

**IQTIG**

Institut für  
Qualitätssicherung  
und Transparenz im  
Gesundheitswesen

# **Konstruktion und Auswertungs- methodik für Qualitätsindikatoren der Patientenbefragung**

15. Dezember 2023

# Inhaltsverzeichnis

Tabellenverzeichnis .....	3
Abbildungsverzeichnis .....	3
Abkürzungsverzeichnis .....	4
1 Konzeption von Qualitätsindikatoren der Patientenbefragung .....	5
1.1 Formativ-reflektives Modell zur Konstruktion von Qualitätsindikatoren der Patientenbefragung .....	5
1.2 Vorgehen bei der Aggregation mehrerer Items und Qualitätsmerkmalen zu Qualitätsindikatorergebnissen.....	6
1.2.1 Zweistufiges Vorgehen bei der Aggregation .....	6
1.2.2 Skalierung der Antwortalternativen.....	7
1.3 Statistische Auswertungsmethodik .....	8
1.3.1 Zusammenfassung der Auswertungsmethodik.....	8
1.3.2 Grundlegende Annahmen .....	10
1.3.3 Definition des Indikators für einen Leistungserbringer .....	13
1.4 Referenzbereiche .....	14
1.5 Konsistenz und Homogenität von Items eines Qualitätsindikators .....	15
1.5.1 Bedeutung für den reflektiven Messansatz .....	16
1.5.2 Bedeutung für den formativen Messansatz.....	16
1.5.3 Statistische Kennzahlen .....	17
2 Biometrische Methodik zur Auswertung der Qualitätsindikatoren .....	19
2.1 Detaillierte Darstellung der statistischen Methodik .....	19
2.1.1 Berechnung des Indikatorwerts eines Leistungserbringers .....	19
2.1.2 Spezialfall nur eines binären Items.....	20
2.1.3 Allgemeiner Fall.....	23
2.2 Quantifizierung von Unsicherheit bei der Bestimmung der Indikatorwerte.....	30
2.3 Quantitative Auffälligkeitseinstufung.....	31
2.4 R-Implementation der statistischen Auswertungsmethodik.....	32
2.5 Diskussion .....	32
3 Exemplarische Ergebnisberechnung für einen befragungsbasierten Qualitätsindikator ....	35

Literatur .....39  
Impressum.....42

## Tabellenverzeichnis

Tabelle 1: Beispiele für die Skalierung der Antwortalternativen .....8  
Tabelle 2: Notation für die Erläuterung der Auswertungsmethodik ..... 11  
Tabelle 3: Fragen (Items) und Qualitätsmerkmal eines befragungsbasierten  
Qualitätsindikators .....35  
Tabelle 4: Beispielszenario für einen fiktiven Leistungserbringer mit 5 Fragebögen .....36  
Tabelle 5: Zwischenergebnis pro Item, nachdem erreichte und maximale Punkte ermittelt  
wurden .....36  
Tabelle 6: Berechnungsergebnisse der Parameter  $a$  und  $b$  sowie des Erwartungswerts für  
jeden Merkmalswert ..... 37

## Abbildungsverzeichnis

Abbildung 1: Hierarchische Struktur im formativ-reflektiven Messmodell beispielhaft für  
einen Qualitätsindikator .....6  
Abbildung 2: Graphisches Modell für den  $i$ 'ten Leistungserbringer für einen Indikator .....12  
Abbildung 3: Konjugierte A-priori-Verteilung und resultierende A-posteriori-Verteilung .....22  
Abbildung 4: Kategoriewahrscheinlichkeiten (Kategorie-WK) im HW-Multinomialmodell in  
Abhängigkeit des zugrunde liegenden Parameters  $\theta_{im}$  ..... 27

## Abkürzungsverzeichnis

Abkürzung	Bedeutung
DeQS-RL	Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung
HW	Hardy-Weinberg
MNAR	<i>missing not at random</i>
QI	Qualitätsindikator
QM	Qualitätsmerkmal
QS	Qualitätssicherung
QS PCI	QS-Verfahren <i>Perkutane Koronarintervention und Koronarangiographie</i>
SGB V	Fünftes Buch Sozialgesetzbuch

# 1 Konzeption von Qualitätsindikatoren der Patientenbefragung

Patientenbefragungen haben zum Ziel, Aussagen über die Qualität der Versorgung von Leistungserbringern im Gesundheitswesen, bspw. Krankenhäusern oder Praxen, aus Sicht von Patientinnen und Patienten zu ermöglichen. Fokus der Qualitätsindikatoren der Patientenbefragung ist es, die Versorgungsqualität leistungserbringerbezogen zu messen und so Qualitätsvergleiche zwischen Leistungserbringern zu ermöglichen (SGB V § 137a). Damit liegt das Erkenntnisinteresse in der Schätzung der Kompetenz eines Leistungserbringers in einem bestimmten Versorgungsbereich.<sup>1</sup> Im Folgenden werden die methodische Konzeption des Messmodells und die darauf aufbauende statistische Auswertungsmethodik von Qualitätsindikatoren dargelegt. Diese gelten sowohl für neu entwickelte als auch für in der Vergangenheit entwickelte und ggf. bereits bestehende Patientenbefragungen.

## 1.1 Formativ-reflektives Modell zur Konstruktion von Qualitätsindikatoren der Patientenbefragung

Die Konzeption der Qualitätsindikatoren von Patientenbefragungen orientiert sich an der hierarchischen Struktur der Qualitätsmodelle mit Qualitätsmerkmalen und übergeordneten Qualitätsaspekten (IQTIG 2022: Kapitel 11). Die Qualitätsaspekte sind in Form von bestimmten Qualitätsanforderungen, den Qualitätsmerkmalen, konkretisiert. Diese Qualitätsmerkmale werden durch ein oder mehrere Items operationalisiert.

Die Messung von Qualität anhand von Qualitätsindikatoren folgt gemäß den „Methodischen Grundlagen“ des IQTIG einem formativen Messansatz. Hierbei wird angenommen, dass die Indikatoren das Konstrukt „Qualität der Gesundheitsversorgung“ definieren bzw. *formen* (IQTIG 2022: Abschnitt 4.3). In diesem Sinne formen bei den Qualitätsindikatoren der Patientenbefragung die zugeordneten Qualitätsmerkmale den jeweiligen Qualitätsindikator. Im oberen Teil von Abbildung 1 führen die Pfeile daher von den Qualitätsmerkmalen (QM) zum Qualitätsindikator (QI): Die Ergebnisse zu den zugeordneten Qualitätsmerkmalen werden zu einem Indikatorergebnis aggregiert.

---

<sup>1</sup> Damit liegt das Erkenntnisinteresse nicht auf der Schätzung der Kompetenz auf Ebene einzelner Patientinnen und Patienten. Zur Diskussion über die Anwendbarkeit klassischer psychometrischer Ansätze, welche für den Einsatz von Messinstrumenten zur Untersuchung von Unterschieden zwischen einzelnen Individuen (bspw. Testpersonen oder Patientinnen und Patienten) entwickelt wurden, auf Patientenbefragungen mit Patient-Reported Experience Measures (PREMs) siehe Sizmur et al. (2020).

Auf Ebene der Items, welche die Qualitätsmerkmale operationalisieren, wird ein reflektiver Messansatz verwendet. Alle Items eines Qualitätsmerkmals stellen Messungen der zugrunde liegenden Qualitätsanforderung dar und bilden somit das Qualitätsmerkmal als einheitliches, unidimensionales Konstrukt ab bzw. *reflektieren* dieses. Im unteren Teil von Abbildung 1 führen die Pfeile daher von den Qualitätsmerkmalen zu den jeweiligen Items.

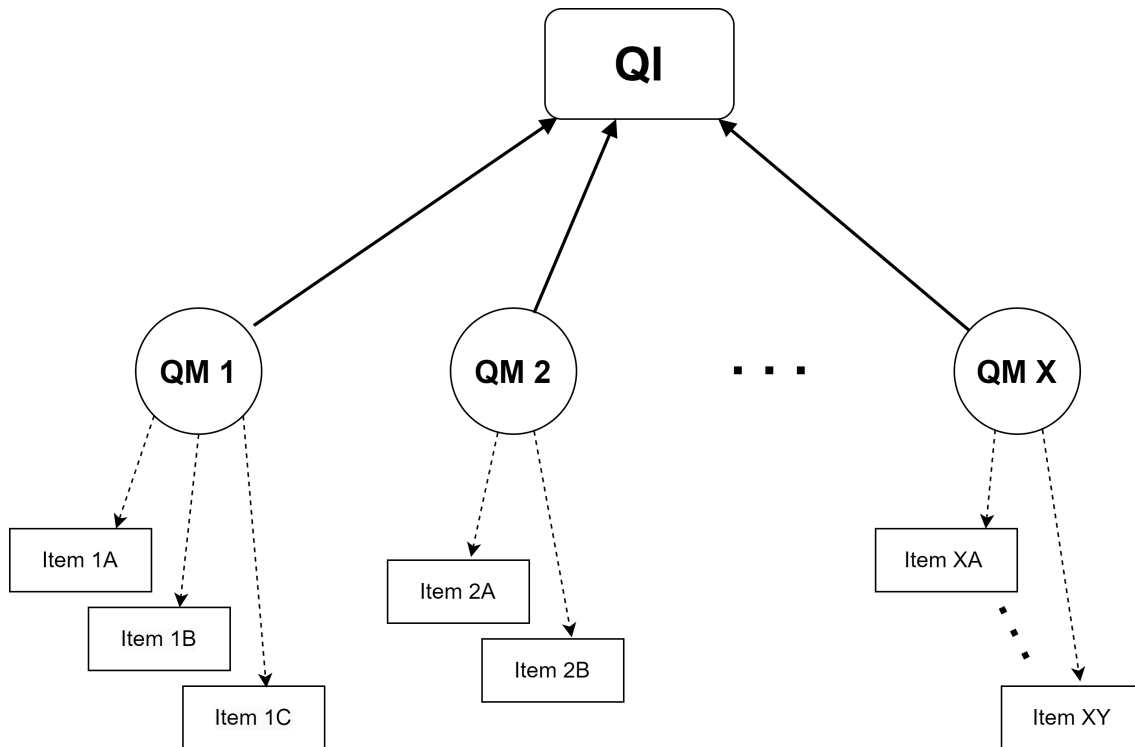


Abbildung 1: Hierarchische Struktur im formativ-reflektiven Messmodell beispielhaft für einen Qualitätsindikator

Qualitätsindikatoren der Patientenbefragung können aus einem oder mehreren Qualitätsmerkmalen bestehen. Die Zuordnung der Qualitätsmerkmale zu den Qualitätsindikatoren erfolgt dabei nach Gesichtspunkten der Inhaltsvalidität und der Anforderung, möglichst kompakte Sets an Qualitätsindikatoren zu entwickeln. Die Qualitätsmerkmale, welche einem Indikator zugeordnet sind, werden über ein oder mehrere Items operationalisiert.

## 1.2 Vorgehen bei der Aggregation mehrerer Items und Qualitätsmerkmalen zu Qualitätsindikatorergebnissen

### 1.2.1 Zweistufiges Vorgehen bei der Aggregation

Zur Bestimmung des Ergebnisses eines Leistungserbringers in einem Qualitätsindikator geht das IQTIG, entsprechend des in Abschnitt 1.1 vorgestellten formativ-reflektiven Modells, folgendermaßen vor:

- Auf Basis der gültigen bzw. substanziellen Antworten (Abschnitt 1.2.2) der von einem Leistungserbringer behandelten Patientinnen und Patienten zu den zugehörigen Items eines Qualitätsmerkmals wird der Kompetenzparameter eines jeden Leistungserbringers für dieses Qualitätsmerkmal geschätzt (Abschnitt 1.3).<sup>2</sup> Der Kompetenzparameter eines Leistungserbringers bemisst dabei die zugrunde liegende Behandlungsqualität des Leistungserbringers bezüglich des Qualitätsmerkmals und bezeichnet somit die konkrete Operationalisierung der zugrunde liegenden Kompetenz eines Leistungserbringers bezüglich eines Qualitätsmerkmals im Rahmen eines Qualitätsindikators (z. B. innerhalb der Definition von Rechenregeln) (IQTIG 2022: Abschnitt 20.1.1).
- Im zweiten Schritt wird auf Basis der Kompetenzparameter aller Qualitätsmerkmale dieses Qualitätsindikators dann das Indikatorergebnis berechnet.<sup>3</sup> Die Aggregation der Kompetenzparameterwerte der Qualitätsmerkmale erfolgt unter der Annahme des formativen Messansatzes. Das bedeutet, dass aus den Kompetenzparametern ein Indexwert gebildet wird.

Da die Qualitätsmerkmale als thematische Ausdifferenzierungen der Qualitätsaspekte gleichwertig behandelt werden, werden alle Qualitätsmerkmale dabei gleichgewichtig aggregiert (Shwartz et al. 2015). So ist das (zugrunde liegende) Indikatorergebnis als ungewichteter Mittelwert der Kompetenzparameter definiert. Die Aggregation der Items erfolgt ebenfalls gleichgewichtet, so dass innerhalb eines Qualitätsindikators eine implizite Gewichtung aller zugehörigen Items vorhanden ist, sofern die eingehenden Qualitätsmerkmale eine unterschiedliche Anzahl an Items aufweisen.

## 1.2.2 Skalierung der Antwortalternativen

Für die Beantwortung der Items werden unterschiedliche Antwortskalen verwendet. In der Regel kommen dichotome Antwortkategorien (z. B. „Ja“ und „Nein“) oder ordinalskalierte Antwortskalen mit vier Antwortkategorien (z. B. „Immer“, „Meistens“, „Selten“, „Nie“) zum Einsatz. Für die Berechnung der Qualitätsindikatoren werden diese gültigen bzw. substanziellen Antworten verwendet, da sie Hinweise zur Qualität der Versorgung geben. Nicht gültige bzw. nicht substanzielle Antwortkategorien wie „Weiß nicht“ bzw. „Weiß nicht mehr“ werden als neutral behandelt und gehen nicht in die Berechnung ein. Antworten, die als Ausweichkategorie dienen, wie „Ich wollte/brauchte das nicht“ oder „Trifft auf mich nicht zu“, sollen Leistungserbringern weder positiv noch negativ zugerechnet werden und werden als ungültige Antwortkategorien ebenfalls nicht für die Berechnung der Indikatoren verwendet.

---

<sup>2</sup> Falls ein Qualitätsmerkmal nur durch ein Item operationalisiert wird, werden pro Leistungserbringer nur die gültigen Antworten zu diesem einen Item aggregiert.

<sup>3</sup> Sofern ein Qualitätsindikator nur durch ein Qualitätsmerkmal mit einem oder mehreren Items repräsentiert wird, entfällt die formative Beziehung zwischen den Merkmalen und dem Indikator. Vielmehr ist der QI identisch zu dem einen zugrundeliegenden Qualitätsmerkmal bzw. dessen latentem Konstrukt. In diesem Fall ist das Messmodell auch als rein reflektives Modell zu verstehen, da die zugrundeliegenden Items Messungen dieses einen latenten Konstrukts darstellen.

Zur Kodierung der Antwortkategorien für die Patientenbefragung wird – in Anlehnung an den National Health Service (NHS) – ein Punktschema verwendet (NHS England Analytic Team 2014). Dabei wird jeder gültigen Antwortalternative ein Wert von 0 (Minimum) bis 100 (Maximum) zugewiesen, wobei höhere Werte eine bessere Bewertung der Qualität widerspiegeln. Die Punktwerte werden dabei gleichmäßig auf die Antwortkategorien verteilt. Beispiele für die Kodierung unterschiedlicher Antwortskalen finden sich in Tabelle 1. Dieses Verfahren impliziert, dass die Abstände zwischen den einzelnen Antwortkategorien eines Items als „gleich“ angenommen werden. Diese Annahme von äquidistanten Antwortkategorien ermöglicht die Interpretation der Punktwerte auch bei mehreren aggregierten Items mit unterschiedlichen Antwortskalen. Da sich die Qualitätsindikatoren durch dieses Vorgehen auf der gleichen Skala wie Prozentangaben bewegen, ist davon auszugehen, dass die Interpretation von Indikatorergebnissen auf dieser Skala erleichtert wird.

Tabelle 1: Beispiele für die Skalierung der Antwortalternativen

Dichotome Antwortalternative		Vierstufige Antwortalternative	
Antwortkategorien	Punkte	Antwortkategorien	Punkte
Ja	100	Immer	100
		Meistens	67
Nein	0	Selten	33
		Nie	0

### 1.3 Statistische Auswertungsmethodik

Die statistische Auswertungsmethodik für die Patientenbefragung folgt dem in Abschnitt 1.1 dargestellten formativ-reflektiven Modell und ist in die statistische Auswertungsmethodik aus den „Methodischen Grundlagen“ (IQTIG 2022: Abschnitt 20.1) eingebettet.<sup>4</sup> Dabei wurden Festlegungen für Zielsetzung, Stichprobenart, Berechnungsart und Bewertungsart getroffen (IQTIG 2022: Abschnitt 20.1).

#### 1.3.1 Zusammenfassung der Auswertungsmethodik

Die Auswertung erfolgt nach einer sogenannten *analytischen Zielsetzung*. Die analytische Zielsetzung hebt hervor, dass es das Ziel der Datenauswertung ist, Aussagen über die zugrunde liegenden Versorgungsprozesse zu generieren. Somit ist die Berücksichtigung von statistischer Unsicherheit eine Kernkomponente der Auswertungsmethodik: Neben dem Schätzwert für die

<sup>4</sup> Die statistische Auswertungsmethodik wurde bereits in den Abschlussberichten zur Entwicklung der Patientenbefragungen der QS-Verfahren *Perkutane Koronarintervention und Koronarangiographie (QS PCI)* (IQTIG 2018b), *Versorgung von volljährigen Patienten und Patientinnen mit Schizophrenie, schizotypen und wahnhaften Störungen (QS Schizophrenie)* (IQTIG 2018a), *Nierenersatztherapie bei chronischem Nierenversagen (NET) einschließlich Pankreastransplantationen (QS NET)* (IQTIG 2021b), *Ambulante Psychotherapie (QS AmbPT)* (IQTIG 2021a) und *Entlassmanagement (QS ENTLASS)* (IQTIG 2023) dargestellt.



zugrunde liegende Größe wird auch ein zugehöriges Unsicherheitsintervall angegeben. Außerdem wird Unsicherheit bei der quantitativen Leistungserbringerbewertung explizit mitberücksichtigt.

Bezüglich der *Stichprobenart* handelt es sich um eine Vollerhebung auf Ebene der Leistungserbringer. Auf Basis der QS-pflichtigen Fälle handelt es sich um eine Vollerhebung (bis 200 Fälle des Leistungserbringers) bzw. eine Zufallsstichprobe von 200 Fällen pro Leistungserbringer, wenn dieser im Erfassungszeitraum mehr als 200 Fälle versorgt hat. Die Stichprobe wird durch die Versendestelle gezogen.

Diese Festlegungen bezüglich Zielsetzung und Stichprobenart führen zu einer *Berechnungsart* als „Patientenbefragungs-Index“, welcher im Folgenden genauer erläutert wird. Weil die Teilnahme an der Befragung (Ausfüllen des Fragebogens) freiwillig ist, ist der adäquate Umgang mit fehlenden Patientenantworten eine weitere wichtige Anforderung an die Auswertungsmethodik bei Patientenbefragungen. Daraus resultiert bei der Operationalisierung – in Übereinstimmung mit dem Rahmenkonzept für die statistische Auswertungsmethodik (IQTIG 2022: Abschnitt 20.1) – ein hierarchischer Bayes-Ansatz. Dieser berücksichtigt die in Abschnitt 1.1 dargelegte hierarchische Struktur, dass jedes Fragebogenitem einem Qualitätsmerkmal zugeordnet ist und ein Qualitätsindikator aus der gemeinsamen Betrachtung eines oder mehrerer Qualitätsmerkmale entsteht. Für jedes Fragebogenitem werden Punktwerte von 0 bis 100 Punkten für die substantiellen Antwortmöglichkeiten der Patientinnen und Patienten festgelegt (Abschnitt 1.2.2). Die erhaltenen Punkte in den ausgefüllten Fragebögen werden entsprechend der Hierarchie von Fragebogenitems und Qualitätsmerkmalen des betrachteten Qualitätsindikators über alle Antworten der entsprechenden Auswertungsebene (z. B. Leistungserbringerebene) inferenztechnisch verrechnet, sodass Schlüsse über den kombinierten Kompetenzparameter auf der Auswertungsebene gezogen werden können. Je größer die Anzahl an tatsächlich ausgewerteten Fragebögen dabei ist, desto größer ist die Sicherheit bei der Bestimmung des Qualitätsindikators. Konkret wird der Qualitätsindikator gemäß des formativ-reflektiven Modells aus Abschnitt 1.1 als A-posteriori-Verteilung über den kombinierten Kompetenzparameter auf der Skala von 0 bis 100 Punkten operationalisiert. Der Erwartungswert dieser A-posteriori-Verteilung wird als Punktschätzer für den kombinierten Kompetenzparameter gewählt und stellt den sogenannten *Indikatorwert* dar. Zweiseitige Unsicherheitsintervalle können als entsprechende Quantile der A-posteriori-Verteilung bestimmt werden.

Bei der *Bewertungsart* handelt es sich, wie im Rahmenkonzept für die statistische Auswertungsmethodik in den „Methodischen Grundlagen“ (IQTIG 2022: Abschnitt 20.1) beschrieben, um eine binäre, statistische Klassifikation des zugrunde liegenden Kompetenzparameters anhand des Referenzwerts. Ziel ist es, die Qualität der Leistungserbringer anhand der vorliegenden Daten auf Auffälligkeiten zu prüfen und im Falle von Auffälligkeiten ein Stellungnahmeverfahren nach § 17

der Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung (DeQS-RL)<sup>5</sup> einzuleiten. Die quantitative Auffälligkeitseinstufung gilt als Operationalisierung des Begriffs „Hinreichender Hinweis für ein Qualitätsdefizit“: Ist genügend statistische Evidenz für das Nichterreichen des Referenzbereichs vorhanden, führt dies im Rahmen eines zweiten qualitativen Schritts zu einer fachlichen Klärung (IQTIG 2022: Abschnitt 20.1). Konkret wird bei der quantitativen Auffälligkeitseinstufung geprüft, ob genügend Masse der A-posteriori-Verteilung auf Punktwerten außerhalb des Referenzbereichs liegt. Dies ist äquivalent zum Vergleich, ob die entsprechende Intervallgrenze des dualen Unsicherheitsintervalls innerhalb oder außerhalb des Referenzbereichs liegt. Details zur Auffälligkeitseinstufung finden sich in Abschnitt 2.3.

### 1.3.2 Grundlegende Annahmen

Die Definition der Indikatoren der Patientenbefragung sowie die Berechnungen der Indikatorenergebnisse einschließlich der quantitativen Einstufung basieren auf bestimmten Grundannahmen hinsichtlich der Struktur und der Zusammenhänge der zugrunde liegenden Qualitätsaspekte, Qualitätsmerkmale und Fragebogenitems, welche durch das in Abschnitt 1.1. dargelegte formativ-reflektive Modell vorgeben werden.

#### Notation

In Tabelle 2 wird die Notation für die folgende Erläuterung der Auswertungsmethodik eingeführt. Diese Notation orientiert sich an gängiger statistischer Methodik für hierarchische Modelle (vgl. z. B. Skrondal und Rabe-Hesketh 2004, Carlin und Louis 2009).

Um die Klarheit der Darstellung zu erhöhen, werden Zufallsvariablen (mit Ausnahme von Verteilungsparametern<sup>6</sup>) mit Großbuchstaben gekennzeichnet, wohingegen die Realisierungen mit den entsprechenden Kleinbuchstaben benannt sind. Vektorielle Größen sind fett gedruckt. Schätzer einer Größe, z. B. von  $\theta$ , werden im Folgenden mit einem Dach, d. h. im Beispiel  $\hat{\theta}$ , bezeichnet.

---

<sup>5</sup> Richtlinie des Gemeinsamen Bundesausschusses zur datengestützten einrichtungsübergreifenden Qualitätssicherung. In der Fassung vom 19. Juli 2018, zuletzt geändert am 15. Dezember 2022, in Kraft getreten am 1. Januar 2023. URL: <https://www.g-ba.de/informationen/richtlinien/105/> (abgerufen am: 24.11.2023).

<sup>6</sup> Verteilungsparameter können im Bayesianischen Kontext auch Zufallsvariablen darstellen.

Tabelle 2: Notation für die Erläuterung der Auswertungsmethodik

Notation	Erläuterung
$i = 1, \dots, I$	Index für die Leistungserbringer
$j = 1, \dots, J_i$	Index für die Patientinnen und Patienten von Leistungserbringer $i$ , die die Grundgesamtheitsbedingungen des Indikators erfüllen und einen Fragebogen zurückgesendet haben
$QI_i$	Indikator für Leistungserbringer $i$ (zugrunde liegende Größe)
$Y_{ijml}, l = 1, \dots, L_m$	Antwort von Patientin oder Patient $j$ behandelt von Leistungserbringer $i$ auf Item $l$ von Qualitätsmerkmal $m$ (Zufallsvariable)
$K_{ml}$	Anzahl Antwortkategorien für Item $l$ von Qualitätsmerkmal $m$
$A \vee B$	A oder B (logischer Operator)
$A := B$	A wird definiert als B
$\text{Log}(x)$	Natürlicher Logarithmus von $x$
$E(X)$	Erwartungswert der Zufallsvariable $X$
$P(X \geq x)$	Wahrscheinlichkeit, dass Zufallsvariable $X$ einen Wert größer/gleich $x$ annimmt
$X \perp Y$	Zufallsvariablen $X$ und $Y$ sind statistisch unabhängig
$X   y$	Zufallsvariable $X$ bedingt darauf, dass Zufallsvariable $Y$ den Wert $y$ annimmt
$X \sim F$	Zufallsvariable $X$ folgt der Wahrscheinlichkeitsverteilung $F$
$X_i \stackrel{\text{ind.}}{\sim} F_i, i = 1, \dots, n$	Zufallsvariablen $X_1, \dots, X_n$ sind unabhängig verteilt und folgen den jeweiligen Wahrscheinlichkeitsverteilungen $F_i$
$X_i \stackrel{\text{i.i.d.}}{\sim} F, i = 1, \dots, n$	Zufallsvariablen $X_1, \dots, X_n$ sind unabhängig und identisch verteilt und folgen der Wahrscheinlichkeitsverteilung $F$
$\text{Ber}(\pi)$	Bernoulli-Verteilung mit Erfolgswahrscheinlichkeit $\pi \in [0, 1]$
$\text{Binom}(n, \pi)$	Binomialverteilung mit Anzahl an Versuchen $n \in \mathbb{N}_+$ und Erfolgswahrscheinlichkeit $\pi \in [0, 1]$
$\text{Mult}(n, \boldsymbol{\pi} = (\pi_1, \dots, \pi_k)')$	Multinomialverteilung mit Anzahl an Versuchen $n \in \mathbb{N}_+$ und Ereigniswahrscheinlichkeiten $\pi_1, \dots, \pi_k \in [0, 1], \sum_{i=1}^k \pi_i = 1$
$\text{Beta}(a, b)$	Betaverteilung mit Parametern $a > 0, b > 0$
$N(\mu, \sigma^2)$	Normalverteilung mit Erwartungswert $\mu$ und Varianz $\sigma^2$
$\text{Logistic}(\alpha, \beta)$	Logistische Verteilung mit Parametern $\alpha, \beta > 0$

Abbildung 2 zeigt das formativ-reflektive Modell zur Konzeption der Qualitätsindikatoren der Patientenbefragung als graphisches Modell für den  $i$ 'ten Leistungserbringer; dabei gilt es zu beachten, dass die Pfeile in der Abbildung lediglich die formative bzw. reflektive Beziehung im Messmodell darstellen und keine Zusammenhänge geschätzt bzw. berechnet werden. Auf unterster Ebene stehen die Patientenantworten  $Y_{ijml}$  bezüglich der einzelnen Items, wobei die Umrahmungen der untersten Ebene die Antworten der einzelnen Patientinnen und Patienten symbolisieren. Das bedeutet, jede Patientin bzw. jeder Patient wird in dem graphischen Modell durch ein Plättchen dargestellt. Die Items stellen Operationalisierungen der Qualitätsmerkmale dar. Somit liegt allen Patientenantworten des  $m$ 'ten Qualitätsmerkmals der gemeinsame, unbekannte Parameter  $\theta_{im}$  zugrunde. Dieser Parameter kann als Kompetenz des Leistungserbringers bezüglich des im  $m$ 'ten Qualitätsmerkmal abgebildeten Kompetenzfeldes interpretiert werden.

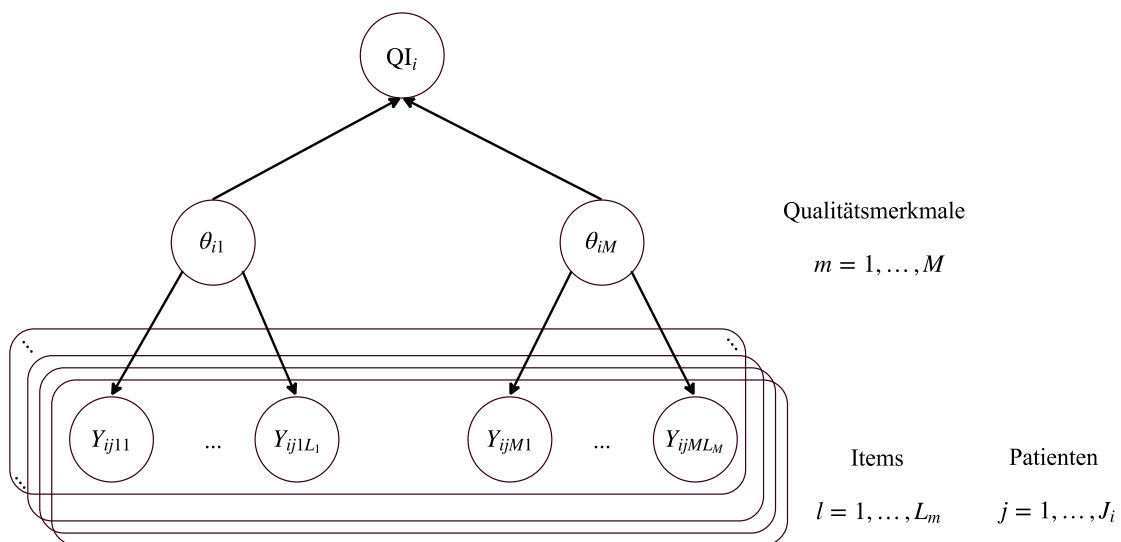


Abbildung 2: Graphisches Modell für den  $i$ 'ten Leistungserbringer für einen Indikator

Beispielsweise könnte es sich dabei um die zugrunde liegende Aufklärungsrate über Behandlungsalternativen handeln, wenn im Qualitätsmerkmal die Aufklärung über Behandlungsalternativen betrachtet wird. Welche Antwort eine Patientin bzw. ein Patient auf die Items zugehörig zum  $m$ 'ten Qualitätsmerkmal gibt, hängt laut formativ-reflektiven Modell somit von der Kompetenz des Leistungserbringers bezüglich des im entsprechenden Qualitätsmerkmal abgebildeten Kompetenzfeldes ab. Diese Annahme rechtfertigt, dass aus den Patientenantworten Schlüsse auf die Kompetenz des Leistungserbringers gezogen werden können. Gemäß des formativ-reflektiven Messmodells wird an dieser Stelle der reflektive Messansatz der Items zu den Qualitätsmerkmalen abgebildet.

Aus dem graphischen Modell lässt sich zudem ablesen, dass Unabhängigkeit der Antworten auf Items desselben Qualitätsmerkmals bedingt auf den Parameter auf Qualitätsmerkmalsebene angenommen wird (Abschnitt 2.5 zur Diskussion dieser Annahme), d. h. mit der mathematischen Notation

$$Y_{ijml} \perp Y_{ij'm'l'} \mid \theta_{im}, \text{ sofern } j \neq j' \vee l \neq l'$$

Weiterhin wird angenommen, dass die Patientenantworten auf Items, die zu unterschiedlichen Qualitätsmerkmalen gehören, unabhängig sind, d. h.

$$Y_{ijml} \perp Y_{ijm'l} \text{ für } m \neq m'$$

Diese beiden Unabhängigkeitsannahmen beziehen sich sowohl auf Antworten derselben Patientin bzw. desselben Patienten als auch auf Antworten zwischen verschiedenen Patientinnen und Patienten.

Der Indikator selbst,  $QI_i$ , wird als Kombination der unbekanntes qualitätsmerkmalspezifischen Parameter definiert. Im Gegensatz zum unteren Teil des graphischen Modells, bei dem die Pfeile und somit die Kausalitätsrichtung von oben nach unten gerichtet sind, liegt der Indikator den Parametern  $\theta_{i1}, \dots, \theta_{iM}$  nicht zugrunde, sondern wird aus ihnen formativ gebildet. Somit bildet sich entsprechend des grundlegenden formativ-reflektiven Modells die Beziehung zwischen der Ebene der Qualitätsmerkmale und ihren Operationalisierungen, den Items, über einen reflektiven Messansatz ab, wohingegen die Beziehung zwischen Qualitätsmerkmalen und Qualitätsindikator formativ ist.

### 1.3.3 Definition des Indikators für einen Leistungserbringer

Wie aus dem graphischen Modell in Abbildung 2 hervorgeht, wird der Indikator für Leistungserbringer  $i$  als Kombination der qualitätsmerkmalspezifischen Parameter  $\theta_{im}, m = 1, \dots, M$ , definiert. Eine naheliegende Möglichkeit, die Qualitätsmerkmale in einem Indikator zusammenzufassen, besteht darin, die zugehörigen Parameter zu mitteln. Dabei wird von einer Gleichgewichtung der Qualitätsmerkmale und somit der Parameter ausgegangen (siehe auch Abschnitt 1.2). Der Indikator für Leistungserbringer  $i$  wird somit wie folgt definiert:

$$QI_i := \frac{1}{M} \sum_{m=1}^M \theta_{im}$$

Die Parameter  $\theta_{im}$  sind dabei unbekannte Größen, die, wie in Abschnitt 1.3.2 beschrieben, als Kompetenz des Leistungserbringers bezüglich des  $m$ 'ten Qualitätsmerkmals aufgefasst werden können. Wie in Abbildung 2 dargestellt, wird angenommen, dass die Parameter  $\theta_{im}$  den Patientenantworten auf die Items des  $m$ 'ten Qualitätsmerkmals zugrunde liegen. Genauer wird angenommen, dass die Patientenantworten bei Leistungserbringer  $i$  auf Items des  $m$ 'ten Qualitätsmerkmals einer Wahrscheinlichkeitsverteilung  $F$  mit Parameter  $\theta_{im}$  folgen, d. h.

$$Y_{ijml} \mid \theta_{im} \stackrel{\text{i.i.d.}}{\sim} F(\theta_{im})$$

Demnach wird davon ausgegangen, dass, abgesehen von stochastischen Komponenten, die Antworten der Patientinnen und Patienten die Kompetenz des Leistungserbringers bezüglich des jeweiligen Qualitätsmerkmals widerspiegeln.

Das Interesse liegt also in einer Funktion von unbekanntes Parametern. Anhand der Patientenantworten wird Inferenz für die Parameter  $\theta_{im}$  und somit für den Indikator betrieben. Dabei gehen

alle vorliegenden Patientenantworten auf Items des  $m$ 'ten Qualitätsmerkmals in die Inferenz für  $\theta_{im}$  ein. Die konkrete Bestimmung des Indikatorwerts eines Leistungserbringers wird in Abschnitt 2.1.2 zunächst anhand eines einfachen Spezialfalls und dann im Allgemeinen beschrieben. Die entscheidende Annahme ist dabei, dass die zugrunde liegenden Kompetenzparameter der Qualitätsmerkmale eines Qualitätsindikators als unabhängig voneinander angesehen werden können.

## 1.4 Referenzbereiche

Qualitätsindikatoren erlauben einen Rückschluss auf die Versorgungsqualität eines Leistungserbringers über den Abgleich der oberen Grenze des zum Indikatorwert gehörigen Unsicherheitsintervalls eines Leistungserbringers mit dem Referenzbereich des jeweiligen Indikators. Im Rahmen von QS-Verfahren, die Qualitätsverbesserungen durch externen Vergleich und Fördermaßnahmen anstreben, definieren Referenzbereiche Entscheidungsschwellen, bei deren Unterschreitung gemäß der DeQS-RL qualitätssteigernde Maßnahmen eingeleitet werden sollen.

Gemäß den „Methodischen Grundlagen“ des IQTIG werden – wann immer möglich – feste Referenzbereiche definiert (IQTIG 2022: Kapitel 16). Feste Referenzbereiche erlauben eine von der Versorgungsqualität der anderen Leistungserbringer unabhängige Bewertung der Versorgungsqualität. Die indikatorspezifische Ableitung solcher Standards erfolgt nach den „Methodischen Grundlagen“ des IQTIG auf Basis einer Einschätzung, welche Indikatorwerte erreichbar wären, und unter Berücksichtigung von Leitlinien, wissenschaftlichen Studien sowie medizinisch-ethischen oder rechtlichen Grundlagen (IQTIG 2022: Abschnitt 16.3). Verteilungsbasierte Referenzbereiche werden dagegen immer dann definiert, wenn die Setzung eines festen Referenzbereichs auf Basis der vorliegenden Informationen (noch) nicht ausreichend ist oder relative Qualitätsaussagen getroffen werden sollen.

### Feste Referenzbereiche

Vor allem für Prozesse, deren Durchführung immer indiziert ist und die durch die Leistungserbringer beeinflussbar sind, sollte der erreichbare Indikatorwert üblicherweise bei 100 % bzw. für die Bewertung der Indikatoren der Patientenbefragung bei 100 Punkten liegen. In der Versorgungspraxis finden sich jedoch immer wieder Versorgungssituationen, die durch die Qualitätsindikatoren (und ggf. durch eine zusätzliche Risikoadjustierung) nicht adäquat abgebildet werden und daher selbst bei bester Versorgungsqualität eine vollständige Erfüllung des Qualitätsziels erschweren. Aus diesem Grund wurden die Referenzbereiche auf  $\geq 95$  Punkte festgelegt, um auch solchen besonderen Versorgungssituationen gerecht zu werden. Dieser Abschlag von 5 Punkten berücksichtigt pauschal, dass nicht jede Versorgungskonstellation im Qualitätsindikator adäquat abgebildet wird und nicht jeder Abweichung vom erreichbaren Indikatorwert nachgegangen werden muss. Ein höherer Abschlag wird nicht empfohlen, da die Auffälligkeitseinstufung für die Qualitätsindikatoren auf Basis von Patientenbefragungen zusätzlich unter Berücksichtigung statistischer Unsicherheit anhand entsprechender Vertrauensintervalle der Leistungserbringerergebnisse erfolgt. Diese Vertrauensintervalle sind abhängig von der Fallzahl und erlauben daher mit

abnehmender Fallzahl eine größere Abweichung des Indikatorergebnisses vom Referenzbereich, bevor festgestellt wird, dass das Ergebnis eines Leistungserbringers den Referenzbereich nicht erreicht hat.

### **Verteilungsbasierte Referenzbereiche**

Vor allem für befragungsbasierte Ergebnisindikatoren, deren Qualitätsziele ein möglichst hohes Maß an gesundheitsbezogener Lebensqualität oder eine möglichst niedrige Symptombelastung fordern, kann es sinnvoll sein, verteilungsbasierte Referenzbereiche zu definieren, da sich oft a priori keine erwartbaren Standards aus wissenschaftlichen Studien oder Leitlinien ableiten lassen. Auch für bestimmte Prozessvariablen können verteilungsbasierte Referenzbereiche besser geeignet sein als feste Referenzbereiche, z. B. bei Indikatoren, die sich auf die Häufigkeit von wiederkehrenden erwünschten Ereignissen im Behandlungsprozess über einen längeren Zeitraum beziehen.

Es gibt zur Auffälligkeitseinstufung verschiedene Vorgehensweisen. So kann über perzentilbasierte Referenzbereiche der Referenzwert so gestaltet werden, dass bspw. ca. 5 % der Leistungserbringerergebnisse auffällig werden, wenn diese in den jeweiligen Qualitätsindikatoren die wenigsten Punkte erreicht haben. Es wurde am IQTIG eine entsprechende Methodik entwickelt (IQTIG 2022: Abschnitt 20.1.4), die hierfür sowohl den Indikatorwert als auch die durch Stochastizität verursachte Unsicherheit des Ergebnisses berücksichtigt. Alternativ zu einem bestimmten Perzentil kann der Referenzwert auch anhand einer anderen Charakteristik der Verteilung, wie z. B. des Mittelwerts über alle Leistungserbringerergebnisse, gesetzt werden. Liegt das Ergebnis einer Einrichtung signifikant unter diesem Referenzwert, kann dies zur Interpretation für niedrige Qualität herangezogen werden. Bei einer solchen Vorgehensweise würden dann beispielhaft nur die Leistungserbringerergebnisse auffällig, die einer Qualität entsprechen, die signifikant unter der durchschnittlichen Qualität aller Leistungserbringer läge.

## **1.5 Konsistenz und Homogenität von Items eines Qualitätsindikators**

Aus dem formativ-reflektiven Messmodell für die Qualitätsindikatoren der Patientenbefragung und der entsprechenden statistischen Auswertungsmethodik ergeben sich Anforderungen daran, in welchem Maße sich die Items eines Indikators auf ein gemeinsames Konstrukt beziehen sollen. Dabei werden die zwei Ebenen in der hierarchischen Struktur des formativ-reflektiven Messmodells getrennt betrachtet:

- Die Konsistenz und Homogenität der Items innerhalb eines Merkmals: Diese wird betrachtet, um zu prüfen, ob die entwickelten Items der Annahme eines reflektiven Messansatzes und einem damit einhergehenden unidimensionalen Konstrukt auf Ebene der Qualitätsmerkmale entsprechen.

- Die Konsistenz und Homogenität aller Items eines Qualitätsindikators als Maß für die inhaltliche Verbundenheit der zugeordneten Merkmale im Sinne des formativen Messansatzes: Diese Betrachtung dient dazu sicherzustellen, dass die Items gleichgerichtet sind und eine thematische Grundlage bzw. konzeptuelle Einheit für den Qualitätsindikator vorhanden ist.

### 1.5.1 Bedeutung für den reflektiven Messansatz

Im Sinne der klassischen psychometrischen Testtheorie werden zur Prüfung, ob Items der Annahme eines unidimensionalen Konstrukts entsprechen, statische Kennzahlen der internen Konsistenz verwendet und entsprechende Schwellenwerte festgelegt (Abschnitt 1.5.3). Entsprechen die Items eines Merkmals nicht den Anforderungen eines einheitlichen Konstruktes, muss geprüft werden, inwieweit durch die Modifikation des Qualitätsmerkmals bspw. durch Aufspaltung des Qualitätsmerkmals statistisch konsistentere Items pro Qualitätsmerkmal erreicht werden können.

### 1.5.2 Bedeutung für den formativen Messansatz

Im Rahmen des formativen Messansatzes wird die Anforderung an die Homogenität der Items als Maß für die inhaltliche Verbundenheit der Qualitätsmerkmale herangezogen. Aufgrund des formativen Messansatzes, welcher für die Beziehung zwischen einem Qualitätsindikator und seinen Qualitätsmerkmalen angenommen wird, kann die Homogenität der zugehörigen Items bezogen auf den gesamten Qualitätsindikator niedriger ausfallen, als dies bei reflektiv operationalisierten Konstrukten üblicherweise angestrebt wird. Dennoch sollen alle Items, die zur Erfassung eines Qualitätsindikators herangezogen werden, positiv interkorreliert sein. Positive Interkorrelationen der Items untereinander sowie positive Korrelationen der Items mit dem Gesamtwert des Indikators ermöglichen die Interpretation eines Qualitätsindikators. Je stärker die einzelnen Items untereinander und mit dem Gesamtwert assoziiert sind, desto eindeutiger ist das Indikatorergebnis zu interpretieren.

Sehr niedrige oder negative Interkorrelationen der Items führen dazu, dass hohe Werte auf einem Item niedrige Werte auf einem anderen Item ausgleichen. Aufgrund der hierarchischen Struktur des formativ-reflektiven Messmodells werden diese Kompensationseffekte auf Ebene der Qualitätsmerkmale eines Indikators übertragen und können somit zu einer reduzierten Variabilität der Indikatorergebnisse führen. Der Qualitätsindikator ist dadurch weniger gut geeignet, gute und schlechte Leistungserbringerergebnisse voneinander zu trennen. Dies erschwert auch die Detektion von Qualitätsdefiziten. Im ungünstigsten Fall könnte dies dazu führen, dass Leistungserbringer stets nur einen beliebigen Teil der Inhalte eines Qualitätsindikators erfüllen müssen, ohne dass ihre Ergebnisse auffällig werden.

Ergeben sich aus der Betrachtung der Homogenität aller Items eines Qualitätsindikators Hinweise auf ein geringes Maß an Homogenität, wird mit Blick auf die dem betroffenen Qualitätsindikator zugrunde liegenden Qualitätsmerkmale überprüft, inwieweit eine Modifikation in der Zusammenstellung der Qualitätsmerkmale des Indikators notwendig bzw. sinnvoll ist (z. B. Aufteilung der zusammengefassten Qualitätsmerkmale auf mehrere Qualitätsindikatoren). Dies ist aus inhaltlicher



Sicht insbesondere dann der Fall, wenn negative Korrelationen von Items eines Qualitätsindikators vorliegen. Weiterhin gilt es zu berücksichtigen, dass aufgrund einer deutlichen inhaltlich-thematischen Verbundenheit der Qualitätsmerkmale zum Qualitätsindikator trotz geringer, aber positiver statistischer Kennzahlen der Homogenität und internen Konsistenz (siehe Abschnitt 1.5.3) auch die Möglichkeit besteht, keine Anpassungen in den Kompositionen der Indikatoren vorzunehmen. Dies ist dann der Fall, wenn im Rahmen der Qualitätsförderung angenommen werden kann, dass die Leistungserbringer erst im Verlauf der Zeit die entsprechenden Kompetenzen entwickeln, um die Qualitätsanforderungen zu erfüllen.

### 1.5.3 Statistische Kennzahlen

Um die oben dargestellte Bedeutung der Konsistenz und Homogenität für den reflektiven und den formativen Messansatz betrachten und ggf. Anpassungen an der Zusammenstellung der Items bzw. Qualitätsmerkmale eines Indikators vornehmen zu können, sollen statistische Kennzahlen herangezogen und passende Schwellenwerte definiert werden. Dabei gilt es zu beachten, dass sich das IQTIG zum Ziel gesetzt hat, kompakte Messinstrumente zu entwickeln, die anhand von möglichst wenigen Items (Datensparsamkeit), die dabei alle Facetten eines Konstrukts abdecken, zum zuverlässigen Vergleich von Einrichtungen herangezogen werden können.

#### **Statistische Kennzahlen für den reflektiven Messansatz**

Um die Annahme eines unidimensionalen Konstrukts, welches die Items eines Qualitätsmerkmals abbilden, zu prüfen, sollen statistische Kennzahlen zur internen Konsistenz herangezogen werden. Im klassischen psychometrischen Testansatz wird häufig Cronbachs Alpha als Maß für die interne Konsistenz einer Antwortskala herangezogen (Bühner 2011). Cronbachs Alpha liegt allerdings die Annahme zugrunde, die für Konstrukte mit relativ heterogenen Items nur begrenzt geeignet sind (Tavakol und Dennick 2011, Zinbarg et al. 2005). Zudem ist Cronbachs Alpha nicht robust gegenüber nicht normalverteilten Daten – was im Kontext der externen Qualitätssicherung eher die Regel als eine Ausnahme darstellt – und wird dadurch tendenziell unterschätzt (Sheng und Sheng 2012, Trizano-Hermosilla und Alvarado 2016). Als robusteres Maß mit weniger restriktiven Annahmen wird daher McDonalds Omega als Koeffizient für die interne Konsistenz berechnet, wobei die Interpretation des Koeffizienten äquivalent zu Cronbachs Alpha erfolgt (Trizano-Hermosilla und Alvarado 2016). Welcher Wert als ausreichend definiert wird, hängt dabei u. a. davon ab, ob die Konstruktausprägung einzelner Individuen gemessen werden soll oder ob aggregierte Messwerte auf Gruppenebene erfasst werden sollen. Für die Erfassung von Konstrukten auf Gruppenebene – worauf auch der Einsatz der entwickelten Fragebögen abzielt – werden generell niedrigere Werte als ausreichend angesehen. Als Hinweis auf eine hinreichend homogene Messung sollte aus Sicht des IQTIG die interne Konsistenz einen Wert von 0,6 für McDonalds Omega nicht unterschreiten.

### **Statistische Kennzahlen für den formativen Messansatz**

Als Maß für den Zusammenhang eines einzelnen Items mit dem Gesamtwert eines Indikators verwendet das IQTIG die korrigierte Item-Skala-Korrelation (Revelle 2018). Die korrigierte Item-Skala-Korrelation wird in erster Linie herangezogen, um sicherzustellen, dass Items nicht negativ mit dem Gesamtwert korreliert sind. Darüber hinaus wird für jedes Itemset, das zur Messung eines Qualitätsindikators herangezogen wird, Loevingers  $H$  als Maß für die Homogenität berechnet (Molenaar und Sijtsma 1984). Loevingers  $H$  beschreibt das Ausmaß, in dem Patientinnen und Patienten mit gleichem Summenscore gleiche Antwortmuster zeigen, was konzeptuell der durchschnittlichen Korrelation zwischen den Items entspricht (Goeman und De Jong 2018). Für Qualitätsindikatoren mit nur 2 Items kann der  $\text{Chi}^2$ -Test und das Zusammenhangsmaß Cramers  $V$  berechnet werden. Das IQTIG geht von einer ausreichenden Homogenität der Items eines Qualitätsindikators aus, wenn Loevingers  $H$  größer als 0,3 ist (Stochl et al. 2012) und die korrigierte Item-Skala-Korrelationen für jedes Item positiv ausfallen.

## 2 Biometrische Methodik zur Auswertung der Qualitätsindikatoren

Im Folgenden wird die für die Patientenbefragung entwickelte statistische Methodik im Detail beschrieben. Aufbauend auf der in Kapitel 1 erläuterten methodischen Herangehensweise, den grundlegenden Annahmen und der formalen Definition eines Indikators werden im Folgenden die konkrete Berechnung der Indikatorwerte und der dazugehörigen Unsicherheitsintervalle sowie die quantitative Einstufung beschrieben.

### 2.1 Detaillierte Darstellung der statistischen Methodik

#### 2.1.1 Berechnung des Indikatorwerts eines Leistungserbringers

Der Indikatorwert des  $i$ 'ten Leistungserbringers ergibt sich als Schätzwert des im Abschnitt 1.3.3 definierten Indikators,  $QI_i$ , basierend auf den für den Indikator relevanten<sup>7</sup> Patientenantworten für diesen Leistungserbringer.

Die Patientenantworten liegen bei dieser Patientenbefragung in der Anzahl an Kategorien entsprechenden, gleichabständigen Punkten zwischen 0 und 100 vor, wobei einheitlich für alle Items 100 Punkte die bestmögliche Antwort im Sinne der Qualitätssicherung abbildet (Abschnitt 1.2.2). Aufgrund statistisch-methodischer Aspekte basiert die Berechnung der Indikatorwerte und entsprechend auch der dazugehörigen Unsicherheitsintervalle nicht auf den Punkten selbst, sondern auf den entsprechenden gleichabständigen Werten zwischen 0 und 1. Insbesondere werden in der Berechnung nicht die gerundeten Punkte 0, 33, 67 und 100 im Beispiel von 4 Kategorien verwendet, sondern das Intervall von 0 bis 1 wird gleichabständig unterteilt in  $0, \frac{1}{3}, \frac{2}{3}, 1$ . Für die Darstellung der Ergebnisse werden für eine bessere Verständlichkeit wieder Punkte zwischen 0 und 100 angegeben. Mehr Details dazu finden sich in Abschnitt 2.1.3.

Zur Berechnung der Indikatorwerte für die Leistungserbringer wird ein Bayesianischer Modellierungsansatz verwendet (z. B. Carlin und Louis 2009). Bayesianische Verfahren haben den Vorteil gegenüber einer frequentistischen Herangehensweise, dass sie sehr flexibel, leicht interpretierbar und gut erweiterbar sind und auch bei niedrigen Fallzahlen gut funktionieren.

In der Bayesianischen Modellierung wird die A-priori-Information über die Modellparameter mittels der Informationen aus den vorliegenden Daten aufdatiert und die daraus resultierende A-

---

<sup>7</sup> Es werden im Allgemeinen pro Indikator nur Antworten von Patientinnen und Patienten berücksichtigt, die die Grundgesamtheitsbedingungen des Indikators erfüllen. Eine Bedingung für den Einschluss in die Berechnung eines Indikatorwerts kann bspw. sein, dass eine bestimmte Prozedur durchgeführt wurde oder dass es sich bei dem betrachteten Eingriff um einen stationären Eingriff handelte. Folglich gehen neben den Patientenantworten auf Items, die zur direkten Berechnung des Indikatorwerts benötigt werden, indirekt auch Antworten auf Items ein, die in der Definition der Grundgesamtheitsbedingung verwendet werden.

posteriori-Verteilung der Parameter für die Inferenz über die Parameter verwendet. Je mehr Informationen aus den Daten gewonnen werden können, desto weniger fällt das Vorwissen über die Parameter ins Gewicht. Bei sehr kleinen Fallzahlen spielt die A-priori-Verteilung hingegen eine vergleichsmäßig große Rolle.

A-priori- und A-posteriori-Verteilungen, die für eine gegebene Verteilung der Daten den gleichen Verteilungstyp besitzen, werden als konjugierte Verteilungen bezeichnet. Die A-priori-Verteilung wird auch zur Verteilung der Daten konjugierte A-priori-Verteilung genannt. Die Verwendung konjugierter Verteilungen bringt insbesondere zwei Vorteile mit sich. Erstens lässt sich die A-posteriori-Verteilung analytisch darstellen und muss nicht mittels numerischer Verfahren approximiert werden, um Inferenz für die Parameter zu betreiben. Dies erlaubt eine effiziente Berechnung, bei der die Berechnungszeit im Vergleich zu nicht konjugierten Verteilungen wesentlich verringert wird. Zweitens liefern konjugierte Verteilungen häufig eine intuitive Interpretation dafür, wie sich die A-posteriori-Verteilung durch das Aufdatieren der A-priori-Verteilung ergibt. Aus diesen Gründen werden für die Bestimmung der Indikatorwerte der Leistungserbringer konjugierte A-priori-Verteilungen verwendet.

Im folgenden Abschnitt wird die konkrete Bestimmung der Indikatorwerte zunächst im Spezialfall nur eines binären Items im Detail erläutert. Darauf folgt die Darstellung für den allgemeinen Fall in Abschnitt 2.1.3.

## 2.1.2 Spezialfall nur eines binären Items

Betrachtet wird im Folgenden ein Indikator, in den nur ein Qualitätsmerkmal einfließt, welches durch ein einziges binäres Item operationalisiert ist. In der verwendeten Notation (Abschnitt 1.3.2) bedeutet dies  $M = L_1 = 1$  und  $K_{11} = 2$ . Vorausgesetzt, eine vom  $i$ 'ten Leistungserbringer behandelte Patientin bzw. ein behandelter Patient  $j$  fällt in die Grundgesamtheit des Indikators, so kann eine Antwort  $Y_{ij11}$  entweder den Wert 0 oder 1 annehmen bzw. 0 oder 100 Punkte ergeben. Es ist auch möglich, dass die Antwort auf das Item ein fehlender Wert ist, falls die Patientin oder der Patient die Frage unbeantwortet gelassen oder keine gültige Antwort gegeben hat. In diesem Fall würde die Patientin oder der Patient nicht in die Berechnung dieses Indikators eingehen, da der Indikator in diesem Spezialfall nur aus diesem einen Item besteht, d. h., der Indikator definiert ist als  $Q_i = \theta_{i1}$ . Da es sich um ein binäres Item handelt, kann der Indikator in diesem Spezialfall als die zugrunde liegende „Erfolgswahrscheinlichkeit“, d. h. die Wahrscheinlichkeit für eine Antwort mit dem Wert 1, aufgefasst werden.

Einen Indikatorwert für den  $i$ 'ten Leistungserbringer erhält man in diesem Spezialfall, indem man einen Wert für den unbekannt Parameter  $\theta_{i1}$  aus den vorliegenden Patientenantworten des  $i$ 'ten Leistungserbringers schätzt. Die Schätzung von  $\theta_{i1}$  basiert auf der Annahme, dass die (binären) Patientenantworten  $Y_{ij11}$  Bernoulli-verteilte Zufallsvariablen mit Erfolgswahrscheinlichkeit  $\theta_{i1}$  sind und bedingt auf  $\theta_{i1}$  als unabhängig voneinander betrachtet werden können (Abschnitt 1.3.2). Die Wahrscheinlichkeitsverteilung  $F$  aus Abschnitt 1.3.3 entspricht somit einer Bernoulli-

Verteilung und es gilt  $Y_{ij11} | \theta_{i1} \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(\theta_{i1})$ . Daraus folgt, dass die Summe über die vorhandenen Patientenantworten (bedingt auf  $\theta_{i1}$ ) binomial verteilt ist. Es bezeichne  $\delta_{ij11}$  einen Indikator dafür, ob für die  $j$ 'te Patientin bzw. den  $j$ 'ten Patienten eine Antwort für das Item vorhanden ist ( $\delta_{ij11} = 1$ ) oder nicht ( $\delta_{ij11} = 0$ ) und sei  $J_{i11}^* = \sum_{j=1}^{J_i} \delta_{ij11}$  die Anzahl an vorliegenden Patientenantworten. Damit gilt

$$\sum_{j=1}^{J_i} Y_{ij11} \delta_{ij11} | \theta_{i1} \sim \text{Binom}(J_{i11}^*, \theta_{i1})$$

Als A-priori-Verteilung für Parameter  $\theta_{i1}$  wird die zur Binomialverteilung konjugierte Beta-Verteilung gewählt, die als eine Wahrscheinlichkeitsverteilung über die möglichen Erfolgswahrscheinlichkeiten angesehen werden kann. Diese häufig vorkommende Kombination aus binomialverteilten Daten und einer entsprechenden Beta-Verteilung als A-priori-Verteilung wird auch als konjugiertes Beta-Binomialmodell bezeichnet (vgl. z. B. Carlin und Louis 2009). Die beiden Parameter der Beta-Verteilung werden mit  $a$  und  $b$  bezeichnet, d. h.

$$\theta_{i1} \sim \text{Beta}(a, b)$$

Im Spezialfall eines Qualitätsindikators mit nur einem Qualitätsmerkmal werden diese Parameter auf  $a = b = \frac{1}{2}$  gesetzt. Mit dieser Wahl entspricht die Verteilung der um Wahrscheinlichkeit  $\frac{1}{2}$  symmetrischen, nicht informativen Jeffreys' A-priori-Verteilung (Carlin und Louis 2009).

Aus der Wahl der konjugierten A-priori-Verteilung als Beta-Verteilung folgt, dass auch die A-posteriori-Verteilung, auf der basierend die Inferenz für  $\theta_{i1}$  betrieben wird, einer Beta-Verteilung entspricht. Die resultierende A-posteriori-Verteilung mit aufdatierten Parametern hat dann die Form

$$\theta_{i1} | y_{i111}, \dots, y_{ij_{i11}11} \sim \text{Beta} \left( a + \sum_{j=1}^{J_i} y_{ij11} \delta_{ij11}, b + J_{i11}^* - \sum_{j=1}^{J_i} y_{ij11} \delta_{ij11} \right)$$

Der erste Parameter der Beta-Verteilung kann in Zusammenhang mit der Binomialverteilung als Anzahl an Erfolgen (Patientenantworten mit Wert 1) und der zweite als Anzahl an Misserfolgen (Patientenantworten mit Wert 0) angesehen werden. Somit lässt sich zum Beispiel die Wahl der Parameter der A-priori-Verteilung bei  $a = b = \frac{1}{2}$  derart interpretieren, dass je ein halber Fall zu den beiden Antwortkategorien hinzugespielt wird. Abbildung 3 veranschaulicht die A-priori- und A-posteriori-Verteilung von  $\theta_{i1}$ . Die linke Grafik zeigt dabei die A-priori-Verteilung Beta ( $\frac{1}{2}, \frac{1}{2}$ ). Die mittlere und die rechte Grafik zeigen, wie sich das Aufdatieren der A-priori-Verteilung mittels der Informationen aus den Patientenantworten auswirkt. Beide Grafiken zeigen die A-posteriori-Verteilungen im Fall einer beobachteten Erfolgswahrscheinlichkeit von 0,7. Der Unterschied zwischen der mittleren und der rechten Grafik entsteht durch die Anzahl an vorliegenden Patientenantworten. In die mittlere Grafik gehen nur 10 Antworten ein, wohingegen für die rechte Grafik 100 Patientenantworten berücksichtigt werden. Dadurch erhalten im ersten Fall die Daten mehr und

die A-priori-Informationen über den Parameter entsprechend weniger Gewicht als im zweiten Fall.

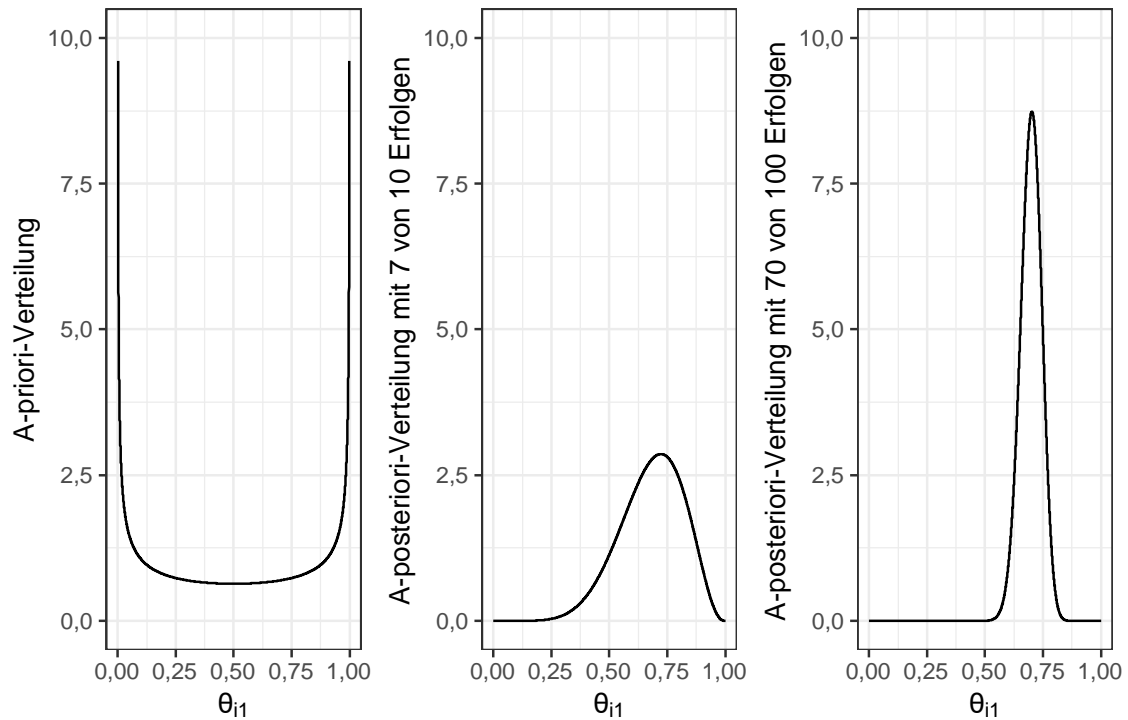


Abbildung 3: Konjugierte A-priori-Verteilung und resultierende A-posteriori-Verteilung; links: Beta-Verteilung mit Parametern  $a = b = 1/2$ ; Mitte: beispielhaft resultierende A-posteriori-Verteilung bei 7 von 10 Erfolgen: Beta(7,5, 3,5); rechts: beispielhaft resultierende A-posteriori-Verteilung bei 70 von 100 Erfolgen: Beta(70,5, 30,5)

Ist die A-posteriori-Verteilung einmal ermittelt, so lassen sich basierend auf ihr Rückschlüsse auf den Parameter  $\theta_{i1}$  und somit auf den Indikator  $Q_i = \theta_{i1}$  ziehen. Der Indikatorwert des  $i$ 'ten Leistungserbringers,  $qi_i$ , wird als A-posteriori-Erwartungswert von  $\theta_{i1}$  basierend auf den Patientantworten ermittelt, d. h.

$$qi_i := \hat{Q}_i = E(\theta_{i1} | y_{i111}, \dots, y_{ij11})$$

Der Erwartungswert einer Beta-Verteilung mit Parametern  $a$  und  $b$  ist gegeben als  $a/(a + b)$  (vgl. Carlin und Louis 2009). Somit berechnet sich der Indikatorwert für den  $i$ 'ten Leistungserbringer als

$$\begin{aligned} qi_i = \hat{Q}_i &= \left( a + \sum_{j=1}^{J_i} y_{ij11} \delta_{ij11} \right) / \left( a + \sum_{j=1}^{J_i} y_{ij11} \delta_{ij11} + b + J_{i11}^* - \sum_{j=1}^{J_i} y_{ij11} \delta_{ij11} \right) \\ &= \left( a + \sum_{j=1}^{J_i} y_{ij11} \delta_{ij11} \right) / (J_{i11}^* + a + b) \end{aligned}$$

Die Indikatorwerte für die Beispiele in Abbildung 3 liegen jeweils bei 0,682 (Mitte) und 0,698 (rechts). Damit liegt der Indikatorwert der A-posteriori-Verteilung, die auf mehr Patientenantworten beruht, näher an der beobachteten Erfolgswahrscheinlichkeit von 0,7, wohingegen der Indikatorwert der A-posteriori-Verteilung, in die wenige Patientenantworten eingehen, dem Erwartungswert der A-priori-Verteilung, der  $\frac{1}{2}$  beträgt, näher ist. Dieser Effekt erzielt eine Stabilisierung der Ergebnisse bei niedrigen Fallzahlen, d. h., wenn nur wenig Information vorliegt.

In diesem Spezialfall mit nur einem binären Item lässt sich der Schätzer auch einfach mit dem (frequentistischen) Maximum-Likelihood-Schätzer für  $\theta_{i1}$  vergleichen. Dieser ergibt sich für ein binäres Item als Anteil an Erfolgen (Antworten mit Wert 1), d. h.

$$\hat{\theta}_{i1}^{\text{ML}} = \left( \sum_{j=1}^{J_i} y_{ij11} \delta_{ij11} \right) / J_{i1}^*$$

Der Vergleich mit  $\hat{\theta}_{i1}^{\text{ML}}$  veranschaulicht erneut das Hinzufügen je eines halben Falls zu den beiden Antwortkategorien durch die gewählte A-priori-Verteilung. Es lässt sich allgemein für das konjugierte Beta-Binomialmodell zeigen, dass sich der A-posteriori-Erwartungswert als gewichtetes Mittel aus dem Maximum-Likelihood-Schätzer  $\hat{\theta}_{i1}^{\text{ML}}$  und dem A-priori-Erwartungswert ergibt (vgl. Carlin und Louis 2009), d. h.

$$E(\theta_{i1} \mid y_{i111}, \dots, y_{iJ_{i1}11}) = w_i \hat{\theta}_{i1}^{\text{ML}} + (1 - w_i) \frac{a}{a + b}$$

wobei die Gewichte gegeben sind als  $w_i = J_{i1}^* / (J_{i1}^* + a + b)$ . Daran sieht man, dass je größer die Anzahl an eingehenden Beobachtungen  $J_{i1}^*$  ist, desto weniger die A-priori-Information über  $\theta_{im}$  ins Gewicht fällt.

Das beschriebene Vorgehen ergibt sich als Spezialfall der Methodik, die im allgemeineren Fall verwendet wird. Die allgemeine Methode wird im folgenden Abschnitt vorgestellt.

### 2.1.3 Allgemeiner Fall

Allgemein kann sich ein Indikator bei der Patientenbefragung aus  $M$  Qualitätsmerkmalen zusammensetzen, die je durch potenziell mehrere zwei- oder mehrkategoriale Items operationalisiert werden. Somit ergibt sich die Verallgemeinerung vom einfachen Fall zum allgemeinen Fall aus drei wesentlichen Erweiterungen:

1. Qualitätsmerkmale können durch mehr als ein Item operationalisiert werden.
2. Items können mehr als zwei Antwortkategorien aufweisen.
3. Indikatoren können sich aus mehr als einem Qualitätsmerkmal zusammensetzen.

Die erste Erweiterung erfordert für den allgemeinen Fall keine größere methodische Erweiterung, unter der Bedingung, dass alle Items eines Qualitätsmerkmals die gleiche Skala aufweisen und für die Antwortmöglichkeiten die gleichen Punkte vergeben werden. Sei dazu  $\theta_{im}$  der Kompetenzparameter zum  $m$ 'ten Qualitätsmerkmal. Der einzige Unterschied ist nun, dass für die Inferenz über

$\theta_{im}$  anstelle der Patientenantworten nur eines Items, das das  $m$ 'te Qualitätsmerkmal operationalisiert, die Patientenantworten aller Items verwendet werden.

Die zweite Erweiterung wirkt sich insbesondere auf die geeignete Annahme einer Verteilung für die Daten und eine A-priori-Verteilung für die zugrunde liegenden Parameter aus. Im allgemeinen Fall können die Antworten daher nicht mehr nur die Werte 0 und 1 annehmen. Die Antworten auf ein Item mit  $K_{ml}$  Antwortkategorien können die folgenden Werte annehmen:

$$y_{ijml} \in \left\{ 0, \frac{1}{K_{ml}-1}, \dots, \frac{K_{ml}-2}{K_{ml}-1}, 1 \right\}$$

Für  $K_{ml} = 3$  Antwortkategorien können somit die Werte 0,  $\frac{1}{2}$ , 1 und für  $K_{ml} = 4$  Kategorien die Werte 0,  $\frac{1}{3}$ ,  $\frac{2}{3}$ , 1 erreicht werden. Basierend auf diesen Werten und einem speziellen Multinomialmodell lässt sich der allgemeine Fall wieder auf ein konjugiertes Beta-Binomialmodell zurückführen, das nachfolgend näher beschrieben wird.

Betrachtet wird im Folgenden ein Indikator für den  $i$ 'ten Leistungserbringer, in den  $M$  Qualitätsmerkmale einfließen, die je durch  $L_m$  Items mit  $K_m := K_{ml}$  Antwortkategorien operationalisiert werden. Es wird somit vorausgesetzt, dass alle Items eines Qualitätsmerkmals gleich viele Antwortkategorien aufweisen. Weiterhin wird zur einfachen Darstellung angenommen, dass die Kategorien mit Kategorie 1 beginnen und derart geordnet sind, dass die höchste Kategorie 100 Punkten bzw. dem Wert 1 entspricht. Im Gegensatz zum Spezialfall nur eines Items gehen die Antworten einer Patientin bzw. eines Patienten auch dann in die Berechnung des Indikatorwerts ein, wenn für diese Patientin bzw. diesen Patienten Antworten auf einzelne Items des Indikators fehlen, da alle vorliegenden und für den Indikator informativen Patientenantworten für die Schätzung des zugrunde liegenden Indikators verwendet werden. Eine solche Available-case-Analyse hat den Vorteil gegenüber Complete-case-Analysen, dass alle vorliegenden Antworten genutzt werden. Grundvoraussetzung für unverzerrte Ergebnisse unter Verwendung von Available-case-Ansätzen in Zusammenhang mit der verwendeten Bayesianischen Methodik ist, dass es sich bei dem Mechanismus, der zu den fehlenden Werten führt, nicht um *missing not at random* (MNAR) handelt (vgl. z. B. Molenberghs et al. 2008). Diese Annahme wird im Folgenden getroffen.

Aus der dritten Erweiterung auf den allgemeinen Fall folgt zudem, dass der Indikatorwert für den  $i$ 'ten Leistungserbringer als Kombination der A-posteriori-Verteilungen der qualitätsmerkmals-spezifischen Parameter  $\theta_{im}$ ,  $m = 1, \dots, M$ , bestimmt wird. Entsprechend der in Abschnitt 1.3.3 eingeführten Definition bestimmt sich der Indikatorwert als Mittelwert der Verteilungen der merkmals-spezifischen Parameter, d. h.

$$QI_i = \frac{1}{M} \sum_{m=1}^M \theta_{im}$$

Insbesondere stellt der Indikatorwert  $QI_i$  im Bayesianischen Kontext damit selbst eine Zufallsverteilung dar.



Im Folgenden wird im Detail erläutert, welche Verteilung für die Daten verwendet wird, wie die entsprechend konjugierte A-priori-Verteilung aussieht und wie die A-posteriori-Verteilungen der Parameter kombiniert werden.

Die Verallgemeinerung der Bernoulli-Verteilung auf mehr als zwei Antwortkategorien stellt die Multinomialverteilung mit nur einer Antwort dar (vgl. z. B. Agresti (2013) für eine nähere Beschreibung der Multinomialverteilung)<sup>8</sup>. Im allgemeinen Fall wird daher angenommen, dass die Wahrscheinlichkeitsverteilung  $F$  aus Abschnitt 1.3.3 einer Multinomialverteilung entspricht, d. h. die Antwort der  $j$ 'ten Patientin bzw. des  $j$ 'ten Patienten auf das  $l$ 'te Item des  $m$ 'ten Qualitätsmerkmals aus einer Multinomialverteilung stammt. Im Folgenden werden die Wahrscheinlichkeiten für das Zutreffen der  $K_m$  Antwortkategorien mit  $\boldsymbol{\pi}_{im} = (\pi_{im}^1, \dots, \pi_{im}^{K_m})'$  bezeichnet, wobei sich diese zu 1 aufsummieren, d. h.  $\sum_{k=1}^{K_m} \pi_{im}^k = 1$ . Aufgrund der Annahme, dass die Patientenantworten von der zugrunde liegenden Kompetenz des Leistungserbringers  $\theta_{im}$  abhängen (Abschnitt 1.3.2), lassen sich die Wahrscheinlichkeiten  $\boldsymbol{\pi}_{im}$  als Funktion von  $\theta_{im}$  darstellen, d. h.

$$\boldsymbol{\pi}_{im}(\theta_{im}) = (\pi_{im}^1(\theta_{im}), \dots, \pi_{im}^{K_m}(\theta_{im}))'$$

Diese Annahme einer den Wahrscheinlichkeiten für die Antwortkategorien zugrunde liegenden, latenten Größe ist eng verwandt mit dem Vorgehen in Modellen der Item-Response-Theory (IRT-Modellen; vgl. Bühner 2011). Somit gilt für die Antwort der  $j$ 'ten Patientin bzw. des  $j$ 'ten Patienten auf das  $l$ 'te Item des  $m$ 'ten Qualitätsmerkmals

$$Y_{ijml} \mid \theta_{im} \stackrel{\text{i.i.d.}}{\sim} \text{Mult}(\mathbf{1}, \boldsymbol{\pi}_{im}(\theta_{im}))$$

Die Antworten der unterschiedlichen Patientinnen und Patienten auf das  $l$ 'te Item von Qualitätsmerkmal  $m$  bedingt auf die Kompetenz des Leistungserbringers  $\theta_{im}$  werden als unabhängig und identisch verteilt angenommen (Abschnitte 1.3.2 und 1.3.3). Daraus ergibt sich für die Summe über die vorhandenen Patientenantworten (bedingt auf  $\theta_{im}$ ) wieder eine Multinomialverteilung. Es bezeichne  $\delta_{ijml}$  einen Indikator dafür, ob für die  $j$ 'te Patientin bzw. den  $j$ 'ten Patienten eine Antwort auf das  $l$ 'te Item von Qualitätsmerkmal  $m$  vorliegt ( $\delta_{ijml} = 1$ ) oder nicht ( $\delta_{ijml} = 0$ ). Des Weiteren bezeichne  $J_{iml}^* = \sum_{j=1}^{J_i} \delta_{ijml}$  die Anzahl an vorliegenden Antworten für das  $l$ 'te Item des  $m$ 'ten Qualitätsmerkmals. Für die Summe über die Patientenantworten gilt somit aufgrund der Unabhängigkeitsannahme der Patientenantworten auf Items unterschiedlicher Qualitätsmerkmale

$$\sum_{j=1}^{J_i} Y_{ijml} \delta_{ijml} \mid \theta_{im} \stackrel{\text{ind.}}{\sim} \text{Mult}(J_{iml}^*, \boldsymbol{\pi}_{im}(\theta_{im}))$$

Gemäß den Annahmen in Abschnitt 1.3.2 liegt den Patientenantworten aller  $L_m$  Items des  $m$ 'ten Qualitätsmerkmals der gemeinsame, unbekannte Parameter  $\theta_{im}$  zugrunde. Somit können die Antworten auf alle Items des  $m$ 'ten Qualitätsmerkmals für die Inferenz über  $\theta_{im}$  herangezogen werden. Es bezeichne hierfür  $J_{im}^* = \sum_{l=1}^{L_m} J_{iml}^*$  die Gesamtanzahl an vorliegenden Antworten für

<sup>8</sup> Da die Multinomialverteilung eigentlich eine multivariate Verteilung ist, wird zur Abgrenzung dieses Spezialfalls mit nur einer Antwort auch der Begriff „kategoriale Verteilung“ verwendet.

das  $m$ 'te Qualitätsmerkmal über alle Items und Patientinnen und Patienten hinweg. Damit ergibt sich die folgende Verteilungsannahme für die Summe über alle Patientenantworten auf Items des  $m$ 'ten Qualitätsmerkmals

$$\sum_{j=1}^{J_i} \sum_{l=1}^{L_m} Y_{ijml} \delta_{ijml} \mid \theta_{im} \stackrel{\text{ind.}}{\sim} \text{Mult}(J_{im}^*, \boldsymbol{\pi}_{im}(\theta_{im}))$$

Für die Inferenz über den Kompetenzparameter  $\theta_{im}$  basierend auf den vorliegenden Patientenantworten wird eine Annahme darüber benötigt, wie die Kategoriewahrscheinlichkeiten  $\boldsymbol{\pi}_{im}$  von dem zugrunde liegenden  $\theta_{im}$  abhängen. Da es sich um ein ordinalskaliertes Outcome handelt, ist davon auszugehen, dass für steigende Werte von  $\theta_{im}$  die Wahrscheinlichkeit für das Zutreffen der Kategorien mit höheren Punktwerten steigt. Entsprechend der gleichabständigen Punktevergabe für die Antwortkategorien (Abschnitt 1.2.2) wird angenommen, dass sich die Wahrscheinlichkeit für das Zutreffen der  $k$ 'ten Antwortkategorie wie folgt aus  $\theta_{im}$  ergibt:

$$\pi_{im}^k(\theta_{im}) = \binom{K_m-1}{k-1} \theta_{im}^{k-1} (1 - \theta_{im})^{K_m-1-(k-1)}, \text{ für } k = 1, \dots, K_m$$

Die Summe über die Kategoriewahrscheinlichkeiten ergibt sich entsprechend mithilfe des binomischen Lehrsatzes als

$$\sum_{k=1}^{K_m} \pi_{im}^k(\theta_{im}) = \sum_{k=0}^{K_m-1} \binom{K_m-1}{k} \theta_{im}^k (1 - \theta_{im})^{K_m-1-k} = (\theta_{im} + (1 - \theta_{im}))^{K_m-1} = 1$$

Somit wird das Multinomialmodell mit den  $K_m$  unbekanntem Parametern auf ein Modell mit einem einzigen Parameter,  $\theta_{im}$ , reduziert, welches auch die Ordinalskalierung der Antworten berücksichtigt. Dies ermöglicht, dass die Inferenz über  $\theta_{im}$  auf die Inferenz des Parameters einer Binomialverteilung zurückgeführt werden kann, wie im Folgenden gezeigt wird.

Der angenommene Zusammenhang zum latenten Parameter  $\theta_{im}$  lässt sich auf verschiedene Weisen veranschaulichen. Grundsätzlich ergibt sich jede Kategoriewahrscheinlichkeit als Wahrscheinlichkeit einer bestimmten Kombination aus Erfolgen und Misserfolgen bezüglich der latenten Erfolgswahrscheinlichkeit  $\theta_{im}$ . Aufgrund der Anordnung der Kategorien, bei der die höchste Kategorie dem Wert von 1 entspricht, soll bei hohen Kategorien die Erfolgswahrscheinlichkeit und bei niedrigen Kategorien entsprechend die Gegenwahrscheinlichkeit dominieren. Betrachtet man die  $K_m$  möglichen Kombinationen aus Erfolgen und Misserfolgen, wenn die Reihenfolge der Kombination keine Rolle spielt, so ergeben sich beispielsweise im Fall von  $K_m = 3$  Kategorien die folgenden drei Kombinationsmöglichkeiten

- {Misserfolg, Misserfolg} mit Wahrscheinlichkeit  $(1 - \theta_{im})(1 - \theta_{im}) = (1 - \theta_{im})^2$
- {Erfolg, Misserfolg} mit Wahrscheinlichkeit  $\theta_{im}(1 - \theta_{im})$
- {Erfolg, Erfolg} mit Wahrscheinlichkeit  $\theta_{im}\theta_{im} = \theta_{im}^2$

Diese Kombinationen werden dann entsprechend mit der Anzahl an Möglichkeiten, zu der Kombination zu gelangen, gewichtet, woraus sich für  $K_m = 3$  die folgenden Kategoriewahrscheinlichkeiten ergeben

$$\boldsymbol{\pi}_{im}(\theta_{im}) = ((1 - \theta_{im})^2, 2\theta_{im}(1 - \theta_{im}), \theta_{im}^2)'$$

Die resultierenden Kategoriewahrscheinlichkeiten sind in der Populationsgenetik auch als Hardy-Weinberg-Anteile (HW-Anteile) für einen polyploiden Chromosomensatz mit zwei Allelen bekannt, die zur Berechnung von Allelfrequenzen in idealen Populationen verwendet werden (vgl. Yang und Nakaya 2006 für mehr Details zum Hintergrund der HW-Anteile). Aus diesem Grund wird die Multinomialverteilung mit den HW-Anteilen als Kategoriewahrscheinlichkeiten im Folgenden als HW-Multinomialmodell bezeichnet.

Der Fall eines binären Items ( $K_m = 2$ ), der in Abschnitt 2.1.2 behandelt wurde, ergibt sich als direkter Spezialfall des HW-Multinomialmodells. Die Kategoriewahrscheinlichkeiten reduzieren sich in diesem Fall auf die latente Erfolgswahrscheinlichkeit bzw. die Gegenwahrscheinlichkeit der Binomialverteilung für eine Antwort mit dem Wert 1 bzw. mit dem Wert 0

$$\boldsymbol{\pi}_{im}(\theta_{im}) = ((1 - \theta_{im}), \theta_{im})'$$

Die Kategoriewahrscheinlichkeiten für  $K_m \in \{2, 3, 4, 5\}$  in Abhängigkeit des zugrunde liegenden Parameters  $\theta_{im}$  sind in Abbildung 4 dargestellt. Man sieht, dass passend zur Ordinalskala des Outcomes mit steigendem latenten Parameter die Kategorienwahrscheinlichkeit für höhere Kategorien zunimmt. Des Weiteren lässt sich ablesen, dass sich für jeden Wert von  $\theta_{im}$  die Wahrscheinlichkeiten der verschiedenen Antwortkategorien zu 1 aufsummieren.

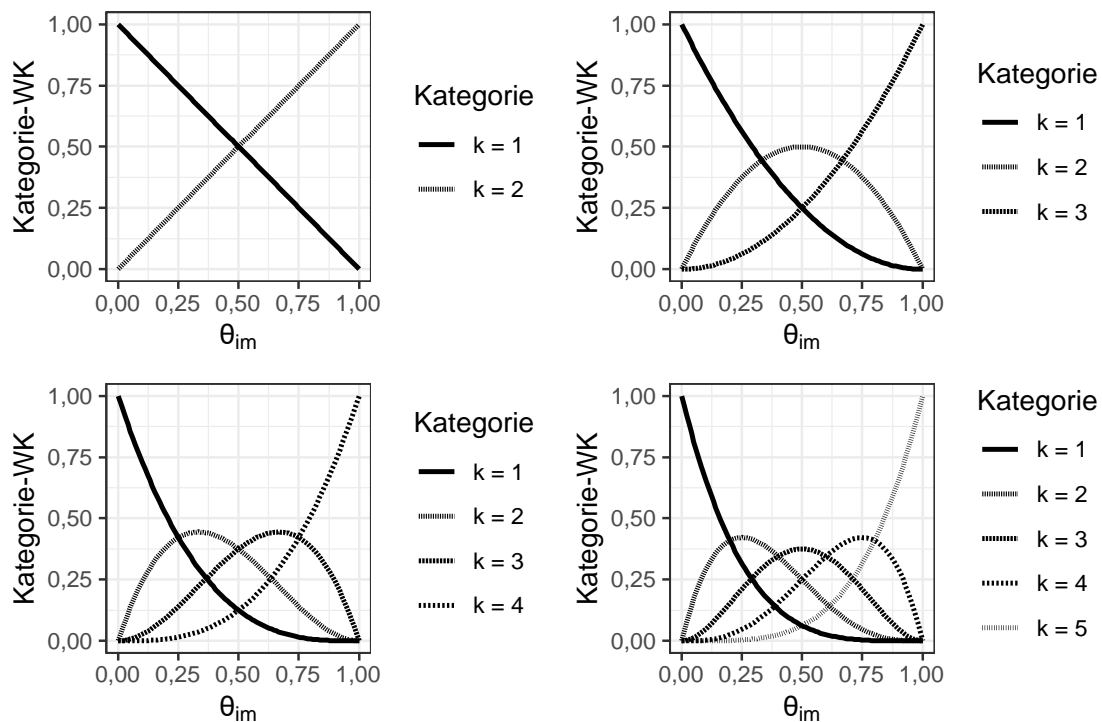


Abbildung 4: Kategoriewahrscheinlichkeiten (Kategorie-WK) im HW-Multinomialmodell in Abhängigkeit des zugrunde liegenden Parameters  $\theta_{im}$ ; oben links: für zwei Antwortkategorien; oben rechts: für drei Antwortkategorien; unten links: für vier Antwortkategorien; unten rechts: für fünf Antwortkategorien

Es bezeichne  $I(y_{ijml} = k)$  die Indikatorfunktion für die Auswahl der  $k$ 'ten Kategorie durch Patientin bzw. Patient  $j$  von Leistungserbringer  $i$  für das  $l$ 'te Item des  $m$ 'ten Qualitätsmerkmals. Diese Funktion nimmt den Wert 1 an, wenn die  $k$ 'te Kategorie ausgewählt wird, und sonst den Wert 0. Bezeichne weiterhin mit  $y_{i \cdot m}^k$  die Anzahl, mit der über alle Patientinnen und Patienten und über alle Items des  $m$ 'ten Qualitätsmerkmals hinweg die  $k$ 'te Kategorie ausgewählt wurde, d. h.

$$y_{i \cdot m}^k := \sum_{j=1}^{J_i} \sum_{l=1}^{L_m} I(y_{ijml} = k)$$

Damit lässt sich zeigen, dass das HW-Multinomialmodell pro Qualitätsmerkmal sich auf ein Binomialmodell mit  $L_m J_{im}^* (K_m - 1)$  Versuchen, von denen  $\sum_{k=1}^{K_m} (k - 1) y_{i \cdot m}^k$  Erfolge sind, zurückgeführt werden kann, bei dem die Erfolgswahrscheinlichkeit  $\theta_{im}$  ist (vgl. IQTIG 2018b: Anhang, Kapitel 12). Die Anzahl, mit der die  $k$ 'te Kategorie ausgewählt wurde,  $y_{i \cdot m}^k$ , wird somit im – dem HW-Multinomialmodell entsprechenden – Binomialmodell mit Gewicht  $(k - 1)$  als Erfolg gezählt. Passend zur Anordnung der Kategorien zählen Antworten in niedrigen Kategorien folglich weniger als Erfolg als Antworten in hohen Kategorien. Im Beispiel für  $K_m = 3$  Kategorien erhält man als Gewichte 0, 1, 2 und bei  $K_m = 4$  Kategorien entsprechend 0, 1, 2, 3. Dividiert man diese Gewichte durch  $(K_m - 1)$ , d. h., bezieht man sie auf  $J_{im}^*$  Versuche anstelle der  $J_{im}^* (K_m - 1)$  Versuche in der Binomialverteilung, so erhält man für  $K_m = 3$  Kategorien Gewichte von 0,  $\frac{1}{2}$ , 1 und für  $K_m = 4$  Kategorien entsprechend 0,  $\frac{1}{3}$ ,  $\frac{2}{3}$ , 1. Diese Gewichte entsprechen den – aus den Punkten abgeleiteten – Werten für die Antwortkategorien (Abschnitt 1.2.2). Die Anzahl an Erfolgen  $\sum_{k=1}^{K_m} (k - 1) y_{i \cdot m}^k$  geht, in Analogie zu Abschnitt 2.1.2, in die Parameter der A-posteriori-Verteilung des konjugierten Beta-Binomialmodells ein, das im Folgenden auf den allgemeinen Fall erweitert wird.

Kurz zusammengefasst: Für die Summe der Patientenantworten auf das  $m$ 'te Qualitätsmerkmal wird die folgende Verteilungsannahme getroffen

$$\sum_{j=1}^{J_i} \sum_{l=1}^{L_m} Y_{ijml} \delta_{ijml} \mid \theta_{im} \stackrel{\text{ind.}}{\sim} \text{Mult}(J_{im}^*, \boldsymbol{\pi}_{im}(\theta_{im})), \text{ mit}$$

$$\boldsymbol{\pi}_{im}(\theta_{im}) = \left( \binom{K_m - 1}{0} \theta_{im}^0 (1 - \theta_{im})^{K_m - 1}, \dots, \binom{K_m - 1}{K_m - 1} \theta_{im}^{K_m - 1} (1 - \theta_{im})^0 \right)'$$

Als A-priori-Verteilung für die einzelnen Parameter  $\theta_{im}$  wird jeweils die zur Binomialverteilung und damit auch zum HW-Multinomialmodell konjugierte Beta-Verteilung gewählt. Die A-priori-Parameter der Beta-Verteilungen  $a$  und  $b$  werden dabei jeweils für alle Parameter  $\theta_{im}$  identisch gewählt und es wird angenommen, dass die Parameter  $\theta_{im}$  unabhängig voneinander sind (Abschnitt 1.3.2). Somit sind die A-priori-Verteilungen für die  $M$  Parameter gegeben durch

$$\theta_{im} \stackrel{\text{i.i.d.}}{\sim} \text{Beta}(a, b), m = 1, \dots, M$$

Im allgemeinen Fall von Qualitätsindikatoren mit mehreren Qualitätsmerkmalen werden die Parameter  $a$  und  $b$  abhängig von der Anzahl der Qualitätsmerkmale  $M$  gesetzt, und zwar  $a = b = 0.5/M$ . Für den oben beschriebenen Fall eines Qualitätsindikators mit nur einem Merkmal ent-

spricht diese Setzung nach wie vor der Jeffreys' Prior. Mit steigender Merkmalsanzahl  $M$  allerdings verliert die A-priori-Verteilung auf Merkmalsebene durch diese Parametersetzung zunehmend an Gewicht. Dies soll dem Effekt entgegenwirken, dass auf Ebene der Qualitätsindikatoren die A-priori-Verteilung des Indikators  $QI_i$ , die aus dem Durchschnitt der  $M$  merkmalspezifischen A-priori-Verteilungen resultiert, mit zunehmender Merkmalsanzahl an Gewicht zunimmt.<sup>9</sup> Dies würde bei Indikatoren mit vielen Merkmalen zu einem besonders starken Shrinkage-Effekt hin zum Mittelwert 0,5 führen. Diese merkmalsabhängige Setzung der A-priori-Parameter geht insbesondere zurück auf Analysen zur theoretischen Sensitivität und Spezifität der statistischen Auffälligkeitsbestimmung, auch in Anbetracht verschiedener Indikator- und Fallzahlsszenarien. Für die Auswertung der Qualitätsindikatoren der Patientenbefragung wurde die hier dargestellte Setzung der A-priori-Parameter verwendet.

Die Unabhängigkeit der Parameter  $\theta_{im}$  ergibt sich aus der Annahme, dass die Patientenantworten auf Items unterschiedlicher Qualitätsmerkmale voneinander unabhängig sind (Abschnitt 1.3.2). Aufgrund der getroffenen Unabhängigkeitsannahmen sowie der Konjugiertheit des HW-Multinomialmodells und der Beta-Verteilung als A-priori-Verteilung entsprechen die A-posteriori-Verteilungen der zugrunde liegenden Parameter  $\theta_{im}$  gegeben den Patientenantworten wieder bedingt unabhängigen Beta-Verteilungen mit entsprechend aufdatierten Parametern. Bezeichne  $\mathbf{y}_{ijm} = (y_{ijm1}, \dots, y_{ijmL_m})'$  den Vektor an realisierten Antworten der  $j$ 'ten Patientin bzw. des  $j$ 'ten Patienten von Leistungserbringer  $i$  auf die Items des  $m$ 'ten Qualitätsmerkmals. Die A-posteriori-Verteilung für den  $m$ 'ten Parameter hat damit die Form

$$\theta_{im} \mid \mathbf{y}_{i1m}, \dots, \mathbf{y}_{ijm} \stackrel{\text{ind.}}{\sim} \text{Beta} \left( a + \sum_{k=1}^{K_m} (k-1) y_{i-m}^k, b + J_{im}^* - \sum_{k=1}^{K_m} (k-1) y_{i-m}^k \right)$$

wobei der erste Parameter  $a$  der A-priori-Verteilung wie im einfachen Fall um die Anzahl an Erfolgen bezüglich des latenten Parameters  $\theta_{im}$  erweitert wird und der zweite Parameter  $b$  entsprechend um die Anzahl an Misserfolgen aufdatiert wird (vgl. z. B. Held und Bové 2014).

Sind die A-posteriori-Verteilungen, d. h. die aufdatierten Parameter der Beta-Verteilungen, für alle  $M$  Parameter ermittelt, lässt sich daraus der Indikatorwert  $qi_i$  als Erwartungswert der A-posteriori-Verteilung des Indikators  $QI_i$  bestimmen, welche sich als Verteilung des Mittels der unabhängigen A-posteriori-Verteilungen der Parameter  $\theta_{im}$  ergibt (vgl. Indikatordefinition in Abschnitt 1.3.3)

$$\begin{aligned} qi_i &= \widehat{QI}_i = E(QI_i \mid \mathbf{y}_{i11}, \dots, \mathbf{y}_{ij1}, \dots, \mathbf{y}_{i1M}, \dots, \mathbf{y}_{ijM}) \\ &= E \left( \frac{1}{M} \sum_{m=1}^M \theta_{im} \mid \mathbf{y}_{i11}, \dots, \mathbf{y}_{ij1}, \dots, \mathbf{y}_{i1M}, \dots, \mathbf{y}_{ijM} \right) \end{aligned}$$

<sup>9</sup> Konkret nimmt auf Ebene der Qualitätsindikatoren die A-priori-Varianz des Indikators  $QI_i$  um den Mittelwert mit zunehmender Merkmalsanzahl ab, falls die Parameter der merkmalspezifischen Beta-a-priori-Verteilungen konstant bei  $a = b = 0,5$  gesetzt wären.

Aufgrund der Linearität des Erwartungswerts kann der Indikatorwert auch als Mittel der A-posteriori-Erwartungswerte berechnet werden, d. h.

$$q_i = \frac{1}{M} \sum_{m=1}^M E(\theta_{im} | \mathbf{y}_{i1m}, \dots, \mathbf{y}_{ij_i m}) = \frac{1}{M} \sum_{m=1}^M \frac{\sum_{k=1}^{K_m} (k-1) \sum_{j=1}^{J_i} \sum_{l=1}^{L_m} I(y_{ijml} = k) + a}{J_{im}^* + a + b}$$

Im Gegensatz zum einfachen Fall ist die Herleitung des Maximum-Likelihood-Schätzer im allgemeinen Fall komplexer, da hierfür die Verteilung des Mittels der HW-Multinomialverteilungen benötigt wird. Unter der getroffenen Unabhängigkeitsannahme der Antworten von Patientinnen und Patienten auf Items unterschiedlicher Qualitätsmerkmale ließe sich der Maximum-Likelihood-Schätzer zumindest numerisch bestimmen. Im Gegensatz zum Bayesianischen Ansatz ist eine Erweiterung auf abhängige Qualitätsmerkmale jedoch schwierig.

## 2.2 Quantifizierung von Unsicherheit bei der Bestimmung der Indikatorwerte

Zusätzlich zum Indikatorwert eines Leistungserbringers wird ein (zweiseitiges) Bayesianisches Kreditabilitätsintervall bestimmt, das die Unsicherheit bezüglich der aus den Daten gewonnenen Informationen über den zugrunde liegenden Indikatorwert quantifiziert. Bayesianische Kreditabilitätsintervalle geben einen Bereich von Werten für den Indikator an, die nach der Kombination der A-priori-Information und den aus den Patientenantworten gewonnenen Informationen am plausibelsten sind. Genauer wird ein  $(1 - \alpha) \cdot 100\%$ -Kreditabilitätsintervall  $[l, u]$  für den Indikator derart bestimmt, dass die Fläche unter der A-posteriori-Verteilung des Indikators für diese Werte  $1 - \alpha$  entspricht, d. h.

$$\int_l^u f(q_i | \mathbf{y}_{i11}, \dots, \mathbf{y}_{ij_i 1}, \dots, \mathbf{y}_{i1M}, \dots, \mathbf{y}_{ij_i M}) (q_i) \, dq_i = 1 - \alpha$$

Die dafür benötigte Dichte der A-posteriori-Verteilung des Indikators

$$f(q_i | \mathbf{y}_{i11}, \dots, \mathbf{y}_{ij_i 1}, \dots, \mathbf{y}_{i1M}, \dots, \mathbf{y}_{ij_i M})$$

ergibt sich entsprechend der Indikatordefinition als Verteilung des Mittels der unabhängigen A-posteriori-Verteilungen der zugrunde liegenden Parameter  $\theta_{im}$ , welche jeweils Beta-Verteilungen entsprechen (Abschnitt 2.1.3). Da die Verteilung des Mittels von unabhängigen Beta-Verteilungen nicht in analytischer Form darstellbar ist und keiner bekannten Verteilung entspricht, wird die A-posteriori-Verteilung des Indikators mittels Monte-Carlo-Simulation numerisch bestimmt (vgl. Robert und Casella 2004). Dafür wird aus den  $M$  A-posteriori-Verteilungen der Parameter  $\theta_{im}$  eine große Anzahl an Realisierungen aus den jeweiligen Beta-Verteilungen generiert, die dann jeweils entsprechend dem Indikator gemittelt werden. Somit erhält man eine Approximation der gewünschten A-posteriori-Verteilung des Indikators. Im einfachen Spezialfall eines binären Items wird keine Monte-Carlo-Simulation benötigt, da die A-posteriori-Verteilung des Indikators einer Beta-Verteilung entspricht (Abschnitt 2.1.2).

Für die Patientenbefragung werden sogenannte Equal-tailed-Kredibilitätsintervalle verwendet (vgl. z. B. Carlin und Louis 2009). Diese werden basierend auf den Quantilen der A-posteriori-Verteilung des Indikators bestimmt. Die untere und die obere Intervallgrenze für ein  $(1 - \alpha) \cdot 100\%$ -Kredibilitätsintervall sind durch das  $\alpha/2$ - bzw. das  $(1 - \alpha/2)$ -Quantil gegeben. Dabei ist zu beachten, dass die Intervalle nicht zwingend symmetrisch um den Indikatorwert sind. Es lässt sich zeigen, dass die verwendeten Equal-tailed-Kredibilitätsintervalle im konjugierten Beta-Binomialmodell mit einer Jeffreys' A-priori-Verteilung eng verwandt sind mit Mid-P-Clopper-Pearson-Konfidenzintervallen (vgl. Brown et al. 2001). Letztere kommen in der Methodik für planungsrelevante Qualitätsindikatoren zum Einsatz (vgl. IQTIG 2016). Für die Patientenbefragung werden 95%-Kredibilitätsintervalle verwendet.

Die Unsicherheitsintervalle dienen zum einen der Information über die statistische Unsicherheit bezüglich des Indikatorwerts und werden zum anderen für die quantitative Auffälligkeitseinstufung herangezogen, wie im folgenden Abschnitt beschrieben wird.

### 2.3 Quantitative Auffälligkeitseinstufung

Ziel der quantitativen Auffälligkeitseinstufung ist es, Leistungserbringer basierend auf ihren rechnerischen Ergebnissen für die weitergehende Überprüfung auszuwählen. Dies erfolgt im Vergleich mit dem Referenzbereich des Indikators. Referenzbereiche können gemäß den „Methodischen Grundlagen“ des IQTIG fest oder verteilungsabhängig sein (IQTIG 2022: Abschnitt 16.2). Die festen Referenzbereiche bei der Patientenbefragung liegen in Form von Punkten zwischen 0 und 100 vor. Die verteilungsabhängigen Referenzbereiche werden aus der Verteilung der mit 100 multiplizierten Ergebnisse der Leistungserbringer bestimmt.

Um Fehlentscheidungen aufgrund zufälliger Einflüsse und Konstellationen zu vermeiden (Abschnitt 1.3), wird anstelle des Indikatorwerts eines Leistungserbringers das dazugehörige Bayesianische Unsicherheitsintervall, das im vorherigen Abschnitt beschrieben wurde, für den Vergleich mit dem Referenzbereich und somit für die quantitative Einstufung verwendet. Hierfür werden zunächst die Grenzen des Unsicherheitsintervalls mit 100 multipliziert, damit sie auf der Punkteskala liegen, auf der auch der Referenzbereich angegeben ist.

Für einen Indikator mit einem festen Referenzbereich, z. B. dem Intervall [95, 100] Punkte, wird die obere Grenze des 95%-Unsicherheitsintervalls mit dem Referenzwert  $t$  (im Beispiel  $t = 95$  Punkte) verglichen. Liegt die obere Grenze des Unsicherheitsintervalls unter 95 Punkten, so wird das Leistungserbringerergebnis für diesen Indikator als quantitativ auffällig eingestuft. Ansonsten wird es als quantitativ unauffällig bewertet. Die Einstufung über die Bayesianischen Unsicherheitsintervalle entspricht der Verwendung eines Bayesianischen Tests mit der Nullhypothese, dass der zugrunde liegende Indikatorwert des  $i$ 'ten Leistungserbringers im Referenzbereich liegt (Carlin und Louis 2009), d. h.

$$H_0: Q|_i \geq t \text{ vs. } H_1: Q|_i < t$$

Unter Verwendung des Bayesianischen Tests wird das Ergebnis von Leistungserbringer  $i$  genau dann als quantitativ auffällig eingestuft, wenn die Nullhypothese bezüglich der A-posteriori-Verteilung eine zu niedrige Wahrscheinlichkeit hat. Genauer wird das Ergebnis von Leistungserbringer  $i$  als quantitativ auffällig eingestuft, wenn die A-posteriori-Wahrscheinlichkeit für die Nullhypothese kleiner ist als die vorgegebene Irrtumswahrscheinlichkeit  $\alpha/2$ , d. h.

$$P(H_0 \mid \mathbf{y}_{i11}, \dots, \mathbf{y}_{ij1}, \dots, \mathbf{y}_{i1M}, \dots, \mathbf{y}_{ijM}) \leq \alpha/2$$

Die Irrtumswahrscheinlichkeit von  $\alpha/2$  im Vergleich zum  $(1 - \alpha) \cdot 100\%$ -Kreditintervall ergibt sich dadurch, dass es sich bei diesem Test um einen einseitigen Test handelt, das Kreditintervall jedoch zweiseitig ist. Dieses Vorgehen ist sehr eng verwandt mit der Bestimmung statistischer Auffälligkeit für planungsrelevante Indikatoren (vgl. IQTIG 2016).

Für Indikatoren mit verteilungsbasierten Referenzbereichen wird bei der Einstufung genauso vorgegangen. Allerdings wird der Referenzwert  $t$ , der die oben genannte Nullhypothese  $H_0$  definiert, zunächst anhand der Menge der Leistungserbringerergebnisse bestimmt und ist nicht a priori vorgegeben.

## 2.4 R-Implementation der statistischen Auswertungsmethodik

Zur Erhöhung der Transparenz der methodischen Arbeit und um die Anwendbarkeit der Auswertungsmethodik zu gewährleisten, wurde ein R-Paket ‚iqtigprm‘ (prm = *patient related measures*) entwickelt, das eine Implementation der in diesem Abschnitt dargestellten statistischen Methodik enthält, und unter dem github-Account des IQTIG veröffentlicht:

<https://github.com/iqtigorg/iqtigprm>

Das Paket bezieht sich alleinig auf die implementationsmäßige Umsetzung der in den Abschnitten 2.1 bis 2.3 angegebenen mathematischen Formel zur Indikatorberechnung und quantitativen Auffälligkeitseinstufung. Ebenso enthält das Paket simulierte Daten aus der Patientenbefragung des QS-Verfahrens *QS PCI*, um die methodische Vorgehensweise beispielhaft zu illustrieren.

## 2.5 Diskussion

Die vorgeschlagene Bayesianische Auswertungsmethodik liefert Indikatorwerte, zugehörige Unsicherheitsintervalle sowie eine quantitative Einstufung, deren Bestimmung nachvollziehbar ist und auf einem methodisch etablierten Vorgehen beruht, und ist in das entsprechende Rahmenkonzept für die statistische Auswertungsmethodik des IQTIG eingebettet (IQTIG 2022: Abschnitt 20.1). Die getroffenen Annahmen über den datengenerierenden Prozess wurden explizit dargestellt. Die Methodik ermöglicht die Ergebnisberechnung für sowohl kleine als auch große Anzahlen an vorliegenden Patientenantworten. Für die quantitative Auffälligkeitseinstufung wird statistische Unsicherheit berücksichtigt, was die Einstufung unter den getroffenen Annahmen belastbar macht und zulässt, Konsequenzen abzuleiten. Durch die Verwendung eines hierarchischen Bayesianischen Ansatzes wird der hierarchischen Struktur der Indikatoren, Qualitätsmerkmale



und Items Rechnung getragen. Somit erfüllt die vorgeschlagene Auswertungsmethodik den Großteil der in Abschnitt 1.3 formulierten Anforderungen.

Wie in Abschnitt 1.3 erläutert, wird bei der Modellierung angenommen, dass die Patientenantworten auf Items unterschiedlicher Qualitätsmerkmale und dementsprechend die Parameter  $\theta_{im}$ ,  $m = 1, \dots, M$ , voneinander unabhängig sind. Darüber hinaus wird Unabhängigkeit der Antworten einer Patientin oder eines Patienten auf Items eines Qualitätsmerkmals, bedingt auf den zugrunde liegenden Parameter  $\theta_{im}$ , angenommen. Diese Unabhängigkeitsannahmen sind zunächst notwendig, um eine nachvollziehbare Auswertungsmethodik bereitstellen zu können, die auch bei großen Datenmengen funktioniert. Die Unabhängigkeitsannahmen sind gerechtfertigt, falls es keine patientenseitigen Einflüsse wie beispielsweise positives oder negatives Antwortverhalten gibt und die Kompetenz der Leistungserbringer bezüglich der verschiedenen Qualitätsmerkmale voneinander unabhängig ist. Werden diese Annahmen verletzt, so spiegelt sich dies in einer zu geringen Breite der Unsicherheitsintervalle wider. Zum aktuellen Zeitpunkt kann noch keine Aussage über die Verletzungen der Annahmen bezüglich der Patientenantworten in der Routineauswertung getroffen werden. Im Abschlussbericht zur Patientenbefragung des QS-Verfahrens QS PCI (IQTIG 2018b: Abschnitt 12.5) werden Möglichkeiten zur Einbettung der vorgeschlagenen Methodik in einen generellen, in den Sozialwissenschaften etablierten Modellierungsrahmen diskutiert und Ansätze zur Lockerung der Unabhängigkeitsannahmen skizziert, falls sich dies als notwendig herausstellt.

Neben den Unabhängigkeitsannahmen wird bei der vorgeschlagenen Auswertungsmethodik davon ausgegangen, dass es sich bei dem Mechanismus, der zu fehlenden Werten führt, nicht um MNAR handelt (Abschnitt 2.1.3), was sich im Allgemeinen nicht überprüfen lässt (Bennett 2001).

Um die Zuschreibbarkeit sicherzustellen, kann bei einzelnen Indikatoren prinzipiell eine Risikoadjustierung hinsichtlich patientenseitiger Faktoren notwendig sein. Aufgrund der hierarchischen Struktur der Indikatoren, Qualitätsmerkmale und Items und der Definition der Indikatoren als Mittel der zugrunde liegenden Parameter auf Qualitätsmerkmalsebene muss die derzeit verwendete Methodik für die Risikoadjustierung (wie sie beispielsweise für die planungsrelevanten Qualitätsindikatoren verwendet wird) weiterentwickelt werden. Die Einbettung der vorgeschlagenen Auswertungsmethodik in den generellen Modellierungsrahmen ermöglicht, dass etablierte Verfahren zur Risikoadjustierung angewendet werden können, falls dies zu einem späteren Zeitpunkt relevant wird.

Entsprechend der vorgeschlagenen Methodik werden die Qualitätsmerkmale bei der Definition des Indikators gleichgewichtet (Abschnitt 1.3.3). Eine bewusste Konsequenz davon ist, dass sich die Qualitätsmerkmale innerhalb eines Indikators kompensieren können. Das bedeutet, dass ein Leistungserbringer eine niedrige Punktezahl für ein Qualitätsmerkmal durch eine hohe Punktezahl für ein anderes Qualitätsmerkmal desselben Qualitätsindikators möglicherweise ausgleichen kann. Sollten für zukünftige Patientenbefragungen keine Kompensationseffekte erwünscht sein,

muss nicht von der vorgeschlagenen Methodik abgewichen werden, da sich diese entsprechend erweitern lässt (Fayers und Hand 2002).

Die Empfehlung des IQTIG ist, die Konsequenzen der Annahmen des in den Abschnitten 2.1 bis 2.3 dargestellten Bayesianischen Beta-Binomialmodells im Rahmen des Regelbetriebs der Patientenbefragung bei Qualitätsindikatoren zu überprüfen und ggf. die Auswertungsmethodik zu erweitern. Jedoch sind die durch eine Erweiterung gewonnenen Vorteile gegen die Verständlichkeit, die Kommunizierbarkeit und die Praktikabilität eines erweiterten Modellansatzes abzuwägen. Untersuchungen zur Praktikabilität und Relevanz können jedoch erst im Rahmen der Daten des Regelbetriebs durchgeführt werden. Die vorgestellte Auswertungsmethodik stellt somit nicht nur einen verständlichen, flexiblen und methodisch gut begründeten Ansatz dar, sondern garantiert auch, dass das Verfahren selbst bei den im Regelbetrieb zu erwartenden großen Datenmengen funktionieren wird.

### 3 Exemplarische Ergebnisberechnung für einen befragungsbasierten Qualitätsindikator

Im Folgenden wird die statistische Auswertungsmethodik an einem hypothetischen Beispiel eines Qualitätsindikators für einen Leistungserbringer illustriert. Der betrachtete Qualitätsindikator beruht auf vier Qualitätsmerkmalen, welche jeweils mit einer Frage (Item) operationalisiert werden.

Die konkreten Fragen und Antwortmöglichkeiten sind in der folgenden Tabelle 3 aufgelistet. Die Antwort „Ja“ wird jeweils mit 100 Punkten, die Antwort „Nein“ mit 0 Punkten bewertet. Die Antwortmöglichkeiten „Brauchte ich nicht“ sowie „Weiß nicht“ werden für den Leistungserbringer als neutral gewertet, d. h., sie gehen weder positiv noch negativ in die Berechnung ein.

Tabelle 3: Fragen (Items) und Qualitätsmerkmal eines befragungsbasierten Qualitätsindikators

Qualitätsmerkmal mit entsprechendem Item	Antwortmöglichkeiten
Qualitätsmerkmal 1	
Item A	Ja (100 Punkte) Nein (0 Punkte) Weiß nicht (neutral)
Qualitätsmerkmal 2	
Item B	Ja (100 Punkte) Nein (0 Punkte) Weiß nicht (neutral)
Qualitätsmerkmal 3	
Item C	Ja (100 Punkte) Nein (0 Punkte) Brauchte ich nicht (neutral) Weiß nicht (neutral)
Qualitätsmerkmal 4	
Item D	Ja (100 Punkte) Nein (0 Punkte) Brauchte ich nicht (neutral) Weiß nicht (neutral)

### Beispielszenario

Für einen konkreten Leistungserbringer sind insgesamt 5 ausgefüllte Fragebögen von Patientinnen und Patienten zurückgekommen, die in Tabelle 4 mit den jeweiligen Antworten dargestellt sind.

Tabelle 4: Beispielszenario für einen fiktiven Leistungserbringer mit 5 Fragebögen

Patientenpseudonym	Punkte			
	Item A	Item B	Item C	Item D
PA	100	100	0	100
PB	0	100	100	0
PC	100	0	0	100
PD	0	0	0	100
PE	0	0	0	0

Für jedes Item (A, B, C und D) wird zunächst über die vorliegenden 5 Fragebögen gezählt, wie viele Punkte maximal hätten erreicht werden können und wie viele tatsächlich erreicht worden sind (Tabelle 5). Für jedes Merkmal wird wiederum über die zugehörigen Items aufsummiert, wie viele Punkte erreicht wurden und maximal erreicht werden konnten. Für das Qualitätsmerkmal 1 mit Item A des Indikators hat der Leistungserbringer in diesem Beispiel 200 von 500 Punkten erreicht. Dieser Wert kann für das jeweilige Merkmal zur Schätzung des merkmalspezifischen Parameters  $\theta_m$  des Leistungserbringers verwendet werden, wobei hier  $m$  das Merkmal 1 indiziert.<sup>10</sup>

Tabelle 5: Zwischenergebnis pro Item, nachdem erreichte und maximale Punkte ermittelt wurden

Qualitätsmerkmal mit Item	Maximale Punkte	Erreichte Punkte
Qualitätsmerkmal 1 – A	500	200
Qualitätsmerkmal 2 – B	500	200
Qualitätsmerkmal 3 – C	500	100
Qualitätsmerkmal 4 – D	500	300

Da die statistische Inferenz in einem Bayesianischen Kontext geschieht, wird die sogenannte A-priori-Verteilung der Merkmalsparameter ( $\theta_m$ ) mittels der Informationen aus den vorliegenden Daten aufdatiert und die daraus resultierende A-posteriori-Verteilung der Merkmalsparameter für die Inferenz verwendet. Konkret wird bei diesem Qualitätsindikator mit insgesamt vier Merkmalen angenommen, dass a priori die Merkmalsparameter einer Beta-Verteilung mit den beiden Parametern  $a = b = \frac{1}{8}$  folgen ( $\theta_m \stackrel{i.i.d.}{\sim} \text{Beta}(\frac{1}{8}, \frac{1}{8})$ ). Da die A-posteriori-Verteilungen der Merkmale

<sup>10</sup> Es wird im Folgenden auf den Leistungserbringerindex  $i$  verzichtet, da in diesem Beispiel nur ein Leistungserbringer betrachtet wird.

ebenfalls Beta-Verteilungen sind, können die Erwartungswerte und Unsicherheitsintervalle direkt daraus bestimmt werden.

Um die aufdatierten Parameter der A-posteriori-Beta-Verteilung zu erhalten, müssen die Punkte auf durchgehend ganzzahlige Werte von 0 bis N skaliert werden, wobei auch die möglichen Zwischenwerte berücksichtigt werden müssen. In diesem Fall waren alle Punktwerte von 0, 100, 200 usw. bis 500 möglich. Diese werden auf die Werte 0, 1, 2 usw. bis 5 skaliert.<sup>11</sup> Die im Beispiel vom Leistungserbringer erreichte Punktzahl von 200 für das erste Qualitätsmerkmal entspricht einem re-skalierten Punktwert 2 von 5.

Die Parameter der A-posteriori-Beta-Verteilung werden durch Hinzuaddieren der Punktwerte aufdatiert, d. h., es werden bei den Parametern  $a$  und  $b$  jeweils die erreichten Punkte bzw. die zum Maximalwert fehlenden Punkte hinzuaddiert (siehe dazu die folgende Tabelle 6).

Tabelle 6: Berechnungsergebnisse der Parameter  $a$  und  $b$  sowie des Erwartungswerts für jeden Merkmalswert

Merkmal	$a$	$b$	Erwartungswert des Merkmalwerts: $E(\theta_m)$
1	$\frac{1}{8} + 2 = 2,125$	$\frac{1}{8} + (5 - 2) = 3,125$	$2,125 / (2,125 + 3,125) = 0,4047$
2	$\frac{1}{8} + 2 = 2,125$	$\frac{1}{8} + (5 - 2) = 3,125$	$2,125 / (2,125 + 3,125) = 0,4047$
3	$\frac{1}{8} + 1 = 1,125$	$\frac{1}{8} + (5 - 1) = 4,125$	$1,125 / (1,125 + 4,125) = 0,2142$
4	$\frac{1}{8} + 3 = 3,125$	$\frac{1}{8} + (5 - 3) = 2,125$	$3,125 / (3,125 + 2,125) = 0,5952$

Die konkrete A-posteriori-Verteilung für den Indikatorwert sowie der entsprechende Erwartungswert und das Unsicherheitsintervall lassen sich wie oben dargestellt bestimmen. Im Falle von Qualitätsindikatoren mit mehreren Merkmalen resultiert der Indikatorwert aus dem Mittelwert der Merkmalparameter, z. B. beim vorliegenden Indikator mit vier Merkmalen mittels  $\theta = \frac{1}{4}\theta_1 + \frac{1}{4}\theta_2 + \frac{1}{4}\theta_3 + \frac{1}{4}\theta_4$ . Somit ist der Indikatorwert (d. h. der Punktschätzer) für diesen konkreten Leistungserbringer der Mittelwert der Erwartungswerte der vier Merkmalsparameter und liegt hier bei 0,4047 bzw. bei 40,47 Punkten auf der Punkteskala.

Das Unsicherheitsintervall ergibt sich nun aus dem 2,5%- und 97,5%-Quantil der A-posteriori-Verteilung, die in diesem Fall als die gemittelte Verteilung der A-posteriori-Verteilungen der vier Merkmale berechnet wird. Konkret ergibt dies ein Intervall von 22,61 bis 59,29 Punkten.

Hätte der Indikator einen festen Referenzbereich mit dem Intervall [95, 100] Punkte, wird die obere Grenze des 95%-Unsicherheitsintervalls mit dem Referenzwert  $t$  (im Beispiel  $t = 95$  Punkte) ver-

<sup>11</sup> Hätten die Items mehr als zwei informative Kategorien ermöglicht, z. B. Antworten mit Punktwerten 0, 50 und 100, so wären entsprechend mehr Zwischenstufen möglich gewesen: 0, 50, 100, 150 usw. bis 500. Diese wären auf die Werte 0, 1, 2, 3 usw. bis 10 skaliert worden.

glichen. Liegt die obere Grenze des Unsicherheitsintervalls unter 95 Punkten, so wird der Leistungserbringer für diesen Indikator als quantitativ auffällig eingestuft. Ansonsten wird er als quantitativ unauffällig bewertet. Da die obere Grenze des Intervalls für das konkrete Leistungserbringerergebnis bei 59,29 Punkten liegt und somit kleiner als 95 Punkte ist, wird das Leistungserbringerergebnis in der Auswertung als quantitativ auffällig eingestuft.

## Literatur

- Agresti, A (2013): *Categorical Data Analysis*. Third Edition. (Wiley Series in Probability and Statistics). Hoboken, US-NJ: Wiley. ISBN: 978-0-470-46363-5.
- Bennett, DA (2001): How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health* 25(5): 464-469. DOI: 10.1111/j.1467-842X.2001.tb00294.x.
- Brown, LD; Cai, TT; DasGupta, A (2001): Interval Estimation for a Binomial Proportion. *Statistical Science* 16(2): 101-117.
- Bühner, M (2011): *Einführung in die Test- und Fragebogenkonstruktion*. 3., aktualisierte und erweiterte Auflage. München [u. a.]: Pearson Studium. ISBN: 978-3-86894-033-6.
- Carlin, BP; Louis, TA (2009): *Bayesian Methods for Data Analysis*. Third Edition. (Texts in Statistical Science). Boca Raton, US-FL [u. a.]: CRC Press. ISBN: 978-1-58488-697-6.
- Fayers, PM; Hand, DJ (2002): Causal variables, indicator variables and measurement scales: an example from quality of life. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 165(2): 233-253. DOI: 10.1111/1467-985X.02020.
- 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.
- Held, L; Bové, DS (2014): *Applied Statistical Inference: Likelihood and Bayes*. Heidelberg [u. a.]: Springer. ISBN: 978-3-642-37886-7.
- 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/dateien/berichte/2016/IQTIG\\_Planungsrelevante-Qualitätsindikatoren\\_Abschlussbericht.pdf](https://iqtig.org/dateien/berichte/2016/IQTIG_Planungsrelevante-Qualitätsindikatoren_Abschlussbericht.pdf) (abgerufen am: 05.12.2018).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018a): Entwicklung einer Befragung von Patienten und Patientinnen mit Schizophrenie, schizotypen und wahnhaften Störungen. *Entwicklung einer Patientenbefragung im Rahmen der Aktualisierung und Erweiterung des QS-Verfahrens Versorgung von volljährigen Patienten und Patientinnen mit Schizophrenie, schizotypen und wahnhaften Störungen*. Abschlussbericht. Stand: 15.12.2018. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2018/IQTIG\\_Patientenbefragung\\_QS-Verfahren-Schizophrenie\\_Abschlussbericht\\_2018-12-15\\_barrierefrei.pdf](https://iqtig.org/downloads/berichte/2018/IQTIG_Patientenbefragung_QS-Verfahren-Schizophrenie_Abschlussbericht_2018-12-15_barrierefrei.pdf) (abgerufen am: 25.10.2021).

- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018b):  
Entwicklung von Patientenbefragungen im Rahmen des Qualitätssicherungsverfahrens  
*Perkutane Koronarintervention und Koronarangiographie*. Abschlussbericht. Stand: 15.12.2018.  
Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2018/IQTIG\\_Patientenbefragung\\_QS-PCI\\_Abschlussbericht\\_2018-12-15\\_barrierefrei.pdf](https://iqtig.org/downloads/berichte/2018/IQTIG_Patientenbefragung_QS-PCI_Abschlussbericht_2018-12-15_barrierefrei.pdf) (abgerufen am: 26.10.2021).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2021a):  
Entwicklung einer Patientenbefragung für das Qualitätssicherungsverfahren zur ambulanten  
psychotherapeutischen Versorgung gesetzlich Krankenversicherter. Abschlussbericht.  
Stand: 15.12.2021. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2021/IQTIG\\_Patientenbefragung-QS-Verfahren-Ambulante-Psychotherapie\\_Abschlussbericht\\_2021-12-15.pdf](https://iqtig.org/downloads/berichte/2021/IQTIG_Patientenbefragung-QS-Verfahren-Ambulante-Psychotherapie_Abschlussbericht_2021-12-15.pdf) (abgerufen am: 21.02.2023).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2021b):  
Entwicklung von Patientenbefragungen für das Qualitätssicherungsverfahren  
*Nierenersatztherapie*. Abschlussbericht. Stand: 24.06.2021. Berlin: IQTIG. URL:  
[https://iqtig.org/downloads/berichte/2021/IQTIG\\_QS-NET\\_Befragung\\_Abschlussbericht\\_2021-06-24.pdf](https://iqtig.org/downloads/berichte/2021/IQTIG_QS-NET_Befragung_Abschlussbericht_2021-06-24.pdf) (abgerufen am: 20.11.2023).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2022):  
Methodische Grundlagen. Version 2.0. Stand: 27.04.2022. Berlin: IQTIG. URL:  
[https://iqtig.org/downloads/berichte-2/meg/IQTIG\\_Methodische-Grundlagen\\_Version-2.0\\_2022-04-27.pdf](https://iqtig.org/downloads/berichte-2/meg/IQTIG_Methodische-Grundlagen_Version-2.0_2022-04-27.pdf) (abgerufen am: 25.05.2022).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2023):  
Entwicklung einer Patientenbefragung für das Qualitätssicherungsverfahren  
Entlassmanagement. Abschlussbericht. Stand: 28.04.2023. Berlin: IQTIG. URL:  
[https://iqtig.org/downloads/berichte/2023/IQTIG\\_Patientenbefragung-QS-Verfahren-Entlassmanagement\\_Abschlussbericht\\_2023-04-28.pdf](https://iqtig.org/downloads/berichte/2023/IQTIG_Patientenbefragung-QS-Verfahren-Entlassmanagement_Abschlussbericht_2023-04-28.pdf) (abgerufen am: 13.11.2023).
- Molenaar, IW; Sijtsma, K (1984): Internal Consistency and Reliability in Mokken's Nonparametric  
Item Response Model. *Tijdschrift voor Onderwijsresearch* 9(5): 257-269. URL:  
<http://objects.library.uu.nl/reader/index.php?obj=1874-208777&lan=en> (abgerufen am:  
07.09.2020).
- Molenberghs, G; Beunckens, C; Sotto, C; Kenward, MG (2008): Every missingness not at random  
model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical  
Society: Series B (Statistical Methodology)* 70(2): 371-388. DOI: 10.1111/j.1467-  
9868.2007.00640.x.
- NHS England Analytic Team (2014): Methods, Reasoning and Scope. Statement of Methodology  
for the Overall Patient Experience Scores (Statistics). [Stand:] December 2014. Wakefield, GB:  
NHS England. URL: [https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/  
2014/04/Methods-statement\\_20150420.pdf](https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2014/04/Methods-statement_20150420.pdf) (abgerufen am: 04.12.2023).



- Revelle, W (2018): Package 'psych'. Procedures for Personality and Psychological Research. Version 1.8.4. [Stand:] 30.04.2018. Evanston, US-IL: Northwestern University. URL: <https://cran.r-project.org/web/packages/psych/psych.pdf> (abgerufen am: 04.12.2023).
- Robert, C; Casella, G (2004): Monte Carlo Statistical Methods. Second Edition. New York, US-NY: Springer. ISBN: 978-1-4419-1939-7.
- Sheng, Y; Sheng, Z (2012): Is coefficient alpha robust to non-normal data? *Frontiers in Psychology* 3(34). DOI: 10.3389/fpsyg.2012.00034.
- 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.
- Sizmur, S; Graham, C; Bos, N (2020): Psychometric evaluation of patient-reported experience measures: is it valid? *International Journal for Quality in Health Care* 32(3): 219-220. DOI: 10.1093/intqhc/mzaa006.
- Skrondal, A; Rabe-Hesketh, S (2004): Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models. (Interdisciplinary Statistics Series). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC. ISBN: 978-1-58488-000-4.
- Stochl, J; Jones, PB; Croudace, TJ (2012): Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT method in empirical research for applied health researchers. *BMC: Medical Research Methodology* 12(1): 74. DOI: 10.1186/1471-2288-12-74.
- Tavakol, M; Dennick, R (2011): Making sense of Cronbach's alpha. *International Journal of Medical Education* 2: 53-55. DOI: 10.5116/ijme.4dfb.8dfd.
- Trizano-Hermosilla, I; Alvarado, JM (2016): Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. *Frontiers in Psychology* 7: 769. DOI: 10.3389/fpsyg.2016.00769.
- Yang, W; Nakaya, J (2006): Statistical Applications for SNPs Analysis. *Chem-Bio Informatics Journal* 6(2): 55-68. DOI: 10.1273/cbij.6.55.
- Zinbarg, RE; Revelle, W; Yovel, I; Li, W (2005): Cronbach's  $\alpha$ , Revelle's  $\beta$ , and McDonald's  $\omega_H$ : their relations with each other and two alternative conceptualizations of reliability. *Psychometrika* 70(1): 123-133. DOI: 10.1007/s11336-003-0974-7.

# Impressum

## HERAUSGEBER

---

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

Telefon: (030) 58 58 26-0

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

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