesetzlich verankerte Kernaufgabe des IQTIG ist die Schaffung von Transparenz über die Qualität der Gesundheitsversorgung. Dahinter steht der Wille des Gesetzgebers, Qualitätsinformationen für die Öffentlichkeit zugänglich zu machen und so Transparenz über die Qualität der Gesundheitsversorgung herzustellen (siehe Begründung zum GKV-FQWG, BR-Drucksache 151/14 vom 11.04.2014: 25). Dieses Kapitel bezieht sich auf die Schaffung von Transparenz durch Veröffentlichung der Indikator- und Kennzahlergebnisse der Leistungserbringer.

Gemäß §§ 136 ff. SGB V erfolgt eine Veröffentlichung von Indikatorergebnissen in den Qualitätsberichten der Krankenhäuser nach den Regelungen des G-BA (Regelungen zum Qualitätsbericht der Krankenhäuser, Qb-R)<sup>40</sup>. Zu diesem Zweck spricht das IQTIG jährlich eine Empfehlung an den G-BA aus, für welche Qualitätsindikatoren und für welche Kennzahlen die Leistungserbringerergebnisse veröffentlicht werden sollen. Die Grundsätze für die Empfehlungen zur Veröffentlichung sind im Folgenden beschrieben.

### Transparenz über die Qualität der Gesundheitsversorgung

Transparenz über die Qualität der Gesundheitsversorgung gilt als ein gesellschaftlicher Wert an sich, da die Leistungserbringer im Gesundheitswesen damit ihrer 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 IQTIG 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 Schaffung von Transparenz in diesem Sinne unterscheidet das IQTIG von der gezielten Bereitstellung leistungserbringerbezogener Qualitätsinformationen als Steuerungsinstrument zur Steigerung der Versorgungsqualität (z. B. durch die Ermöglichung von qualitätsorientierten Auswahlentscheidungen zwischen Leistungserbringern, siehe Abschnitt 2.4).<sup>41</sup> Die Entwicklung solcher Steuerungsinstrumente und von Kriterien für ihren Einsatz sind nicht Gegenstand dieses Kapitels.

---

<sup>40</sup> Regelungen des Gemeinsamen Bundesausschusses gemäß § 137 Abs. 3 Satz 1 Nr. 4 SGB V über Inhalt, Umfang und Datenformat eines strukturierten Qualitätsberichts für nach § 108 SGB V zugelassene Krankenhäuser In der Neufassung vom 16. Mai 2013, zuletzt geändert am 16. Dezember 2021, in Kraft getreten am 15. Februar 2022. URL: <https://www.g-ba.de/richtlinien/39/> (abgerufen am 21.03.2022).

<sup>41</sup> Eine bloße Veröffentlichung von leistungserbringerbezogenen Qualitätsinformationen ist für diese Zwecke nicht ausreichend, sondern dies erfordert zusätzlich eine adressatengerechte Aufbereitung der Qualitätsinformationen, beispielsweise durch Aggregation, Visualisierung und adressatengerechte Sprache.

### **Kriterien zur Veröffentlichung von Qualitätsindikatorergebnissen**

Für die Veröffentlichung von Informationen über die Versorgungsqualität aus Transparenzgründen ist aus Sicht des IQTIG die einzige fachliche Voraussetzung, dass diese Informationen tatsächlich Qualität abbilden. Dies wird anhand der Eignungskriterien für Qualitätsmessungen beurteilt (siehe Kapitel 13). Demnach sieht das IQTIG die Ergebnisse aller Qualitätsindikatoren, die als geeignet für die gesetzlich verpflichtende Qualitätssicherung beurteilt wurden, auch als geeignet für die Veröffentlichung an. Das IQTIG empfiehlt daher, grundsätzlich die Ergebnisse aller Qualitätsindikatoren leistungserbringerbezogen zu veröffentlichen, die in der gesetzlich verpflichtenden Qualitätssicherung eingesetzt werden, sofern nicht bei sehr geringen Fallzahlen der Datenschutz einzelner Patientinnen oder Patienten dem entgegensteht.

### **Veröffentlichung der Ergebnisse neuer Qualitätsindikatoren**

Für neue Qualitätsindikatoren wird die Eignung durch die Berücksichtigung der Eignungskriterien (siehe Kapitel 13) im Entwicklungsprozess sichergestellt. Teilweise können die Eignungskriterien jedoch erst anhand der Erfahrung aus den ersten Jahren ihres Einsatzes abschließend beurteilt werden. Das Gleiche trifft für Indikatoren zu, die derart weitgehenden Änderungen unterzogen wurden, dass Auswirkungen auf die Aussagekraft des Indikators nicht ausgeschlossen werden können. In beiden Fällen werden ihre Ergebnisse vorerst nicht zur Veröffentlichung empfohlen. Für diese Indikatoren wird in den Folgejahren wiederkehrend geprüft, ob die Eignungskriterien hinreichend sicher als erfüllt gelten und die Indikatorergebnisse damit auch zur Veröffentlichung empfohlen werden können.

### **Veröffentlichung von weiteren Kennzahlergebnissen**

Neben den Qualitätsindikatoren werden in der externen Qualitätssicherung weitere Kennzahlen verwendet, die Versorgungsaspekte beschreiben, die aber – weil für sie kein Referenzbereich festgelegt ist – zur Bewertung von Versorgungsqualität nicht herangezogen werden. Gleichwohl kann es sinnvoll sein, bestimmte Kennzahlen im Zusammenhang mit Indikatorergebnissen zu veröffentlichen, wenn sie z. B. die Interpretation von Indikatorergebnissen unterstützen oder per se für die Öffentlichkeit im Zusammenhang mit der Versorgungsqualität einer Einrichtung von Interesse sind. Daher empfiehlt das IQTIG ggf. auch weitere Kennzahlergebnisse zusätzlich zu den Indikatorergebnissen zur Veröffentlichung.

## 16 Bewertungskonzepte und Referenzbereiche

Entscheidungen auf Grundlage von Qualitätsmessungen, z. B. über den Einsatz von Maßnahmen zur Qualitätsverbesserung (Steuerungsinstrumente, siehe Abschnitt 2.4), benötigen Vergleichsmaßstäbe, anhand derer die Messergebnisse den möglichen Handlungsoptionen zugeordnet werden. Diese Vergleichsmaßstäbe und die Regeln zu ihrer Anwendung werden als Bewertungskonzept für eine Qualitätsmessung bezeichnet. In diesem Kapitel wird beschrieben, wie Bewertungskonzepte für Qualitätsmessungen mittels Qualitätsindikatoren festgelegt werden.

### 16.1 Funktion von Referenzbereichen

Um mittels Qualitätsindikatoren qualitätsbezogene Entscheidungen zu treffen, beispielsweise, ob Fördermaßnahmen eingeleitet werden sollten, werden ein oder mehrere Vergleichswerte auf der Indikatorkala als Bewertungsmaßstab benötigt (JCAHO 1990: 85 ff., Goodwin 1996, Goldhammer und Hartig 2012: 175, Kane 2017: 16). Die Vergleichswerte grenzen Bereiche auf der Indikatorkala voneinander ab, die bestimmten Handlungsoptionen zugeordnet sind, und die als *Referenzbereich* des Indikators für die jeweilige Handlungsoption bezeichnet werden. Der Referenzbereich stellt die konkreten Anforderungen (das Soll) in Bezug auf ein operationalisiertes Qualitätsmerkmal dar und erlaubt damit einen Soll-Ist-Abgleich zwischen Indikatorergebnissen und konkreter Anforderung. Da der Einsatz unterschiedlicher Steuerungsinstrumente der Qualitätssicherung an unterschiedlich hohe Anforderungen geknüpft werden kann, können für einen Indikator auch mehrere Referenzbereiche definiert werden.

Indikatorergebnis und Referenzbereich werden über ein statistisches Verfahren miteinander verglichen, das damit ebenfalls Teil des Bewertungskonzepts ist. Dadurch wird berücksichtigt, dass bei Qualitätsmessungen das beobachtete Indikatorergebnis eines Leistungserbringers vom interessierenden, zugrunde liegenden Wert des Leistungserbringers abweichen kann und es sich bei Qualitätsbewertungen um Entscheidungen unter Unsicherheit handelt (siehe Abschnitt 20.1.4). Die Ergebnisse der Bewertung hängen daher nicht nur vom Referenzbereich, sondern auch von weiteren Festlegungen wie z. B. dem Signifikanzniveau für den statistischen Vergleich ab. Um mithilfe von Indikatoren eindeutige Bewertungen vornehmen zu können, muss die Qualitätsrichtung eines Indikators monoton sein, d. h. höhere Indikatorwerte sollen entweder bessere Qualität bedeuten (bei positiv gepolten Indikatoren) oder schlechtere Qualität bedeuten (bei negativ gepolten Indikatoren).

### 16.2 Arten von Referenzbereichen

Vergleichswerte für die Bewertung von Messergebnissen können grundsätzlich entweder kriteriumsbezogen oder verteilungsbezogen festgelegt werden (Glass 1978, Goodwin 1996, Goldhammer und Hartig 2012: 175, Popham und Husek 1969). Sollen qualitätsbezogene Entscheidungen daran geknüpft werden, ob ein bestimmter Standard erreicht wurde oder nicht, wird ein von den Ergebnissen anderer Leistungserbringer unabhängiger Referenzbereich benötigt (Norcini und Shea 1997, Popham und Husek 1969). *Kriteriumsbezogene Referenzbereiche*

werden daher durch einen festen Wert auf der Indikatorskala abgegrenzt und auch als *feste Referenzbereiche* bezeichnet.

Kriteriumsbezogene Referenzbereiche erlauben eine Aussage darüber, ob ein Leistungserbringer die zuvor anhand *externer Kriterien* festgelegten Anforderungen erreicht hat (Goldhammer und Hartig 2012: 182 f., Popham und Husek 1969). Diese Kriterien können zum einen auf empirischen Daten beruhen (Shepard 1980: 456 ff., Kane 2017: 18 ff.). Beispielsweise kann das Wissen über den Status quo der Versorgungssituation die Erwartung begründen, dass ein bestimmter Erfüllungsgrad der Anforderungen erreicht werden kann (z. B. dass die Sterblichkeit bei der Behandlung von Pneumonien eine bestimmte Rate nicht überschreiten sollte). Zum anderen kann auch aus einem bestehenden fachlichen oder gesellschaftlichen Konsens abgeleitet werden, dass bestimmte Anforderungen erfüllt werden sollen, z. B. dass Patientinnen und Patienten umfassend informiert und beteiligt werden. Kriteriumsbezogene Referenzbereiche können daher für die Leistungserbringer beispielsweise sichtbar machen, wann hohe und wann geringe Versorgungsqualität vorliegt (Liao et al. 2016). Die Methodik, wie kriteriumsbezogene Referenzbereiche bei fallbasierten Indikatortypen (Prozess- und Ergebnisindikatoren) entwickelt werden, ist in Abschnitt 16.3 dargestellt.

*Verteilungsbezogene Referenzbereiche* beziehen demgegenüber das Ergebnis eines Leistungserbringers auf die Ergebnisse aller Leistungserbringer während desselben Messzyklus (in der externen Qualitätssicherung typischerweise des Erfassungsjahres). Dazu wird beispielsweise ein Wert aus der Verteilung (z. B. ein Perzentil) der Indikatorwerte aller Leistungserbringer als Referenzbereichsgrenze gesetzt. Beispielsweise kann ein verteilungsbezogener Referenzbereich die Anforderung stellen, dass ein Indikatorergebnis nicht schlechter als die 90 % besten Leistungserbringerergebnisse sein soll.

Verteilungsbezogene Referenzbereiche machen die Höhe der Anforderungen von dem Ergebnis anderer Leistungserbringer abhängig (vgl. Norcini und Shea 1997: 43). Sie erlauben also keine Aussage über die Einhaltung von Standards, sondern erlauben lediglich den relativen Vergleich mit anderen Leistungserbringern (Goldhammer und Hartig 2012: 175). Dies geht mit dynamischen, im Zeitverlauf wechselnden Bereichsgrenzen einher und führt bei perzentilbasierten Referenzbereichen zwangsläufig dazu, dass ein bestimmter Prozentsatz an Leistungserbringern als vom Referenzbereich abweichend klassifiziert wird. Im Beispielfall des 90. Perzentils als Referenzbereichsgrenze erreichen in der Regel<sup>42</sup> 10 % der Leistungserbringer den Referenzbereich nicht. Verteilungsabhängige Referenzbereiche werden vom IQTIG nur unter folgenden Bedingungen empfohlen:

- Wenn nur ein relativer Vergleich der Leistungserbringer untereinander vorgenommen und keine Aussage über das Erreichen von Standards getroffen werden soll
- Wenn es nicht möglich erscheint, einen kriteriumsbezogenen Referenzbereich zu begründen

Letzteres kann bei neuen Qualitätsindikatoren der Fall sein, wenn noch keine ausreichenden empirischen Informationen verfügbar sind, um die Erreichbarkeit der Qualitätsanforderungen

---

<sup>42</sup> Der tatsächliche Anteil kann aus verschiedenen mathematischen Gründen davon abweichen, z. B. wenn der Indikator diskrete Merkmale wie Anteilswerte abbildet oder wenn das statistische Verfahren Unsicherheit berücksichtigt.

sicher einzuschätzen (siehe Abschnitt 16.3.1). Sobald eine ausreichende Datenbasis aus dem Regelbetrieb zur Verfügung steht, empfiehlt das IQTIG eine Überführung in einen kriteriumsbezogenen Referenzbereich.

### 16.3 Entwicklung kriteriumsbezogener Referenzbereiche

Die Festlegung von Referenzbereichen vereint immer zwei Elemente:

- eine fachliche Beurteilung des Erreichbaren (fachliches Element) sowie
- eine normative Setzung des Geforderten (normatives Element).

Welche Indikatorwerte erreichbar sind, hängt davon ab, wie gut der Indikator die Versorgungsqualität abbildet, also von der Operationalisierung des Qualitätsmerkmals durch den Indikator. Bei einer entsprechenden Abweichung vom theoretischen Minimum bzw. Maximum handelt es sich also nicht um ein Qualitätsdefizit. Die bei Umsetzung der fachlichen Standards von den Leistungserbringern erreichbaren Indikatorwerte schätzt das IQTIG auf Grundlage fachlicher Überlegungen ein (siehe Abschnitt 16.3.1).

Demgegenüber beziehen sich die Qualitätssicherungsmaßnahmen, für die die Indikatorergebnisse herangezogen werden, auf Qualitätsdefizite. Qualitätsdefizite sind vom Leistungserbringer zu verantwortende Unterschiede zwischen dem Erreichbaren und dem Erreichten. Wie viel vom Erreichbaren gefordert wird und welche Abweichung noch toleriert wird, ist keine fachliche Entscheidung, sondern eine normative Setzung (siehe Abschnitt 16.3.2).

Das IQTIG betrachtet diese beiden Elemente getrennt, um Transparenz über die fachliche Basis der von ihm empfohlenen Referenzbereiche zu schaffen und deren angemessene Berücksichtigung beim Beschluss von Referenzbereichen durch den G-BA als Normgeber zu ermöglichen.

#### 16.3.1 Fachliche Beurteilung des Erreichbaren

Neben der Qualität der Versorgung können auch von den Leistungserbringern nicht zu verantwortende Faktoren zu einer systematischen Abweichung der Indikatorergebnisse vom theoretischen Minimum bzw. Maximum führen (Baker und Yendro 2018). Diese Faktoren können bekannt oder unbekannt sein.<sup>43</sup> Bei einem Indikator zur Sterblichkeit nach Femurfraktur beispielsweise könnten solche Faktoren z. B. bestimmte Begleiterkrankungen, ein höheres Alter oder die genaue Frakturlokalisierung sein. Bei diesem Indikator ist das theoretische Minimum 0 %, aber selbst bei bester Behandlungsqualität würde nicht mit 0 % Sterblichkeit nach Versorgung einer Femurfraktur gerechnet werden.

Nur ein Teil solcher Einflussfaktoren kann bereits über die Berechnungsvorschrift des Indikators berücksichtigt werden, indem der Indikator möglichst genau auf die interessierende Population von Patientinnen und Patienten eingeschränkt wird. Dazu werden die Fälle, für die die Anforderung des Qualitätsindikators nicht gilt oder bei denen unklar ist, ob die Qualitätsanforderung des

---

<sup>43</sup> Zusätzlich zu diesen gerichteten Einflüssen auf den Indikatorwert wirken auf die Messung mit Indikatoren zufällige Einflüsse ein, die zu einer ungerichteten Abweichung der empirisch bestimmten Indikatorwerte eines Leistungserbringers von seinem zugrunde liegenden Indikatorwert führen (siehe Kapitel 15). Diese müssen mittels statistischer Verfahren berücksichtigt werden.

Indikators angemessen ist, aus der Indikatorberechnung ausgeschlossen. Für manche Einflussfaktoren ist ein solcher Ausschluss dagegen aus den folgenden Gründen nicht sinnvoll oder nicht möglich:

- Einige dieser Faktoren erhöhen nur die Wahrscheinlichkeit für ein ungünstigeres Indikatorergebnis (z. B. Begleiterkrankungen der Patientinnen und Patienten), aber ihr Vorliegen soll nicht zu einem vollständigen Ausschluss der betreffenden Patientenfälle aus der interessierenden Population führen, da sich die Qualitätsanforderung des Indikators auch auf diese Fälle erstreckt. Beispielsweise erhöht eine chronische Bettlägerigkeit die Sterblichkeit durch eine Lungenentzündung zwar, dieser Faktor beeinflusst jedoch nicht in dem Maße das Ergebnis, dass der Leistungserbringer für dieses nicht mehr verantwortlich wäre.
- In manchen Fällen würde die Erhebung der relevanten Daten einen unverhältnismäßig hohen Aufwand bedeuten (Meyer 2004: 26). Zum Beispiel kann es für eine nach Leitlinie empfohlene medikamentöse Behandlung eine Vielzahl von möglichen Kontraindikationen geben, deren umfassende Dokumentation jedoch nicht praktikabel wäre.
- Ein Teil der Faktoren, die das erreichbare Indikatorergebnis im Mittel über alle Leistungserbringer hinweg systematisch (in eine bestimmte Richtung) beeinflussen, sind unbekannt. Beispielsweise ist bei schweren Krankheitsbildern wie einem Herzinfarkt oder einem rupturierten Bauchaortenaneurysma keine Überlebensrate von 100 % zu erwarten. Wegen der Charakteristika der Erkrankung oder ihrer Behandlungsmöglichkeiten kann eine Besserung oder Genesung nur mit einer bestimmten Wahrscheinlichkeit erzielt werden, ohne dass alle ursächlichen Faktoren hierfür vollständig erklärt sind.

Die fachliche Einschätzung, welche Einflussfaktoren zu berücksichtigen sind und welcher Erfüllungsgrad des Qualitätsindikators bei Erfüllung aller fachlichen Standards überhaupt im Mittel erreicht werden kann, ist daher eine wichtige Voraussetzung für die Festlegung eines angemessenen Referenzbereichs.

### **Einschätzung des Erreichbaren**

Welcher Wert eines Indikators bei einer Versorgung, die die fachlichen Standards umsetzt, erreichbar wäre, kann nicht direkt gemessen werden, da in empirischen Messergebnissen Qualitätsdefizite bzw. Verbesserungspotenziale enthalten sind, deren Ausmaß unbekannt ist. Das IQTIG nimmt daher unter Berücksichtigung empirischer Daten eine fachliche Einschätzung vor, welche Versorgungsqualität bisher erreicht wird (Status quo) und welche Versorgungsqualität erreichbar wäre. Dazu werden in der wissenschaftlichen Literatur Hinweise zu folgenden Fragen recherchiert:

- Welche Werte werden für die vom Indikator abgebildeten Merkmale bisher erreicht? (Beispielsweise: Mit welcher Häufigkeit werden Komplikationen bei einer Behandlung berichtet oder wie häufig werden erwartete Versorgungsprozesse eingehalten?)
- Liegen Interventionsstudien zum Ausmaß möglicher Verbesserungen vor?

Falls entsprechende Daten vorliegen (z. B. bei Entwicklungen von Qualitätsindikatoren auf Basis von Sozialdaten bei den Krankenkassen oder bei der Überprüfung bestehender Indikatoren),

werden die o. g. Fragestellungen auch anhand eigener Auswertungen durch das IQTIG untersucht.

Bei der Interpretation der Ergebnisse wird das IQTIG durch Fachexpertinnen und -experten insbesondere dazu beraten, welche nicht von den Leistungserbringern zu verantwortenden Einflüsse auf das Qualitätsmerkmal zu berücksichtigen sind und ob der Qualitätsindikator auf bestimmte Behandlungsfälle oder Konstellationen in der Versorgung nicht anwendbar ist.

Auf Grundlage der vorgenannten Informationen wird der Indikatorwert, der die erreichbare Qualität repräsentiert, durch das IQTIG nach folgenden Kriterien eingeschätzt:

- Die erreichbare Qualität ist in der Regel besser als die Ergebnisse für die Gesamtpopulation aus Literatur und aus eigenen Auswertungen (d. h. es liegt näher am Ende der Indikatorskala), da empirisch bestimmte Werte auch die Ergebnisse von Leistungserbringern mit suboptimaler Qualität einschließen.
- Das Ausmaß, in dem ein nicht in der Berechnungsvorschrift des Indikators berücksichtigter Faktor die Indikatorwerte beeinflusst, kann für manche Einflussfaktoren eindeutig angegeben werden. Dies trifft auf Einflussfaktoren zu, bei deren Vorliegen der jeweilige Behandlungsfall aus der Grundgesamtheit des Indikators ausgeschlossen worden wäre.<sup>44</sup> Würde beispielsweise ein Indikator zur Durchführung einer Thromboseprophylaxe auch Patientinnen und Patienten mit Kontraindikation in seinem Nenner umfassen und wäre bekannt, dass im Mittel bei 5 % der Fälle eine Kontraindikation vorliegt, so liegt der bei Einhaltung der fachlichen Standards im Mittel erreichbare Indikatorwert höchstens bei 95 % (statt bei 100 %).
- Sofern eine weitere Eingrenzung der Hinweise aus Literatur und eigenen Auswertungen auf fachlicher Basis erforderlich ist, werden die beratenden Expertinnen und Experten um ihre Einschätzung der erreichbaren Indikatorwerte unter Praxisbedingungen gebeten.
- Sind auf Ebene der einzelnen Behandlungsfälle die dem Indikator zugrunde liegenden Qualitätsanforderungen leicht zu erfüllen und ist nicht von Faktoren auszugehen, die außerhalb der Verantwortung der Leistungserbringer liegen, so sind Indikatorwerte nahe am Ende der Indikatorskala erreichbar.

Die Einschätzung des Erreichbaren durch das IQTIG orientiert sich an den vorgenannten Kriterien. Ist die Informationsgrundlage nicht ausreichend für eine Einschätzung, wird auf einen kriteriumsbezogenen Referenzbereich verzichtet und stattdessen ein verteilungsbezogener Referenzbereich (siehe Abschnitt 16.2) verwendet.

---

<sup>44</sup> Für die Beurteilung der erreichbaren Qualität ist es unerheblich, ob für bestimmte Einflussfaktoren eine Risikoadjustierung der Leistungserbringerergebnisse vorgenommen wird oder nicht. Durch die Risikoadjustierung wird der Indikatorwert eines Leistungserbringers lediglich standardisiert und damit besser vergleichbar mit den Werten anderer Leistungserbringer. Die in der Gesamtpopulation erreichbare Qualität wird durch eine Risikoadjustierung nicht verändert.

### 16.3.2 Geforderter Erreichungsgrad

Neben der Berücksichtigung des Erreichbaren drückt der Referenzbereich auch aus, wieviel vom Erreichbaren gefordert wird bzw. in welchem Maß Qualitätsdefizite toleriert werden. Zwar ist immer bestmögliche Versorgung wünschenswert, die Abwesenheit von Varianz in der Versorgungsqualität ist jedoch unrealistisch. Es wird also – innerhalb eines akzeptablen Bereichs – ein gewisses Maß an Varianz in der Fähigkeit der Einrichtungen zur Erbringung guter Versorgungsleistungen toleriert.

Der Einsatz von Steuerungsinstrumenten zur Qualitätsverbesserung ist zudem erst ab einem bestimmten Ausmaß der Abweichung vom Erreichbaren sinnvoll, da den erwarteten positiven Effekten die mit dem Steuerungsinstrument verbundenen Nebenwirkungen gegenüberstehen, beispielsweise der mit einer Maßnahme verbundene Aufwand, die Auswirkungen auf andere Aspekte der Versorgung oder die Folgen für die beteiligten Gesundheitsprofessionen. Bei geringen Abweichungen vom Erreichbaren sind jedoch nur geringe positive Effekte durch die Entscheidung oder Maßnahme zu erwarten. Die Festlegung des bei Erfüllung der fachlichen Standards erreichbaren Indikatorwerts als Anforderung ist daher für die meisten Verwendungszwecke von Indikatoregebnissen nicht sinnvoll.

### Größe des geforderten Erreichungsgrads

Anders als die fachliche Einschätzung des Erreichbaren handelt es sich beim geforderten Erreichungsgrad um eine normative Setzung. Für diese Setzung ist von Bedeutung, für welche Handlungsanschlüsse (siehe Abschnitt 2.4) die Bewertungsergebnisse verwendet werden sollen (Doran 2014, Kane 2017: 26 f.). So kann der Einsatz unterschiedlicher Handlungsanschlüsse an unterschiedlich hohe Anforderungen geknüpft werden.

### 16.3.3 Festlegung der empfohlenen Referenzbereichsgrenze

In die Festlegung der Referenzbereichsgrenze eines kriteriumsbezogenen Referenzbereichs gehen die Einschätzung des Erreichbaren sowie der geforderte Erreichungsgrad ein. Abbildung 16 zeigt das Vorgehen, wenn (vereinfacht) angenommen wird, dass die Auswirkung der systematisch wirkenden, nicht vom Leistungserbringer zu verantwortenden Faktoren (siehe Abschnitt 16.3.1) auf die empirisch bestimmten Indikatorwerte konstant ist:<sup>45</sup> Diese Faktoren führen dazu, dass sich der im Mittel bestenfalls erreichbare Indikatorwert vom Skalenende unterscheidet. Die empfohlene Größe des Referenzbereichs ergibt sich dann als Summe der Auswirkung dieser Faktoren sowie einer zusätzlichen, vom geforderten Erreichungsgrad abhängigen Abweichung.

---

<sup>45</sup> In welchem Maße diese nicht vom Leistungserbringer zu verantwortenden Einflussfaktoren bei der Patientenstichprobe eines Leistungserbringers vorliegen, unterliegt zufälligen (stochastischen) Einflüssen. Dies kann durch eine statistische Modellierung genauer berücksichtigt werden.

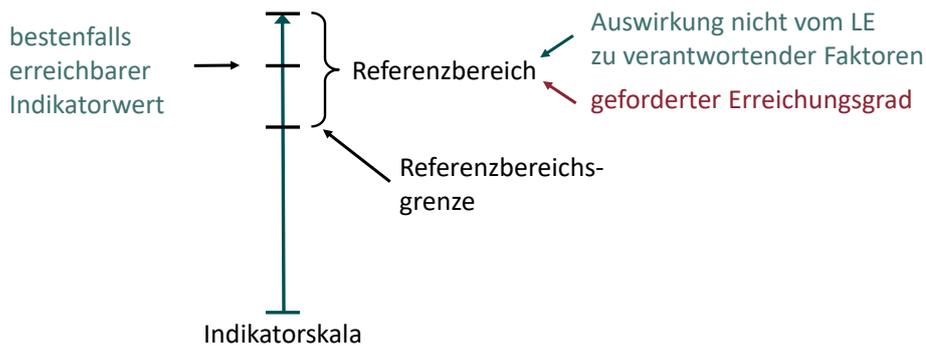


Abbildung 16: Komponenten eines kriteriumsbezogenen Referenzbereichs; vereinfachte Darstellung ohne Berücksichtigung von Stochastizität

### 16.3.4 Prüfung der Angemessenheit

Nach der initialen Bestimmung der Referenzbereichsgrenze sollte deren Angemessenheit anhand empirischer Daten beurteilt werden (Doran 2014, Kane 2017: 24 ff.), sofern entsprechende Informationen verfügbar sind (z. B. wenn der Referenzbereich eines Indikators angepasst wird, der bereits im Regelbetrieb genutzt wird). Dabei wird z. B. anhand einer Proberechnung ermittelt, wie viele Leistungserbringer den Referenzbereich voraussichtlich erreichen würden (vgl. Berk 1996: 230, Norcini und Shea 1997: 44). Dies gibt einen Hinweis darauf, inwieweit das Anforderungsniveau durch die Leistungserbringer erfüllt werden kann und ob es für den vorgesehenen Handlungsanschluss passend ist. Eine Proberechnung ermöglicht auch eine Beurteilung, wie die Lage des Referenzbereichs und das statistische Vorgehen bei der Klassifikation der Indikatorergebnisse zusammenwirken.

## 16.4 Sentinel Events

Sentinel Events sind eine besondere Form von Bewertungskonzept für Prozess- und Ergebnisindikatoren. Dabei wird das Indikatorergebnis eines Leistungserbringers bereits bei einem einzigen aufgetretenen Ereignis, das der Merkmalsdefinition des Indikators entspricht, als „auffällig“ klassifiziert. Fallbezogene Einflussfaktoren wie patientenseitige Risikofaktoren und stochastische Einflüsse (siehe Abschnitt 20.1) werden bei diesem Bewertungskonzept nicht berücksichtigt. Da die Anzahl der Ereignisse und damit der Grad, in dem die Anforderung des Indikators erfüllt wird, für dieses Vorgehen nicht relevant ist, erübrigt sich auch die Festlegung eines Referenzbereichs für den Vergleich mit dem Indikatorwert.<sup>46</sup>

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 [2016]). Diese Sicht hebt die Vermeidbarkeit der Ereignisse als definierendes

<sup>46</sup> Formal könnte ein Sentinel Event als ein Referenzbereich von 0 % mit einem einfachen mathematischen Größer-Vergleich als Berechnungsvorschrift beschrieben werden.

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 das Indikatorergebnis weitestgehend durch die Einrichtung beeinflussbar sei und fallbezogene Einflussfaktoren vernachlässigbar seien (vgl. Abschnitt 13.2.5).

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 Indikatorergebnisses bzw. eine vollständige Vermeidbarkeit der Ereignisse durch die Leistungserbringer unterstellt werden kann. Beispiele für solche Indikatoren sind Todesfälle bei Lebendorganspendern. Das Vorliegen eines Sentinel Events erlaubt hier ohne ergänzende Beurteilung keine Aussage über die Versorgungsqualität, sondern wird lediglich als Auslöser für nachfolgende, genauere Untersuchungen verwendet, beispielsweise ein Stellungnahmeverfahren, in dem eine umfassende, vom Indikatorergebnis unabhängige Qualitätsbewertung vorgenommen wird, oder eine Ursachenanalyse (Hooker et al. 2019). Die Nutzung der Indikatorergebnisse als Auslösekriterium und der Verzicht auf eine quantitative Qualitätsbewertung begründet sich bei diesen Indikatoren dadurch, dass voraussichtlich nur wenige Ereignisse auftreten und eine detaillierte Datenerhebung für alle Behandlungsfälle zu möglichen, nicht vom Leistungserbringer zu verantwortenden Gründen für das Auftreten der Ereignisse einen unverhältnismäßig großen Aufwand bedeuten würde. 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.

## 17 Bewertung von Indikatorergebnissen auf Systemebene

Die Indikatorergebnisse der externen Qualitätssicherung erlauben über die Qualitätsdarstellung für einzelne Leistungserbringer hinaus auch eine Betrachtung der Versorgungsqualität im jeweiligen Themenbereich auf Systemebene. Die Benennung „besonderer Handlungsbedarf“ durch das IQTIG stellt eine solche Bewertung der Indikatorergebnisse auf Systemebene dar mit dem Ziel, die Institutionen des Gesundheitswesens auf Themenbereiche aufmerksam zu machen, in denen ein zusätzlicher Bedarf an Qualitätssicherungsmaßnahmen besteht. Die Institutionen können daraufhin beispielsweise eine Ursachenanalyse für die Qualitätsdefizite oder eine Entwicklung von Handlungsempfehlungen zu deren Behebung einleiten.

### 17.1 Definition des besonderen Handlungsbedarfs

Im Gegensatz zur Bewertung von Indikatorergebnissen auf Ebene der Leistungserbringer (siehe Kapitel 16) bewertet das IQTIG Indikatorergebnisse auf Systemebene bundesweit übergreifend über alle Leistungserbringer und Patientinnen und Patienten. Zeigt sich dabei ein ausgeprägtes Qualitätsdefizit oder bleiben die angestrebten Verbesserungen der Versorgungsqualität aus (persistierendes Qualitätsdefizit), wird dies als „besonderer Handlungsbedarf“ für diesen Qualitätsindikator bezeichnet.

Die Bewertung erfolgt auf Grundlage der Daten und Indikatorergebnisse der externen Qualitätssicherung und für jeden Indikator einzeln. Sie stellt keine umfassende Analyse der Systemqualität im Themenbereich dar. Informationen aus anderen Quellen können zwar zur Interpretation (z. B. Analyse der Ursachen für ein Qualitätsdefizit) ergänzend herangezogen werden, fließen aber nicht unmittelbar in die Feststellung eines besonderen Handlungsbedarfs ein.

### 17.2 Bewertungskriterien

Um die Ergebnisse für alle Indikatoren möglichst objektiv zu bewerten, werden einheitliche Leitkriterien auf die Ergebnisse der Qualitätsindikatoren angewandt. Bereits die Erfüllung eines dieser Leitkriterien ist hinreichend, um auf einen besonderen Handlungsbedarf für den betreffenden Indikator hinzuweisen. Sollte die Bewertung durch das IQTIG von den Leitkriterien abweichen, wird dies begründet.

#### 17.2.1 Leitkriterien

Die Bewertung der Indikatorergebnisse auf Systemebene erfolgt aus zwei Perspektiven: Zum einen wird der Anteil der Leistungserbringer, der die Ziele des Qualitätsindikators erreicht, betrachtet und zum anderen, inwieweit die Versorgungsziele des Qualitätsindikators für die Patientenpopulation als Ganzes erreicht werden.

**Kriterium 1: Anteil von Leistungserbringern, die die Qualitätsanforderungen erfüllen**

Es werden zwei Prüfgrößen unabhängig voneinander herangezogen:

- Anteil der Leistungserbringer, deren Indikatorwert statistisch signifikant<sup>47</sup> vom Referenzbereich abweicht: Das IQTIG definiert einen Anteil von > 10 % der Leistungserbringer (mindestens aber 5 Leistungserbringer) im jüngsten Auswertungszeitraum als Hinweis auf ein ausgeprägtes Qualitätsdefizit unter der Vorstellung, dass Qualitätsdefizite in mehr als jeder zehnten Einrichtung in besonderem Maße Aktivitäten zur Verbesserung erfordern.<sup>48</sup> Voraussetzung für die Anwendung dieser Prüfgröße ist, dass der Referenzbereich des jeweiligen Indikators ein verteilungsunabhängiges Anforderungsniveau benennt (kriteriumsbezogener Referenzbereich, siehe Abschnitt 16.2).
- Anteil der Leistungserbringer, die im Stellungnahmeverfahren als qualitativ auffällig (nach DeQS-RL) bzw. als qualitativ unzureichend (nach plan. QI-RL) in diesem Indikator bewertet wurden: Hier verwendet das IQTIG einen Anteil von > 5 % der Leistungserbringer (mindestens aber 5 Leistungserbringer) im letzten Stellungnahmeverfahren als Hinweis auf ein ausgeprägtes Qualitätsdefizit, da im Stellungnahmeverfahren eine Bewertung als qualitativ auffällig bzw. als qualitativ unzureichend bisher deutlich zurückhaltender erfolgt als anhand der Indikatorwerte. Bereits ein niedrigerer Anteil von Leistungserbringern, die dieses Kriterium erfüllen, weist daher auf ein ausgeprägtes Qualitätsdefizit hin.

**Kriterium 2: bundesweites Indikatorergebnis**

Das bundesweite Indikatorergebnis auf Grundlage aller Behandlungsfälle (Bundeswert) beschreibt die Erfüllung der Versorgungsziele bezogen auf alle betreffenden Patientinnen und Patienten. Weicht dieser Wert stark vom erreichbaren Optimum ab, stellt dies ebenfalls ein ausgeprägtes Qualitätsdefizit dar. Dies kann – analog zur Bewertung der Indikatorergebnisse der Leistungserbringer mittels Referenzbereichen (siehe Kapitel 16) – als Vergleich mit einem Bundesreferenzbereich verstanden werden. Da eine Festlegung solcher Referenzbereiche für die Bewertung auf Bundesebene<sup>49</sup> bisher noch nicht systematisch erfolgt ist, werden anhand eines einzelnen Bundeswerts derzeit keine Hinweise auf besonderen Handlungsbedarf generiert.

---

<sup>47</sup> Bei der Bewertung anhand der rechnerischen Indikatorergebnisse wird statistische Unsicherheit mittels des Verfahrens berücksichtigt, das auch für planungsrelevante Qualitätsindikatoren eingesetzt wird (einseitiger exakter Binomial- bzw. Poisson-Test mit mid-p-Werten, mit einem Signifikanzniveau von 5 %, vgl. IQTIG 2016).

<sup>48</sup> Dieser Grenzwert berücksichtigt außerdem, dass es auch zufallsbedingt – selbst wenn keinerlei Qualitätsdefizite auf Leistungserbringerebene vorliegen – bei bis zu 5 % der Leistungserbringer zu einer statistisch signifikanten Abweichung vom Referenzbereich kommen kann, wenn diese Abweichung jeweils auf einem Signifikanzniveau von 5 % geprüft wird.

<sup>49</sup> Da die Auslösung des Strukturierten Dialogs nach der früheren Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern (QSKH-RL) zufällige Einflüsse nicht berücksichtigte, wurden die bisherigen Referenzbereiche für einzelne Leistungserbringer in der Regel entsprechend toleranter festgelegt, als dies für eine Anwendung auf Bundesebene sinnvoll ist.

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 14. Mai 2020, in Kraft getreten am 14. Mai 2020.

URL: <https://www.g-ba.de/informationen/richtlinien/38/> (abgerufen am 21.03.2022). Die Richtlinie ist am 1. Januar 2021 außer Kraft getreten.

### Bewertung von Zeitverläufen bei den Leitkriterien

Die drei oben genannten Prüfgrößen werden auch im Zeitverlauf betrachtet, um persistierende Qualitätsdefizite zu identifizieren. Zeigt die Analyse, dass es trotz Verbesserungsbedarfs über die letzten 3 Jahre eindeutig zu keiner Verbesserung der Versorgungsqualität gekommen ist, wird dies als Hinweis auf ein persistierendes Qualitätsdefizit gewertet. Voraussetzung für diese Bewertung ist, dass der Verlauf in der Zusammenschau aller Informationen plausibel auf Unterschiede bzw. ausbleibende Unterschiede in der Versorgungsqualität und nicht z. B. auf geändertes Dokumentationsverhalten zurückzuführen ist.

Bei Qualitätsindikatoren mit Risikoadjustierung werden der Bundeswert und dessen Zeitverlauf ggf. durch die Risikoadjustierung beeinflusst. Bei diesen Indikatoren wird der Zeitverlauf der Bundeswerte daher nicht für die Feststellung eines persistierenden Qualitätsdefizits herangezogen.

#### 17.2.2 Weitere Kriterien

Neben den oben genannten Kriterien nimmt das IQTIG gegebenenfalls weitere Analysen der Indikatorergebnisse vor und berücksichtigt diese bei der Bewertung. Beispielsweise können deutliche Unterschiede in der Versorgungsqualität zwischen verschiedenen Patientengruppen oder zwischen verschiedenen Regionen Anlass für die Feststellung eines besonderen Handlungsbedarfs sein.

Für folgende Indikatoren trifft das IQTIG keine Aussage zu besonderem Handlungsbedarf, da sie keine eindeutige und sichere Qualitätsaussage zulassen:

- Qualitätsindikatoren in der Erprobungsphase (mindestens das erste Anwendungsjahr nach Entwicklung oder nach einer erheblichen Änderung)
- Qualitätsindikatoren, die nicht zwischen Versorgungsqualität und Dokumentationsqualität unterscheiden.<sup>50</sup>

---

<sup>50</sup> 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). Die Art des Qualitätsdefizits (z. B. Versorgungsproblem vs. Dokumentationsproblem) kann bei diesen Indikatoren erst im Stellungnahmeverfahren mit einzelnen Leistungserbringern ermittelt werden.

### 17.3 Beratung durch Expertinnen und Experten

Bei der Bewertung der Indikatorergebnisse wird das IQTIG durch die Fachexpertinnen und -experten für die jeweiligen QS-Verfahren zu folgenden Fragestellungen beraten:

- Einschätzung des medizinisch-fachlich Erreichbaren für das Qualitätsziel eines Indikators
- Interpretation von Zeitverläufen im Hinblick auf Verbesserungspotenzial der Indikatorergebnisse
- Beurteilung möglicher Einschränkungen der Validität der Messung, z. B. aufgrund von Problemen der Datenqualität
- sofern auf Basis vorliegender Informationen ersichtlich: Hinweise auf Ursachen der Qualitätsdefizite und ggf. auf Möglichkeiten zu deren Behebung

## 18 Stichprobenziehung und Vollerhebung

Im Rahmen der Entwicklung von Qualitätssicherungsmaßnahmen und deren Durchführung kommen an verschiedenen Stellen Datenerhebungsverfahren zum Einsatz, die sowohl von Teilerhebungen (Stichproben) als auch von Vollerhebungen Gebrauch machen. In diesem Kapitel werden die grundlegenden Prinzipien dargelegt, nach denen das IQTIG Stichprobenziehungen oder Vollerhebungen selbst durchführt, etwa im Rahmen von Entwicklungsleistungen, oder dem G-BA empfiehlt, beispielsweise für den Regelbetrieb von indikatorbasierten Qualitätssicherungsmaßnahmen. Andere Formen der Datenerhebung, wie etwa bei der Prüfung der Datenqualität von QS-Dokumentationsdaten durch Abgleich mit der Referenzquelle (siehe Kapitel 19), bei denen auch Stichprobenziehungen vorgenommen werden, sind nicht Gegenstand dieses Kapitels.

### 18.1 Ebenen der Stichprobenziehung und Vollerhebung

Da sich Qualitätsmessungen, deren Bewertung und daran anschließende Qualitätssicherungsmaßnahmen (im Sinne des Rahmenmodells der Aufgaben von Qualitätssicherung gemäß Abschnitt 2.5) grundsätzlich auf verschiedene Ebenen beziehen können, stellt sich auch die Frage nach Stichprobenziehung oder Vollerhebung jeweils auf verschiedenen Ebenen. Mögliche Ebenen sind im Kontext der Qualitätssicherung zum einen die Leistungserbringer und zum anderen deren Behandlungsfälle. Denkbar sind aber als weitere Ebenen auch z. B. Bundesländer, Fachabteilungen innerhalb von Leistungserbringern, Ärzte und Ärztinnen innerhalb von Fachabteilungen/Praxen sowie Patientinnen und Patienten, die mehrmals behandelt werden.

Die Datenerhebung auf Ebene der Patientinnen und Patienten dient vor allem der Beschreibung der Leistungserbringer als Aggregateinheiten (O'Brien 1990). Dabei erfolgt die Berechnung von Indikatorwerten für die Leistungserbringer meist durch Aggregation der Daten der Behandlungsfälle eines bestimmten Zeitraums (siehe Kapitel 20), um quantitative Aussagen über die Qualität der Versorgung durch den Leistungserbringer zu ermöglichen. Dieses Kapitel bezieht sich nur auf Stichprobenziehungen auf Ebene der Patientinnen und Patienten bzw. Behandlungsfälle.

Als Grundgesamtheit (= Zielpopulation) der Datenerhebung auf Patientenebene werden jene Patientinnen und Patienten bezeichnet, deren Daten für die jeweilige Qualitätsmessung der Leistungserbringer relevant sind. Die Grundgesamtheit wird in Abhängigkeit von der jeweiligen Fragestellung der Datenerhebung eindeutig vom IQTIG definiert. Bevor eine Stichprobe gezogen oder eine Vollerhebung durchgeführt werden kann, muss die sogenannte Auswahlgesamtheit (engl. *sampling frame*) definiert werden, die der Grundgesamtheit möglichst gut entsprechen soll. Insofern kann die Auswahlgesamtheit als Operationalisierung der Grundgesamtheit verstanden werden, wobei nur in dem Ausmaß, in dem die Grundgesamtheit eindeutig beschrieben ist, die Definition der Auswahlgesamtheit möglich ist.

Die Auswahlgesamtheit stellt die tatsächlich verfügbare Grundlage zur Stichprobenziehung dar (Schnell et al. 2013, Lohr 2010). Im Kontext der 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 (zufällig oder systematisch) gewonnene Auswahl von Behandlungsfällen aus dieser Auswahlgesamtheit bezeichnet. Werden von allen Behandlungsfällen der Auswahlgesamtheit Daten erhoben, spricht man von einer Vollerhebung.

Die Auswahlgesamtheit bzw. die mittels der softwarebasierten QS-Filter erfassbaren Behandlungsfälle müssen nicht deckungsgleich mit der interessierenden Grundgesamtheit sein. Beispielsweise könnte eine interessierende Grundgesamtheit nur unscharf mittels einer Kombination aus verschiedenen Informationen abgebildet werden, die für den QS-Filter zugänglich sind (z. B. ICD- und OPS-Kodes, Alter). Diese Abbildung kann also interessierende Fälle ausschließen („Undercoverage“) sowie nicht interessierende Fälle einschließen („Overcoverage“; Schnell et al. 2013). Diese Problematik betrifft sowohl Stichprobenziehungen als auch Vollerhebungen. Under- und Overcoverage beschränken sich dabei nicht nur auf die Grundgesamtheit bzw. deren Verhältnis zur Auswahlgesamtheit, sondern können auch dazu führen, dass Behandlungsfälle oder Ereignisse fälschlicherweise in die Berechnung eines Qualitätsindikators ein- oder davon ausgeschlossen werden (Badgery-Parker et al. 2020). Diese Situation besteht grundsätzlich bei allen Datenerhebungen, bei denen keine vollständige Liste der Grundgesamtheit der Merkmalsträger vorliegt, und wird im Rahmen der Entwicklung neuer Qualitätssicherungsmaßnahmen durch das IQTIG bei der Definition des QS-Filters sowie von Zähler und Nenner der Qualitätsindikatoren durch möglichst exakte Operationalisierung adressiert (siehe Kapitel 6 und 7). Kann das Ausmaß von Over- und Undercoverage für alle für einen Qualitätsindikator relevanten Mengen (beispielsweise die interessierenden Fälle und interessierenden Ereignisse) quantifiziert werden, werden diese Aspekte bei der statistischen Auswertung eines Qualitätsindikators berücksichtigt.

Sofern Stichprobenziehungen durchgeführt werden und ein inferenzstatistischer Verallgemeinerungsschluss auf die Grundgesamtheit angestrebt wird, sind Zufallsstichproben oder Varianten von Zufallsstichproben (z. B. stratifizierte Zufallsstichproben) indiziert. Diese erlauben den inferenzstatistischen Verallgemeinerungsschluss auf die Grundgesamtheit (sogenannte Repräsentativität; Schnell et al. 2013, Lohr 2010).

## 18.2 Kriterien für Stichprobenziehung und Vollerhebungen

Durch § 299 Abs. 1 SGB V wird vorgegeben, dass Stichprobenziehungen auf der Ebene von Fällen bzw. von Patientinnen und Patienten die Regel darstellen sollen und Vollerhebungen nur bei Vorliegen gewichtiger medizinisch-fachlicher oder gewichtiger methodischer Gründe zum Einsatz kommen sollen. Innerhalb dieser gesetzlichen Rahmenbedingungen prüft das IQTIG, ob Gründe gegen eine Stichprobenziehung vorliegen. Die Kriterien für diese Prüfung werden im Folgenden dargelegt.

### 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.

### **Ressourcenbedarf**

Der Ressourcenbedarf ist für Vollerhebungen fast immer deutlich höher als für Stichprobenziehungen.

### **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 erschwert die Ziehung einer Zufallsstichprobe.

### **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. Die Dokumentationsgüte kann aber auch bei Vollerhebungen größer sein, wenn mehr und kontinuierlichere „Übung“ im Umgang mit dem Dokumentationsbogen vorliegt.

### **Seltene Ereignisse**

Sollen mit der Datenerhebung seltene Ereignisse erfasst werden, können Stichproben kontraindiziert sein. Beispielsweise sind für die Erfassung seltener Ereignisse, bei denen schon dem einmaligen Auftreten eines Ereignisses (z. B. Versterben nach Hüftendoprothesen-Erstimplantation bei geringer Sterbewahrscheinlichkeit) in jedem Fall nachgegangen werden soll (sogenannte Sentinel Events), Zufallsstichproben nicht geeignet, da diese nicht gewährleisten können, jedes Auftreten zu erfassen.

## **18.3 Stichprobengrößenberechnungen**

Sofern eine Stichprobenziehung als zielführend beurteilt wird, führt das IQTIG eine Fallzahlberechnung durch, um die benötigte Mindestzahl an Behandlungsfällen 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 (siehe Abschnitt 20.1.2). 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 es in Fällen großer Varianz innerhalb der Grundgesamtheit nur mit verhältnismäßig großen Stichproben möglich ist. Die im Regelbetrieb benötigte Bruttostichprobengröße 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 20). 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. Das IQTIG erläutert seine Überlegungen und Empfehlungen zu benötigten Fallzahlen in seinen Entwicklungsberichten.

## 19 Prüfung der Qualität von QS-Dokumentationsdaten

Die Qualität der zugrunde liegenden Daten stellt eine wichtige Voraussetzung für aussagekräftige Qualitätsmessungen dar (siehe Abschnitt 13.3.2). Bei Qualitätsmessungen auf Basis der Dokumentation der Leistungserbringer liegen unterschiedliche Gründe nahe, weshalb es zu Fehldokumentationen kommen kann (vgl. Abschnitt 2.4). Bei Maßnahmen der Qualitätsförderung wird primär von einer intrinsischen Motivation zur Qualitätssteigerung der Leistungserbringer ausgegangen. Die Leistungserbringer haben daher vermutlich ein besonderes Interesse an einer guten Datenbasis und Fehldokumentationen gehen vermutlich eher auf unbeabsichtigte Fehler zurück. Beim Einsatz von anreizbasierten und auswahlbasierten Qualitätssicherungsmaßnahmen stellen dagegen die externen Anreize und Konsequenzen, die an die Qualitätsmessung geknüpft sind, Fehlanreize bei der Dokumentation dar und können damit zu schlechterer Datenqualität führen. Beispielsweise könnten Leistungserbringer durch Fehldokumentationen, z. B. einer Underdokumentation von unerwünschten Ereignissen (vgl. Döbler et al. 2016), Auswahlentscheidungen zu ihren Gunsten oder Vorteile bei vergütungsbasierten Anreizsystemen erzielen. Aus diesen unterschiedlichen Gründen für potenzielle Fehldokumentationen ergeben sich unterschiedliche Ansätze, um eine hohe Datenqualität sicherzustellen (z. B. durch verständliche und eindeutige Spezifikation der Instrumente zur Dokumentation, Förderung eigenständiger Prüfungen durch die Leistungserbringer, externe Prüfungen, Anreize zur korrekten Dokumentation, Einsatz von nicht dokumentationsbasierten Datenquellen im Rahmen von anreiz- und auswahlbasierten Qualitätssteuerungsinstrumenten).

Ein wichtiger Ansatz zur Sicherstellung einer hohen Datenqualität ist die Prüfung der dokumentierten Daten. Das vorliegende Kapitel beschreibt die Grundsätze, nach denen das IQTIG die Qualität von Daten prüft, die von den Leistungserbringern zum Zweck der externen Qualitätssicherung erfasst werden (sogenannte QS-Dokumentation).<sup>51</sup> Unter Berücksichtigung der Rahmenbedingungen und Anforderungen der externen Qualitätssicherung prüft das IQTIG die folgenden Aspekte der Datenqualität (Ehling et al. 2007, Nonnemacher et al. 2014, Chen et al. 2014, Arts et al. 2002):

- **Vollständigkeit:** alle erforderlichen Angaben zu einem einzelnen Datensatz (z. B. 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/Übereinstimmung:** Übereinstimmung mit einer Referenzquelle

---

<sup>51</sup> Diese Prüfung wird in den Richtlinien des G-BA als „Datenvalidierung“ bezeichnet.

Zur Überprüfung und Steigerung der Datenqualität stehen grundsätzlich drei Methoden zur Verfügung. Es können Prüfungen der Daten schon bei der Dateneingabe und innerhalb der übermittelten Daten vorgenommen werden. Außerdem kann ein Abgleich des zu prüfenden Datensatzes mit einer externen Referenzquelle vorgenommen werden (Horenkamp-Sonntag et al. 2014). Die Prüfung anhand einer externen Referenzquelle gilt dabei zwar als Goldstandard (Schubert et al. 2010, Nonnemacher et al. 2014), ist jedoch häufig aufgrund von zeitlichen, finanziellen, organisatorischen und datenschutzrechtlichen Gründen nicht möglich (Hartmann et al. 2016). Die Prüfungen bei der Dateneingabe und innerhalb des zu prüfenden Datensatzes ist dagegen aufwandsärmer als die Prüfung anhand einer externen Referenzquelle und ermöglicht so die Prüfung einer größeren Datenmenge. Für die Prüfung und Steigerung der Qualität der Dokumentationsdaten in der externen Qualitätssicherung verwendet das IQTIG daher die folgenden drei Elemente:

- eine Prüfung bei Dateneingabe und -annahme
- eine Prüfung innerhalb der dokumentierten Daten nach Datenannahme
- ein Datenabgleich mit Referenzquelle nach Datenannahme

### **19.1 Prüfung bei Dateneingabe und -annahme**

Bei der automatisierten Prüfung bei Dateneingabe werden die Daten schon im Rahmen der Dateneingabe und -erfassung beim Leistungserbringer nach bestimmten Kriterien durch die Software automatisiert auf Fehler geprüft. Die Spezifikationen des IQTIG für die Software zur Datenerfassung und -übermittlung für die externe Qualitätssicherung beinhalten solche Kriterien zur Prüfung der Vollständigkeit der Datensätze und der Plausibilität der Daten. Die Erfassungssoftware ist beispielsweise so konfiguriert, dass fehlende Werte bei verpflichtend auszufüllenden Datenfeldern bei der Dateneingabe durch die Leistungserbringer nicht akzeptiert werden. Ein Datensatz kann erst bei Vollständigkeit abgeschlossen und an die Datenannahmestelle übermittelt werden. Außerdem wird die Erfassungssoftware so spezifiziert, dass unzulässige Werteeingaben und nicht plausible Kombinationen von Werten über Datenfelder hinweg nicht eingegeben werden können und dass Warnungen bei der Eingabe nicht plausibler Werte angezeigt werden.

Darüber hinaus wird nochmals bei der Datenannahme durch die zuständigen Datenannahmestellen automatisiert geprüft, ob die Daten in der spezifizierten Form übermittelt wurden.

### **19.2 Statistische Prüfung nach Datenannahme**

Nach der Datenannahme werden die gemäß DeQS-RL von den Leistungserbringern gelieferten Daten durch das IQTIG einer weiteren Prüfung anhand sogenannter Auffälligkeitskriterien unterzogen. Auffälligkeitskriterien sind Kennzahlen, die auf Mängel in der Dokumentationsqualität hinweisen. Ähnlich wie mit Qualitätsindikatoren werden die Daten jedes Leistungserbringers anhand dieser Auffälligkeitskriterien auf Auffälligkeiten in den dokumentierten Daten geprüft.

Im Rahmen der Entwicklung und Weiterentwicklung von indikatorbasierten QS-Verfahren empfiehlt das IQTIG dem G-BA neben den eigentlichen Qualitätsindikatoren zur Abbildung der Versorgungsqualität auch Auffälligkeitskriterien zur Prüfung und Sicherstellung der Qualität der den dokumentationsbasierten Indikatoren zugrunde liegenden Daten.

### **Auffälligkeitskriterien der Vollzähligkeit**

Die Prüfung auf Vollzähligkeit erfolgt mittels Abgleichs der gelieferten Daten mit einer sogenannten Sollstatistik, die auf den Abrechnungsdaten der Leistungserbringer beruht. Diese Sollstatistik zählt alle Fälle, die in den Abrechnungsdaten des Erfassungsjahres eines Leistungserbringers als dokumentationspflichtig eingeordnet wurden.

Außerdem wird die Anzahl der gelieferten sogenannten Minimaldatensätze untersucht. Ein Minimaldatensatz enthält nur einige wenige Datenfelder und kann ausnahmsweise anstelle des vorgesehenen Datensatzes 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.

### **Auffälligkeitskriterien zur Plausibilität**

Die Auffälligkeitskriterien zur Plausibilität beziehen sich auf die inhaltliche Plausibilität der Angaben innerhalb der von einem Leistungserbringer gelieferten Datensätze. Mittels dieser Kriterien werden unwahrscheinliche oder widersprüchliche Werteverteilungen und Kombinationen von Werten einzelner Datenfelder geprüft (z. B. selten Komplikationen bei hoher Verweildauer). Auffälligkeitskriterien zur Plausibilität können zum Anlass für weitergehende Überprüfungen der Datenqualität eines Leistungserbringers genommen werden.

### **Entwicklung von Auffälligkeitskriterien**

Das IQTIG entwickelt die von ihm empfohlenen Auffälligkeitskriterien zur Prüfung der Dokumentationsqualität nach folgenden Prinzipien:

#### **▪ Vorrang für Plausibilitätsregeln in der Spezifikation**

Wenn eine mögliche Gefährdung der Datenqualität durch eine Plausibilitätsregel in der Spezifikation für die Dokumentationssoftware adressiert werden kann, wird dies einem Auffälligkeitskriterium vorgezogen. Während über Plausibilitätsregeln in der Spezifikation bereits bei der Dateneingabe die Eingabe nicht plausibler Angaben verhindert werden und so die Datenqualität sichergestellt werden kann, können durch Auffälligkeitskriterien Angaben nur nach der Datenannahme geprüft werden. Durch den Vorrang von Plausibilitätsregeln vor Auffälligkeitskriterien soll der Notwendigkeit für Auffälligkeitskriterien entgegengewirkt und sollen Aufwände für die Prüfung der Datenqualität reduziert werden.

#### **▪ Relevanz für Qualitätsindikatoren**

Insbesondere die Datenfelder, die unmittelbar für die Berechnung eines Qualitätsindikators relevant sind, werden bei der Entwicklung von Auffälligkeitskriterien berücksichtigt.

- **Brauchbarkeit für die Steigerung der Dokumentationsqualität**

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 Dokumentationsqualität beitragen.

- **Berücksichtigung von Fallzahlen**

Auffälligkeitskriterien, die die Verbesserung der Dokumentationsprozesse zum Ziel haben, werden nach einer analytischen Zielsetzung unter Berücksichtigung von Stochastizität ausgewertet (vgl. Abschnitt 20.1.1). Auffälligkeitskriterien, die die Korrektheit der Daten abbilden, werden nach einer enumerativen Zielsetzung ausgewertet. Bei einer Vollerhebung ist in diesem Fall keine Berücksichtigung von Stochastizität notwendig. Dies gilt insbesondere für Auffälligkeitskriterien zur Vollzähligkeit.

- **Evaluation von Auffälligkeitskriterien**

Bei der Evaluation von Auffälligkeitskriterien prüft das IQTIG, ob sich anhand der Auffälligkeitskriterien tatsächlich Dokumentationsfehler identifizieren lassen. Dabei wird etwa der Anteil der bestätigten Fehldokumentationen an der Anzahl der durch das Auffälligkeitskriterium identifizierten möglichen Fehldokumentationen herangezogen.

### 19.3 Datenabgleich mit Referenzquelle

Als Referenzquelle zur Überprüfung der für die gesetzliche Qualitätssicherung dokumentierten Daten können verschiedene Datenquellen herangezogen werden (ärztliche Originaldokumentationen, Befragungsdaten, Abrechnungsdaten, amtliche Statistiken oder Registerdaten) (Horenkamp-Sonntag et al. 2014, Benchimol et al. 2011). Der Goldstandard ist jedoch die ärztliche Originaldokumentation in der Patientenakte (Arts et al. 2002, Stange et al. 1998).

Ein Abgleich mit der Originaldokumentation wird aufgrund des hohen Aufwands häufig anhand von Stichproben durchgeführt. Eine statistische Methodik zur Auswertung von Ergebnissen aus solchen Stichprobenüberprüfungen hat das IQTIG in IQTIG (2020: Kapitel 2) entwickelt. Das Vorgehen sowie mögliche Konsequenzen bei unzureichender Datenqualität regelt der G-BA in seinen Richtlinien.

Für einen Abgleich mit anderen, dem IQTIG vorliegenden Datenquellen ist die Einschränkung auf eine Stichprobe unter Umständen nicht notwendig. Hier können beispielsweise Auffälligkeitskriterien definiert werden, die die Konkordanz der Datenquellen prüfen.

## 20 Methoden der Datenanalyse

Für die Wahl einer statistischen Auswertungsmethodik ist von zentraler Bedeutung, welches Ziel mit der Datenanalyse verfolgt wird. Das im Folgenden vorgestellte Rahmenkonzept für die statistische Auswertungsmethodik (Abschnitt 20.1) ermöglicht die Ableitung der adäquaten datenanalytischen Methoden für den jeweiligen Anwendungszweck aus dem Ziel der Datenanalyse sowie aus den interessierenden epidemiologischen Größen. Es integriert die unterschiedlichen Arten von Datenanalysen der externen Qualitätssicherung, wie etwa die Auswertung von Qualitätsindikatoren oder Analysen zur Datenqualität, in einem einheitlichen Rahmen, aus welchem die konkreten statistischen Analysemethoden abgeleitet und begründet werden können. Im Folgenden liegt der Fokus auf der Auswertung von Qualitätsindikatoren, da deren Entwicklung und Auswertung einen zentralen Bestandteil der Aufgaben des IQTIG darstellt. In Abschnitt 20.2 (Risikoadjustierung) und Abschnitt 20.3 (Auswertung von Follow-up-Indikatoren) wird auf zwei spezielle Aspekte der statistischen Datenauswertung eingegangen.

### 20.1 Rahmenkonzept für die statistische Auswertungsmethodik

Ein zentrales Ziel von Qualitätsmessungen nach §§ 136 ff. SGB V ist die Messung und vergleichende Darstellung von Versorgungsqualität (vgl. auch Kapitel 2). Jede Qualitätsmessung erfordert dafür die Festlegung

- der für die Auswertung betrachteten Ebene sowie
- des zu betrachtenden Zeitraums.

Im Rahmen der externen Qualitätssicherung werden Auswertungen auf mehreren Auswertungsebenen vorgenommen, z. B. auf Ebene der Leistungserbringer, der Landes- und der Bundesebene. Um Aussagen über die Versorgungsqualität auf der jeweiligen Ebene, z. B. der der Leistungserbringer, treffen zu können, werden für fallbasierte Qualitätsindikatoren die Messergebnisse auf der Fallebene aggregiert.<sup>52</sup> Die Fallebene stellt dabei die unterste Ebene dar, je nach interessierendem Ereignis beispielsweise die Prozedur- oder Patientenebene.

Der zu betrachtende Zeitraum wiederum definiert den Zeitraum, auf den sich die Aussagen der Qualitätsmessung beziehen und innerhalb dessen die relevanten Fälle in der betrachteten Auswertungsebene in die Qualitätsmessung eingehen. Er kann beispielsweise einem Erfassungsjahr oder einem Quartal entsprechen. Für die Definition des betrachteten Zeitraums ist zudem eine Festlegung erforderlich, wie Fälle einem Zeitraum zugeordnet werden (z. B. über das Datum der Aufnahme, Operation oder Entlassung; siehe auch Abschnitt 20.3).

---

<sup>52</sup> Andere Formen von Qualitätsindikatoren kommen z. B. bei Einrichtungsbefragungen oder Strukturerhebungen vor. In diesen Erhebungssituationen stellt der betrachtete Leistungserbringer direkt die unterste Auswertungsebene dar.

Die Art und Weise, wie Informationen einzelner Messungen aggregiert und welche Aussagen anhand der aggregierten Informationen getroffen werden sollen, kann anhand von vier Komponenten beschrieben werden, die gemeinsam das Rahmenkonzept für die statistische Auswertungsmethodik darstellen:

- **Zielsetzung:** Sollen Aussagen über die beobachteten Daten selbst getroffen oder sollen aus den Daten Rückschlüsse auf den zugrunde liegenden Prozess gezogen werden?
- **Stichprobenart:** Welche Art einer Stichprobe liegt vor und welche Konsequenzen ergeben sich daraus für die Berücksichtigung stochastischer Effekte?
- **Berechnungsart:** Welche interessierende epidemiologische Größe liegt dem Indikator zugrunde (z. B. Rate<sup>53</sup>, risikoadjustierte Rate, Follow-up-Rate)?
- **Bewertungsart:** Nach welcher Klassifikationsmethode wird ein beobachtetes Ergebnis in ein Bewertungsschema eingeordnet (vgl. auch Kapitel 16)?

### 20.1.1 Zielsetzung

Im Folgenden werden zwei prinzipielle Zielsetzungen für eine Auswertung von Qualitätsindikatoren – die analytische sowie die enumerative Zielsetzung – beschrieben (Deming 1953). Beide unterscheiden sich wesentlich in den Konsequenzen für die Wahl der statistischen Methodik.

#### Analytische Zielsetzung

Bei einer analytischen Zielsetzung sollen von beobachteten Ergebnissen von Qualitätsmessungen Rückschlüsse auf den datengenerierenden Prozess gezogen werden, welcher selbst nicht direkt beobachtbar ist. So soll von den beobachteten Ergebnissen eines Qualitätsindikators darauf geschlossen werden, welchen Einfluss die Versorgungsgestaltung durch den Leistungserbringer auf das vom Qualitätsindikator abgebildete Qualitätsmerkmal hat. Die Strukturen und Handlungen eines Leistungserbringers, die ein Qualitätsmerkmal beeinflussen, werden zusammen als dessen zugrunde liegende Kompetenz bzgl. dieses Merkmals bezeichnet. Dabei muss es sich nicht um eine spezifische, benennbare Kompetenz handeln. Vielmehr abstrahiert dieser Begriff das komplexe Zusammenspiel vieler Eigenschaften des Leistungserbringers im Kontext des jeweiligen Qualitätsmerkmals. Der Begriff „Kompetenzparameter“ bezeichnet die konkrete Operationalisierung der zugrunde liegenden Kompetenz eines Leistungserbringers bezüglich eines Merkmals im Rahmen eines Qualitätsindikators (z. B. innerhalb der Definition von Rechenregeln).

Für die analytische Zielsetzung ist die Annahme charakteristisch, dass die Ergebnisse von Qualitätsmessungen nicht allein von der zugrunde liegenden Kompetenz eines Leistungserbringers abhängen, sondern dass weitere Faktoren existieren können, die die Qualitätsmessungen beeinflussen: messbare oder nicht messbare Patientencharakteristika sowie externe Faktoren, die weder durch den Leistungserbringer noch die Patientin oder den Patient beeinflusst werden.<sup>54</sup>

<sup>53</sup> Aus historischen Gründen wird in der externen stationären Qualitätssicherung der Begriff „Rate“ für einen Anteilswert bzw. eine Quote verwendet. Es handelt sich also nicht um eine direkt auf eine Zeitspanne bezogene Größe.

<sup>54</sup> Beispiele für solche externen Einflüsse, die die Leistungserbringer nicht zu vertreten haben, sind überlastende, unvorhersehbare Situationen, etwa große Krankheitsausbrüche oder Naturkatastrophen. Auch

Da immer Einflüsse auf Qualitätsmessungen bestehen, die nicht erfasst werden bzw. erfasst werden können, werden alle nicht gemessenen Einflüsse unter dem Konzept der Stochastizität subsumiert, um sie explizit als stochastische Komponente zu berücksichtigen. Aus den beobachteten Daten sind daher keine unmittelbaren Rückschlüsse auf den einem Qualitätsindikator zugrunde liegenden, datengenerierenden Prozess möglich.

Die analytische Zielsetzung ist insbesondere dann methodisch angemessen, wenn Verallgemeinerungen über das Geschehene hinaus von Interesse sind (Deming 1953). Diese Zielsetzung findet in vielen Arbeiten zur medizinischen Qualitätssicherung Anwendung (siehe z. B. Spiegelhalter et al. (2012), Ash et al. (2012))<sup>55</sup>, um aus den beobachteten Ergebnissen Rückschlüsse auf die Versorgungsgestaltung eines Leistungserbringers zu ziehen. Daher wird bei der analytischen Zielsetzung statistische Unsicherheit aufgrund von Stochastizität bei der Bewertung von Leistungserbringerergebnissen berücksichtigt, selbst wenn für alle Fälle eines Leistungserbringers im betrachteten Erfassungsjahr Daten im Sinne einer Vollerhebung vorliegen.

### **Enumerative Zielsetzung**

Im Gegensatz zur analytischen Zielsetzung liegt das Interesse bei der enumerativen Zielsetzung auf einer interessierenden Größe in einer endlichen Grundgesamtheit. Es wird angenommen, dass die interessierende Größe in der Grundgesamtheit deterministisch ist, d. h., dass es keine unbekannt oder nicht erfassten Faktoren gibt, die die Qualitätsmessung beeinflussen. Die enumerative Zielsetzung ist also dann indiziert, wenn das Ziel der Auswertung darin besteht, Aussagen über die erhobenen Daten selbst zu treffen und die interessierende, deterministische Größe für eine vorgegebene, endliche Population zu beschreiben.

#### **20.1.2 Stichprobenart**

Die Art der Stichprobe, aus welcher die erhobenen Daten stammen, hat ebenfalls Konsequenzen für die Wahl der statistischen Methodik. Im Kontext der enumerativen Zielsetzung ist es sinnvoll zu unterscheiden, ob es sich bei der Stichprobenart um eine Vollerhebung handelt oder nicht (siehe Kapitel 18). Im Fall einer Vollerhebung muss bei der Berechnung des Indikatorergebnisses und dem Vergleich mit einem Referenzwert keine statistische Unsicherheit berücksichtigt werden. Sollen Ergebnisse dagegen von einer Stichprobe auf eine Grundgesamtheit verallgemeinert werden, ist es methodisch angemessen, die Stochastizität der Stichprobenziehung über inferenzstatistische Methoden – auch bei einer enumerativen Zielsetzung – zu berücksichtigen (Kauermann und Küchenhoff 2011). Bei der analytischen Zielsetzung sollte statistische Unsicherheit unabhängig von der Stichprobenart immer berücksichtigt werden; für die konkrete Art und Weise der Berücksichtigung ist dabei auch die Stichprobenart von Relevanz.

Die Stichprobenart hängt von der Auswertungsebene ab. Liegt beispielsweise eine Vollerhebung der Leistungserbringer vor, bei denen jeweils eine bestimmte Anzahl an deren Fällen als Stichprobe zufällig mit gleicher Wahrscheinlichkeit gezogen wird, werden die ausgewählten Fälle als

---

können sporadische Produktionsfehler bei Medizinprodukten oder Eigenschaften von Spenderorganen bei Transplantationen teilweise als solche weiteren Einflüsse gesehen werden.

<sup>55</sup> Insbesondere kann immer von einer analytischen Zielsetzung ausgegangen werden, wenn trotz Vollerhebung im klassischen Sinn statistische Unsicherheit berücksichtigt wird.

für den jeweiligen Leistungserbringer repräsentative Stichprobe bezeichnet (Kauermann und Küchenhoff 2011). Auf Bundesebene wiederum ist diese Stichprobe nicht zwingend repräsentativ für alle behandelten Fälle. Wenn beispielsweise die Anzahl an Fällen zwischen den Leistungserbringern variiert, führt dies dazu, dass die Auswahlwahrscheinlichkeiten der Fälle nicht mehr identisch sind. Solange die Auswahlwahrscheinlichkeiten jedes Falls für die Stichprobe bekannt sind, können auch über eine entsprechende Gewichtung der Stichprobenfälle oder über mehrstufige oder modellbasierte Stichprobenverfahren auf Bundesebene repräsentative Aussagen getroffen werden (Kauermann und Küchenhoff 2011).

### 20.1.3 Berechnungsart

Die Berechnungsart eines Indikators wird durch die interessierende epidemiologische Größe definiert. Diese legt fest, mit welcher Maßzahl Qualität gemessen werden soll und welche durch das Studiendesign verursachte Verzerrungen zusätzlich berücksichtigt werden müssen. Oft ist in der Qualitätssicherung das interessierende Outcome ein binäres Ereignis (z. B. das Eintreten einer bestimmten Komplikation). Die Berechnungsart bestimmt, ob die Häufigkeit dieser Ereignisse absolut oder relativ betrachtet werden soll und welche die Bezugspopulation ist (siehe Kapitel 3 in Rothman et al. 2008).

Oft stellt die interessierende Zielgröße beispielsweise die Inzidenz eines interessierenden Ereignisses innerhalb einer definierten Grundgesamtheit und innerhalb eines bestimmten Zeitraums dar. Für deren Berechnung ergeben sich die Notwendigkeit einer Mengendefinition von Grundgesamtheit und interessierenden Ereignissen, dem betrachteten Zeitraum inklusive der Regel, nach welcher Fälle dem Zeitraum zugeordnet werden, sowie der Berechnungsart als Anteilswert; in der Tradition der Qualitätssicherung werden diese Indikatoren „Ratenindikatoren“ genannt. Ein typisches Beispiel für die Festlegung des relevanten Auswertungszeitraums ist das sogenannte Erfassungsjahr, welches über ein definierendes Ereignis (z. B. die Entlassung) eine zeitliche Zuordnung der Fälle in verschiedene Auswertungszeiträume sichert. Die daraus resultierende Auswertung besteht dann z. B. aus der Inzidenz, welche als Anteil an Fällen mit interessierenden Ereignissen an allen Fällen in der Grundgesamtheit im betrachteten Zeitraum definiert ist (Rothman 2012). Qualitätsindikatoren zu leitlinienkonformer Indikation sind ein Beispiel für Indikatoren, die Mengendefinitionen zur Grundgesamtheit sowie dem Zeitpunkt und der Art von interessierenden Ereignissen vornehmen und die Berechnungsart „Rate“ haben.

Manchmal kann es notwendig sein, die interessierende epidemiologische Größe über Populationen zu standardisieren, damit die Maßzahl besser vergleichbar wird. Beispielsweise verwendet man eine Risikoadjustierung für patientenseitige Faktoren, um die epidemiologische Maßzahl für unterschiedliche Patientenkollektive bei den Leistungserbringern mittels einer Standardpopulation zu standardisieren. In der Regel wird dafür eine indirekte Standardisierung verwendet, bei der die beobachtete mit der erwarteten Anzahl interessierender Ereignisse verglichen wird.<sup>56</sup> Die relevante epidemiologische Größe ist dann das standardisierte Mortalitäts- oder

---

<sup>56</sup> Die erwartete Anzahl interessierender Ereignisse wird dabei für die tatsächlich beobachteten Fälle eines Leistungserbringers anhand der Rate einer Standardpopulation berechnet und nicht, wie bei der direkten Standardisierung, für eine hypothetische Standardpopulation an Fällen anhand der Rate des

Morbiditätsverhältnis (SMR, auch oft einfach nur  $O / E$ <sup>57</sup> genannt). Neben Indikatoren mit Berechnungsart „Rate“ und „ $O / E$ “ finden auch weitere Berechnungsarten Verwendung, wie beispielsweise die Kaplan-Meier-adjustierte Rate (siehe Abschnitt 20.3), das standardisierte Inzidenzverhältnis (siehe Abschnitt 20.2) und verschiedene Index-Konstruktionen.

#### 20.1.4 Bewertungsart

Auf Auswertungsebene der Leistungserbringer umfasst die statistische Auswertung von Qualitätsindikatoren meist auch den Vergleich der Ergebnisse des Leistungserbringers mit einem Referenzbereich und eine daraus resultierende Zuordnung in ein Klassifikationsschema (siehe Kapitel 16). Die statistische Methodik für diesen Vergleich wird hier als Bewertungsart bezeichnet und hängt wesentlich von der Zielsetzung, der Stichproben- sowie der Berechnungsart eines Indikators ab.

Liegt eine Vollerhebung vor und wird die enumerative Zielsetzung gewählt, kann das durch die Berechnungsart eines Indikators ermittelte Indikatorergebnis direkt, d. h. ohne Berücksichtigung statistischer Unsicherheit, mit dem definierten Referenzbereich verglichen werden. Die Klassifikation erfolgt dann danach, ob das Indikatorergebnis innerhalb oder außerhalb des Referenzbereichs liegt.

Bewertungen unter Verwendung einer Stichprobe bei enumerativer Zielsetzung sowie Bewertungen bei analytischer Zielsetzung erfordern jedoch die Berücksichtigung von statistischer Unsicherheit beim Vergleich eines Indikatorergebnisses mit einem Referenzbereich und stellen damit Entscheidungen unter Unsicherheit dar. Dabei ist in der Regel von Interesse, ob der Kompetenzparameter für den jeweiligen Qualitätsindikator im Referenzbereich liegt. Da der Kompetenzparameter unbekannt ist und nur über die beobachteten Daten im statistischen Sinne „geschätzt“ werden kann, erfolgt die Bewertung mithilfe einer Entscheidungsregel, die definiert, welche Ausprägungen der beobachteten Daten (für ratenbasierte Indikatoren beispielsweise welche Fallzahlen und Anzahl an interessierenden Ereignissen) welchen Kategorien des Bewertungsschemas zugeordnet werden.

Die Entscheidungsregel für die Klassifikation beobachteter Daten wird dabei über eine Verlustfunktion hergeleitet. Im Falle einer binären Klassifikation wird dafür der Verlust, der mit falsch positiven und falsch negativen Klassifikationen einhergeht, festgelegt.<sup>58</sup> Der mit einer Klassifikation einhergehende erwartete Verlust ergibt sich dann aus den mit den jeweiligen Fehlklassifikationen assoziierten Verlusten sowie deren Wahrscheinlichkeiten, welche wiederum anhand der beobachteten Daten geschätzt werden (Jensen und Nielsen 2007). Beispielsweise entspricht

---

Leistungserbringers (Rothman 2012). Speziell bei Leistungserbringern mit kleinen Fallzahlen stellt die indirekte Standardisierung deswegen oft das robustere Verfahren dar.

<sup>57</sup> Verhältnis der beobachteten Anzahl (*observed*,  $O$ ) eines interessierenden Ereignisses zur erwarteten Anzahl (*expected*,  $E$ ).

<sup>58</sup> Wenn das Entscheidungskriterium in der Unterscheidung besteht, ob ein zugrunde liegender Parameter innerhalb oder außerhalb eines Referenzbereichs liegt, liegt eine falsch positive Klassifikation dann vor, wenn ein zugrunde liegender Parameter als außerhalb des Referenzbereichs eingestuft wird, obwohl dieser innerhalb liegt. Eine falsch negative Klassifikation beschreibt wiederum den Fall, in dem ein zugrunde liegender Parameter, der außerhalb des Referenzbereichs liegt, als innerhalb des Referenzbereichs klassifiziert wird.

eine Verlustfunktion, bei der der Verlust nur davon abhängt, ob der Kompetenzparameter die durch den Referenzbereich definierten Anforderungen verfehlt, einer Klassifikationsregel, bei der auf eine statistisch signifikante Abweichung vom Referenzbereich geprüft wird. Alternativ ist es denkbar, den Verlust für falsch negative Entscheidungen von der Fallzahl abhängig zu machen oder das Ausmaß der Abweichung vom Referenzbereich in der Verlustfunktion selbst zu berücksichtigen. Diese unterschiedlichen Arten, eine Verlustfunktion zu definieren, führen zu unterschiedlichen Entscheidungsregeln. Da die Konsequenzen von Fehlklassifikationen kontextabhängig sein können, wird die Entscheidungsregel für die Bewertung beobachteter Daten kontextabhängig so gewählt, dass der aus den Entscheidungen resultierende erwartete Verlust minimiert wird. Statistische Details der Herleitung einer Entscheidungsregel werden in Anhang A anhand eines Einflussdiagramms detaillierter verdeutlicht, aus welchem hervorgeht, wie die Komponenten der Bewertung zusammenhängen und wie anhand dieses Modells eine Entscheidungsregel hergeleitet werden kann. Anschließend werden im Anhang drei konkrete Regeln der Auffälligkeitseinstufung basierend auf unterschiedlichen Verlustfunktionen hergeleitet. Die Auffälligkeitseinstufung aufgrund von statistischer Signifikanz wird in Anhang B für einen ratenbasierten Qualitätsindikator beispielhaft illustriert.

Da die Konsequenzen von Fehlklassifikationen im praktischen Rahmen der Qualitätssicherung eine Vielzahl an Akteuren (wie Patientinnen und Patienten, Leistungserbringer, Planungsbehörden und andere) betreffen, wird eine Quantifizierung und Abwägung der jeweiligen Folgen in der Praxis schwer durchführbar sein. Wenn die Herleitung einer konkreten Entscheidungsregel anhand einer solchen Abwägung nicht möglich ist und stattdessen beispielsweise anhand einer normativen Setzung erfolgt, dient das hier vorgestellte Konzept dazu, die aus der Setzung implizit resultierende Quantifizierung der Entscheidungskonsequenzen transparent zu machen. Das Konzept der Verlustfunktion ist daher für Bewertungen unter statistischer Unsicherheit von zentraler Bedeutung, da Entscheidungsregeln dadurch explizit hergeleitet oder die impliziten Annahmen normativer Setzungen transparent gemacht werden können.

## 20.2 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. Dabei sollen nur solche Unterschiede berücksichtigt werden, die direkt oder indirekt die Wahrscheinlichkeit eines indikatorspezifischen Ereignisses beeinflussen und die nicht vom Leistungserbringer selbst im Rahmen der zu bewertenden Leistung beeinflussbar sind. Die entsprechenden Eigenschaften der Fälle werden dabei als Risikofaktoren bezeichnet (vgl. Iezzoni 2013). Ein Beispiel sind patientenseitige Einflüsse wie Multimorbidität, da für multimorbide Patientinnen und Patienten häufig weniger positive Behandlungsergebnisse erreichbar sind als für weniger stark belastete Patientinnen und Patienten (Iezzoni 2012) und die Multimorbidität vom Leistungserbringer selbst im Rahmen der zu bewertenden Leistung nicht beeinflussbar ist. Aber auch Faktoren wie Alter und Geschlecht der Patientinnen und Patienten sind Variablen, die häufig in relevantem und signifikantem Zusammenhang mit dem Behandlungsergebnis stehen (Iezzoni 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 indikatorspezifischen Ereignissen 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 Risikoadjustierung trägt auch dazu bei, Fehlanreize im Sinne der Patientenselektion zu minimieren.

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. Insgesamt ist jede Risikoadjustierung mit einer Abwägung verbunden zwischen der Erfassung von wichtigen Variablen, die eine faire Gestaltung der vergleichenden Darstellung der Leistungserbringer ermöglichen, und dem damit verbundenen erhöhten Aufwand bei der Dokumentation durch die Leistungserbringer.

Im Folgenden werden die 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).

### **20.2.1 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, bzgl. 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 4).

Tabelle 4: 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 überdurchschnittlich viele Notfallpatientinnen und -patienten, ist daher auch eine insgesamt erhöhte Sterblichkeitsrate bei dieser Einrichtung zu erwarten. 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 bzgl. der Stratifizierungsvariablen unverzerrten Vergleich zwischen verschiedenen 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 16) 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, während die Stratifizierung immer kategoriale Risikofaktoren voraussetzt.

### 20.2.2 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 nicht risikoadjustierten 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  $\leq 60$  mmHg sowie ein Alter  $\geq 65$  Jahre. Für jeden bei einer Patientin oder einem Patienten auftretenden Risikofaktor wird ein Punkt vergeben, sodass in der Summe für jede Patientin und jeden Patienten zwischen 0 Punkte (geringster Schweregrad) und 4 Punkte (höchster Schweregrad) vergeben werden können.

Im nächsten Schritt kann dann ein risikoadjustierter Vergleich der Leistungserbringer anhand der ermittelten Punktwerte bzw. nach einer Einteilung der Punktwerte in Risikoklassen erfolgen, wobei hierzu Stratifizierungsverfahren verwendet werden. So werden für Patientinnen und Patienten mit ambulant erworbener Pneumonie die Ergebnisse zur Frühmobilisation 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 der Übertragbarkeit des Scores auf den Kontext der Qualitätssicherung genutzt werden.

### 20.2.3 Statistische Modellierung

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 die Eintrittswahrscheinlichkeit  $P$  für das Ereignis in Abhängigkeit der Risikofaktoren modelliert. Das Modell wird in der Regel auf einem Datensatz geschätzt, der als Referenzpopulation dient. Dieser Datensatz kann z. B. aus den QS-Daten des aktuellen oder des vorangegangenen Erfassungsjahres bestehen. Zur Vergrößerung der Datenbasis können auch zusätzlich Daten aus weiter zurückliegenden Verfahrensjahren bei der Schätzung verwendet werden.

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. Eine auf jede Fragestellung oder Situation passende, allgemeingültige Beurteilung der Modellgüte ist jedoch nicht sinnvoll.

Aus den fallspezifischen Wahrscheinlichkeiten ergibt sich für jeden Leistungserbringer eine *erwartete* Anzahl an Ereignissen (*expected*, E) 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 (*observed*, 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 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. Dem gegenüber steht ein erhöhter Aufwand bei der Durchführung der QS-Verfahren, wenn Regressionskoeffizienten jährlich neu berechnet werden und dementsprechend Anpassungen der Rechenregeln notwendig sind.

## 20.3 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. Beispielsweise ist ein mögliches Qualitätsziel für Follow-up-Indikatoren das möglichst seltene Auftreten von Revisionen nach Hüftendoprothesen-, Knieendoprothesen- und Herzschrittmacherimplantation innerhalb eines definierten Zeitraums nach einem zu bewertenden Ereignis (dem Ersteingriff).

### 20.3.1 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. Bei aktivem Follow-up wird (innerhalb festgelegter Zeiträume nach interessierendem Eingriff) das interessierende Merkmal der Patientin / des Patienten (z. B. Überleben) durch den Leistungserbringer nachverfolgt; dadurch liegen Informationen meist vollständig und unabhängig von der Ausprägung des interessierenden Merkmals vor. Beim passiven Follow-up ist dagegen von der Ausprägung des Merkmals der Patientin / des Patienten abhängig, ob Informationen zum weiteren Verlauf vorliegen (z. B. dadurch, dass eine erneute Operation im Rahmen eines stationären oder ambulanten Aufenthalts dokumentiert wurde) oder nicht. Weitere Beispiele des passiven Follow-up stellen die Verknüpfung von Daten aus den QS-Verfahren *Geburtshilfe* und *Neonatalogie* sowie sozialdatengestützte Follow-up-Indikatoren dar. Darüber hinaus sind zukünftig auch Follow-up-Indikatoren basierend auf weiteren Datenquellen wie Krebsregister denkbar.

Die beiden Arten von Follow-up-Indikatoren unterscheiden sich im Zeitpunkt, zu dem die Follow-up-Informationen für eine Auswertung zur Verfügung stehen. Bei Indikatoren mit aktivem Follow-up liegen die relevanten Informationen zu Ereignissen im Follow-up-Zeitraum 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 Follow-up-Zeitraums komplett vorliegen. Im Folgenden werden neben der Darstellungsweise von Follow-up-Daten (Abschnitt 20.3.2) die methodischen Grundlagen für zwei Auswertungsstrategien von Follow-up-Indikatoren beschrieben (Kohorten- und Periodensichtweise, Abschnitt 20.3.3). Diese unterscheiden sich hinsichtlich der Wahl der Grundgesamtheit sowie der statischen Auswertungsmethodik und in Bezug auf die Follow-up-Zeiträume. Beide Strategien können prinzipiell bei der Auswertung von Indikatoren mit passivem Follow-up angewandt werden. Indikatoren mit aktivem Follow-up sind dagegen generell auf die Kohortensichtweise begrenzt. Bei der Wahl der Auswertungsstrategie sind darüber hinaus Auswirkungen von Datenlieferfristen zu berücksichtigen.

### 20.3.2 Lexis-Diagramme

Die Daten eines Follow-up-Indikators lassen sich mittels eines sogenannten Lexis-Diagramms (Lexis 1875) illustrieren. Abbildung 17 zeigt beispielhaft die beobachteten Verläufe eines Follow-up-Indikators mit einem Follow-up-Zeitraum von einem Jahr in einem solchen Lexis-Diagramm. Auf der x-Achse ist die Kalenderzeit und auf der y-Achse die Follow-up-Zeit seit dem interessierenden Ereignis (in Tagen) dargestellt. In diesem Kontext wird die zeitliche Dimension eines Verlaufs durch Erst- und Folgeeingriff definiert; die Verbindungslinie zwischen den Zeitpunkten dieser Ereignisse, auch Lebenslinie genannt, stellt den Verlauf im Koordinatensystem dar. In diesem Beispiel wird dabei angenommen, dass die Datenerhebung nur bis zum 31. Dezember 2017 erfolgt ist, d. h., dass keine Verläufe 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 Follow-up-Zeitraum vorliegt. Die Rechtszensierung der Follow-up-Zeit 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 Follow-up-Zeiträume (durch Rechtszensierung, Linkstrunkierung oder konkurrierende Ereignisse) und stellen daher einen geeigneten Rahmen für die statistische Auswertungsmethodik von Follow-up-Indikatoren dar.

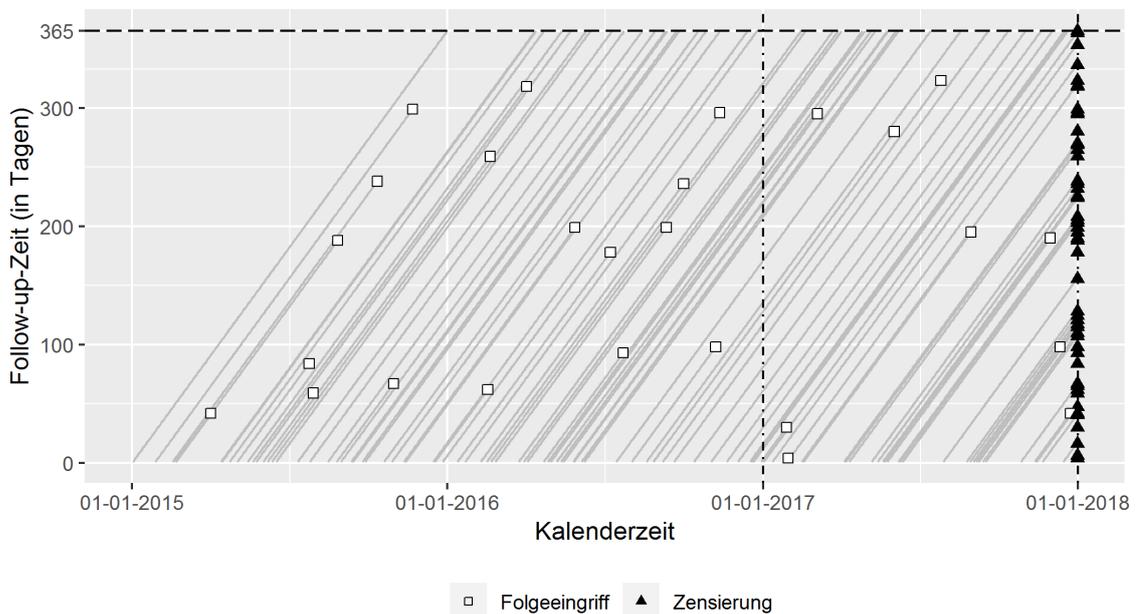


Abbildung 17: Beispielhafte Darstellung der Daten eines 1-Jahres-Follow-up-Indikators mit 100 Fällen in Form von Lebenslinien in einem Lexis-Diagramm

### 20.3.3 Auswertung nach Kohortensichtweise und nach Periodensichtweise

Grundsätzlich unterscheidet das IQTIG zwei Auswertungsstrategien für Follow-up-Indikatoren: Die einmalige Auswertung des gesamten Follow-up-Zeitraums nach dessen vollständiger Beobachtung (siehe Abbildung 18: Kohortensichtweise) sowie die Auswertung der Daten, die im aktuellen Erfassungsjahr beobachtet wurden (siehe Abbildung 18: Periodensichtweise).

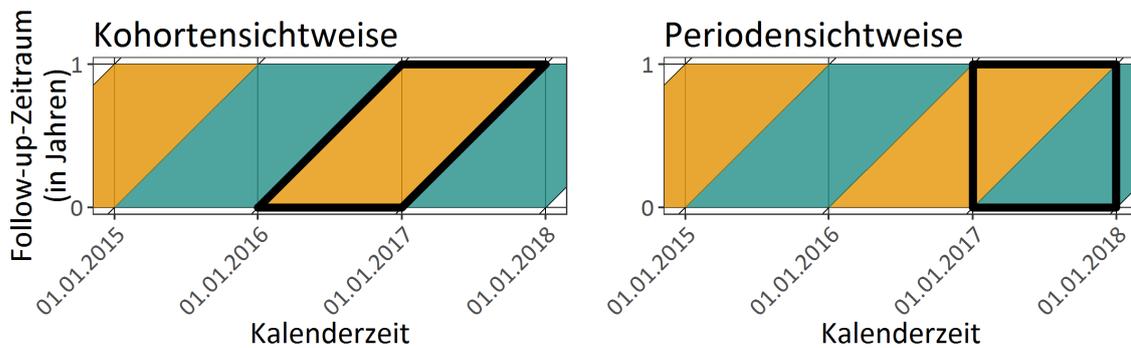


Abbildung 18: Lexis-Diagramme für einen 1-Jahres-Follow-up-Indikator in Kohortensichtweise und in Periodensichtweise

Die x-Achsen in Abbildung 18 stellen die kalendarische Zeit und die y-Achsen die vergangene Zeit nach dem interessierenden Ereignis (Follow-up-Zeitraum) dar. Die Lexis-Diagramme verdeutlichen die Kohorten- und die Periodensichtweise. Die möglichen Unterschiede des Risikos für das Eintreten eines Ereignisses in Abhängigkeit der Follow-up-Zeit kann als patientenseitige Eigenschaft gesehen werden und wird daher immer gleich in der Berechnung der Maßzahl des Qualitätsindikators berücksichtigt werden. Dies entspricht einer Risikoadjustierung für die Follow-up-Zeit.

Die Kohorten- und Periodensichtweisen unterscheiden sich darin, dass die Auswertung 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 18 sind die beiden betrachteten Zeiträume der Sichtweisen jeweils als schwarz umrandete Flächen dargestellt, alle Verläufe, die (vollständig oder teilweise) in dieser Fläche liegen, werden analysiert.

Bei Indikatoren, bei denen die Auswertungsstrategie der Kohortensichtweise entspricht, werden in der Regel alle interessierenden Ereignisse des aktuellsten Erfassungsjahres in die Analyse eingeschlossen, für die der definierte Follow-up-Zeitraum komplett beobachtet wurde. Diese Strategie hat den Vorteil, dass alle interessierenden Ereignisse aus einem einzigen Erfassungsjahr stammen. Der wesentliche Nachteil ist, dass der Follow-up-Zeitraum zum Zeitpunkt der Auswertung bereits vollständig beobachtet sein muss. Indikatoren mit langen Follow-up-Zeiträumen 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

gesamten aktuellen Erfassungsjahres als Follow-up-Zeitraum. Dies hat den Vorteil, dass alle im 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 eindeutig einem Erfassungsjahr zuordnen lassen.

Bei Indikatoren mit aktivem Follow-up werden Informationen zum Follow-up-Zeitraum erst nach dessen Ablauf von den Leistungserbringern übermittelt. Daher ist eine periodenweise Auswertungsstrategie, in die auch unvollständige Beobachtungszeiträumen eingehen, für diese Indikatoren in der Regel 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 Follow-up-Zeitraums indikatorspezifisch getroffen, da Follow-up-Indikatoren jeweils verschiedene Qualitätsziele haben können und damit auch unterschiedliche Auswertungsstrategien für unterschiedliche Indikatoren sinnvoll sein können. Für Follow-up-Indikatoren mit langem Follow-up-Zeitraum hat die Auswertungsstrategie basierend auf der Periodensichtweise den Vorteil, Daten auswerten zu können, ohne den gesamten Follow-up-Zeitraum abwarten zu müssen (wie z. B. bei Qualitätsindikatoren zur Aggregatlaufzeit bei Herzschrittmacher und Defibrillatoren). Bei Indikatoren mit kürzerem Follow-up-Zeitraum kann dagegen die Auswertungsstrategie basierend auf der Kohortensichtweise sinnvoller sein, beispielsweise, wenn die Daten der Indexeingriffe eines einzigen Erfassungsjahres ausgewertet werden sollen (wie z. B. bei Qualitätsindikatoren zu ungeplanten Revisionseingriffen bei Knie- und Hüftgelenksendoprothesen). Ist der Follow-up-Zeitraum wiederum sehr kurz (z. B. 30 oder 90 Tage), hat die Auswertungsstrategie basierend auf der Periodensichtweise den Vorteil, dass für die Auswertung kein ganzes Erfassungsjahr abgewartet werden muss, bis für die gesamte Kohorte der Follow-up-Zeitraum vollständig beobachtet wurde.

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 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 andere 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 der Notwendigkeit der 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 (2018) beschrieben.

## Glossar

### **Datenqualität**

Grad, in dem die zur Berechnung des Indikatorwerts verwendeten Daten den Vorgaben des Indikators hinsichtlich Vollzähligkeit, Vollständigkeit und Dateninhalten entsprechen.

### **Datenquelle**

Stelle, an der Daten für die Qualitätssicherung generiert oder erfasst werden. Zurzeit stehen folgende Datenquellen zur Verfügung: Dokumentation der Leistungserbringer (fallbezogen und einrichtungsbezogen), Sozialdaten bei den Krankenkassen und Befragung von Patientinnen und Patienten.

### **Datenvalidität**

Siehe Datenqualität

### **Erfassungsjahr**

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**

Punktschätzer, der durch Anwendung der Rechenregeln eines Qualitätsindikators auf die jeweiligen Daten resultiert.

### **Indikatorergebnis**

Das Ergebnis, das durch Anwendung der Rechenregeln eines Qualitätsindikators auf die jeweiligen Daten resultiert, einschließlich beobachtetem Punktschätzer (Indikatorwert) und Vertrauensbereichen.

### **Interessenkonflikt**

Gegebenheiten, die ein Risiko dafür schaffen, dass professionelles Urteilsvermögen oder Handeln einer Person, 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), die Patientinnen und Patienten medizinisch-pflegerisch versorgen.

**Messmodell (formatives, reflektives)**

Logische Beziehung zwischen einem theoretischen Konstrukt und den Indikatoren, die dieses abbilden sollen. Bei einem formativen Messmodell bestimmen die Indikatoren die Zusammensetzung des Konstrukts, bei einem reflektiven Messmodell spiegeln die gemessenen Indikatorwerte das Konstrukt als einzige, gemeinsame Ursache wider.

**Operationalisierung**

Festlegung, wie ein theoretisches Konzept messbar gemacht werden soll. Beispielsweise werden Qualitätsmerkmale durch Qualitätsindikatoren operationalisiert, also messbar gemacht.

**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 die Qualität der Gesundheitsversorgung in einen bestimmten Bereich beschreiben lässt.

Im Gegensatz zu Qualitätsdimensionen handelt es sich bei Qualitätsaspekten in der Methodik des IQTIG nicht um generelle Kategorien, sondern um spezifisch für den jeweiligen Versorgungsbereich abgeleitete Themen.<sup>59</sup>

**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).

---

<sup>59</sup> Der Begriff ist nicht gleichbedeutend mit den elf „grundlegenden Qualitätsaspekten“ der DIN EN 15224, die nach der Methodik des IQTIG Qualitätsdimensionen entsprechen.

**Qualitätsindikator**

Quantitative Größe, die Aussagen über die Erfüllung konkreter Qualitätsanforderungen mittels eines Messverfahrens und eines Bewertungskonzepts ermöglicht. Qualitätsindikatoren umfassen die drei Komponenten Qualitätsziel, Messverfahren und Bewertungskonzept.

**Qualitätsindikatorensatz**

Summe der Qualitätsindikatoren, die Versorgungsqualität in einem gemeinsamen Themenbereich erfassen sollen.

**Qualitätsmerkmal**

Eigenschaft der Versorgung (z. B. Erlangen von Gefähigkeit, Information über alternative Behandlungsmöglichkeiten), die mit Anforderungen (z. B. ein möglichst hoher Anteil gefähiger Patientinnen und Patienten nach Operation) verknüpft ist.

**Qualitätsmessung**

Empirische Beschreibung der Erfüllung von Qualitätsanforderungen, z. B. anhand von Qualitätsindikatoren.

**Qualitätsmodell**

Qualitätsaspekte (siehe Qualitätsaspekte) innerhalb eines Themenbereichs der Gesundheitsversorgung, für die Qualitätsindikatoren entwickelt werden sollen.

**Qualitätssicherung**

Bezeichnet im Gesundheitswesen alle Prozesse und Maßnahmen, durch die die Qualität der Versorgung von Patientinnen und Patienten gemessen, dargestellt und gewährleistet oder verbessert werden soll.

**Qualitätssicherung, externe**

Prozesse und Maßnahmen, durch die für ausgewählte Themenbereiche der Gesundheitsversorgung 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 Prozessen und Maßnahmen, die in Bezug auf einen bestimmten Themenbereich der Gesundheitsversorgung die Versorgungsqualität sichern und steigern sollen.

**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**

Bereich auf der Messskala eines Qualitätsindikators, der eine konkret zu erfüllende Anforderung für ein Qualitätsmerkmal darstellt (z. B. „Es wird eine Komplikationsrate unter 5 % gefordert.“).

**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.

**Steuerungsinstrument (Qualitätssteuerungsinstrument)**

Maßnahmen zur Durchsetzung bestimmter Qualitätsanforderungen an die Versorgung, die vom IQTIG anhand dreier grundsätzlicher Wirkmechanismen klassifiziert werden, über die die Ergebnisse von Qualitätsmessungen zu einer Steigerung der Versorgungsqualität führen können: durch Förderung (z. B. wie in Verfahren nach der DeQS-RL), durch qualitätsorientierte Auswahl von Leistungserbringern (z. B. durch Patientinnen und Patienten) und durch Anreize (z. B. Vergütungswegfall beim Nichterreichen von Mindestanforderungen).

**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.

**Handlungsanschlüsse für Qualitätsmessungen**

Entscheidungen und Handlungen, die auf Grundlage der Ergebnisse der Qualitätsmessungen eines QS-Verfahrens erfolgen sollen. Dabei kann es sich um Entscheidungen auf Mesoebene (z. B. über den Einsatz von Qualitätssicherungsmaßnahmen) oder auf Mikroebene (z. B. Auswahlentscheidungen von Patientinnen und Patienten) handeln.

**Versorgungsqualität**

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

# Anhang

## Anhang A: Datenbasierte Klassifikation von Leistungserbringerergebnissen

Im Folgenden werden drei Operationalisierungen der in Abschnitt 20.1.4 beschriebenen datenbasierten Klassifikationsverfahren hergeleitet. Dafür werden in Anhang A.1 zunächst die Komponenten der Bewertungsart betrachtet und anschließend Entscheidungsregeln basierend auf verschiedenen Annahmen hergeleitet (Anhang A.2) sowie deren Anwendung illustriert (Anhang A.3). Abschließend werden in Anhang A.4 die Annahmen und Limitationen des entscheidungstheoretischen Ansatzes beschrieben.

### Anhang A.1: Komponenten der Bewertungsart

Im Folgenden soll anhand eines ratenbasierten Qualitätsindikator im Rahmen der analytischen Zielsetzung illustriert werden, wie diese Komponenten zur Methodik der Bewertungsart beitragen. Abbildung 19 bildet die zugrunde liegenden Annahmen über die kausalen Zusammenhänge dieser Komponenten grafisch als Einflussdiagramm (Jensen und Nielsen 2007) ab:

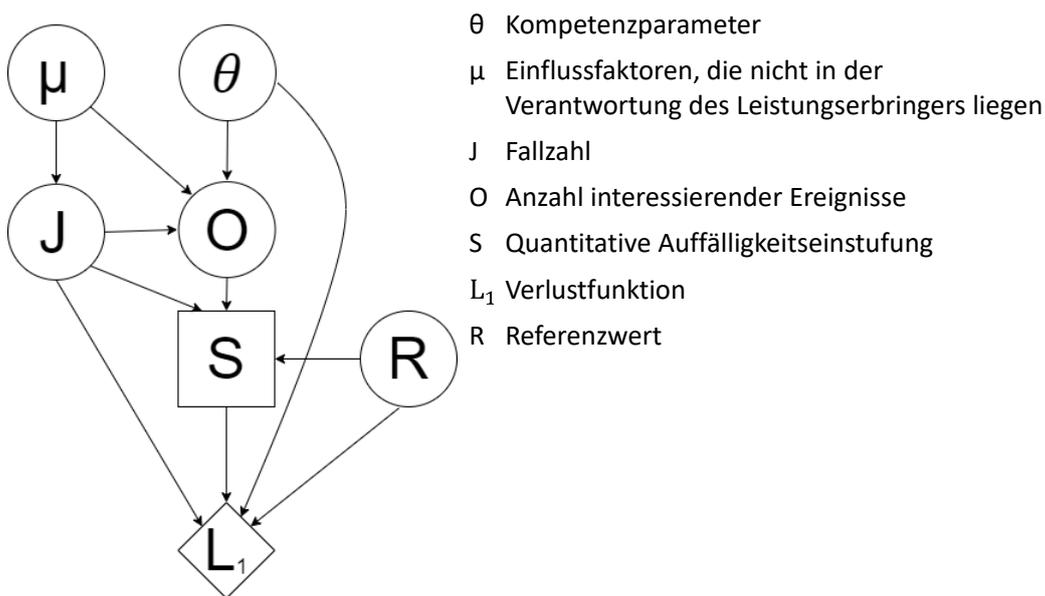


Abbildung 19: Vereinfachtes Einflussdiagramm für einen ratenbasierten Qualitätsindikator

In der Darstellung bezeichnet dabei  $\theta$  den Kompetenzparameter eines Leistungserbringers,  $\mu$  Einflussfaktoren, die nicht in der Verantwortung des LE liegen,  $O$  die Anzahl interessierender Ereignisse,  $J$  die Fallzahl,  $R$  den Referenzwert und  $S$  das Ergebnis der quantitativen (d. h. der datenbasierten) Klassifikation. Im einfachsten Fall ist eine Entscheidung zwischen zwei Klassifikationskategorien notwendig. Die Pfeile symbolisieren bedingte Abhängigkeiten. In Abbildung 19 wird beispielsweise davon ausgegangen, dass  $J$ ,  $\mu$  und  $\theta$  die Variable  $O$  probabilistisch festlegen und dass  $J$  und  $\theta$  voneinander unabhängig sind.

Für die Klassifikationsentscheidung  $S$  ist neben der zugrunde liegenden Kompetenz  $\theta$  und dem Referenzbereich eines Indikators eine Abwägung von Sensitivität und Spezifität notwendig. Diese Abwägung der Relevanz von falsch positiven und falsch negativen Klassifikationsentscheidungen wird im Rahmen einer sogenannten Verlustfunktion (in Abbildung 19 mit  $L_1$  bezeichnet) quantifiziert. Eine Möglichkeit, zu einer Klassifikationsentscheidung zu kommen, besteht darin, diejenige Kategorie zu wählen, mit welcher der kleinste erwartete Verlust (im Sinne der Verlustfunktion) verbunden ist (Jensen und Nielsen 2007). Bei Festlegung einer Klassifikationsregel für Entscheidungen unter Unsicherheit wird also immer explizit oder implizit auch eine Verlustfunktion festgelegt.

## Anhang A.2: Herleitung von Entscheidungsregeln anhand der Verlustfunktion

Ausgehend vom statistischen Rahmenkonzept aus Abschnitt 20.1 und einer expliziten Formulierung der in Abschnitt 20.1.4 eingeführten Verlustfunktion  $L_1$  werden im Folgenden drei mögliche Methoden für die Klassifikation von Leistungserbringerergebnissen in ein binäres Klassifikationsschema vorgestellt:

- die rechnerische Klassifikation
- die statistisch signifikante Klassifikation
- die statistisch relevante Klassifikation

Diese stellen unter unterschiedlichen Annahmen über das Zustandekommen der Ergebnisse, die Verlustfunktion und die Zielsetzung der Klassifikation jeweils die theoretisch optimale Vorgehensweise zur Klassifikation dar. Dafür wird im Folgenden beispielhaft von einem ratenbasierten Qualitätsindikator mit festem Referenzwert ausgegangen, der anhand der Daten eines Erfassungsjahres ausgewertet werden soll. Anhand dieses Qualitätsindikators werden der aus jeder Entscheidung resultierende Aufwand als Funktion des Kompetenzparameters, des Referenzwerts und der binären Klassifikation analysiert. Um die Parameter zu finden, die entsprechend der Verlustfunktion zu theoretisch optimalen Entscheidungen führen, wird die zugehörige Verlustfunktion nach dem vorgegebenen Zielkriterium minimiert (Jensen und Nielsen 2007). Im Folgenden wird als Zielkriterium die Minimierung des Erwartungswerts der Verlustfunktion verwendet. Es können aber auch andere Zielkriterien wie z. B. Minimierung des Worst-Case-Verlusts von Relevanz sein (siehe z. B. Parmigiani und Inoue 2009: Kapitel 7).

Im Rahmen einer entscheidungstheoretischen Betrachtungsweise führt die Minimierung des erwarteten Verlusts zur sogenannten optimalen Bayes-Klassifikation (Berger 2010). Diese stellt die theoretisch optimale Klassifikationsentscheidung in dem Sinne dar, dass jede andere Entscheidungsregel unter gleichen Annahmen zu den gleichen oder größeren erwarteten Verlusten führt. Im Rahmen eines Ratenindikators und unter den in Abschnitt 20.1.4 dargelegten Annahmen an den datengenerierenden Prozess ist die optimale Bayes-Klassifikation definiert als Lösung folgendes Optimierungsproblems:

$$s(o, J, R) = \operatorname{argmin}_{s \in S} E(L_1(s, J, R, \theta)) = \operatorname{argmin}_{s \in S} \int_0^1 L_1(s, J, R, \theta) f(\theta|o, J) d\theta,$$

wobei  $S$  in  $\{\text{ja, nein}\}$  und der Erwartungswert über alle Zufallsgrößen in der Verlustfunktion gebildet wird, d. h. über alle zum Zeitpunkt der Entscheidung unbekanntes Größen in der Verlustfunktion. Im konkreten Fall von Abbildung 19 wird nur über den stetigen Kompetenzparameter  $\theta$  integriert, für welchen zur expliziten Berechnung des Integrals eine Wahrscheinlichkeitsverteilung  $f(\theta|o, J)$  spezifiziert werden muss. Da für die Wahl der optimalen Entscheidung bei der quantitativen Klassifikation genau zwei Entscheidungsalternativen miteinander kontrastiert werden, gilt entsprechend, dass

$$s(o, J, R) = I\{E[L_1(s = \text{ja}, J, R, \theta)] < E[L_1(s = \text{nein}, J, R, \theta)]\},$$

wobei  $I(\cdot)$  die Indikatorfunktion bezeichnet. Hier wird die Konvention gewählt, dass bei gleichem erwartetem Verlust der beiden Entscheidungsalternativen die Entscheidung zugunsten des Leistungserbringers für die Alternative „nein“ getroffen wird.

Ausgangspunkt für alle weiteren entscheidungstheoretischen Betrachtungen ist die Formulierung unterschiedlicher Verlustannahmen anhand einer Vier-Felder-Tafel. Die in Tabelle 5 dargestellte Vier-Felder-Tafel stellt für jede Kombination der wahren Werte des Leistungserbringers und der Klassifikationsentscheidung die damit einhergehenden Aufwände dar. Die Befüllung dieser Vier-Felder-Tafel mit konkreten Zahlen für diese Aufwände ist nicht trivial, weil viele Akteure mit unterschiedlichen Interessen involviert sind. Wie diese konkreten Akteure und ihre Präferenzen zu berücksichtigen sind, hängt auch mit dem konkreten Anwendungskontext der Klassifikation zusammen. Im Rahmen der Qualitätsförderung fallen beispielsweise Aufwände für die Durchführung von Stellungnahmeverfahren und Maßnahmen der Qualitätsverbesserung an, anders als für den Anwendungsfall der Berichterstattung von Ergebnissen. Unterschiedliche Aufwände in dieser Vier-Felder-Tafel führen zu unterschiedlichen Entscheidungsregeln. Dies wird im Folgenden erläutert.

Tabelle 5: Vier-Felder-Tafel möglicher Aufwandsannahmen des binären Klassifikationsproblems

		wahrer Wert des Leistungserbringers	
		$\theta \leq R$	$\theta > R$
<b>Ergebnis der quantitativen Klassifikation</b>	$s(o, J, R) = \text{nein}$	$A_{0,0}(J, \theta, R)$	$A_{0,1}(J, \theta, R)$
	$s(o, J, R) = \text{ja}$	$A_{1,0}(J, \theta, R)$	$A_{1,1}(J, \theta, R)$

$A_{0,0}(J, \theta, R)$  sind dabei die Aufwände, die bei einer korrekt-negativen Klassifikation in Abhängigkeit von  $J, \theta$  und  $R$  entstehen,  $A_{1,0}(J, \theta, R)$  jene einer falsch positiven Klassifikation usw. Für den Kontext der Qualitätsförderung wird im Folgenden dargestellt, wie basierend auf Annahmen für die Aufwände in Tabelle 6 Entscheidungsregeln ohne Berücksichtigung von Unsicherheit (rechnerische Klassifikation) und mit Unsicherheitsberücksichtigung (statistisch signifikante Klassifikation) hergeleitet werden können. Die statistisch relevante Klassifikation wird anschließend als Alternative zur statistisch signifikanten Klassifikation hergeleitet, bei der ebenfalls statistische Unsicherheit berücksichtigt wird.

Für den Kontext der Qualitätsförderung, in dem eine Klassifikation als auffällig im Rahmen der datenbasierten Klassifikation in der Regel eine qualitative Bewertung nach sich zieht, wird sowohl für falsch positive ( $A_{1,0}$ ) als auch für richtig positive Klassifikationsentscheidungen ( $A_{1,1}$ ) angenommen, dass für die Durchführung der qualitativen Bewertung und möglicher Maßnahmen der Qualitätsverbesserung Aufwände entstehen, die im Folgenden als  $\zeta$  bezeichnet werden. Von diesen wird vereinfachend angenommen, dass sie unabhängig von der Anzahl der behandelten Fälle  $J$ , dem zugrunde liegenden Parameter  $\theta$  und dem Referenzwert  $R$  sind. Während für eine richtig negative Klassifikation keine Aufwände angenommen werden, führt eine falsch negative Klassifikation eines Leistungserbringers dazu, dass Aufwände (im Folgenden  $fn$ ) entstehen, beispielsweise durch nicht getroffene Maßnahmen zur Qualitätsförderung. Weiterhin wird angenommen, dass  $fn > \zeta$ , d. h., die Aufwände für eine falsch negative Klassifikation sind höher als die eines Stellungnahmeverfahrens. Mit  $A_{1,0} = 0$ ,  $A_{1,0} = A_{1,1} = \zeta$  und  $A_{0,1} = fn$  ergeben sich folgende Aufwandsannahmen:

Tabelle 6: Vier-Felder-Tafel möglicher Aufwandsannahmen des binären Klassifikationsproblems im Kontext der Qualitätsförderung

		wahrer Wert des Leistungserbringers	
		$\theta \leq R$	$\theta > R$
<b>Ergebnis der quantitativen Klassifikation</b>	$s(o, J, R) = \text{nein}$	0	$fn$
	$s(o, J, R) = \text{ja}$	$\zeta$	$\zeta$

Unterschiedliche Aufwandsannahmen für falsch negative Klassifikationen (d. h. nicht entdeckte Qualitätsdefizite) führen zu unterschiedlichen optimalen Klassifikationsmethoden. Im folgenden Abschnitt wird gezeigt, dass die sogenannte rechnerische Klassifikation entscheidungstheoretisch optimal ist, unter der Annahme, dass die Schätzung des Kompetenzparameters  $\theta$  keiner Unsicherheit unterliegt (enumerative Zielsetzung, also  $\theta = o/J$ ) und dass der Aufwand  $fn$  unabhängig von der Fallzahl  $J$  des Leistungserbringers und dem Ausmaß der Überschreitung ( $\theta - R$ ) des Referenzbereichs ist. Unter den gleichen Aufwandsannahmen, aber mit Berücksichtigung statistischer Unsicherheit (analytische Zielsetzung) wird in Anhang A.2.2 gezeigt, dass die statistisch signifikante Klassifikation die optimale Klassifikationsmethode darstellt. Schließlich wird in Anhang A.2.3 betrachtet, welches die optimale Klassifikationsmethode ist, unter der Annahme, dass der Aufwand  $fn$  sowohl vom Ausmaß der Überschreitung des Referenzbereichs als auch von der Fallzahl des Leistungserbringers abhängt.

### Anhang A.2.1: Rechnerische Klassifikation

Betrachtet wird ein ratenbasierter Qualitätsindikator mit  $o$  als beobachtete Anzahl an interessierenden Ereignissen,  $J$  als Fallzahl an behandelten Patientinnen und Patienten im betrachteten Zeitraum und  $R$  als Referenzwert des Indikators, d. h. mit Referenzbereich  $[0, R]$ . Die rechnerische Klassifikation ist definiert als Entscheidungsstrategie

$$s_{\text{rech}}(o, J, R) = I\left(\frac{o}{J} > R\right),$$

wobei  $I(A)$  die Indikatorfunktion darstellt, welche 1 (bzw. „ja“) ist, falls das Boolesche Argument  $A$  erfüllt ist<sup>60</sup>, und 0 (bzw. „nein“) ist, falls  $A$  nicht erfüllt ist.

Sofern unter der Annahme einer enumerativen Zielsetzung Stochastizität keine Rolle spielt, gilt

$$P\left(\theta > R \mid \theta = \frac{o}{J}\right) = I\left(\frac{o}{J} > R\right),$$

d. h., die Kompetenz des Leistungserbringers ist entweder sicher über dem Referenzwert  $I\left(\frac{o}{J} > R\right) = 1$  oder sicher unterhalb des Referenzwerts  $I\left(\frac{o}{J} > R\right) = 0$ . Für diesen Fall der enumerativen Zielsetzung ist die optimale Entscheidungsregel daher:

$$s_{\text{rech}}\left(o, J, R \mid \theta = \frac{o}{J}\right) = I\left(\frac{o}{J} > R\right).$$

Mit anderen Worten ist die rechnerische Klassifikation optimal, wenn es keine Unsicherheit bei der Bestimmung des Kompetenzparameters gibt, d. h. im Rahmen des vorgestellten Rahmenkonzepts bei enumerativer Zielsetzung mit Vollerhebung.

### Anhang A.2.2: Statistisch signifikante Klassifikation

Bei der statistisch signifikanten Klassifikation wird, anders als bei der rechnerischen Klassifikation, davon ausgegangen, dass die beobachtete Anzahl an interessierenden Ereignissen die Realisation einer Zufallsvariable und der zugrunde liegende Parameter  $\theta$  eines Leistungserbringers nicht direkt beobachtbar ist, sondern anhand der beobachteten Daten im inferenzstatistischen Sinne geschätzt werden muss.

Ausgehend von den Aufwandsannahmen in der Vier-Felder-Tafel aus Tabelle 6 werden die gleichen Aufwandsannahmen wie für die rechnerische Klassifikation gestellt. Dabei ist  $fn > 0$  wieder der Aufwand einer falsch negativen Klassifikation,  $\zeta > 0$  der Aufwand eines Stellungnahmeverfahrens. Wieder wird davon ausgegangen, dass der Aufwand einer Fehlklassifikation des Kompetenzparameters immer gleich ist, d. h. nicht von Eigenschaften des Leistungserbringers, wie z. B. der Fallzahl und dem Ausmaß der Überschreitung des Referenzbereichs, abhängen. Zusätzlich wird, anders als bei der rechnerischen Klassifikation, davon ausgegangen, dass der Parameter  $\theta$  nicht direkt beobachtbar ist, sondern nur aus den beobachteten Daten (und möglichen Vorinformationen) inferenzstatistisch geschätzt werden kann. Diese Schätzung ist mit Unsicherheit behaftet, weshalb die Verteilung der Wahrscheinlichkeiten für den Parameter  $\theta$  von Interesse ist. Für die Klassifikationsentscheidung zeigt sich, dass die Wahrscheinlichkeit  $P(\theta \leq R \mid o, J, \theta)$ , d. h. die Evidenz für das Vorliegen erwartbarer Qualität, die relevante Größe für die Entscheidung ist. Aus dieser und den oben angenommenen Aufwänden für Fehlklassifikationen ergeben sich die beiden erwarteten Verluste als:

$$E[l_{\text{stat.sig}}(o, J, R, \theta, s = \text{ja})] = \zeta$$

und

$$E[l_{\text{stat.sig}}(o, J, R, \theta, s = \text{nein})] = fn \cdot P(\theta > R).$$

<sup>60</sup> Im obigen Fall ist das Boolesche Argument  $A$  gleich  $o/J > R$ .

Daraus ergibt sich

$$s_{\text{stat.sig}}(o, J, R, \alpha) = I(\zeta < fn \cdot (1 - P(\theta \leq R))) = I\left(P(\theta \leq R) < \frac{fn - \zeta}{fn}\right).$$

Bei der statistisch signifikanten Klassifikation wird also die Wahrscheinlichkeit für die Einhaltung des Referenzbereichs  $P(\theta \leq R|o, J, \theta)$  mit dem Aufwandsverhältnis  $\frac{fn - \zeta}{fn}$  verglichen. Wird die Entscheidungsregel  $I(P(\theta \leq R|o, J, \theta) < \alpha)$  mit dem Schwellenwert  $\alpha \geq 0$  verwendet, dann ist diese Regel für die oben angenommene Verlustfunktion optimal, wenn der Schwellenwert  $\alpha = (fn - \zeta)/fn$  gewählt wird. Die Wahl von  $\alpha$  im Rahmen der statistisch signifikanten Klassifikation kann daher auch so interpretiert werden, dass über eine Festlegung des Aufwandsverhältnisses zwischen dem Aufwand eines Stellungnahmeverfahrens und dem Aufwand einer falsch negativen Klassifikation ein  $\alpha$  festgelegt wird. Der Fall  $\alpha = 0,05$  impliziert daher beispielsweise, dass das Verhältnis des Aufwands eines Stellungnahmeverfahrens gegenüber einer falsch negativen Klassifikation genau 19 zu 20 ist. Mit dieser Wahl von  $\alpha = 0,05$  werden somit dem Übersehen eines Qualitätsdefizits nur wenig mehr Aufwände zugeschrieben als dem Aufwand eines Stellungnahmeverfahrens. Wird der Aufwand für das Übersehen eines Qualitätsdefizits dagegen beispielsweise als doppelt so hoch bewertet wie der Aufwand eines Stellungnahmeverfahrens, so führt ein  $\alpha$  von  $(2\zeta - \zeta)/(2\zeta) = 0,5$  zur optimalen Entscheidungsregel.

### Anhang A.2.3: Statistisch relevante Klassifikation

Im vorigen Abschnitt wurde angenommen, dass der Aufwand für eine falsch negative Klassifikation eines Leistungserbringers fix ist und nicht von weiteren Faktoren abhängt, wie z. B. der behandelten Anzahl an Patienten und Patientinnen oder der Größe der Abweichung vom Referenzwert. Alternativ kann angenommen werden, dass die falsch negative Klassifikation eines Leistungserbringers, der im nächsten Beobachtungszeitraum viele Fälle behandeln wird, aus Sicht der Patienten und Patientinnen höhere Aufwände verursacht als jene eines Leistungserbringers mit wenigen Fällen. Gleiches gilt auch für die Größe der Abweichung des Kompetenzparameters vom Referenzwert bei zwei Leistungserbringern mit gleicher Fallzahl: Ist die Abweichung vom Referenzwert sehr groß, dann ist die Anzahl an betroffenen Patienten und Patientinnen größer als bei einem Leistungserbringer mit gleicher Fallzahl, aber mit geringer Abweichung vom Referenzwert. In beiden Fällen liegt die Annahme zugrunde, dass das Stellungnahmeverfahren zu einer Aufdeckung von möglichen Qualitätsdefiziten führt und damit zu einer Verbesserung der Behandlungsqualität für alle nachfolgend behandelten Patientinnen und Patienten. Die „Aufdeckung“ eines Qualitätsdefizits bei Leistungserbringern mit großer Fallzahl wäre in dieser Betrachtungsweise dabei wichtiger als die „Aufdeckung“ eines (gleich großen) Qualitätsdefizits bei Leistungserbringern mit kleiner Fallzahl.

Eine alternative Entscheidungsregel in dieser Betrachtungsweise beruht nicht allein auf der statistischen Evidenz für die Einhaltung des Referenzbereichs, d. h. auf  $P(\theta \leq R|o, J, \theta)$ , sondern auf der *Relevanz* der Abweichung des Kompetenzparameters vom Referenzbereich für Patienten und Patientinnen. Die Relevanz hängt von der Größe der Abweichung vom Referenzwert sowie der Anzahl an betroffenen Patienten und Patientinnen beim Leistungserbringer in der Zukunft ab. Selbst wenn für die Anzahl betroffener Patientinnen und Patienten nur die Fallzahl im

nächsten Beobachtungszeitraum, d. h. die Fallzahl im nächsten Erfassungsjahr als Metrik genommen wird, ist diese zum Zeitpunkt der Klassifikationsentscheidung unbekannt. Vereinfachend wird daher im Folgenden angenommen, dass die Fallzahl  $J$  im aktuellen Beobachtungszeitraum ein guter Schätzer für die Fallzahl im nächsten Beobachtungszeitraum ist. Die mit einer Fehlklassifikation verbundenen Aufwände aus patientenzentrierter Sichtweise lassen sich dann wie in untenstehender Matrix als abgewandelte Version der Vier-Felder-Tafel in Tabelle 6 quantifizieren:

Tabelle 7: Aufwandsannahmen für die statistisch relevante Klassifikationsmethode

		wahrer Wert des Leistungserbringers	
		$\theta \leq R$	$\theta > R$
<b>Ergebnis der quantitativen Klassifikation</b>	$s_{\text{stat.rel}}(o, J, R) = \text{nein}$	0	$fn \cdot (\theta - R)_+ \cdot J$
	$s_{\text{stat.rel}}(o, J, R) = \text{ja}$	$\zeta$	$\zeta$

In dieser Darstellung ist  $(\theta - R)_+ \cdot J$  die Anzahl an Fällen mit unerwünschtem Ereignis, die über die vom Referenzwert  $R$  tolerierte Anzahl an Fällen mit unerwünschtem Ereignis hinausgeht, mit  $(\theta - R)_+ := \max(0, \theta - R)$ . Des Weiteren ist  $fn > 0$  der mit jedem dieser Fälle assoziierte Aufwand und  $\zeta > 0$  wiederum der Aufwand eines Stellungnahmeverfahrens. Vereinfachend kann die Verlustfunktion auch als  $fn \cdot (\theta - R)_+ \cdot J$  bei  $s_{\text{stat.rel}} = \text{nein}$  und  $\zeta$  bei  $s_{\text{stat.rel}} = \text{ja}$  dargestellt werden.

Die optimale Entscheidungsregel lässt sich wieder aus den beiden Verlustfunktionen herleiten als Entscheidungsregel

$$s_{\text{stat.rel}}(o, J, R, \theta) = \mathbb{I} \left( \zeta < \int fn \cdot (\theta - R)_+ \cdot J \cdot f(\theta|o, J) d\theta \right),$$

mit  $f(\theta|o, J)$  als Dichtefunktion der Wahrscheinlichkeitsverteilung von  $\theta$  (gegeben die beobachteten Daten  $o, J$ ). Das heißt, es wird der Aufwand für die Durchführung eines Stellungnahmeverfahrens  $\zeta$  dem erwarteten Aufwand für die Anzahl an Patientinnen und Patienten mit unerwünschtem Ereignis, die über die vom Referenzwert  $R$  tolerierte Anzahl hinausgeht, gegenübergestellt. Im Folgenden wird dabei vereinfachend  $fn = 1$  gesetzt, weil die optimale Klassifikationsmethode lediglich vom Aufwandsverhältnis von  $\zeta/fn$  abhängt. Wird die Grenze für quantitative Klassifikation als  $\zeta = 1$  gesetzt, werden daher alle Leistungserbringer quantitativ auffällig, bei denen mindestens ein Fall mit unerwünschtem Behandlungsergebnis über die vom Referenzwert tolerierte Anzahl hinaus erwartet wird.

### Anhang A.3: Illustration der Konsequenzen unterschiedlicher Entscheidungsregeln zur Klassifikation

Die Konsequenzen der drei oben dargestellten Entscheidungsregeln werden im Folgenden beispielhaft anhand eines ratenbasierten Qualitätsindicators mit Referenzbereich  $\leq 15\%$  illustriert.

#### Anhang A.3.1: Schwellenwerte der Entscheidungsregeln

Die Entscheidungsregeln werden dabei zunächst hinsichtlich ihrer Schwellenwerte, die zu einer Bewertung als auffälliges Leistungserbringerergebnis führen, verglichen. Um die Fallzahlabhängigkeit der Entscheidungsregeln explizit darzustellen, wird bei vorgegebenen  $J$  und Referenzwert  $R$  das kleinste  $o$  berichtet, welches zu einer Einstufung als auffällig führt, d. h.

$$o_{\min}(J, R) = \operatorname{argmin}_{0 \leq o \leq J} \{s(o, J, R) = \text{ja}\}.$$

In Abbildung 20 sind diese Schwellenwerte als Funktion von  $J$  bei festgelegtem Wert  $R = 15\%$  in einem sogenannten Funnelpplot dargestellt (Spiegelhalter et al. 2012) – jeweils für die rechnerische, die statistisch signifikante ( $\alpha = 5\%$ ) sowie die statistisch relevante ( $\zeta = 2$ ) Klassifikation.

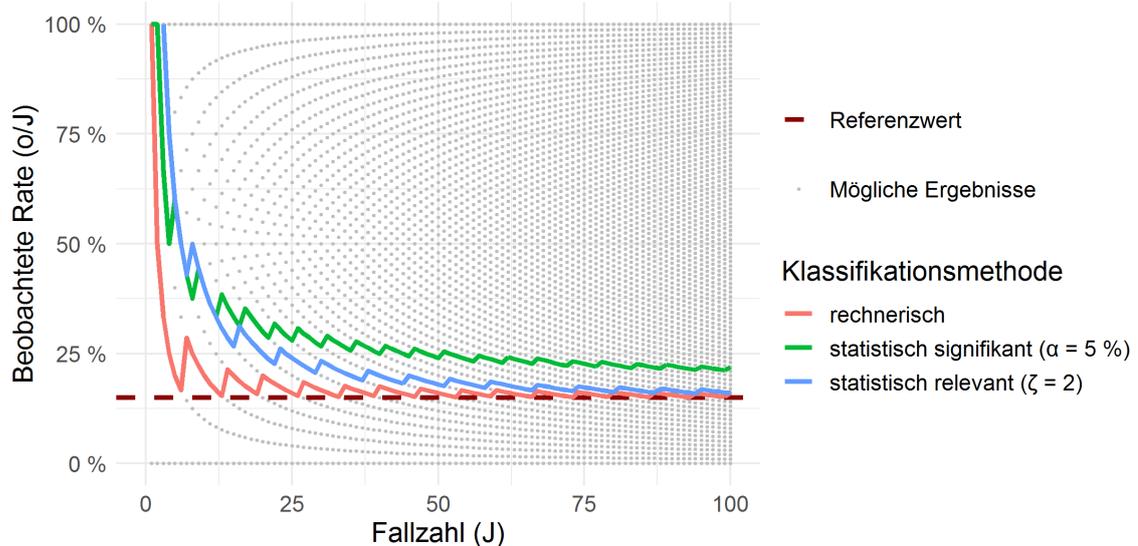


Abbildung 20: Funnelpplot für die Klassifikation der drei Entscheidungsregeln für einen Indikator mit Referenzbereich [0–15%]

Für eine gegebene Fallzahl  $J$  sind dabei nur beobachtete Raten  $0, 1/J, 2/J, \dots, 1$  möglich, die als Punkte in der Grafik dargestellt werden. Liegt ein beobachtetes Leistungserbringerergebnis auf oder oberhalb der entsprechenden Linie für den Schwellenwert der Methode, wird dieser Leistungserbringer als quantitativ auffällig in dem entsprechenden Indikator klassifiziert. Aus der Abbildung wird deutlich, dass sowohl die rechnerische als auch die statistisch signifikante Klassifikation mit steigender Fallzahl gegen den Referenzwert  $R$  konvergieren. Die Schwellenwerte der rechnerischen Klassifikation liegen dabei durchgängig unter jenen der statistisch signifikanten Klassifikation; dies ist aber nicht grundsätzlich so, sondern von der Wahl von  $\alpha$  abhängig.

Die statistisch relevante Klassifikation ist bei dem betrachteten Indikator mit  $R = 15\%$  bis zu einer Fallzahl von  $J = 8$  gleich oder weniger streng als die statistisch signifikante Klassifikation. Ab dieser Fallzahl sinken die Schwellenwerte der statistisch relevanten Klassifikation jedoch schneller als jene der statistisch signifikanten Klassifikation, d. h., für diese Fallzahlen führt die statistisch relevante Klassifikation zu strengeren Schwellenwerten. Der Grund hierfür liegt in der Wahl der Verlustfunktion: Für die statistisch signifikante Klassifikation relevant ist lediglich eine möglichst sichere Klassifikation des Kompetenzparameters des Leistungserbringers, d. h. ob  $\theta \leq R$  oder  $\theta > R$ , egal, wie groß eine potenzielle Abweichung vom Referenzwert ist bzw. wie viele Patienten und Patientinnen davon betroffen sind. Die statistisch relevante Klassifikation gewichtet dagegen die Relevanz der Abweichung für Patienten und Patientinnen höher.<sup>61</sup>

### Anhang A.3.2: Sensitivität und Spezifität der Entscheidungsregeln

Im Folgenden werden die theoretische Sensitivität und Spezifität der Entscheidungsregeln anhand des beispielhaften Qualitätsindikators illustriert. Als Sensitivität wird dabei die Wahrscheinlichkeit bezeichnet, dass ein Leistungserbringer mit  $\theta > R$  auch tatsächlich als quantitativ auffällig mit der verwendeten Methode eingestuft wird, basierend auf den beobachteten Daten des Leistungserbringers. Spezifität bezeichnet wiederum die Wahrscheinlichkeit, dass ein Leistungserbringer mit  $\theta \leq R$  auch tatsächlich als quantitativ unauffällig mit der jeweiligen Methode eingestuft wird.

Für einen detaillierteren Vergleich von Sensitivität und Spezifität der Entscheidungsregeln werden im Folgenden vier Szenarien angenommen, bei denen ein Leistungserbringer jeweils die zugrunde liegende Rate von  $\theta = 10\%$ ,  $\theta = 15\%$ ,  $\theta = 20\%$  und  $\theta = 25\%$  hat. Für diese vier Szenarien wird die Wahrscheinlichkeit einer quantitativ auffälligen Einstufung für die oben diskutierten Klassifikationsmethoden berechnet, jeweils für Fallzahlen zwischen  $J = 1, \dots, 1000$ . Die oberen beiden Darstellungen in Abbildung 21 zeigen dabei die Wahrscheinlichkeit falsch positiver Klassifikationen ( $1 - \text{Spezifität}$ ) der Entscheidungsregeln, da die wahren zugrunde liegenden Raten innerhalb des Referenzbereichs liegen. Die unteren beiden Abbildungen zeigen die Wahrscheinlichkeit richtig positiver Klassifikationen (Sensitivität). Verglichen werden wieder die rechnerische Klassifikation, die statistisch signifikante Klassifikation mit  $\alpha = 5\%$  sowie die statistisch relevante Klassifikation mit  $\zeta = 2$ .<sup>62</sup>

<sup>61</sup> Konsequenzen der Klassifikationsmethoden bei anderen Wahlen von  $R$ ,  $\alpha$  oder  $\zeta$  sowie eine Tabelle von  $\sigma_{\min}(J, R)$  als Funktion der Fallzahl lässt sich mit der unter <https://igtig.shinyapps.io/funnelplot/> abrufbaren interaktiven Funnelplot-App illustrieren.

<sup>62</sup> Eine interaktive Version der Abbildung, bei der sowohl  $R$ ,  $\alpha$  als auch  $\zeta$  variiert werden können, ist als Shiny-App unter [https://igtig.shinyapps.io/sensitivitaet\\_spezifitaet/](https://igtig.shinyapps.io/sensitivitaet_spezifitaet/) zu finden.

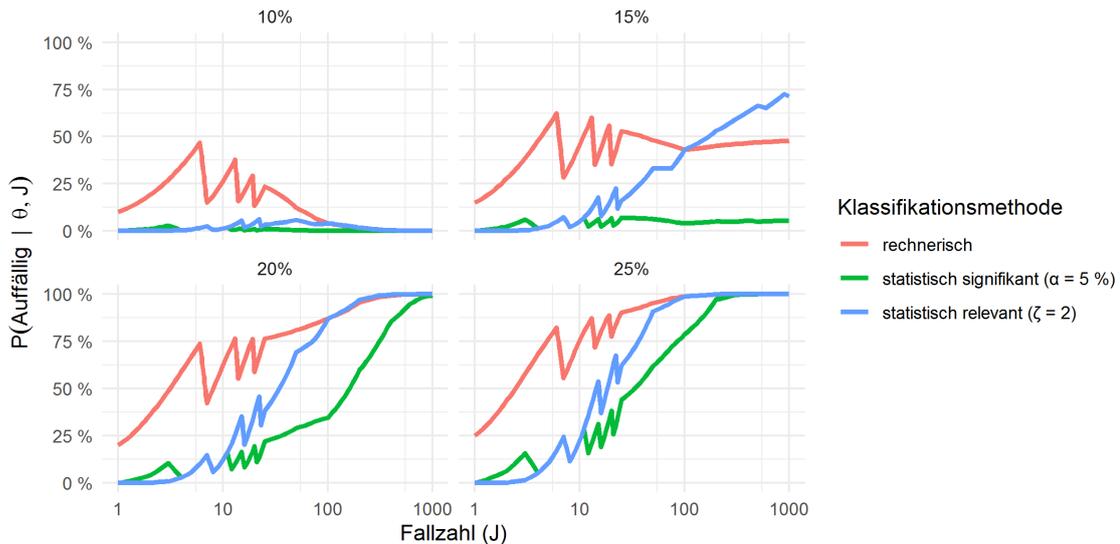


Abbildung 21: Vergleich von Sensitivität und 1-Spezifität für die drei Methoden der Klassifikation. Die Fallzahl auf der x-Achse ist dabei auf einer logarithmischen Skala dargestellt.

Der Fall, in dem die wahre zugrunde liegende Rate genau auf dem Referenzwert liegt, d. h.  $\theta = 15\%$ , ist insofern besonders, als die Methodik der statistisch signifikanten Klassifikation so ausgelegt ist, dass in diesem Fall (im Mittel über alle Fallzahlen  $J$ ) die Wahrscheinlichkeit einer auffälligen Einstufung genau  $\alpha$  entspricht, d. h., mit  $\alpha = 5\%$  liegt diese bei  $5\%$ . Die Spezifität ist dabei wie auch im Fall  $\theta = 10\%$  im Mittel über alle Fallzahlen bei der statistisch signifikanten Klassifikation sehr groß. Bei der statistisch relevanten Klassifikation dagegen ist die Spezifität bei  $\theta = 10\%$  ebenfalls groß, bei  $\theta = 15\%$  jedoch besonders bei großen Fallzahlen sehr klein und konvergiert mit steigender Fallzahl gegen 0. Umgekehrt hat die statistisch signifikante Klassifikation bei großen Fallzahlen die geringste Sensitivität der hier verglichenen Methoden. Der Grund für die schlechte Spezifität der statistischen Relevanz bei  $\theta = 15\%$  für Leistungserbringer mit großen Fallzahlen ist, dass die Leistungserbringer in diesem Grenzfall bei der statistischen relevanten Klassifikation schnell eine relevante Anzahl an Patienten und Patientinnen haben können, die über den Referenzwert hinaus ein unerwünschtes Qualitätsergebnis hatten. Hat ein solcher Leistungserbringer jedoch einen Kompetenzparameter, der nur leicht unter  $R$  liegt, dann ist auch hier die Spezifität gut. Mit anderen Worten: Es spielt bei der statistischen Relevanz weniger eine Rolle, ob nun genau  $\theta \leq R$  ist oder nicht, sondern der mit der Fallzahl gewichtete Abstand zum Referenzwert ist von Interesse.

## Anhang A.4: Limitationen

Grundsätzlich fordert der oben vorgestellte Verlustfunktionsansatz eine explizite Festlegung des Aufwandsverhältnisses von Stellungnahmeverfahren gegenüber nicht entdeckten Qualitätsdefiziten. Die Komplexität des Bewertungsprozesses und die vielen Akteure mit unterschiedlichen Aufwandsperspektiven stellen hier eine Herausforderung für die Festlegung dar. Gerade deshalb führt der entscheidungstheoretische Rahmen mit seinen Vereinfachungen zu Transparenz über die getroffenen Annahmen und kann z. B. bei der Interpretation der Schwellenwerte der Klassifikationsmethoden helfen. Konkrete Festlegungen für Werte von  $\alpha$  bzw.  $\zeta$  werden zudem erschwert, wenn sich der datenbasierten Klassifikation ein zweiter Bewertungsschritt anschließt, wie z. B. ein Stellungnahmeverfahren. In diesem Fall hängt die Wahl der Parameter von  $\alpha$  und  $\zeta$  nicht nur von der Abwägung der Aufwände ab, sondern auch von Faktoren wie der Güte eines Qualitätsindikators. Beispielsweise kann es bei einem Indikator, bei dem aufgrund einer ungenauen Datenerfassung oder Operationalisierung viele im ersten Schritt als auffällig klassifizierte Ergebnisse im anschließenden Schritt anders bewertet werden, sinnvoll sein, dies bei der Wahl von  $\alpha$  bzw.  $\zeta$  im ersten Schritt zu berücksichtigen (Raats und Moors 2003).

## Anhang B: Illustration der statistisch signifikanten Klassifikation als Bewertungsart

Im Folgenden wird das Vorgehen bei der Berücksichtigung von Unsicherheit anhand der statistisch signifikanten Klassifikation (im Vergleich zur Nichtberücksichtigung bei der rechnerischen Klassifikation) illustriert. Dies geschieht beispielhaft anhand eines an einen aus dem QS-Verfahren *Hüftendoprothesenversorgung* angelehnten Ratenindikator für einen Leistungserbringer.

Zähler, Nenner und Referenzbereich des Qualitätsindikators sind in Tabelle 8 dargestellt.

*Tabelle 8: Rechenregeln eines Beispielindikators, angelehnt an den Qualitätsindikator „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“*

<b>Zähler</b>	Eingriffe, bei denen die Operation später als 48 Stunden nach der Aufnahme erfolgte
<b>Nenner</b>	alle Eingriffe zur endoprothetischen Versorgung einer hüftgelenknahen Femurfraktur
<b>Referenzbereich</b>	$\leq 15 \%$

Gegenüber der Methode der rechnerischen Klassifikation wird auf Basis der Beobachtungen von „Anzahl Fälle im Zähler“ und „Anzahl Fälle im Nenner“ für die statistisch signifikante Klassifikation eine Wahrscheinlichkeit dafür ermittelt werden, dass die wahre, nicht direkt beobachtbare Rate eines Leistungserbringers innerhalb des Referenzbereichs liegt. Damit soll dem Sachverhalt Rechnung getragen werden, dass gleiche rechnerische Ergebnisse „Anzahl Fälle im Zähler durch Anzahl Fälle im Nenner“ von Leistungserbringern mit unterschiedlicher Nenner-Fallzahl nicht gleichbedeutend sind; Ergebnisse mit größerer Nenner-Fallzahl liefern nämlich stärkere statistische Evidenz dafür, dass die zugrunde liegende Rate eines Leistungserbringers im Qualitätsindikator inner- bzw. außerhalb des Referenzbereichs liegt. Beispielsweise überschreiten die Ergebnisse 50/100 und 1/2 beide den Referenzbereich  $\leq 15 \%$  um 35 Prozentpunkte. Die Evidenz für ein Qualitätsdefizit bei einer Fallzahl von 100 ist jedoch größer als bei einer Fallzahl von 2, bei der das Ergebnis auch allein aufgrund zufälliger Variation zustande gekommen sein kann. Für die statistisch signifikante Klassifikation wird für einen Leistungserbringer im ersten Schritt die Wahrscheinlichkeit dafür berechnet, dass die zugrunde liegende Rate im betrachteten Indikator innerhalb des Referenzbereichs liegt. Illustriert wird dies im Folgenden anhand eines hypothetischen Leistungserbringers und dem oben genannten Qualitätsindikator mit:

$$J = 20; o = 5$$

Die beobachtete Rate des Leistungserbringers liegt bei  $o/J = 25 \%$ . Um Rückschlüsse auf die zugrunde liegende, nicht direkt beobachtbare Rate des Leistungserbringers zu ziehen, wird auf Basis der beobachteten Daten und mittels eines bayesianischen Modells die sogenannte A-posteriori-Wahrscheinlichkeit der zugrunde liegenden Rate  $\theta$  berechnet. Die Verteilung der Wahrscheinlichkeit für den oben betrachteten Fall ist in Abbildung 22 dargestellt.

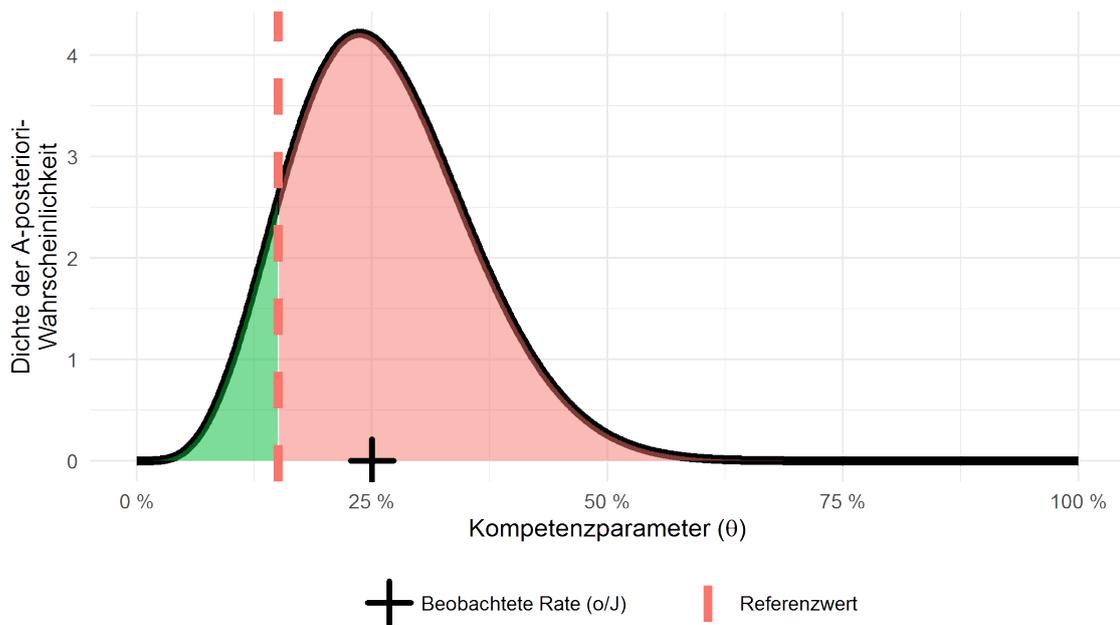


Abbildung 22: Dichte der A-posteriori-Wahrscheinlichkeit für die zugrunde liegende Rate  $\theta$  bei einem Leistungserbringer mit  $J = 20$  Fällen im Nenner und  $o = 5$  Fällen im Zähler. Anhand des Referenzwerts von 15 % wird der Wertebereich von  $\theta$  in einen Bereich innerhalb und außerhalb des Referenzbereichs unterteilt.

Die grüne Fläche kennzeichnet dabei die Dichte der Verteilung innerhalb des Referenzbereichs, die rote Fläche jene außerhalb des Referenzbereichs. Insgesamt summieren sich die beiden Flächen zu dem Wert 1. Die A-posteriori-Wahrscheinlichkeit dafür, dass die zugrunde liegende Rate  $\theta$  innerhalb des Referenzbereichs liegt, kann als Fläche der Dichte innerhalb des Referenzbereichs ausgedrückt werden und liegt in diesem Fall bei 11 %. Unterschreitet diese Wahrscheinlichkeit den vorgegebenen Schwellenwert  $\alpha$  (Signifikanzniveau), wird das Leistungserbringerergebnis als quantitativ auffällig eingestuft, weil die Wahrscheinlichkeit dafür, dass die zugrunde liegende Rate des Leistungserbringers im Referenzbereich liegt, unwahrscheinlicher ist als  $\alpha$ . Ist dies nicht der Fall, wird das Indikatorergebnis als quantitativ unauffällig eingestuft.<sup>63</sup>

<sup>63</sup> Eine interaktive Version der Abbildung, in der sowohl  $\alpha, J$  als auch der Referenzwert sowie das Signifikanzniveau verändert werden können, ist unter [https://iqtig.shinyapps.io/illustration\\_einstufungsmethodik/](https://iqtig.shinyapps.io/illustration_einstufungsmethodik/) zu finden.

## Literatur

- Adams, JL; Mehrotra, A; McGlynn, EA (2010): Estimating Reliability and Misclassification in Physician Profiling [*Technical Report*]. Santa Monica, US-CA: RAND. URL: [https://www.rand.org/content/dam/rand/pubs/technical\\_reports/2010/RAND\\_TR863.pdf](https://www.rand.org/content/dam/rand/pubs/technical_reports/2010/RAND_TR863.pdf) (abgerufen am: 18.08.2017).
- AGREE Next Steps Consortium (2017): Appraisal of Guidelines for Research & Evaluation II. AGREE II Instrument. [Stand:] 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: AHRQ. URL: <https://www.qualitymeasures.ahrq.gov/browse/domain> (abgerufen am: 29.08.2017).
- American Educational Research Association; American Psychological Association; National Council on Measurement in Education (2014): Standards for Educational and Psychological Testing. Washington, US-DC: American Educational Research Association. ISBN: 978-0-935302-35-6.
- Ansmann, L; Hower, K (2020): Eine Frage der Organisation. *SW – Sozialwirtschaft* 30(3): 7-10. DOI: 10.5771/1613-0707-2020-3-7.
- 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, GB: 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](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] ([2016]): 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.
- Armstrong, MJ; Bloom, JA (2017): Patient involvement in guidelines is poor five years after institute of medicine standards: review of guideline methodologies. *Research Involvement and Engagement* 3:19. DOI: 10.1186/s40900-017-0070-2.
- Armstrong, MJ; Mullins, CD; Gronseth, GS; Gagliardi, AR (2018): Impact of patient involvement on clinical practice guideline development: a parallel group study. *Implementation Science* 13:55. DOI: 10.1186/s13012-018-0745-6.
- 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.
- Asch, SE (1955): Opinions and Social Pressure. *Scientific American* 193(5): 31-35.
- Ash, AS; Fienberg, SE; Louis, TA; Normand, SLT; Stukel, TA; Utts, J (2012): Statistical Issues in Assessing Hospital Performance. Commissioned by the Committee of Presidents of Statistical Societies [White paper]. Revised: 27.01.2012. Baltimore, US-MD: Centers for Medicare and Medicaid Services; Committee of Presidents of Statistical Societies. URL: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf> (abgerufen am: 30.10.2020).
- Ash, AS; Schwartz, M; Peköz, EA (2013): Comparing Outcomes Across Providers. Chapter 12. In: Iezzoni, LI; Hrsg.: *Risk Adjustment for Measuring Healthcare Outcomes*. 4. Edition. Chicago, US-IL: Health Admission Press, 297-333. ISBN: 978-1-56793-437-3.
- Badgery-Parker, T; Pearson, S-A; Elshaug, AG (2020): Estimating misclassification error in a binary performance indicator: case study of low value care in Australian hospitals. *BMJ: Quality & Safety* 29(12): 992-999. DOI: 10.1136/bmjqs-2019-010564.
- 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, DW; Yendro, S (2018): Setting Achievable Benchmarks for Value-Based Payments: No Perfect Solution. *JAMA* 319(18): 1857-1858. DOI: 10.1001/jama.2018.2360.
- 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.

- Beaussier, A-L; Demeritt, D; Griffiths, A; Rothstein, H (2020): Steering by their own lights: Why regulators across Europe use different indicators to measure healthcare quality. *Health Policy* 124(5): 501-510. DOI: 10.1016/j.healthpol.2020.02.012.
- Benchimol, El; Manuel, DG; To, T; Griffiths, AM; Rabeneck, L; Guttmann, 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.
- Berger, JO (2010): Statistical Decision Theory and Bayesian Analysis. Second Edition. (Springer Series in Statistics). New York, US-NY [u. a.]: Springer. ISBN: 978-1-4419-3074-3.
- 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. DOI: 10.1097/00005650-199112000-00004.
- 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.
- Bevan, G; Hood, C (2006): What's Measured Is What Matters: Targets and Gaming in The English Public Health Care System. *Public Administration* 84(3): 517-538. DOI: 10.1111/j.1467-9299.2006.00600.x.
- Bevan, G; Fasolo, B (2013): Models of governance of public services: empirical and behavioural analysis of 'econs' and 'humans'. Chapter 2. In: Oliver, A: *Behavioural Public Policy*. Cambridge, GB: Cambridge University Press, 38-62. ISBN: 978-1-107-61737-7.
- Bevan, G; Evans, A; Nuti, S (2019): Reputations count: why benchmarking performance is improving health care across the world. *Health Economics, Policy and Law* 14(2): 141-161. DOI: 10.1017/S1744133117000561.
- Birkmeyer, JD; Kerr, EA; Dimick, JB (2006): Commissioned Paper. Improving the Quality of Quality Measurement. Appendix F. In: Institute of Medicine: *Performance Measurement: Accelerating Improvement*. Washington, US-DC: The National Academies Press, 177-203. DOI: 10.17226/11517
- Black, N; Varaganum, M; Hutchings, A (2014): Relationship between patient reported experience (PREMs) and patient reported outcomes (PROMs) in elective surgery. *BMJ: Quality & Safety* 23(7): 534-542. DOI: 10.1136/bmjqs-2013-002707.

- Blackwood, J; Armstrong, MJ; Schaefer, C; Graham, ID; Knaapen, L; Straus, SE; et al. (2020): How do guideline developers identify, incorporate and report patient preferences? An international cross-sectional survey. *BMC: Health Services Research* 20:458. DOI: 10.1186/s12913-020-05343-x.
- Blomqvist, Å (1991): The doctor as double agent: Information asymmetry, health insurance, and medical care. *Journal of Health Economics* 10(4): 411-432. DOI: 10.1016/0167-6296(91)90023-g.
- Bogner, A; Menz, W (2009): Das theoriegenerierte Experteninterview. Erkenntnisinteresse, Wissenformen, Interaktion. Teil 1. Kapitel 2. In: Bogner, A; Littig, B; Menz, W: *Experteninterviews. Theorie, Methode, Anwendungsfelder*. 3., grundlegend überarbeitete Auflage. Wiesbaden: Springer, 61-98. ISBN: 978-3-531-16259-1.
- Bogner, A; Littig, B; Menz, W (2014): Interviews mit Experten. Eine praxisorientierte Einführung. (Qualitative Sozialforschung). Wiesbaden: Springer. ISBN: 978-3-531-19416-5.
- 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; Diamantopoulos, A (2017): In Defense of Causal-Formative Indicators: A Minority Report. *Psychological Methods* 22(3): 581-596. DOI: 10.1037/met0000056.
- 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](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, US-CA: Jossey-Bass. ISBN: 978-0-7879-7088-8.
- Braithwaite, J (2018): Changing how we think about healthcare improvement. *BMJ* 361: k2014. DOI: 10.1136/bmj.k2014.
- Breckenridge, K; Bekker, HL; Gibbons, E; van der Veer, SN; Abbott, D; Briançon, S; et al. (2015): How to routinely collect data on patient-reported outcome and experience measures in renal registries in Europe: an expert consensus meeting. *Nephrology Dialysis Transplantation* 30(10): 1605-1614. DOI: 10.1093/ndt/gfv209.

- BT-Drucksache 17/10323 vom 10.07.2021. Unterrichtung durch die Bundesregierung. Sondergutachten 2012 des Sachverständigenrates zur Begutachtung der Entwicklung im Gesundheitswesen. Wettbewerb an der Schnittstelle zwischen ambulanter und stationärer Gesundheitsversorgung. URL: <https://dserver.bundestag.de/btd/17/103/1710323.pdf> (abgerufen am: 07.09.2021).
- BT-Drucksache 18/1307 vom 05.05.2014. 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: <https://dserver.bundestag.de/btd/18/013/1801307.pdf> (abgerufen am: 15.07.2021).
- Bühner, M (2011): Einführung in die Test- und Fragebogenkonstruktion. 3., aktualisierte und erweiterte Auflage. München [u. a.]: Pearson. ISBN: 978-3-86894-033-6.
- Cacace, M; Geraedts, M; Berger, E (2019): Public reporting as a quality strategy. Part II, Chapter 13. In: Busse, RK, N.; Panteli, D; Quentin, W: *Improving healthcare quality in Europe. Characteristics, effectiveness and implementation of different strategies*. (Health Policy, Series 53). Kopenhagen, DK: WHO [World Health Organization] [u. a.], 331-355. ISBN: 978-92-890-5175-0. URL: <https://apps.who.int/iris/bitstream/handle/10665/327356/9789289051750-eng.pdf> (abgerufen am: 13.07.2020).
- Callegaro, M; Murakami, MH; Tepman, Z; Henderson, V (2015): Yes-no answers versus check-all in self-administered modes. A systematic review and analyses. *International Journal of Market Research* 57(2): 203-224. DOI: 10.2501/IJMR-2015-014a.
- 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* 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.
- 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.
- CASP [Critical Appraisal Skills Programme] (2013): 10 questions to help you make sense of qualitative research. [Stand:] 31.05.2013. Oxford, GB: CASP. URL: <https://hhs.hud.ac.uk/lqsu/Useful/critap/Qualitative%20Research%20Checklist/CASP-Qualitative-Research-Checklist-31.05.13.pdf> (abgerufen am: 21.09.2020).

- Cattel, D; Eijkenaar, F; Schut, FT (2020): Value-based provider payment: towards a theoretically preferred design. *Health Economics, Policy and Law* 15(1): 94-112. DOI: 10.1017/S1744133118000397.
- Chen, H; Hailey, D; Wang, N; Yu, P (2014): A Review of Data Quality Assessment Methods for Public Health Information Systems. *International Journal of Environmental Research and Public Health* 11(5): 5170-5207. DOI: 10.3390/ijerph110505170.
- CHP/PCOR [Center for Health Policy/Center for Primary Care and Outcomes Research]; Battelle Memorial Institute (2011): Quality Indicator Measure Development, Implementation, Maintenance, and Retirement (Prepared by Battelle, under Contract No. 290-04-0020). [Stand:] May 2011. Rockville, US-MD: AHRQ [Agency for Healthcare Research and Quality]. URL: [http://www.qualityindicators.ahrq.gov/Downloads/Resources/Publications/2011/QI\\_Measure\\_Development\\_Implementation\\_Maintenance\\_Retirement\\_Full\\_5-3-11.pdf](http://www.qualityindicators.ahrq.gov/Downloads/Resources/Publications/2011/QI_Measure_Development_Implementation_Maintenance_Retirement_Full_5-3-11.pdf) (abgerufen am: 30.01.2017).
- Cleary, PD (1999): The increasing importance of patient surveys. Now that sound methods exist, patient surveys can facilitate improvement. *BMJ* 319(7212): 720-721. DOI: 10.1136/bmj.319.7212.720.
- Cochrane Deutschland; AWMF-IMWi [Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften - Institut für Medizinisches Wissensmanagement] (2016): Bewertung des Biasrisikos (Risiko systematischer Fehler) in klinischen Studien: ein Manual für die Leitlinienerstellung. Version 1.0. [Stand:] 04.05.2016. Freiburg [u. a.]: The Cochrane Collaboration [u. a.]. URL: [https://www.cochrane.de/sites/cochrane.de/files/public/uploads/manual\\_biasbewertung.pdf](https://www.cochrane.de/sites/cochrane.de/files/public/uploads/manual_biasbewertung.pdf) (abgerufen am: 24.09.2020).
- 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.
- 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, US-NY [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](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).
- Contandriopoulos, D; Champagne, F; Denis, J-L (2014): The Multiple Causal Pathways Between Performance Measures' Use and Effects. *Medical Care Research and Review* 71(1): 3-20. DOI: 10.1177/1077558713496320.
- Corallo, AN; Croxford, R; Goodman, DC; Bryan, EL; Srivastava, D; Stukel, TA (2014): A systematic review of medical practice variation in OECD countries. *Health Policy* 114: 5-14. DOI: 10.1016/j.healthpol.2013.08.002.

- Coulter, I; Adams, A; Shekelle, P (1995): Impact of varying panel membership on ratings of appropriateness in consensus panels: a comparison of a multi- and single disciplinary panel. *Health Services Research* 30(4): 577-591. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1070076/pdf/hsresearch00048-0076.pdf> (abgerufen am: 23.06.2021).
- Creswell, JW; Plano Clark, VL (2018): *Designing and Conducting Mixed Methods Research*. Third Edition. Los Angeles, US-CA [u. a.]: Sage. ISBN: 9781506386621.
- Cronbach, LJ (1971): Test Validation. Chapter 14. In: Thorndike, RL; Hrsg.: *Educational Measurement*. Second Edition. Washington, US-DC: American Council on Education, 443-507. ISBN: 0-8268-1271-6
- D'Ardenne, J (2015): Developing Interview Protocols. Chapter 5. In: Collins, D: *Cognitive Interviewing Practice*. Los Angeles, US-CA [u. a.]: Sage, 101-125. ISBN: 978-1-4462-5600-5.
- Dawes, RM; Faust, D; Meehl, PE (1989): Clinical Versus Actuarial Judgment. *Science* 243(4899): 1668-1674. DOI: 10.1126/science.2648573.
- De Feo, JA (2017): *Juran's Quality Handbook: The Complete Guide to Performance Excellence*. Seventh Edition. New York, US-NY: McGraw-Hill Education. ISBN: 978-1-259-64361-3.
- de Koning, J; Smulders, A; Klazinga, N (2007): Appraisal of Indicators through Research and Evaluation (AIRE). Versie 2.0. [Stand:] Januari 2007. Amsterdam, NL: 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).
- Delaney, LJ (2018): Patient-centred care as an approach to improving health care in Australia. *Collegian: Journal of the Royal College of Nursing Australia* 25(1): 119-123. DOI: 10.1016/j.colegn.2017.02.005.
- Delnevo, CD; Gundersen, DA; Manderski, MTB; Giovenco, DP; Giovino, GA (2017): Importance of Survey Design for Studying the Epidemiology of Emerging Tobacco Product Use Among Youth. *American Journal of Epidemiology* 186(4): 405-410. DOI: 10.1093/aje/kwx031.
- Deming, WE (1953): On the Distinction Between Enumerative and Analytic Surveys. *Journal of the American Statistical Association* 48(262): 244-255. DOI: 10.1080/01621459.1953.10483470.
- 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.

- den Breejen, EME; Hermens, RPMG; Galama, WH; Willemsen, WNP; Kremer, JAM; Nelen, WLDM (2016): Added value of involving patients in the first step of multidisciplinary guideline development: a qualitative interview study among infertile patients. *International Journal for Quality in Health Care* 28(3): 299-305. DOI: 10.1093/intqhc/mzw020.
- 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).
- 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.
- Döbler, K; Boukamp, KB; Nguyen, TT (2016): Dokumentationsqualität in der externen stationären Qualitätssicherung nach § 136 SGB V. *Gesundheitsökonomie & Qualitätsmanagement* 21(5): 250-257. DOI: 10.1055/s-0042-110822.
- Döbler, K; Geraedts, M (2018): Ausgewogenheit der Qualitätsindikatorensets der externen Qualitätssicherung nach §136 SGB V. *ZEFQ – Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 134: 9-17. DOI: 10.1016/j.zefq.2017.11.004.
- Döbler, K; Schrappe, M; Suske, S; Schmitt, J; Sens, B; Boywitt, D; et al. (2019): Eignung von Qualitätsindikatorensets in der Gesundheitsversorgung für verschiedene Einsatzgebiete - Forschungs- und Handlungsbedarf. *Das Gesundheitswesen* 81(10): 781-787. DOI: 10.1055/a-1007-0811.
- Donabedian, A (2003): An introduction to quality assurance in health care. Oxford, GB [u. a.]: Oxford University Press. ISBN: 0-19-515809-1.
- Doran, T (2014): Targets in pay for performance programmes: lessons from QOF. *BMJ* 348: 20-22. URL: <https://www.bmj.com/bmj/section-pdf/751786?path=/bmj/348/7948/Analysis.full.pdf> (abgerufen am: 14.05.2021).
- 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.
- 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.

- Eckhardt, H; Smith, P; Quentin, W (2019): Pay for Quality: using financial incentives to improve quality of care. Part II, Chapter 14. In: Busse, R; Klazinga, N; Panteli, D; Quentin, W; Hrsg.: *Improving healthcare quality in Europe. Characteristics, effectiveness and implementation of different strategies.* (Health Policy, Series 53). Kopenhagen, DK: WHO [World Health Organization] [u. a.], 357-400. ISBN: 978-92-890-5175-0. URL: <https://apps.who.int/iris/bitstream/handle/10665/327356/9789289051750-eng.pdf> (abgerufen am: 06.08.2020).
- Ehling, M; Körner, T; Bergdahl, M; Elvers, E; Földesi, E; Kron, A; et al. (2007): Handbook on Data Quality Assessment: Methods and Tools. Wiesbaden: European Commission, eurostat. URL: <https://unstats.un.org/unsd/dnss/docs-nqaf/Eurostat-HANDBOOK%20ON%20DATA%20QUALITY%20ASSESSMENT%20METHODS%20AND%20TOOLS%20%20I.pdf> (abgerufen am: 13.11.2018).
- Epstein, RM; Fiscella, K; Lesser, CS; Stange, KC (2010): Why the Nation Needs a Policy Push on Patient-Centered Health Care. *Health Affairs* 29(8): 1489-1495. DOI: 10.1377/hlthaff.2009.0888.
- EUnetHTA [European network for Health Technology Assessment], JA3WP6B2-2 Authoring Team (2017): Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness. Methodological Guidelines. Version 2.0. [Stand:] December 2019. Diemen, NL: EUnetHTA. URL: [https://eunethta.eu/wp-content/uploads/2020/01/EUnetHTA\\_Guideline\\_Information\\_Retrieval\\_v2-0.pdf](https://eunethta.eu/wp-content/uploads/2020/01/EUnetHTA_Guideline_Information_Retrieval_v2-0.pdf) (abgerufen am: 21.09.2020).
- 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.
- Faulbaum, F; Prüfer, P; Rexroth, M (2009): Was ist eine gute Frage?. Die systematische Evaluation der Fragenqualität Wiesbaden: VS Verlag für Sozialwissenschaften. ISBN: 978-3-531-15824-2.
- 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](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, CH: Hans Huber. ISBN: 978-3-456-84374-2.
- Flick, U (2016): *Qualitative Sozialforschung. Eine Einführung*. 7. Auflage. Reinbek bei Hamburg: Rowohlt. ISBN: 978-3-499-55694-4.
- Fox, J (2007): The uncertain relationship between transparency and accountability. *Development in Practice* 17(4-5): 663-671. DOI: 10.1080/09614520701469955.
- Freeman, T (2002): Using performance indicators to improve health care quality in the public sector: a review of the literature. *Health Services Management Research* 15(2): 126-137. DOI: 10.1258/0951484021912897.
- Fritsch, M (2018): Marktversagen infolge von Informationsmängeln. Kapitel 10. In: Fritsch, M: *Marktversagen und Wirtschaftspolitik: Mikroökonomische Grundlagen staatlichen Handelns*. 10., überarbeitete und ergänzte Edition. München: Franz Vahlen, 249-292. ISBN: 978-3-8006-5643-1.
- 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.
- Fullam, F; VanGeest, JB (2014): Surveys of Patient Populations. In: Johnson, TP: *Health Survey Methods*. Hoboken, US-NJ: Wiley, 561-583. ISBN: 978-1-118-59462-9.
- Furnival, J; Walshe, K; Boaden, R (2017): Emerging hybridity: comparing UK healthcare regulatory arrangements. *Journal of Health Organization and Management* 31(4): 517-528. DOI: 10.1108/JHOM-06-2016-0109.
- Furnival, J; Boaden, R; Walshe, K (2018): Assessing improvement capability in healthcare organisations: a qualitative study of healthcare regulatory agencies in the UK. *International Journal for Quality in Health Care* 30(9): 715-723. DOI: 10.1093/intqhc/mzy085.
- Fürstenberg, T; Laschat, M; Zich, K; Klein, S; Gierling, P; Nolting, H-D; et al. (2013): G-DRG-Begleitforschung gemäß §17b Abs. 8 KHG. Endbericht des dritten Forschungszyklus (2008-2010). [Stand:] März 2013. Berlin: IGES [u. a.]. URL: [https://www.g-drg.de/content/download/4217/34090/file/Begleitforschung\\_%c2%a717b\\_Abs8\\_KHG\\_2008\\_2010\\_Bericht.pdf](https://www.g-drg.de/content/download/4217/34090/file/Begleitforschung_%c2%a717b_Abs8_KHG_2008_2010_Bericht.pdf) (abgerufen am: 10.07.2020).
- Gardner, K; Olney, S; Dickinson, H (2018): Getting smarter with data: understanding tensions in the use of data in assurance and improvement-oriented performance management systems to improve their implementation. *Health Research Policy and Systems* 16:125. DOI: 10.1186/s12961-018-0401-2.
- Geraedts, M; Selbmann, H-K; Ollenschlaeger, G (2003): Critical appraisal of clinical performance measures in Germany. *International Journal for Quality in Health Care* 15(1): 79-85. DOI: 10.1093/intqhc/15.1.79.

- Geraedts, M; 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. *Das 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, CH: 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, US-CA: Jossey-Bass. ISBN: 978-0-7879-6220-3.
- Glass, GV (1978): Standards and Criteria. *Journal of Educational Measurement* 15(4): 237-261. DOI: 10.1111/j.1745-3984.1978.tb00072.x.
- Glasziou, P; Haynes, B (2005): The paths from research to improved health outcomes. *Evidence-Based Nursing* 8(2): 36-38. DOI: 10.1136/ebn.8.2.36.
- Goeman, JJ; De Jong, NH (2018): How Well Does the Sum Score Summarize the Test? Summability as a Measure of Internal Consistency. *Educational Measurement: Issues and Practice* 37(2): 54-63. DOI: 10.1111/emip.12181.
- Goldhammer, F; Hartig, J (2012): Interpretation von Testresultaten und Testeichung. Kapitel 8. In: Moosbrugger, H; Kelava, A; Hrsg.: *Testtheorie und Fragebogenkonstruktion*. 2., aktualisierte und überarbeitete Auflage. Berlin [u. a.]: Springer, 173-201. ISBN: 978-3-642-20071-7.
- 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.
- Greenhalgh, J; Dalkin, S; Gibbons, E; Wright, J; Valderas, JM; Meads, D; et al. (2018): How do aggregated patient-reported outcome measures data stimulate health care improvement? A realist synthesis. *Journal of Health Services Research & Policy* 23(1): 57-65. DOI: 10.1177/1355819617740925.
- Grove, WM; Zald, DH; Lebow, BS; Snitz, BE; Nelson, C (2000): Clinical Versus Mechanical Prediction: A Meta-Analysis. *Psychological Assessment* 12(1): 19-30. DOI: 10.1037/1040-3590.12.1.19.
- Häder, M (2000): Die Expertenauswahl bei Delphi-Befragungen. (ZUMA How-to-Reihe, Nr. 5). Mannheim: ZUMA [Zentrum für Umfragen, Methoden und Analysen]. URL: [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/howto/how-to\\_5mh.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/how-to_5mh.pdf) (abgerufen am: 30.01.2017).

- Hamblin, R (2008): Regulation, measurements and incentives. The experience in the US and UK: does context matter? *JRSH – The journal of the Royal Society for the Promotion of Health* 128(6): 291-298. DOI: 10.1177/1466424008096617.
- 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: 978-3-642-20071-7.
- Hartmann, J; Weidmann, C; Biehle, R (2016): Validierung von GKV-Routinedaten am Beispiel von geschlechtsspezifischen Diagnosen. *Das 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. *The Lancet* 348(9024): 383-386. DOI: 10.1016/S0140-6736(96)01073-2.
- Haynes, SN; Richard, DCS; Kubany, ES (1995): Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods. *Psychological Assessment* 7(3): 238-247. DOI: 10.1037/1040-3590.7.3.238.
- 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 [U. S. Food and Drug Administration], CDER [Center for Drug Evaluation and Research]; HHS.gov [U. S. Department of Health and Human Services], FDA [U. S. Food and Drug Administration], CBER [Center for Biologics Evaluation and Research]; HHS.gov [U. S. Department of Health and Human Services], FDA [U. S. Food and Drug Administration], CDRH [Center for Devices and Radiological Health] (2006): Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health and Quality of Life Outcomes* 4:79. DOI: 10.1186/1477-7525-4-79.
- Hibbard, JH; Stockard, J; Tusler, M (2003): Does Publicizing Hospital Performance Stimulate Quality Improvement Efforts? Results from a study in Wisconsin suggest that making performance information public stimulates quality improvement. *Health Affairs* 22(2): 84-94. DOI: 10.1377/hlthaff.22.2.84.
- Hibbard, JH; Stockard, J; Tusler, M (2005): It Isn't Just about Choice: The Potential of a Public Performance Report to Affect the Public Image of Hospitals. *Medical Care Research and Review* 62(3): 358-371. DOI: 10.1177/1077558705275415.
- Hibbard, JH (2008): What Can We Say about the Impact of Public Reporting? Inconsistent Execution Yields Variable Results. *Annals of Internal Medicine* 148(2): 160-161. DOI: 10.7326/0003-4819-148-2-200801150-00011.

- 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](http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_28.pdf) (abgerufen am: 30.01.2017).
- Hitzler, R (1994): Wissen und Wesen des Experten. Ein Annäherungsversuch - zur Einleitung. Kapitel 1. In: Hitzler, R; Hrsg.: *Expertenwissen. Die institutionalisierte Kompetenz zur Konstruktion von Wirklichkeit*. Opladen: Westdeutscher Verlag, 13-30. ISBN: 978-3-531-12581-7.
- Holmstrom, B; Milgrom, P (1991): Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *The Journal of Law, Economics and Organization* 7(Special Issue: [Papers from the Conference on the New Science of Organization, January 1991]): 24-52.
- Hooker, AB; Etman, A; Westra, M; van der Kam, WJ (2019): Aggregate analysis of sentinel events as a strategic tool in safety management can contribute to the improvement of healthcare safety. *International Journal for Quality in Health Care* 31(2): 110-116. DOI: 10.1093/intqhc/mzy116.
- 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.
- Hurley, J (2000): An Overview of the Normative Economics of the Health Sector. Chapter 2. In: Culyer, AJ; Newhouse, JP; Hrsg.: *Handbook of Health Economics*. Amsterdam, NL [u. a.]: Elsevier, 55-118. ISBN: 978-0-444-50470-8. DOI: 10.1016/S1574-0064(00)80161-4.
- Iezzoni, LI (2012): Risk adjustment for performance measurement. Part III, Chapter 3.1. In: Smith, PC; Mossialos, E; Papanicolas, I; Leatherman, S; Hrsg.: *Performance Measurement for Health System Improvement. Experiences, Challenges, and Prospects*. 3rd edition. New York, US-NY: Cambridge University Press, 251-285. ISBN: 978-0-521-11676-3.
- Iezzoni, LI (2013): 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.

- IGZ [Inspectie voor de Gezondheidszorg] (2016): Brochure voor zorgaanbieders: Calamiteiten melden aan de IGZ. [Stand:] 13.12.2016. Den Haag, NL: 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.
- Institute of Medicine, Committee on Quality of Health Care in America, (2001): Crossing the Quality Chasm. A New Health System for the 21st Century. Washington, US-DC: National Academy Press. ISBN: 978-0-309-07280-9.
- IOM [Institute of Medicine] (2009): Conflict of Interest in Medical Research, Education, and Practice. Washington, US-DC: The National Academies Press. DOI: 10.17226/12598.
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2016): Planungsrelevante Qualitätsindikatoren. Abschlussbericht zur Auswahl und Umsetzung. Stand: 31.08.2016. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2016/IQTIG\\_Planungsrelevante-Qualitätsindikatoren\\_Abschlussbericht.pdf](https://iqtig.org/downloads/berichte/2016/IQTIG_Planungsrelevante-Qualitätsindikatoren_Abschlussbericht.pdf) (abgerufen am: 25.08.2017).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2017): Ereigniszeitanalyse-Methodik für die Follow-up-Indikatoren nach QSKH-RL. Stand: 06.04.2017. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2017/IQTIG\\_Ereigniszeitanalyse-Methodik-f%C3%BCr-Follow-up-Indikatoren-nach-QSKH-RL\\_2017-04-06.pdf](https://iqtig.org/downloads/berichte/2017/IQTIG_Ereigniszeitanalyse-Methodik-f%C3%BCr-Follow-up-Indikatoren-nach-QSKH-RL_2017-04-06.pdf) (abgerufen am: 18.08.2017).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018): 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](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] (2020): Kriterien für den gezielten Datenabgleich in der Datenvalidierung nach QSKH-RL. Abschlussbericht. Stand: 31.01.2020. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2020/IQTIG\\_Kriterien-Datenabgleich-Datenvalidierung-QSKH-RL\\_2020-01\\_31.pdf](https://iqtig.org/downloads/berichte/2020/IQTIG_Kriterien-Datenabgleich-Datenvalidierung-QSKH-RL_2020-01_31.pdf) (abgerufen am: 11.06.2021).
- IQWiG [Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen] (2013): Analytic Hierarchy Process (AHP) – Pilotprojekt zur Erhebung von Patienten-präferenzen in der Indikation Depression. Arbeitspapier, Version 1.0. (IQWiG Berichte – Nr. 163). Köln: IQWiG. URL: [https://www.iqwig.de/download/Arbeitspapier\\_Analytic-Hierarchy-Process\\_Pilotprojekt.pdf](https://www.iqwig.de/download/Arbeitspapier_Analytic-Hierarchy-Process_Pilotprojekt.pdf) (abgerufen am: 10.12.2021).
- Jacob, G; Bengel, J (2000): Das Konstrukt Patientenzufriedenheit: Eine kritische Bestandsaufnahme. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie* 48(3): 280-301.

- Jakob, J; Hinzpeter, M; Weiß, C; Weiß, J; Schlüter, M; Post, S; et al. (2010): Qualität der BQS-Dokumentation. Datenevaluation anhand intra- und postoperativer Komplikationen nach Cholezystektomien. *Der Chirurg* 81(6): 563-567. DOI: 10.1007/s00104-009-1827-4.
- Jarvis, CB; Mackenzie, SB; Podsakoff, PM (2003): A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *Journal of Consumer Research* 30(2): 199-218. DOI: 10.1086/376806.
- JCAHO [Joint Commission on Accreditation of Healthcare Organizations] (1989): Characteristics of Clinical Indicators. *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.
- Jensen, FV; Nielsen, TD (2007): Bayesian Networks and Decision Graphs. Second Edition. (Information Science & Statistics). Berlin [u. a.]: Springer. ISBN: 978-0-387-68281-5.
- 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](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. *Emergency Medicine Australasia* 26(2): 113-124. DOI: 10.1111/1742-6723.12195.
- Kane, MT (2001): Current Concerns in Validity Theory. *Journal of Educational Measurement* 38(4): 319-342. DOI: 10.1111/j.1745-3984.2001.tb01130.x.
- Kane, MT (2016): Explicating validity. *Assessment in Education: Principles, Policy and Practice* 23(2): 198-211. DOI: 10.1080/0969594X.2015.1060192.
- Kane, MT (2017): Using Empirical Results to Validate Performance Standards. Chapter 2. In: Blömeke, S; Gustafsson, J-E: *Standard Setting in Education: The Nordic Countries in an International Perspective*. Cham, CH: Springer, 11-29. DOI: 10.1007/978-3-319-50856-6\_2.
- Karlton, A; Sanne, JM; Aase, K; Anderson, JE; Fernandes, A; Fulop, NJ; et al. (2020): Knowledge management infrastructure to support quality improvement: A qualitative study of maternity services in four European hospitals. *Health Policy* 124(2): 205-215. DOI: 10.1016/j.healthpol.2019.11.005.
- Kauermann, G; Küchenhoff, H (2011): Stichproben. Methoden und praktische Umsetzung mit R. Berlin [u. a.]: Springer. ISBN: 978-3-642-12317-7. DOI: 10.1007/978-3-642-12318-4.
- Kaufman, BG; Spivack, BS; Stearns, SC; Song, PH; O'Brien, EC (2017): Impact of Accountable Care Organizations on Utilization, Care, and Outcomes: A Systematic Review. *Medical Care Research and Review* 76(3): 255-290. DOI: 10.1177/1077558717745916.

- 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.
- Kersting, C; Kneer, M; Barzel, A (2020): Patient-relevant outcomes: what are we talking about? A scoping review to improve conceptual clarity. *BMC: Health Services Research* 20:596. DOI: 10.1186/s12913-020-05442-9.
- Kingsley, C; Patel, S (2017): Patient-reported outcome measures and patient-reported experience measures. *BJA: Education* 17(4): 137-144. DOI: 10.1093/bjaed/mkw060.
- 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, US-NY [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, 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, US-DC: National Academy Press. ISBN: 978-0-309-26174-6.
- Kondo, KK; Damberg, CL; Mendelson, A; Motu'apuaka, M; Freeman, M; O'Neil, M; et al. (2016): Implementation Processes and Pay for Performance in Healthcare: A Systematic Review. *Journal of General Internal Medicine* 31(Suppl 1): S61-S69. DOI: 10.1007/s11606-015-3567-0.

- Kosack, S; Fung, A (2014): Does Transparency Improve Governance? *Annual Review of Political Science* 17(1): 65-87. DOI: 10.1146/annurev-polisci-032210-144356.
- 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.
- Kronick, R; Casalino, LP; Bindman, AB (2015): Introduction. Apple Pickers or Federal Judges: Strong versus Weak Incentives in Physician Payment. *Health Services Research* 50(Suppl 2): 2049-2056. DOI: 10.1111/1475-6773.12424.
- 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. ISBN: 978-3-7799-2922-2.
- Kühn, T; Koschel, K-V (2011): Gruppendiskussionen. Ein Praxis-Handbuch. Wiesbaden: VS Verlag für Sozialwissenschaften. ISBN: 978-3-531-16921-7.
- Kuncel, NR; Klieger, DM; Connelly, BS; Ones, DS (2013): Mechanical Versus Clinical Data Combination in Selection and Admissions Decisions: A Meta-Analysis. *Journal of Applied Psychology* 98(6): 1060-1072. DOI: 10.1037/a0034156.
- Lamnek, S (2010): Qualitative Sozialforschung. Lehrbuch. 5. überarbeitete Auflage. Weinheim [u. a.]: Beltz. ISBN: 978-3-621-27770-9.
- Leemans, K; Cohen, J; Francke, AL; Stichele, RV; Claessen, SJJ; Block, LVd; et al. (2013): Towards a standardized method of developing quality indicators for palliative care: protocol of the Quality indicators for Palliative Care (Q-PAC) study. *BMC: Palliative Care* 12:6. DOI: 10.1186/1472-684X-12-6.
- 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.
- Lexis, W (1875): Einleitung in die Theorie der Bevölkerungsstatistik. Straßburg, FR: Trübner. URL: <http://resolver.sub.uni-goettingen.de/purl?PPN527522651> (abgerufen am: 10.06.2021).
- Liao, JM; Fleisher, LA; Navathe, AS (2016): Increasing the Value of Social Comparisons of Physician Performance Using Norms. *JAMA* 316(11): 1151-1152. DOI: 10.1001/jama.2016.10094.

- 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.
- 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.
- Lloyd, H; Jenkinson, C; Hadi, M; Gibbons, E; Fitzpatrick, R (2014): Patient reports of the outcomes of treatment: a structured review of approaches. *Health and Quality of Life Outcomes* 12:5. DOI: 10.1186/1477-7525-12-5.
- Lohr, KN (1990): Medicare: A Strategy for Quality Assurance, Volume I. Washington, US-DC: 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](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.
- Lützner, C; Lange, T; Lützner, J (2017): Grundlagen patientenberichteter Ergebnisse (Patient-reported Outcome – PRO). *Orthopädie und Unfallchirurgie up2date* 12(6): 661-676. DOI: 10.1055/s-0043-110864.
- Maass, C; Schleiz, W; Weyermann, M; Drösler, SE (2011): Krankenhaus-Routinedaten zur externen Qualitätssicherung? Vergleich von Qualitätsindikatoren anhand der Daten der gesetzlichen externen Qualitätssicherung (BQS) und Routinedaten. *Deutsche Medizinische Wochenschrift* 136(9): 409-414. DOI: 10.1055/s-0031-1274523.
- 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.
- Mainz, J (2003): Developing evidence-based clinical indicators: a state of the art methods primer. *International Journal for Quality in Health Care* 15(Suppl. 1): i5-i11. DOI: 10.1093/intqhc/mzg084.

- 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. ISBN: 978-3-407-25730-7.
- Mazza, D; Chapman, A; Michie, S (2013): Barriers to the implementation of preconception care guidelines as perceived by general practitioners: a qualitative study. *BMC: Health Services Research* 13:36. DOI: 10.1186/1472-6963-13-36.
- McClellan, M; McKethan, AN; Lewis, JL; Roski, J; Fisher, ES (2010): A National Strategy To Put Accountable Care Into Practice. *Health Affairs* 29(5): 982-990. DOI: 10.1377/hlthaff.2010.0194.
- McClellan, M; Kent, J; Beales, SJ; Cohen, SIA; Macdonnell, M; Thoumi, A; et al. (2014): Accountable Care Around The World: A Framework To Guide Reform Strategies. *Health Affairs* 33(9): 1507-1515. DOI: 10.1377/hlthaff.2014.0373.
- McClellan, M; Udayakumar, K; Thoumi, A; Gonzalez-Smith, J; Kadakia, K; Kurek, N; et al. (2017): Improving Care And Lowering Costs: Evidence And Lessons From A Global Analysis Of Accountable Care Reforms. *Health Affairs* 36(11): 1920-1927. DOI: 10.1377/hlthaff.2017.0535.
- McCluskey, A; Vratsistas-Curto, A; Schurr, K (2013): Barriers and enablers to implementing multiple stroke guideline recommendations: a qualitative study. *BMC: Health Services Research* 13: 323. DOI: 10.1186/1472-6963-13-323.
- 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. DOI: 10.1097/00005650-200301001-00005
- 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.
- Meehl, PE (1954): Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. Minneapolis, US-MN: University of Minnesota Press. ISBN: 978-0-8058-1858-1. DOI: 10.1037/11281-000.
- 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](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).
- Michie, S; van Stralen, MM; West, R (2011): The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science* 6:42. DOI: 10.1186/1748-5908-6-42.
- 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, GB: TSO [The Stationery Office]. ISBN: 978-0-10-298146-9. URL: <http://www.midstaffpublicinquiry.com/sites/default/files/report/Volume%203.pdf> (abgerufen am: 23.07.2018).
- Milstein, R; Schreyoegg, J (2016): Pay for performance in the inpatient sector: A review of 34 P4P programs in 14 OECD countries. *Health Policy* 120(10): 1125-1140. DOI: 10.1016/j.healthpol.2016.08.009.
- Mooney, G; Ryan, M (1993): Agency in health care: getting beyond first principles. *Journal of Health Economics* 12(2): 125-135. DOI: 10.1016/0167-6296(93)90023-8.
- 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 (2012a): Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). Kapitel 2. In: Moosbrugger, H; Kelava, A; Hrsg.: *Testtheorie und Fragebogenkonstruktion*. 2., aktualisierte und überarbeitete Auflage. Berlin [u. a.]: Springer, 7-26. ISBN: 978-3-642-20072-4.
- Moosbrugger, H; Kelava, A (2012b): *Testtheorie und Fragebogenkonstruktion*. 2. Auflage. Berlin [u. a.]: Springer. ISBN: 978-3-642-20071-7.
- Muche-Borowski, C; Selbmann, HK; Nothacker, M; Müller, W; Kopp, I; Hrsg. (2012): AWMF-Regelwerk Leitlinien. [Stand:] 06.11.2012. Marburg: AWMF [Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften], Ständige Kommission Leitlinien. URL: [http://www.awmf.org/fileadmin/user\\_upload/Leitlinien/AWMF-Regelwerk/AWMF-Regelwerk.pdf](http://www.awmf.org/fileadmin/user_upload/Leitlinien/AWMF-Regelwerk/AWMF-Regelwerk.pdf) (abgerufen am: 30.01.2017).
- Mühlhauser, I; Müller, H (2009): Patientenrelevante Endpunkte und patient-reported outcomes in klinischer Forschung und medizinischer Praxis. In: Klusen, N; Fließgarten, A; Nebling, T: *Informiert und selbstbestimmt. Der mündige Bürger als mündiger Patient*. (Beiträge zum Gesundheitsmanagement, 24). Baden-Baden: Nomos, 34-65. ISBN: 978-3-8329-4474-2.
- Mullen, PM (2003): Delphi: myths and reality. *Journal of Health Organization and Management* 17(1): 37-52. DOI: 10.1108/14777260310469319.

- Murphy, FG; Scheu, C (2010): Warum ist Transparenz so wichtig? Kapitel 5. In: Kuhlen, R; Rink, O; Zacher, J; Hrsg.: *Jahrbuch Qualitätsmedizin 2010*. Berlin: 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.
- NHS [National Health Service], Institute for Innovation and Improvement,; APHO [Association of Public Health Observatories] ([2007]): The Good Indicators Guide: Understanding how to use and choose indicators [Guidance]. First published: June 2007. Page updated: November 2017. Coventry, GB: NHS. URL: <https://www.england.nhs.uk/improvement-hub/wp-content/uploads/sites/44/2017/11/The-Good-Indicators-Guide.pdf> (abgerufen am: 18.12.2020).
- Niederberger, M; Wassermann, S (2015): Methoden der Experten- und Stakeholdereinbindung in der sozialwissenschaftlichen Forschung. Wiesbaden: Springer VS. ISBN: 978-3-658-01686-9.
- Nilsen, P (2015): Making sense of implementation theories, models and frameworks. *Implementation Science* 10:53. DOI: 10.1186/s13012-015-0242-0.
- 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]. DOI: 10.32745/9783954663743.
- 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. [Stand:] April 2015. Washington, US-DC: NQF. URL: <http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=79434> (abgerufen am: 30.01.2017).
- NQF [National Quality Forum] (2016): Attribution: Principles and Approaches. Final Report. [Stand:] December 2016. Washington, US-DC: NQF. ISBN: 978-1-68248-034-2. URL: [https://www.qualityforum.org/Publications/2016/12/Attribution\\_-\\_Principles\\_and\\_Approaches.aspx](https://www.qualityforum.org/Publications/2016/12/Attribution_-_Principles_and_Approaches.aspx) [Download the Publication] (abgerufen am: 23.06.2021).

- 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
- Obbarius, A; Fischer, KI; Fischer, F; Liegl, G; Obbarius, N; Nolte, S; et al. (2018): Empirische Erfassung subjektiver Gesundheitsmerkmale am Beispiel der gesundheitsbezogenen Lebensqualität. *Psychotherapie, Psychosomatik, Medizinische Psychologie* 68(12): 534-547. DOI: 10.1055/a-0764-4691.
- Ollenschläger, G; Wirth, T; Schwarz, S; Trifyllis, J; Schaefer, C (2018): Unzureichende Patientenbeteiligung an der Leitlinienentwicklung in Deutschland – eine Analyse der von der AWMF verbreiteten ärztlichen Empfehlungen. *ZEFQ – Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 135-136: 50-55. DOI: 10.1016/j.zefq.2018.06.006.
- 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.
- Panteli, D; Quentin, W; Busse, R (2019): Understanding healthcare quality strategies: a five-lens framework. Part I, Chapter 2. In: Busse, RK, N.; Panteli, D; Quentin, W: *Improving healthcare quality in Europe. Characteristics, effectiveness and implementation of different strategies*. (Health Policy, Series 53). Kopenhagen, DK: WHO [World Health Organization], 19-30. ISBN: 978-92-890-5175-0. URL: <https://apps.who.int/iris/bitstream/handle/10665/327356/9789289051750-eng.pdf> (abgerufen am: 13.07.2020).
- Parmigiani, G; Inoue, L (2009): *Decision Theory. Principles and Approaches*. (Wiley Series in Probability and Statistics). Chichester, GB: Wiley. ISBN: 978-0-471-49657-1.
- Paschke, A; Dimancesco, D; Vian, T; Kohler, JC; Forte, G (2018): Increasing transparency and accountability in national pharmaceutical systems. *Bulletin of the World Health Organization* 96(11): 782-791. DOI: 10.2471/blt.17.206516.
- 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.
- Pieper, D; Jülich, F; Antoine, S-L; Bächle, C; Chernyak, N; Genz, J; et al. (2015): Studies analysing the need for health-related information in Germany – a systematic review. *BMC: Health Services Research* 15:407. DOI: 10.1186/s12913-015-1076-9.
- Pimperl, A; Schulte, T; Mühlbacher, A; Rosenmöller, M; Busse, R; Groene, O; et al. (2017): Evaluating the Impact of an Accountable Care Organization on Population Health: The Quasi-Experimental Design of the German Gesundes Kinzigtal. *Population Health Management* 20(3): 239-248. DOI: 10.1089/pop.2016.0036.

- Pluut, B (2016): Differences that matter: developing critical insights into discourses of patient-centeredness. *Medicine, Health Care and Philosophy* 19(4): 501-515. DOI: 10.1007/s11019-016-9712-7.
- Pohontsch, NJ; Müller, V; Brandner, S; Karlheim, C; Jünger, S; Klindtworth, K; et al. (2018): Gruppendiskussionen in der Versorgungsforschung – Teil 1: Einführung und Überlegungen zur Methodenwahl und Planung. *Das Gesundheitswesen* 80(10): 864-870. DOI: 10.1055/s-0043-104379.
- Popham, WJ; Husek, TR (1969): Implications of Criterion-Referenced Measurement. *Journal of Educational Measurement* 6(1): 1-9. DOI: 10.1111/j.1745-3984.1969.tb00654.x.
- 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.
- Provost, LPL, G. J. (1998): The Importance of Concepts in Creativity and Improvement. *Quality Progress* 31: 31-38. URL: <https://qi.elft.nhs.uk/wp-content/uploads/2015/12/10-concepts-gp-1998.pdf> (abgerufen am: 13.07.2020).
- Prüfer, P; Rexroth, M (1996): Verfahren zur Evaluation von Survey – Fragen: Ein Überblick. [Stand:] Mai 1996. (ZUMA-Arbeitsbericht, Nr. 96/05). Mannheim: ZUMA [Zentrum für Umfragen, Methoden und Analysen]. URL: [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/zuma\\_arbeitsberichte/96\\_05.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/zuma_arbeitsberichte/96_05.pdf) (abgerufen am: 30.01.2017).
- Prüfer, P; Rexroth, M; Hrsg. (2005): Kognitive Interviews. (ZUMA How-to-Reihe, 15). Mannheim: ZUMA [Zentrum für Umfragen, Methoden und Analysen]. URL: [http://www.ssoar.info/ssoar/bitstream/handle/document/20147/ssoar-2005-prufer\\_et\\_al-kognitive\\_interviews.pdf?sequence=1](http://www.ssoar.info/ssoar/bitstream/handle/document/20147/ssoar-2005-prufer_et_al-kognitive_interviews.pdf?sequence=1) (abgerufen am: 30.01.2017).
- Raats, VM; Moors, JJA (2003): Double-checking auditors: a Bayesian approach. *Journal of the Royal Statistical Society: Series D (The Statistician)* 52(3): 351-365. DOI: 10.1111/1467-9884.00364.
- Raykov, T; Marcoulides, GA (2011): Introduction to Psychometric Theory. London, GB: Routledge. ISBN: 978-0-415-87822-7.
- Reed, JE; Card, AJ (2016): The problem with Plan-Do-Study-Act cycles. *BMJ: Quality & Safety* 25(3): 147-152. DOI: 10.1136/bmjqs-2015-005076.
- Reeves, D; Doran, T; Valderas, JM; Kontopantelis, E; Trueman, P; Sutton, M; et al. (2010): How to identify when a performance indicator has run its course. *BMJ* 340: c1717. DOI: 10.1136/bmj.c1717.
- Reiter, A; Fischer, B; Kötting, J; Geraedts, M; Jäckel, WH (2007): QUALIFY: Ein Instrument zur Bewertung von Qualitätsindikatoren. Düsseldorf: BQS [Bundesgeschäftsstelle Qualitätssicherung]. [Zugesandt vom G-BA auf Nachfrage].

- 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.
- Rhee, CM; Brunelli, SM; Subramanian, L; Tentori, F (2018): Measuring patient experience in dialysis: a new paradigm of quality assessment. *Journal of Nephrology* 31(2): 231-240. DOI: 10.1007/s40620-017-0401-2.
- 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. *Das Gesundheitswesen* 78(Suppl. 1): e110-e119. DOI: 10.1055/s-0042-102347.
- Roland, M; Dudley, RA (2015): How Financial and Reputational Incentives Can Be Used to Improve Medical Care. *Health Services Research* 50(Suppl 2): 2090-2115. DOI: 10.1111/1475-6773.12419.
- Roters (2020): § 137a SGBV 5. Rn. 34. In: Körner, A; Leitherer, S; Mutschler, B; Rolfs, C: *Kasseler Kommentar. Sozialversicherungsrecht*. Stand: März 2020. (Band 2, Ergänzungslieferung 108). München: Beck, 10. ISBN: 978-3-406-74561-4.
- Rothman, KJ; Greenland, S; Lash, TL (2008): *Modern Epidemiology*. Third Edition. Philadelphia, US-PA [u. a.]: Lippincott Williams and Wilkins. ISBN: 978-1-4511-9005-2.
- Rothman, KJ (2012): *Epidemiology: An Introduction*. Second Edition. Oxford, GB [u. a.]: Oxford University Press. ISBN: 978-0-19-975455-7.
- Sackett, DL; Rosenberg, WM; Gray, JA; Haynes, RB; Richardson, WS (1996): Evidence based medicine: what it is and what it isn't. *BMJ* 312: 71-72. DOI: 10.1136/bmj.312.7023.71.
- 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).
- Sanderson, J; Lonsdale, C; Mannion, R (2019): What's Needed to Develop Strategic Purchasing in Healthcare? Policy Lessons from a Realist Review. *International Journal of Health Policy and Management* 8(1): 4-17. DOI: 10.15171/ijhpm.2018.93.
- Schang, L; Blotenberg, I; Boywitt, D (2021): What Makes a Good Quality Indicator Set? A Systematic Review of Criteria. *International Journal for Quality in Health Care* 33(3): mzab107. DOI: 10.1093/intqhc/mzab107.
- 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](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.
- Schreyögg, J; Bäuml, M; Krämer, J; Dette, T; Busse, R; Geissler, A (2014): Forschungsauftrag zur Mengenentwicklung nach § 17b Abs. 9 KHG. Endbericht. [Stand:] Juli 2014. Hamburg: HCHE [Hamburg Center for Health Economics]. URL: [https://www.g-drg.de/content/download/4861/38231/version/2/file/Endbericht+zum+Forschungsauftrag+gem.+%C2%A7+17b+Abs.+9+KHG+%28Druck%29.pdf?pk\\_campaign=Forschungsauftrag&pk\\_kwd=Endbericht%20Druck](https://www.g-drg.de/content/download/4861/38231/version/2/file/Endbericht+zum+Forschungsauftrag+gem.+%C2%A7+17b+Abs.+9+KHG+%28Druck%29.pdf?pk_campaign=Forschungsauftrag&pk_kwd=Endbericht%20Druck) (abgerufen am: 10.07.2020).
- Schubert, I; Ihle, P; Köster, I (2010): Interne Validierung von Diagnosen in GKV-Routinedaten: Konzeption mit Beispielen und Falldefinition. *Das Gesundheitswesen* 72(6): 316-322. DOI: 10.1055/s-0030-1249688.
- 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. URL: [https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/SpotGesPubRep\\_dt\\_final\\_web.pdf](https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/SpotGesPubRep_dt_final_web.pdf) (abgerufen am: 30.01.2017).
- Scott, A; Sivey, P; Ait Ouakrim, D; Willenberg, L; Naccarella, L; Furler, J; et al. (2011): The effect of financial incentives on the quality of health care provided by primary care physicians (Review). *Cochrane Database of Systematic Reviews* (9). Art. No.: CD008451. DOI: 10.1002/14651858.CD008451.pub2.
- 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* 358: j4008. DOI: 10.1136/bmj.j4008.

- Shepard, L (1980): Standard Setting Issues and Methods. *Applied Psychological Measurement* 4(4): 447-467. DOI: 10.1177/014662168000400403.
- Shwartz, M; Cohen, AB; Restuccia, JD; Ren, ZJ; Labonte, A; Theokary, C; et al. (2011): How Well Can We Identify the High-Performing Hospital? *Medical Care Research and Review* 68(3): 290-310. DOI: 10.1177/1077558710386115.
- Shwartz, M; Restuccia, JD; Rosen, AK (2015): Composite Measures of Health Care Provider Performance: A Description of Approaches. *The Milbank Quarterly* 93(4): 788-825. DOI: 10.1111/1468-0009.12165.
- SIGN [Scottish Intercollegiate Guidelines Network] ([kein Datum]): [Search filters:] Systematic Reviews. Edinburgh, GB: SIGN. URL: <https://www.sign.ac.uk/assets/search-filters-systematic-reviews.docx> (abgerufen am: 21.09.2020).
- Sireci, SG; Sukin, T (2013): Test Validity. Chapter 4. In: Geisinger, KF; Bracken, BA; Carlson, JF; Hansen, J-IC; Kuncel, NR; Reise, SP; et al.; Hrsg.: *APA Handbook of Testing and Assessment in Psychology®. Vol. 1: Test Theory and Testing and Assessment in Industrial and Organizational Psychology*. Washington, US-DC: APA [American Psychological Association], 61-84. ISBN: 978-1-4338-1229-3.
- Smith, P (1995): On the Unintended Consequences of Publishing Performance Data in the Public Sector. *International Journal of Public Administration* 18(2-3): 277-310. DOI: 10.1080/01900699508525011.
- 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.
- Spiegelhalter, D; Sherlaw-Johnson, C; Bardsley, M; Blunt, I; Wood, C; Grigg, O (2012): Statistical methods for healthcare regulation: rating, screening and surveillance. *Journal of the Royal Statistical Society: Series A* 175(1): 1-47. DOI: 10.1111/j.1467-985X.2011.01010.x.
- 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.
- Sterne, JAC; Hernán, MA; Reeves, BC; Savović, J; Berkman, ND; Viswanathan, M; et al. (2016): ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 355:i4919. DOI: 10.1136/bmj.i4919.
- Sterne, JAC; Savović, J; Page, MJ; Elbers, RG; Blencowe, NS; Boutron, I; et al. (2019): RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 366:l4898. DOI: 10.1136/bmj.l4898.
- 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, CH: Hans Huber. ISBN: 978-3-456-85500-4.
- SVR Gesundheit [Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen] (2018): Bedarfsgerechte Steuerung der Gesundheitsversorgung. Gutachten 2018. [Stand:] Mai 2018. Bonn: SVR Gesundheit. URL: [https://www.svr-gesundheit.de/fileadmin/user\\_upload/Gutachten/2018/SVR-Gutachten\\_2018\\_WEBSEITE.pdf](https://www.svr-gesundheit.de/fileadmin/user_upload/Gutachten/2018/SVR-Gutachten_2018_WEBSEITE.pdf) (abgerufen am: 10.07.2020).
- Tang, E; Bansal, A; Novak, M; Mucsi, I (2018): Patient-Reported Outcomes in Patients with Chronic Kidney Disease and Kidney Transplant – Part 1. *Frontiers in Medicine* 4. Article 254. DOI: 10.3389/fmed.2017.00254.
- Tausch, A; Menold, N (2015): Methodische Aspekte der Durchführung der Fokusgruppen in der Gesundheitsforschung. Welche Anforderungen ergeben sich aufgrund der besonderen Zielgruppe und Fragestellungen? (GESIS Papers, 2015/12). Mannheim [u. a.]: GESIS [Leibniz-Institut für Sozialwissenschaften]. URL: [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_papers/GESIS-Papers\\_2015-12.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_papers/GESIS-Papers_2015-12.pdf) (abgerufen am: 30.01.2017).
- Taylor, MJ; McNicholas, C; Nicolay, C; Darzi, A; Bell, D; Reed, JE (2014): Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. *BMJ: Quality & Safety* 23(4): 290-298. DOI: 10.1136/bmjqs-2013-001862.
- Tello, JE; Barbazza, E; Waddell, K (2020): Review of 128 quality of care mechanisms: A framework and mapping for health system stewards. *Health Policy* 124(1): 12-24. DOI: 10.1016/j.healthpol.2019.11.006.

- Thompson, C; Sansoni, J; Morris, D; Capell, J; Williams, K (2016): Patient-reported Outcome Measures. An environmental scan of the Australian healthcare sector. Sydney, AU: ACSQHC [Australian Commission on Safety and Quality in Health Care]. ISBN: 978-1-925224-60-3. URL: <https://www.safetyandquality.gov.au/wp-content/uploads/2017/01/PROMs-Environmental-Scan-December-2016.pdf> (abgerufen am: 07.07.2021).
- Thompson, DF (1993): Understanding financial conflicts of interest. *NEJM – The New England Journal of Medicine* 329(8): 573-576. DOI: 10.1056/nejm199308193290812.
- Thompson, DF (2009): The Challenge of Conflict of Interest in Medicine. *ZEFQ – Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 103(3): 136-140. DOI: 10.1016/j.zefq.2009.02.021.
- Tong, A; Lopez-Vargas, P; Howell, M; Phoon, R; Johnson, D; Campbell, D; et al. (2012): Consumer involvement in topic and outcome selection in the development of clinical practice guidelines. *Health Expectations* 15(4): 410-423. DOI: 10.1111/j.1369-7625.2011.00676.x.
- Turpin, RS; Darcy, LA; Koss, R; McMahill, C; Meyne, K; Morton, D; et al. (1996): A Model to Assess the Usefulness of Performance Indicators. *International Journal for Quality in Health Care* 8(4): 321-329. DOI: 10.1093/intqhc/8.4.321.
- Valderas, JM; Alonso, J (2008): Patient reported outcome measures: a model-based classification system for research and clinical practice. *Quality of Life Research* 17: 1125-1135. DOI: 10.1007/s11136-008-9396-4.
- Van Belle, S; Mayhew, SH (2016): What can we learn on public accountability from non-health disciplines: a meta-narrative review. *BMJ: Open* 6: e010425. DOI: 10.1136/bmjopen-2015-010425.
- Van Herck, P; De Smedt, D; Annemans, L; Remmen, R; Rosenthal, MB; Sermeus, W (2010): Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC: Health Services Research* 10. Article number: 247. DOI: 10.1186/1472-6963-10-247.
- Van Slyke, DM (2007): Agents or Stewards: Using Theory to Understand the Government-Non-profit Social Service Contracting Relationship. *Journal of Public Administration Research and Theory* 17(2): 157-187. DOI: 10.1093/jopart/mul012.
- von Elm, E; Schreiber, G; Haupt, CC (2019): Methodische Anleitung für Scoping Reviews (JBI-Methodologie). *ZEFQ – Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 143: 1-7. DOI: 10.1016/j.zefq.2019.05.004.
- Vrangbæk, K; Byrkjeflot, H (2016): Accountability in Health Care. Chapter 21. In: E., F; Montgomery, K; Pedersen, AR: *The Oxford Handbook of Health Care Management*. Oxford, GB: Oxford University Press, 481-495. ISBN: 978-0-19-870510-9.
- Waffenschmidt, S; Janzen, T; Hausner, E; Kaiser, T (2013): Simple search techniques in PubMed are potentially suitable for evaluating the completeness of systematic reviews. *Journal of Clinical Epidemiology* 66(6): 660-665. DOI: 10.1016/j.jclinepi.2012.11.011.

- Weiber, R; Mühlhaus, D (2014): Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS. Zweite, erweiterte und korrigierte Auflage. Berlin [u.a]: Springer Gabler. ISBN: 978-3-642-35011-5.
- Weltärztebund (2017): Deklaration von Genf. Berlin: Bundesärztekammer. URL: [https://www.bundesaerztekammer.de/fileadmin/user\\_upload/downloads/pdf-Ordner/International/Deklaration\\_von\\_Genf\\_DE\\_2017.pdf](https://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/pdf-Ordner/International/Deklaration_von_Genf_DE_2017.pdf) (abgerufen am: 10.07.2020).
- Wessling, M; Gravius, S; Gebert, C; Smektala, R; Günster, C; Harges, J; et al. (2016): Ergebnisqualität in der Revisionsendoprothetik: Eine Analyse von Routinedaten mit dem Vergleich zur externen Qualitätssicherung. *Zeitschrift für Orthopädie und Unfallchirurgie* 154(1): 63-71. DOI: 10.1055/s-0041-107670.
- 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. Genf, CH: WHO. ISBN: 92-4-156198-X. URL: [http://www.who.int/whr/2000/en/whr00\\_en.pdf](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*. Genf, CH: WHO, 182-184. ISBN: 92-4-156243-9. URL: [http://www.who.int/whr/2003/en/whr03\\_en.pdf](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 Services Research* 14:478. DOI: 10.1186/1472-6963-14-478.
- Williams, B (1994): Patient satisfaction: A valid concept? *Social Science & Medicine* 38(4): 509-516. DOI: 10.1016/0277-9536(94)90247-X.
- Willis, GB (2004): Cognitive Interviewing Revisited. A Useful Technique in Theory? Chapter 2. In: Presser, S; Hrsg.: *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, US-NJ: Wiley, 23-43. ISBN: 978-0-471-45841-8.
- Willis, GB (2005): Cognitive Interviewing. A Tool to improve Questionnaire Design. Thousand Oaks, US-CA: Sage. 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. *The 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.

Xyländer, M; Kleineke, V; Jünger, S; Klindtworth, K; Karlheim, C; Steffen, H; et al. (2020): Gruppendiskussionen in der Versorgungsforschung – Teil 2: Überlegungen zum Begriff der Gruppe, zur Moderation und Auswertung von Gruppendiskussionen sowie zur Methode der Online-Gruppendiskussion. *Das Gesundheitswesen* 82(12): 998-1007. DOI: 10.1055/a-0882-6325.

Zeh, S; Christalle, E; Hahlweg, P; Härter, M; Scholl, I (2019): Assessing the relevance and implementation of patient-centredness from the patients' perspective in Germany: results of a Delphi study. *BMJ: Open* 9(12): e031741. DOI: 10.1136/bmjopen-2019-031741.