



Institut für Qualitätssicherung und  
Transparenz im Gesundheitswesen

# **Kriterien für den gezielten Datenabgleich in der Datenvalidierung nach QSKH-RL**

Abschlussbericht

Erstellt im Auftrag des  
Gemeinsamen Bundesausschusses

Stand: 31. Januar 2020

---

# Impressum

**Thema:**

Kriterien für den gezielten Datenabgleich in der Datenvalidierung nach QSKH-RL.  
Abschlussbericht

**Ansprechpartnerin:**

Anna Maria Steinmann

**Auftraggeber:**

Gemeinsamer Bundesausschuss

**Datum des Auftrags:**

17. November 2017

**Datum der Abgabe:**

26. Juli 2018, geänderte Fassung vom 31. Januar 2020

**Herausgeber:**

IQTIG – Institut für Qualitätssicherung  
und Transparenz im Gesundheitswesen

Katharina-Heinroth-Ufer 1  
10787 Berlin

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

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

<https://www.iqtig.org>

# Inhaltsverzeichnis

Tabellenverzeichnis.....	5
Abbildungsverzeichnis.....	6
Abkürzungsverzeichnis.....	7
Glossar.....	9
1 Einleitung.....	12
1.1 Hintergrund.....	12
1.2 Verständnis und Vorgehen der Überarbeitung.....	14
2 Methodik zur Berechnung von Fehldokumentationsraten.....	20
2.1 Berechnung von Fehldokumentationsraten .....	21
2.1.1 Fehldokumentationsrate eines Datenfeldes.....	22
2.1.2 Fehldokumentationsrate über alle Datenfelder hinweg.....	25
2.1.3 Geschichtete Stichproben .....	29
2.1.4 Zusammenfassung.....	29
2.2 Statistische Methodik (Vertrauensintervalle).....	30
2.2.1 Vertrauensintervalle für einfache Zufallsstichproben ohne Zurücklegen....	31
2.2.2 Vertrauensintervalle für Cluster-Stichproben.....	33
2.2.3 Einordnung in das statistische Rahmenkonzept für Auswertungen .....	39
3 Addendum I zum Häufigkeitskriterium .....	42
3.1 Dokumentationsfehler von besonderer Häufigkeit .....	42
3.2 Dokumentationsfehler von erheblicher Häufigkeit .....	44
3.3 Weitere Aspekte zu Kriterien und Prüfmethodik von Dokumentationsfehlern mit besonderer und erheblicher Häufigkeit.....	45
3.4 Prozessablauf bei Dokumentationsfehlern von besonderer oder erheblicher Häufigkeit .....	46
3.5 Limitationen .....	47
4 Addendum II zum Relevanzkriterium .....	49
4.1 Methodisches Vorgehen .....	49
4.2 Definition des Relevanzkriteriums .....	56
4.3 Statistisches Verfahren (Vertrauensintervall).....	62
4.4 Sentinel-Event-Kriterium.....	63

4.5	Prozessablauf bei Unterdokumentation relevanter Ereignisse und Nicht-Dokumentation eines Sentinel Event.....	64
4.6	Limitationen .....	66
5	Addendum III zum Kriterium der wiederholten rechnerischen Auffälligkeit .....	69
5.1	Methodisches Vorgehen .....	69
5.2	Modellrechnungen.....	72
5.3	Herausforderungen zum Kriterium zur wiederholten rechnerischen Auffälligkeit ..	74
5.3.1	Schlussfolgerung des IQTIG.....	75
5.4	Definition und Prozessablauf zum Kriterium zur wiederholten rechnerischen Auffälligkeit .....	76
5.5	Limitationen .....	78
6	Prozessablauf des gezielten Datenabgleichs insgesamt .....	79
7	Bewertung der Dokumentationsqualität .....	83
7.1	Methodisches Vorgehen .....	83
7.2	Bewertungsebenen .....	86
7.3	Synthese.....	89
7.4	Limitationen .....	97
8	Berichterstattung nach § 136b SGB V .....	99
8.1	Berichterstattung nach Durchführung des Stichprobenverfahrens mit Datenabgleich .....	99
8.2	Berichterstattung nach Durchführung des gezielten Datenabgleichs .....	100
8.3	Limitationen .....	101
9	Prüfung nicht übermittelter Fälle.....	103
10	Spezifikation des bundeseinheitlichen Prüfberichts .....	105
11	Entwicklung eines Erfassungstools für die prüfenden Stellen für den Datenabgleich vor Ort.....	107
12	Abschließende Empfehlungen.....	108
	Literatur.....	109

## Tabellenverzeichnis

Tabelle 1: Schätzung der Fehldokumentationen insgesamt ohne und mit Fallzahlgewichtung	25
Tabelle 2: Formeln für die Berechnung der Fehldokumentationsrate eines Datenfeldes und der Gesamtfehldokumentationsrate .....	29
Tabelle 3: Fehldokumentationsrate für besonders häufige Dokumentationsfehler in Abhängigkeit von der Gesamtfallzahl (Simulation).....	43
Tabelle 4: Fehldokumentationsrate für erheblich häufige Dokumentationsfehler in Abhängigkeit von der Gesamtfallzahl (Simulation).....	44
Tabelle 5: QS-Verfahren/QS-Modul und Anzahl der Standorte für den gezielten Datenabgleich 2019 .....	46
Tabelle 6: Beispielhafte Darstellung von unterdokumentierten relevanten Ereignissen.....	53
Tabelle 7: Ergebnisse zur Dokumentation des Ereignisses „Tod“ in 17/1 .....	55
Tabelle 8: : Ergebnisse zur Dokumentation spezifischer Komplikationen (nur im Datenabgleich überprüfte Komplikationen) in 17/1.....	56
Tabelle 9: Gewichtung der für das Indikatorergebnis relevanten Ereignisse invers proportional zu ihrer Prävalenz .....	58
Tabelle 10: Exemplarisch für die Modellrechnung ausgewählte Ereignisse zum Relevanzkriterium in 17/1.....	60
Tabelle 11: Unterdokumentation definierter Ereignisse über alle Krankenhausstandorte hinweg.....	61
Tabelle 12: Anzahl der Auffälligkeitskriterien je QS-Verfahren für das EJ 2018.....	70
Tabelle 13: Wiederholt rechnerisch auffällige Standorte pro QS-Verfahren bezogen auf EJ 2016, EJ 2017 und EJ 2018. ....	73
Tabelle 14: Wiederholt rechnerisch auffällige Standorte pro QS-Verfahren bezogen auf EJ 2017 und EJ 2018. ....	74
Tabelle 15: Betrachtungsebenen der Dokumentationsqualität inklusive Beispiele .....	87
Tabelle 16: Grenzwerte für die Bewertung der Dokumentationsqualität auf Standortebene...	91
Tabelle 17: Grenzwerte für die Bewertung der Dokumentationsqualität auf standortübergreifender Ebene .....	92
Tabelle 18: Synthese der Grenzwerte aus Tabelle 16 und Tabelle 17 und Bewertungsschema für die Standortebene und standortübergreifende Ebene.....	93

## Abbildungsverzeichnis

Abbildung 1: Darstellung der Datenvalidierung nach § 9 QSKH-RL .....	17
Abbildung 2: Schematische Darstellung der Stichprobe gemäß § 9 QSKH-RL .....	20
Abbildung 3: Veranschaulichung der mathematischen Notation in Bezug zu Abbildung 2 .....	21
Abbildung 4: Die Stichprobe gemäß § 9 QSKH-RL als Cluster-Stichprobe: Für die $I$ Standorte in der Standortstichprobe wird jeweils eine Fallstichprobe von i. d. R. 20 Fällen gezogen. .	24
Abbildung 5: Geschichte Stichprobe: Für jeden der $I$ Krankenhausstandorte werden separat sowohl Fälle mit einem bestimmten Merkmal (orange) als auch Fälle ohne dieses Merkmal (grün) gezogen.....	29
Abbildung 6: Überdeckungswahrscheinlichkeiten für die drei diskutierten Vertrauensintervalle für Fehldokumentationsraten eines Datenfeldes bei einem Krankenhausstandort für $n = 20$ und $N = 25$ sowie $N = 100$ . .....	33
Abbildung 7: Überdeckungswahrscheinlichkeiten für verschiedene Varianten von Vertrauensintervallen für die Gesamtfehldokumentationsrate eines Krankenhausstandortes gemäß der durchgeführten Simulation. ....	38
Abbildung 8: Gezielter Datenabgleich nach besonders häufigen Dokumentationsfehlern und sofortige Benennung gegenüber dem Lenkungsgremium bei erheblichen Dokumentationsfehlern.....	47
Abbildung 9: Ergebnisse des Häufigkeits- und Relevanzkriteriums im Vergleich .....	62
Abbildung 10: Gezielter Datenabgleich nach relevanten Dokumentationsfehlern und der Nicht-Dokumentation eines Sentinel Event.....	65
Abbildung 11: Anzahl der Standorte mit mehrjähriger Auffälligkeit je Auffälligkeitskriterium für die Erfassungsjahre 2016, 2017 und 2018 .....	73
Abbildung 12: Gezielter Datenabgleich nach wiederholt rechnerischer Auffälligkeit in der Statistischen Basisprüfung. ....	78
Abbildung 13: Darstellung der Datenvalidierung nach § 9 QSKH-RL: Wege der Krankenhausstandorte in die gezielte Stichprobe über Dokumentationsfehler in den verschiedenen Kriterien.....	79
Abbildung 14: Zeitablauf des Stichprobenverfahrens und gezielten Datenabgleichs .....	82

## Abkürzungsverzeichnis

Abkürzung	Bedeutung
AG DeQS	Arbeitsgruppe zur datengestützten einrichtungsübergreifenden Qualitätssicherung
AG esQS	Arbeitsgruppe Externe stationäre Qualitätssicherung
AK	Auffälligkeitskriterium
ASA	American Society of Anesthesiologists
BGB	Bürgerliches Gesetzbuch
DAS	Datenannahmestelle(n)
DeQS-RL	Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung
DF	Datenfeld
EJ	Erfassungsjahr
GAM-Modell	Generalized Additive Mixed-Modell
G-BA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
GKV-SV	Spitzenverband Bund der Krankenkassen
ICD	Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme ( <i>International Statistical Classification of Diseases and Related Health Problems</i> )
IKNR	Institutionskennzeichennummer
IOM	Institute of Medicine
IQTIG	Institut für Qualitätssicherung und Transparenz im Gesundheitswesen
n.F.	neue Fassung
O / E	Verhältnis aus der beobachteten und der erwarteten Rate ( <i>observed to expected ratio</i> )
plan. QI-RL	Richtlinie zu planungsrelevanten Qualitätsindikatoren
QI	Qualitätsindikator
QS	Qualitätssicherung
QSKH-RL	Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern
SD	Strukturierter Dialog
SGB V	Fünftes Buch Sozialgesetzbuch

<b>Abkürzung</b>	<b>Bedeutung</b>
UA QS	Unterausschuss Qualitätssicherung des G-BA

# Glossar

## **Dokumentationsdefizite**

Dokumentationsdefizite liegen vor, wenn Dokumentationsfehler in besonderer Häufigkeit oder von Relevanz festzustellen sind.

## **Dokumentationsfehler**

Dokumentationsfehler liegen vor, wenn im Rahmen der QS-Dokumentation Datenfelder fehlerhaft dokumentiert und/oder fehlerhaft übermittelt werden. Sie liegen insbesondere vor, wenn die Angaben aus der QS-Dokumentation nicht mit den Angaben aus der Patientenakte übereinstimmen.

## **Dokumentationsmängel**

Die im Auftrag des G-BA erwähnten Dokumentationsmängel werden synonym zu Fehldokumentationen verwendet.

## **Dokumentationsqualität**

Qualität ist der Grad, in dem Merkmale eines Objekts Anforderungen erfüllen (siehe IQTIG (2018b)). In der jährlich stattfindenden Datenvalidierung nach § 9 QSKH-RL wird geprüft, ob die dokumentierten Qualitätssicherungsdaten (QS-Daten) im jeweils betrachteten Erfassungsjahr korrekt, vollständig, vollzählig und plausibel waren. Eine gute Dokumentationsqualität ist Voraussetzung für eine ausreichende Aussagekraft der Qualitätsindikatoren.

## **Erhebliche Dokumentationsmängel**

Sind definiert als *erheblich häufige Dokumentationsfehler* (Häufigkeitskriterium)

## **Fehldokumentation**

Abweichung zwischen den Angaben aus der QS-Dokumentation und der Zweiterfassung (Patientenakte). Eine Fehldokumentation zugunsten eines Krankenhauses liegt vor, wenn QS-Fälle derart fehldokumentiert sind, dass der fälschlich angegebene Wert in der QS-Dokumentation das Qualitätsergebnis des Leistungserbringers verbessert.

## **Fehldokumentationsrate**

Die Fehldokumentationsrate berechnet sich als Anteil aller Datenfelder, die fehldokumentiert sind. Sie kann auf Ebene der Krankenhausstandorte, auf Ebene der Datenfelder oder auf QS-Verfahrenebene berechnet werden. Dabei ist zwischen der (aus dem Datenabgleich bekannten) Fehldokumentationsrate in einer Stichprobe und der (unbekannten) Fehldokumentationsrate bezogen auf alle dokumentierten Fälle eines Krankenhausstandortes zu unterscheiden. Mithilfe von statistischen Verfahren kann von der Fehldokumentationsrate in der Stichprobe auf die Fehldokumentationsrate im Gesamtdatensatz geschlossen werden.

## **Indikatorrelevante/rechenregelrelevante Datenfelder**

Indikatorrelevant sind Datenfelder, wenn über sie der Ein- bzw. Ausschluss von Fällen in den Nenner oder den Zähler von Qualitätsindikatoren bestimmt wird, sie also in den Rechenregeln des Qualitätsindikators genannt werden (rechenregelrelevante Datenfelder).

### **Mängel in der Spezifikation**

Die Spezifikation umfasst eine detaillierte Beschreibung von Anforderungen, die durch das IQTIG spezifiziert werden. Hierzu zählen die Auslösung einer Dokumentationspflicht, die Selektion oder Dokumentation von Daten, der Datenexport und die Datenübermittlung sowie die notwendigen Datenschutzmaßnahmen. Aufgrund der Komplexität des Regelwerkes kann trotz einer Vielzahl von Prüfungen und Kontrollmechanismen nicht ausgeschlossen werden, dass Implausibilitäten, Unschärfen oder Fehler in der Spezifikation enthalten sein können. Diese können wiederum dazu führen, dass die Spezifikation bzw. Teile der Spezifikation Mängel aufweisen.

### **QS-Daten**

Sammelbegriff für alle Daten, die im Zuge eines QS-Verfahrens erhoben und ausgewertet werden.

### **Sensitivität, Spezifität**

In Anlehnung an die entsprechenden Begriffe im Zusammenhang mit der Bewertung statistischer Tests werden die Begriffe Sensitivität und Spezifität hier wie folgt verwendet:

**Sensitivität** bezeichnet den Anteil der Fälle, bei denen sowohl in der QS-Dokumentation als auch in der Patientenakte eine bestimmte Information dokumentiert wurde (Zähler) an allen Fällen, bei denen in den Patientenakten diese Information dokumentiert wurde (Nenner). Die Sensitivität gibt also die bedingte Wahrscheinlichkeit an, dass eine Information (z. B. Lungenembolie, Geschlecht = weiblich) in der QS dokumentiert wurde, gegeben, dass diese Information in der Patientenakte dokumentiert wurde. Durch die Berechnung von  $1 - \text{Sensitivität}$  wird der Anteil an fälschlicherweise nicht in der QS dokumentierten Ereignissen festgelegt und ist das Maß für die Unterdokumentation in der QS-Dokumentation.

**Spezifität** bezeichnet den Anteil der Fälle, bei denen weder in der QS-Dokumentation noch in der Patientenakte eine bestimmte Information dokumentiert wurde (Zähler) an allen Fällen, bei denen in den Patientenakten diese Information nicht dokumentiert wurde (Nenner).

Die Spezifität gibt also die bedingte Wahrscheinlichkeit an, dass eine Information (z. B. Lungenembolie, Geschlecht = weiblich) nicht in der QS dokumentiert wurde, gegeben, dass diese Information nicht in der Patientenakte dokumentiert wurde. Durch die Berechnung von  $1 - \text{Spezifität}$  wird der Anteil der fälschlicherweise für die QS dokumentierten Ereignisse festgelegt und ist das Maß für die Überdokumentation in der QS-Dokumentation.

Der jeweilige Wert in der Patientenakte gilt in dieser Analogie zur Testtheorie also als „Wahrheit“, auf die die QS-Dokumentation „getestet“ wird.

### **Positiv und negativ formulierte Ereignisse**

Positiv formulierte Ereignisse sind erwünschte Ereignisse, wie z. B. die Durchführung einer Antibiotikaprophylaxe oder das Überleben. Negativ formulierte Ereignisse sind unerwünschte Ereignisse wie Komplikationen oder Todesfälle. Aufgrund der Belegung des Begriffs unerwünschte Ereignisse im Arzneimittelbereich wird im hier vorliegendem Bericht von positiv und negativ formulierten Ereignissen gesprochen, deren Unterdokumentation ein relevantes Ereignis abbildet (s. Relevanzkriterium, Kapitel 4)

### **Übereinstimmungsrate**

Die Übereinstimmungsrate ist die Umkehrung der Fehldokumentationsrate (siehe dort): Übereinstimmungsrate = 100 % - Fehldokumentationsrate.

### **Zweiterfassungsdaten**

Datenfelder der QS-Dokumentation, für die anhand der Patientenakte eine Zweiterfassung erfolgt.

# 1 Einleitung

## 1.1 Hintergrund

Das Datenvalidierungsverfahren zur QSKH-RL wurde erstmalig im Rahmen der „Vereinbarung des Gemeinsamen Bundesausschusses gemäß § 137 Abs. 1 SGB V i. V. mit § 135a SGB V über Maßnahmen zur Qualitätssicherung für nach § 108 SGB V zugelassene Krankenhäuser“ am 21. März 2006 beschlossen und trat am 20. April 2006 in Kraft. Diese Vereinbarung ging dann in die QSKH-RL<sup>1</sup> über. In Ergänzung zu den bereits bekannten Verfahrensteilen der Statistischen Basisprüfung und dem Stichprobenverfahren mit Datenabgleich, für den Krankenhausstandorte zufällig ausgewählt werden, sieht § 9 Abs. 3 Satz 1 QSKH-RL ab dem Erfassungsjahr 2018 zusätzlich einen Datenabgleich für eine gezielte Auswahl von Krankenhausstandorten vor. Gezielt ausgewählt werden sollen Krankenhausstandorte, bei denen konkrete Anhaltspunkte für mögliche Dokumentationsmängel vorliegen. Diese konkreten Anhaltspunkte können sich auf Grundlage der folgenden Daten ergeben:

- Ergebnisse des Stichprobenverfahrens mit Datenabgleich (Zufallsstichprobe) des Vorjahres
- Ergebnisse der Statistischen Basisprüfung der Vorjahre

Nach § 9 Abs. 3 QSKH-RL erfolgt ein gezielter Datenabgleich zur Überprüfung der Dokumentationsqualität

1. bei einem Krankenhausstandort, wenn im Vorjahr im Datenabgleich Dokumentationsfehler in besonderer Häufigkeit oder Relevanz festgestellt wurden,
2. bei wiederholten rechnerischen Auffälligkeiten in der Statistischen Basisprüfung oder
3. bei festgestellter Nichtdokumentation eines Sentinel Events im Vorjahr.

Des Weiteren kann der Datenabgleich nach § 9 Abs. 3 Satz 3 QSKH-RL darüber hinaus bei weiteren Auffälligkeiten in der Dokumentationsqualität erfolgen.

Mit Beschluss vom 17. November 2017 beauftragte der Gemeinsame Bundesausschuss (G-BA) das IQTIG bis zum 31. Mai 2018, Kriterien für die Anhaltspunkte zu definieren, wann im Rahmen der Datenvalidierung nach § 9 der Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern (QSKH-RL) die Dokumentationsqualität eines Krankenhausstandortes durch einen gezielten Datenabgleich der QS-Dokumentation mit der Primärdokumentation überprüft werden kann, und Kriterien zu entwickeln, wie die Dokumentationsqualität bewertet werden soll (G-BA 2017).

Der Abschlussbericht zur Beauftragung wurde am 26. Juli 2018 dem Gemeinsamen Bundesausschuss vorgelegt. Die erste Vorstellung des Berichts erfolgte am 18. September 2019 unter Berücksichtigung der Anmerkungen der AG esQS. Es wurde festgestellt, dass zum Auftragsverständnis zwischen IQTIG und G-BA noch Abstimmungsbedarf bestand. Aufgrund der z. T. vorgegebenen Festlegung der Kriterien war die Ausdifferenzierung bereits determiniert, was

---

<sup>1</sup> Richtlinie des Gemeinsamen Bundesausschusses gemäß § 136 Abs. 1 SGB V i. V. m. § 135a SGB V über Maßnahmen der Qualitätssicherung für nach § 108 SGB V zugelassene Krankenhäuser. In der Fassung vom 15. August 2006, zuletzt geändert am 20. Juni 2019, in Kraft getreten am 1. Januar 2020.

eine besondere Herausforderung für die Ausarbeitung bedeutete. Ein wesentlicher Punkt war, dass das Gesamtkonzept und die festzulegenden Kriterien leichter nachvollziehbar sein sollten. Dem IQTIG wurde mitgeteilt, dass deutlicher formulierte Definitionen und Algorithmen sowie Modellrechnungen fehlten.

Zur weiteren Abstimmung wurde ein Expertentreffen eingerichtet, in dem eine methodische und inhaltliche Analyse und Diskussion des eingereichten Berichts erfolgte. Als Ergebnis wurde im G-BA vereinbart, den Bericht in einem mehrstufigen Vorgehen anzupassen. Hierfür schlug das IQTIG vor, 2019 die unterschiedlichen Kriterien (Häufigkeitskriterium, Relevanzkriterium und das Kriterium zur wiederholten rechnerischen Auffälligkeit) jeweils zu überarbeiten und in der AG DeQS zu besprechen. Abschließend würden die zum 31. Januar 2020 überarbeiteten Kriterien in den Abschlussbericht integriert werden. Auf diese Weise könnten einzelne Kriterien, wie z. B. das Häufigkeitskriterium, durch einen gesonderten Beschluss bereits im Jahr 2019 Anwendung finden und auf diesem Weg eine stufenweise Umsetzung des gezielten Datenabgleichs erfolgen.

Somit wurde zunächst das Häufigkeitskriterium angepasst und der ursprüngliche Abschlussbericht in einer geänderten Fassung am 19. Dezember 2018 erneut eingereicht. Das überarbeitete Häufigkeitskriterium wurde in der AG DeQS am 22. Januar 2019 präsentiert und in einem gesonderten Dokument am 5. Februar 2019 als Anlage zum Beschluss für den UA QS eingereicht (Addendum I). Der UA QS des G-BA hat in seiner Sitzung am 6. März 2019 das vom IQTIG eingereichte Kriterium und die Prüfmethodik zu Dokumentationsfehlern von besonderer und erheblicher Häufigkeit beschlossen (Addendum zum Abschlussbericht vom 5. Februar 2019: Beschluss der Abschnitte 2.1, 2.2 und 2.3).

In einem nächsten Schritt erfolgte die Überarbeitung des Relevanzkriteriums. Hierzu hat am 18.07.2019 ein weiteres Expertentreffen stattgefunden, in welchem ausführlich das Kriterium zur Relevanz diskutiert wurde. Das Addendum II zum Relevanzkriterium ist am 26. Juli 2019 dem G-BA vorgelegt worden und wurde vom IQTIG am 5. September 2019 in einer Präsentation erläutert. Die geänderte Fassung vom 22. Oktober 2019 (nach AG-Beratung) wurde dem UA QS zum Beschluss vorgelegt. Der UA QS hat in seiner Sitzung am 4. Dezember 2019 Kapitel 2 des Addendum II beschlossen.

Im dritten Schritt wurde das Addendum III zu den wiederholten rechnerischen Auffälligkeiten in der Statistischen Basisprüfung am 30. September 2019 beim G-BA eingereicht und am 29. Oktober vorgestellt und von der AG DeQS zur Kenntnis genommen. Eine erste Beratung erfolgte am 17. Dezember 2019. Eine Beschlussfassung wird aufgrund der vom IQTIG genannten Limitationen vertagt und die Abgabe des überarbeiteten Abschlussberichts abgewartet.

Alle drei erarbeiteten Addenda wurden in den hier vorliegenden überarbeiteten Abschlussbericht integriert.

Gemäß der QSKH-RL soll die Dokumentationsqualität aufgrund von Dokumentationsfehlern von besonderer Häufigkeit oder Relevanz in der Zufallsstichprobe durch einen gezielten Datenabgleich erneut überprüft werden. Ziel des Auftrags zur Weiterentwicklung war es daher, die vorgegebenen Kriterien inhaltlich auszuarbeiten sowie das Verfahren zur Auslösung des gezielten

Datenabgleichs festzulegen. Der gezielte Datenabgleich ermöglicht damit, bei konkret vorliegenden Anhaltspunkten zu Defiziten in der Dokumentationsqualität diesen nachzugehen und auf die Notwendigkeit einer guten Dokumentationsqualität erneut hinzuweisen. Hierdurch wird auf Standortebene die Möglichkeit geschaffen, die Dokumentationsqualität gezielt zu fördern.

## 1.2 Verständnis und Vorgehen der Überarbeitung

Beratungsergebnis nach Vorstellung des Abschlussberichts (Fassung vom 26. Juli 2018) war wie o. g. eine Überarbeitung in einem mehrstufigen Vorgehen. Hierzu werden nachfolgend alle Gegenstände der Beauftragung genannt und das Verständnis zur Überarbeitung nach erfolgter Beratung dargelegt (oder die bereits erfolgten und beschlossenen Überarbeitungsschritte genannt). Insbesondere die nachfolgend aufgeführten Auftragsgegenstände Nr. 1, 2, 4 und 5 sollten erneut überprüft und adaptiert werden.

Gegenstand der Beauftragung vom 17. November 2017 war die:

1. *Entwicklung von Kriterien für den Anhaltspunkt „wenn bei einem Krankenhausstandort im Vorjahr Dokumentationsfehler in besonderer Häufigkeit oder Relevanz festgestellt wurden“ [...].*
2. *Entwicklung von Kriterien für den Anhaltspunkt „bei wiederholten rechnerischen Auffälligkeiten in der Statistischen Basisprüfung“ [...].*
3. *Entwicklung einer Methodik zur Prüfung nicht übermittelter Fälle [...].*
4. *Entwicklung von Kriterien zu „erheblichen Dokumentationsmängeln“ [...].*
5. *Entwicklung von Bewertungskriterien, die eine Kategorisierung der Ergebnisse des Datenvalidierungsverfahrens für den Bericht nach § 15 Abs. 2 QSKH-RL in Bezug auf die Übereinstimmungsrate, Sensitivität und Spezifität ermöglichen. Die bisher verwendeten Kategorien zur Datenvalidität „verbesserungsbedürftig“, „gut“ und „hervorragend“ sind zu überarbeiten.*
6. *Entwicklung eines Formats für den bundeseinheitlichen Prüfbericht [...].*
7. *Vorschläge für die Veröffentlichung der Ergebnisse des Datenvalidierungsverfahrens im Qualitätsbericht nach § 136b Abs. 1 Satz 1 Nr. 3 SGB V.*
8. *Entwicklung eines Erfassungstools für die prüfenden Stellen für den Datenabgleich vor Ort.*

### Verständnis zu den Auftragsgegenständen Nr. 1, 2 und 4

Folgende Aspekte wurden zu diesen Punkten in der Beauftragung des G-BA konkreter festgelegt:

- Zu.1.: *„Das IQTIG soll Kriterien vorlegen, wie „besondere Häufigkeit“ und „Relevanz“ in Bezug auf Ergebnisse des Datenabgleichs im Vorjahr zu definieren sind. Die Kriterien zur Häufigkeit sollen berücksichtigen, z. B. bei wie vielen Datenfeldern und bei wie vielen Fällen in Bezug auf ein oder mehrere Datenfelder Abweichungen in einer Einrichtung festgestellt wurden [...].*

*Die Kriterien zur Relevanz sollen berücksichtigen, ob beispielsweise Datenfelder von Fehldokumentationen betroffen sind, die besonders relevante Ereignisse erfassen (z. B. Todesfälle oder gravierende Komplikationen) oder die in besonderem Maße Einfluss auf das Ergebnis eines Indikators haben. Die Kriterien zu Häufigkeit und Relevanz können ggf. kombiniert werden“ (G-BA 2017).*

*Zu.2.: „Das IQTIG soll Kriterien vorlegen, wie wiederholt zu definieren ist. Die Auffälligkeitskriterien können die Anzahl der Auffälligkeiten und die Anzahl von Wiederholungen in einer Einrichtung berücksichtigen (z. B. „auffällig bei mindestens x Auffälligkeitskriterien in mindestens x aufeinanderfolgenden Jahren“). Bei der Festlegung sollen vorhandene Daten aus den Berichten zur Datenvalidierung für die Modellierung genutzt werden, um einen effizienten Vorschlag entwickeln zu können“ (G-BA 2017).*

*Zu 4.: „Es ist vorgesehen, dass ein Krankenhausstandort im Rahmen des Strukturierten Dialogs gegenüber dem nach § 14 QSKH-RL verantwortlichen Gremium bei Vorliegen erheblicher Dokumentationsmängel benannt werden soll. Eine Bezugnahme kann hier ggf. auch auf die Kriterien zu „besonderer Häufigkeit“ und „Relevanz“ erfolgen“ (G-BA 2017).*

**Auftragsgegenstand Nr. 1** beinhaltet gemäß Verständnis des IQTIG konkrete Vorschläge ab welcher Ausprägung von Auffälligkeiten ein gezielter Datenabgleich im Folgejahr auf Basis der Ergebnisse des Stichprobenverfahrens mit Datenabgleich für das Kriterium zu Dokumentationsfehlern in besonderer Häufigkeit oder Relevanz erfolgen soll. In der Überarbeitung des Häufigkeits- und Relevanzkriteriums sollten Modellrechnungen die Entscheidungsfindung zu Grenzwertfestlegungen einbezogen werden. Ebenso sollte die statistische Methodik genauer erläutert werden, sowie eine Präzisierung der Definition der Kriterien vorgenommen werden. Zur Verständnisklärung beider Kriterien wurde jeweils ein Expertengremium vom IQTIG einberufen.

In einem ersten Schritt (Addendum I) hat das IQTIG daraufhin die entsprechenden Überarbeitungen für das Häufigkeitskriterium vorgenommen und den Prozess für die Auslösung eines gezielten Datenabgleichs definiert. Das vom IQTIG vorgeschlagene Kriterium zur Häufigkeit wurde daraufhin vom G-BA am 6. März 2019 beschlossen.

In einem weiteren Schritt wurde festgelegt, was unter relevanten Fehldokumentationen zu verstehen ist und ab welcher Ausprägung von Auffälligkeiten ein gezielter Datenabgleich für dieses Kriterium erfolgen soll (Addendum II). Zusätzlich zum Auftrag versteht das IQTIG unter dem Aspekt Relevanz die Unterdokumentation eines Sentinel Events, für welches gemäß QSKH-RL ebenfalls ein gezielter Datenabgleich bei dessen Eintreten erfolgen muss. Aus diesem Grund hat das IQTIG das Sentinel-Event-Kriterium in die Ausarbeitung des Relevanzkriteriums aufgenommen und für die Prozessabläufe zur erneuten Überprüfung mitberücksichtigt. Das vom IQTIG vorgeschlagene Kriterium zur Relevanz wurde daraufhin in Teilen vom G-BA am 4. Dezember 2019 beschlossen.

**Auftragsgegenstand Nr. 2** beinhaltet gemäß Verständnis des IQTIG die Festlegung, ab wann rechnerische Auffälligkeiten der Auffälligkeitskriterien der Statistischen Basisprüfung einen gezielten Datenabgleich auslösen sollen. Hierfür sollte das ehemals vorgeschlagene Verfahren vereinfacht werden und von einer Mindestregelung abgesehen werden. Nach Verständnis des IQTIG wird deshalb eine Betrachtung je Auffälligkeitskriterium gewählt. Das dafür am 30. September 2019 dem G-BA eingereichte Addendum III wurde in diesen Bericht integriert und nochmals angepasst.

**Auftragsgegenstand Nr. 4** beinhaltet gemäß Verständnis des IQTIG konkrete Vorschläge, was unter erheblichen Dokumentationsmängeln zu verstehen ist, die eine sofortige Entanonymisierung gegenüber dem verantwortlichen Lenkungsgremium erfordern. In der Überarbeitung sollten ebenfalls Modellrechnungen einbezogen werden. Das IQTIG hat für die Festlegung erheblicher Dokumentationsmängel zunächst einen Bezug zum Häufigkeitskriterium hergestellt (Addendum I). Das vom IQTIG vorgeschlagene Kriterium zu erheblichen Dokumentationsmängeln als erheblich häufige Dokumentationsfehler wurde daraufhin vom G-BA am 6. März 2019 beschlossen.

Das IQTIG versteht diese drei Gegenstände der Beauftragung darüber hinaus auch als Auftrag, den prozessualen Rahmen als eine Art Gesamtkonzept festzulegen, in welchem die Abläufe definiert sind. Das Gesamtkonzept (Abbildung 1) gemäß Beauftragung und § 9 QSKH-RL sieht demnach vor, bei Vorliegen von Fehldokumentationen in besonderer Häufigkeit oder Relevanz, sowie bei Nicht-Dokumentation eines Sentinel Event und auch bei wiederholter rechnerischer Auffälligkeit in der Statistischen Basisprüfung einen gezielten Datenabgleich im Folgejahr auszulösen.

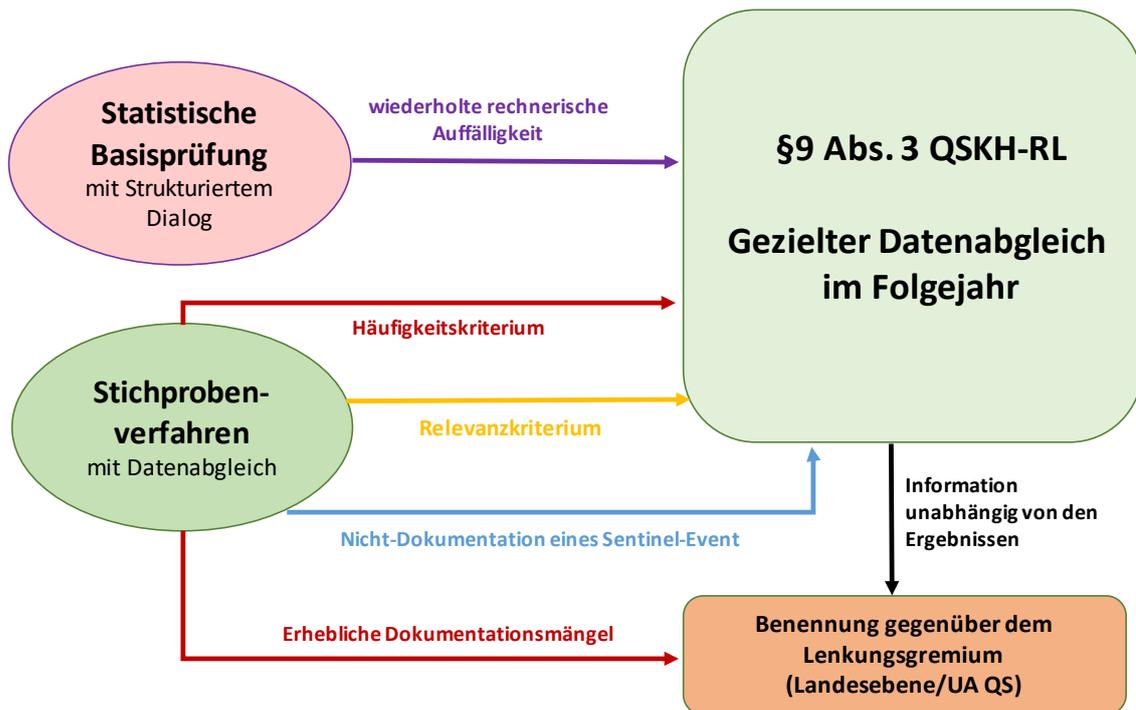


Abbildung 1: Darstellung der Datenvalidierung nach § 9 QSKH-RL: Wege der Krankenhausstandorte in die gezielte Stichprobe über Dokumentationsfehler sowie der Benennung gegenüber dem Lenkungsgremium/Unterausschuss Qualitätssicherung bei Dokumentationsfehlern von erheblicher Häufigkeit und nach gezieltem Datenabgleich

Für dieses Gesamtkonzept mit den einzelnen Wegen in den gezielten Datenabgleich, oder der Benennung gegenüber dem entsprechenden Lenkungsgremium, wurden die jeweiligen Kriterien vom IQTIG überarbeitet und in der Form von drei Addenda beim G-BA eingereicht und beraten. Diese wurden im hier vorliegendem Bericht in den jeweiligen Kapiteln integriert. Der Prozessablauf zu den Kriterien insgesamt wird anschließend gesondert zusammengefasst (vgl. Kapitel 6).

Für das IQTIG besteht zum Stichprobenverfahren mit Datenabgleich und zum gezielten Datenabgleich zusätzlich folgendes Grundverständnis:

Die nach § 108 SGB V zugelassenen Krankenhäuser sind grundsätzlich verpflichtet, die notwendigen Informationen für die Durchführung der in der QSKH-RL zur Sicherung der Qualität festgelegten QS-Verfahren zu dokumentieren (§4 QSKH-RL n. F.). Somit sind die Krankenhäuser dafür verantwortlich, welche Angaben sie in der QS-Dokumentation machen. Dies setzt eine Dokumentationspflicht der Behandlung voraus (§ 630f BGB) und bedeutet nach Einschätzung des IQTIG, dass alle für die QS relevanten Informationen aus der Patientenakte ersichtlich sein müssen. Mit Einführung des gezielten Datenabgleichs sollen die Krankenhausstandorte angehalten werden, die QS-Dokumentation auf Grundlage der Patientenakte korrekt durchzuführen, um im Regelfall nicht bei einem gezielten Datenabgleich erneut überprüft werden zu müssen. Wenn Krankenhausstandorte aufgrund der Ergebnisse des Datenabgleichs im Vorjahr einer erneuten Überprüfung im Folgejahr unterzogen werden, sollten in der dazwischenliegenden Zeit geeignete Verbesserungsmaßnahmen zur Dokumentationsqualität erfolgt sein. Eine Bewertung der

Ergebnisse des gezielten Datenabgleichs oder ggf. auch der Verbesserungsmaßnahmen aufgrund der Ergebnisse des Datenabgleichs im Vorjahr wird dem Gremium nach § 14 QSKH-RL überlassen.

### **Verständnis von Auftragsgegenstand Nr. 3**

Nicht übermittelte Fälle können nicht im Fallabgleich überprüft werden. Gemäß Anmerkungen des GKV-SV nach Einreichung des Abschlussberichts vom 26. Juli 2018 ginge es um die Identifizierung der nicht übermittelten Fälle, also um die Frage, durch wen und zu welchem Zeitpunkt und in welcher Form die Nacherfassung und Nachlieferung erfolgen solle. Da dieser Aspekt eine grundlegende Richtungsänderung bedeutet, nämlich eine Querverbindung zwischen Dokumentation und den darauf aufbauenden QI-Ergebnissen, sollte dieser Punkt nach Verständnis des IQTIG in einen größeren Kontext gestellt werden. Eine Insellösung für den gezielten Datenabgleich erscheint an dieser Stelle nicht sachgerecht, weshalb keine weiteren Anpassungen im hier vorliegenden Abschlussbericht vorgenommen wurden. Grundsätzlich bedeutet eine Querverbindung aufgrund von Fehldokumentationen oder Unterdokumentationen, dass die QI-Ergebnisse verzerrt sein können und theoretisch auch Konsequenzen folgen müssten. Eine Verknüpfung besteht – außer in plan. QI – bisher nicht für den Regelbetrieb. Eine grundlegende Verknüpfung der Dokumentationsqualität mit der medizinischen Qualität wäre begrüßenswert. In diesem Kontext wären dann auch nicht übermittelte Fälle relevant. Perspektivische Überlegungen sind in Kapitel 9 dargestellt.

### **Verständnis von Auftragsgegenstand Nr. 5**

Für die Ergebnisse des Stichprobenverfahrens mit Datenabgleich und die Ergebnisse des gezielten Datenabgleichs sollten Bewertungskriterien in Bezug auf die Übereinstimmungsrate, Sensitivität und Spezifität entwickelt werden. Kritik der zuvor vorgeschlagenen Bewertung „guter“ oder „defizitärer“ Dokumentationsqualität bestand vor dem Hintergrund der Festlegung unter Machbarkeitsgesichtspunkten und in Bezug auf die alleinige Betrachtung der Übereinstimmungsrate.

Überdies sollte die Bewertung nicht nur für die Standortebene vorgenommen werden, sondern eine ergänzende Bewertung auf Bundesebene erarbeitet und ein Bezug zur Sensitivität und Spezifität hergestellt werden. Die Begriffe sind im vorangestellten Glossar genauer definiert. Somit wird insgesamt zu diesem Auftragsgegenstand in der Überarbeitung berücksichtigt (vgl. Kapitel 7), dass ein induktiver Rückschluss von 1. der betrachteten Stichprobe (x Fälle eines Standortes gezogen aus allen behandelten Fälle im betrachteten Erfassungsjahr zum jeweiligen QS-Verfahren) auf den Krankenhausstandort insgesamt getroffen werden kann und 2. von allen überprüften Standorten (Stichprobe) auf alle datenliefernden Standorte des jeweiligen QS-Verfahrens bundesweit zur Dokumentationsqualität erzielt werden kann. Dies erfordert die explizite Berücksichtigung statistischer Unsicherheit bei der Bewertung der Dokumentationsqualität.

**Verständnis von Auftragsgegenstand Nr. 6**

Der bundeseinheitliche Prüfbericht steht in Zusammenhang mit Auftragsgegenstand Nr. 8, für welchen ein eigener Abschlussbericht am 31. Oktober 2018 eingereicht worden ist. Der Prüfbericht befindet sich bereits in Anwendung zum beschlossenen Häufigkeitskriterium. Eine Ergänzung des Prüfberichts für die Darstellung des Relevanzkriteriums wird nach Beschluss von diesem überprüft. Eine grundlegende Änderung des Formats des Prüfberichts ist jedoch nicht vorgesehen. Ergänzende Beschreibungen zum sich bereits in Anwendung befindenden Prüfbericht sind im hier vorliegenden Abschlussbericht unter Kapitel 10 vorgenommen worden.

**Verständnis von Auftragsgegenstand Nr. 7**

Vorschläge für die Veröffentlichung der Ergebnisse des Datenvalidierungsverfahrens im Qualitätsbericht nach § 136b Abs. 1 Satz 1 Nr. 3 SGB V wurden entsprechend der überarbeiteten Kriterien aus den Addenda I bis III sowie in Bezug auf die Bewertung der Dokumentationsqualität nach Auftragsgegenstand Nr. 5 angepasst. Hierbei wurden die genannten Limitationen in Abschnitt 5.5 zum Kriterium wiederholter rechnerischer Auffälligkeiten in der Statistischen Basisprüfung berücksichtigt und damit das Kriterium zur wiederholten rechnerischen Auffälligkeit im hier vorliegenden Veröffentlichungsvorschlag ausgeschlossen.

**Verständnis von Auftragsgegenstand Nr. 8**

Zu diesem Beauftragungspunkt ist ein eigener Abschlussbericht am 31. Oktober 2018 eingereicht worden. Das Erfassungstool befindet sich bereits in Anwendung. Es sind keine Anpassungen im hier vorliegenden Abschlussbericht vorgenommen worden.

## 2 Methodik zur Berechnung von Fehldokumentationsraten

In diesem Kapitel werden Grundlagen für die Berechnung von Fehldokumentationsraten auf Basis der Ergebnisse des gezielten Datenabgleichs gemäß § 9 QSKH-RL dargestellt. Auf ihnen basieren viele der in den folgenden Kapiteln dieses Berichts empfohlenen Bewertungskriterien. Da im Rahmen des gezielten Datenabgleichs i. d. R. keine Vollerhebungen, sondern Stichprobenprüfungen vorgenommen werden, wird hier ferner eine statistische Methodik definiert, wie anhand der Stichprobenerhebungen in den Datenvalidierungsverfahren nach QSKH-RL auf die Gesamtheit aller Fälle eines Krankenhausstandortes geschlossen werden kann.

Das in § 9 QSKH-RL beschriebene Stichprobenverfahren mit Datenabgleich basiert auf einer mehrstufigen Stichprobe (sogenannte Cluster-Stichprobe): In den jeweils zu untersuchenden QS-Verfahren wird zunächst eine einfache Zufallsstichprobe der Krankenhausstandorte gezogen (indirekte Verfahren: 5 % der Standorte pro Bundesland pro Leistungsbereich; direkte Verfahren: 5 % der Standorte pro Leistungsbereich, mind. 4 Standorte mit 40 Fällen). Anschließend wird für diese Krankenhausstandorte jeweils eine Stichprobe von 20 Fällen gezogen (falls weniger als 20 Fälle vorhanden sind: alle Fälle des Krankenhausstandortes). Für diese Fälle wird dann ein Datenabgleich durchgeführt, der Ergebnisse zu einer Vielzahl an Datenfeldern (i. d. R. alle QI-relevanten Datenfelder) liefert. Der Sachverhalt ist schematisch in Abbildung 2 dargestellt.

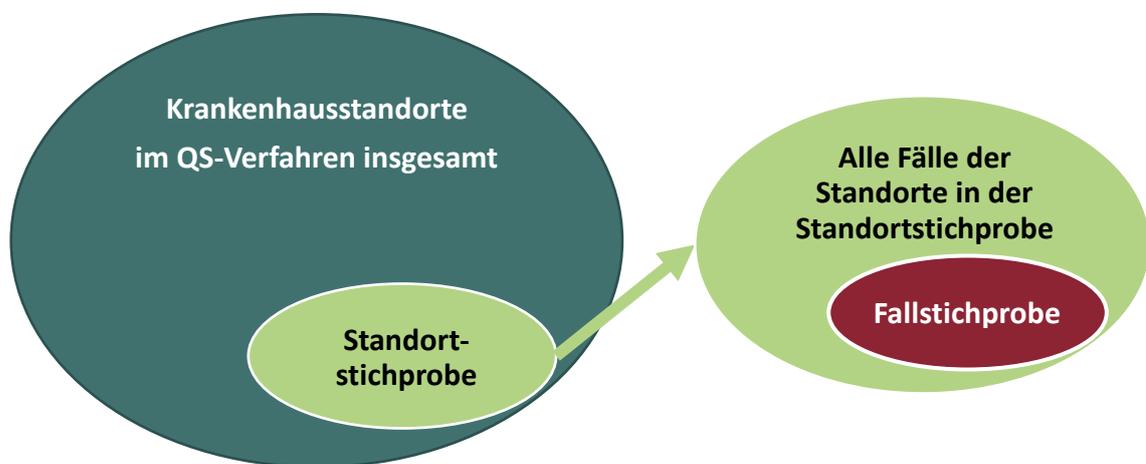


Abbildung 2: Schematische Darstellung der Stichprobe gemäß § 9 QSKH-RL

### Mathematische Notation

Die verwendeten Bestandteile des Datenvalidierungsverfahrens werden zwecks eindeutiger Beschreibung in diesem Kapitel – ergänzend zur textuellen Erläuterung – mittels einer mathematischen Notation beschrieben. Für das jeweils konkret untersuchte QS-Verfahren wird dabei im gesamten Kapitel die folgende grundlegende Notation verwendet:

- der Index  $i = 1, \dots, I$  bezeichnet einen von  $I$  untersuchten Krankenhausstandorten

- der Index  $j = 1, \dots, J_i$  bezeichnet die  $J_i$  Fälle des Krankenhausstandortes  $i$ , wobei die ersten  $J_i^* \leq J_i$  dieser Fälle Teil der Stichprobe sind, die im Datenabgleich überprüft wurden (in der Regel  $J_i^* = 20$ )
- der Index  $k = 1, \dots, K$  bezeichnet die  $K$  Datenfelder, die jeweils für die einzelnen Fälle untersucht werden. Tatsächlich geht aufgrund von (unten spezifizierten) Sonderregeln nicht jedes Datenfeld  $k$  für jeden Fall  $j$  in die Berechnungen ein. In dieser Hinsicht stellt die gewählte Notation eine Vereinfachung dar: Es wird angenommen, dass alle Datenfelder für alle Fälle relevant sind. Eine komplexere Darstellung, bei der nur relevante Datenfelder berücksichtigt werden, würde voraussetzen, dass die Summen jeweils nur über die Konstellationen von Fällen und Datenfeldern summieren, die in die Berechnung eingehen. Dies könnte z. B. mittels einer von  $i, j$  und  $k$  abhängigen Indikatorfunktion formuliert werden. Eine solche Vorgehensweise würde allerdings die Komplexität der Darstellung erhöhen, weshalb hier darauf verzichtet wird. Die Zahl der Datenfelder, die für den Fall  $j$  von Krankenhausstandort  $i$  in eine Berechnung eingehen, wird im jeweiligen Kontext mit  $K_{ij}$  bezeichnet. Dabei ist  $K = \max_{i,j} K_{ij}$ .

Wird nur ein einzelner Krankenhausstandort in einer Beschreibung betrachtet, oder werden alle Fälle als Teil eines einzelnen Krankenhausstandortes aufgefasst, so wird der Index  $i$  der Einfachheit halber weggelassen. In diesem Abschnitt grundsätzlich unberücksichtigt bleibt, dass einige der überprüften Fälle beispielsweise mehrere Prozedurbögen haben können. Eine Erweiterung der Methodik auf diesen Fall ist aber möglich. Abbildung 3 veranschaulicht einen Teil der eingeführten Notation in der schematischen Darstellung von Abbildung 2.

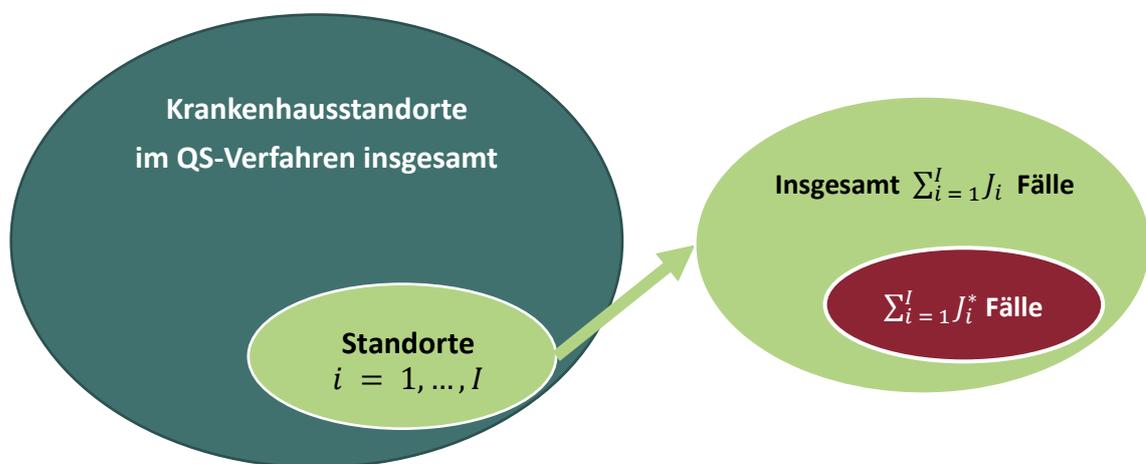


Abbildung 3: Veranschaulichung der mathematischen Notation in Bezug zu Abbildung 2

## 2.1 Berechnung von Fehldokumentationsraten

Die Berechnung von Fehldokumentationsraten kann sowohl **auf Ebene eines Datenfeldes** als auch **über alle untersuchten Datenfelder eines QS-Verfahrens hinweg** erfolgen. Dabei kann die Fehldokumentationsrate jeweils in Bezug auf unterschiedliche Gruppierungseinheiten betrachtet werden: Für einen einzelnen **Krankenhausstandort**, auf **Landesebene**, oder auf **Bundesebene** (und damit für das QS-Verfahren insgesamt).

Die verschiedenen Varianten werden in diesem Abschnitt näher erläutert. Insbesondere wird für die jeweiligen Situationen die statistische Methodik zur Berechnung von Punktschätzern und zugehörigen Vertrauensintervallen diskutiert, die gemeinsam für eine Ergebnisberechnung und eine Bewertung erforderlich sind, welche die Stichprobenunsicherheit berücksichtigt (vgl. auch Abschnitt 5.1 im Vorbericht zur Weiterentwicklung des Strukturierten Dialogs in Krankenhäusern (IQTIG 2019)).

### 2.1.1 Fehldokumentationsrate eines Datenfeldes

Die in einer Fallstichprobe beobachtete **Fehldokumentationsrate** für ein Datenfeld  $k$  ( $k = 1, \dots, K$ ) berechnet sich grundsätzlich als Quotient aus der Anzahl an Fehldokumentationen unter allen  $J^*$  betrachteten Abgleichsergebnissen für dieses Datenfeld in der Stichprobe (d. h. i. d. R. unter allen überprüften Fällen), die alle als aus einem Krankenhausstandort stammend betrachtet werden:

$$F_k^{\text{Stichprobe}} := \frac{\text{Anzahl Fehldokum. von Feld } k \text{ i. d. Stichprobe}}{\text{Anzahl betr. Ergebn. für Feld } k \text{ i. d. Stichprobe}} = \frac{\sum_{j=1}^{J^*} x_{jk}}{J^*} \quad (1)$$

Hierbei ist  $x_{jk} = 1$ , falls das Datenfeld  $k$  für Fall  $j$  fehlerhaft dokumentiert wurde und  $x_{jk} = 0$  andernfalls. Die Anzahl  $J^*$  aller für das Datenfeld  $k$  betrachteter Abgleichsergebnisse entspricht in der Regel der Anzahl an betrachteten Fällen in der Stichprobe mit der folgenden Ausnahme:

Von anderen Datenfeldern abhängige Felder werden nur dann gezählt, wenn sie auch tatsächlich ausfüllbar waren, d. h. dann, wenn das Kriterium des übergeordneten Feldes („Filterfeld“) für den jeweils betrachteten Fall erfüllt ist (in der QS-Dokumentation oder laut Patientenakte). Beispiel: Wenn das übergeordnete Feld Komplikationen = nein korrekt dokumentiert wurde, wird der Wert eines davon abhängigen (hier nicht ausfüllbaren) Feldes „Knochenfraktur“ nicht mitgezählt.

Für Felder mit Mehrfachauswahl („Mehrfachfelder“) gilt zusätzlich die folgende Regel:

Bei Mehrfachfeldern wird statt einer dichotomen Einstufung  $x_{jk} = 0$  (Übereinstimmung) oder  $x_{jk} = 1$  (Fehldokumentation) der Anteil der übereinstimmenden Werte in der Mehrfachauswahl herangezogen:  $x_{jk}$  berechnet sich dann als Wert zwischen 0 und 1 aus dem Anteil der nicht übereinstimmenden Kategorien in der Mehrfachauswahl unter allen unterschiedlichen ausgewählten Kategorien, die in der QS-Dokumentation oder in der Patientenakte für den betrachteten Fall dokumentiert wurden.

#### 2.1.1.1 Bei einem Krankenhausstandort

Die Fehldokumentationsrate eines Datenfeldes bei einem Krankenhausstandort wird bezogen auf die untersuchte Fallstichprobe (i. d. R. 20 Fälle) gemäß Formel (1) berechnet.

Die sich so ergebende Fehldokumentationsrate in der Stichprobe ist zunächst lediglich eine Aussage über die Korrektheit der Dokumentation des Datenfeldes in den (i. d. R. zwanzig) unter-

suchten Fällen. Sofern ausgehend hiervon auf die Fehldokumentationsrate des betreffenden Datenfeldes bei dem Krankenhausstandort insgesamt (d. h. bezogen auf alle Fälle des Krankenhausstandortes, für die das Datenfeld zu dokumentieren war) geschlossen werden soll, kann die beobachtete Fehldokumentationsrate in der (Zufalls-)Stichprobe im inferenzstatistischen Sinne als Schätzer für die (unbeobachtete) Fehldokumentationsrate des Krankenhausstandortes gesehen werden:

$$F_k^{\text{Krankenhaus}} := F_k^{\text{Stichprobe}}$$

Entspricht die Stichprobe der zu betrachteten Grundgesamtheit des Krankenhausstandortes (dies ist der Fall, wenn der Krankenhausstandort 20 oder weniger Fälle hat), so ist die Fehldokumentationsrate in der Stichprobe identisch mit der eigentlich interessierenden Fehldokumentationsrate des Krankenhausstandortes und es ist keine weitere Berücksichtigung von Unsicherheit notwendig. Ist die Stichprobe nur eine Teilmenge der Grundgesamtheit, kann ein  $(1-\alpha)\cdot 100\%$ -Vertrauensintervall genutzt werden, um die Sicherheit dieser Schätzung der Fehldokumentationsrate für die Gesamtpopulation aus der Fehldokumentationsrate in der Stichprobe zu charakterisieren. Je schmaler der Vertrauensbereich, desto belastbarer ist diese Schätzung. Dabei ist  $0 < \alpha < 1$  ein vorher festzulegendes Signifikanzniveau, z. B. wird  $\alpha = 0,05$  gewählt, werden einseitige 95 %-Vertrauensintervalle berichtet (IQTIG 2019, vgl. Abschnitte 2.1.5 sowie 5.5).

Da das Betrachten einer zufälligen Stichprobe von 20 Fällen aus den insgesamt  $J_i$  Fällen eines Krankenhausstandortes  $i$  dem einfachen Ziehen ohne Zurücklegen entspricht, kann eine hypergeometrische Verteilung zugrunde gelegt werden. Das IQTIG empfiehlt hier ein exaktes Vertrauensintervall der hypergeometrischen Verteilung mit mid-p Korrektur. Für eine detaillierte Beschreibung und Diskussion der Methodik siehe Abschnitt 2.2.1.

### 2.1.1.2 Für alle Krankenhausstandorte (bundesweit oder auf Landesebene)

Um für alle Krankenhausstandorte eines QS-Verfahrens (bundesweit oder in einem Bundesland) aggregiert eine Aussage über die Fehldokumentationsrate eines Datenfeldes zu machen, kann entweder eine Aussage über die mittlere Rate pro Krankenhausstandort (Mittelwert über die Standorte) oder über die mittlere Rate pro Fall (Mittelwert über die Fälle) gemacht werden. Da das Ziel ist, Rückschlüsse auf die Fehldokumentationen in den für ein QS-Verfahren ausgewerteten Daten als Ganzes zu treffen, ist in der Regel letzteres, also die mittlere Rate pro Fall, relevant. Dieser Wert entspricht auch dem sog. „Bundesdurchschnitt“ aus den Auswertungen der Qualitätsindikatoren zu den Verfahren nach QSKH-RL.

Prinzipiell wäre es möglich, für jeden untersuchten Krankenhausstandort den Wert  $F_k^{\text{Stichprobe}}$  aus Formel (1) zu berechnen und entweder die mittlere Rate der Krankenhausstandorte zu berechnen oder die Anzahlen in Zähler und Nenner von Formel (1) für alle Krankenhausstandorte aufzusummieren. Letzteres Vorgehen wurde in der Vergangenheit üblicherweise in der Datenvalidierung verwendet, um die Dokumentationsqualität in einem QS-Verfahren insgesamt zu beurteilen (IQTIG 2018a, vgl. Kapitel 3). Diese Betrachtung entspricht der Modellannahme, dass die vorliegende Fallstichprobe einer einfachen Zufallsstichprobe aus allen Fällen, die bundesweit

für das QS-Verfahren dokumentiert wurden, entspricht. Entsprechend könnte für die Berücksichtigung der Stichprobenunsicherheit wieder ein Vertrauensintervall aus der hypergeometrischen Verteilung (siehe Abschnitt 2.2.1) verwendet werden.

Diese Modellannahme ist jedoch nicht korrekt: Tatsächlich handelt es sich im vorliegenden Fall um eine Cluster-Stichprobe (Kauermann und Küchenhoff 2011), in der zunächst eine Zufallsstichprobe von Krankenhausstandorten gezogen wird und dann für jeden Krankenhausstandort aus allen Fällen, für die das Datenfeld  $k$  dokumentiert wurde, eine weitere Zufallsstichprobe von Fällen gezogen wird, siehe Abbildung 4. Die Fallstichprobe hat dabei i. d. R. einen Umfang von 20 Fällen – unabhängig davon, wie viele Fälle der Standort insgesamt im betrachteten QS-Verfahren insgesamt dokumentiert hat.

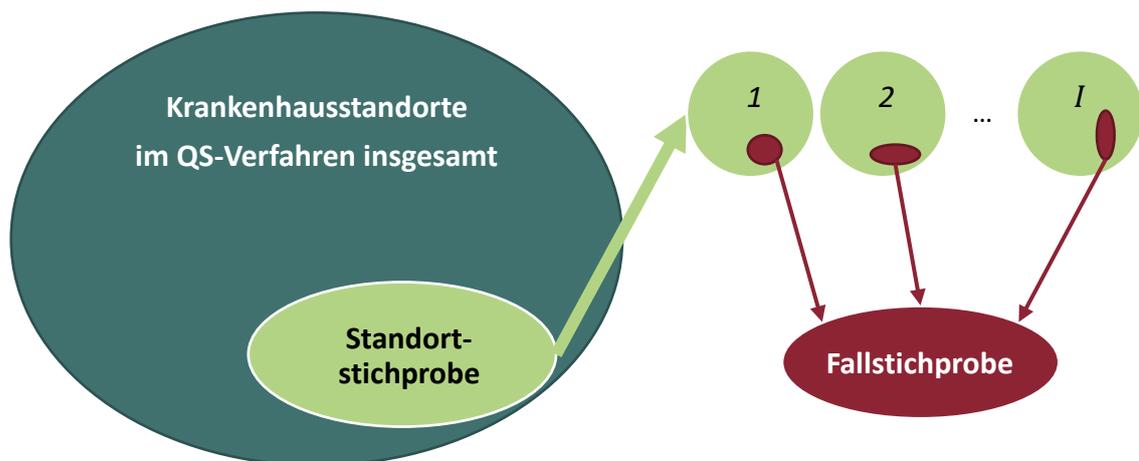


Abbildung 4: Die Stichprobe gemäß § 9 QSKH-RL als Cluster-Stichprobe: Für die  $I$  Standorte in der Standortstichprobe wird jeweils eine Fallstichprobe von i. d. R. 20 Fällen gezogen.

Die Problematik wird am besten an einem kleinen Beispiel deutlich: Betrachte zwei Krankenhausstandorte A und B. Standort A hat 20 Fälle, für die das betreffende Datenfeld immer korrekt dokumentiert wurde. Standort B hat 1.000 Fälle, wobei von den 20 untersuchten Fällen in 10 Fällen eine Fehldokumentation des Datenfeldes festgestellt wurde. Durch gemeinsame Betrachtung der beiden Stichproben wie oben beschrieben ergibt sich für die beiden Standorte zusammen eine Fehldokumentationsrate von  $(0 + 10)/(20 + 20) = 10/40 = 25\%$ .

Darauf basierend könnte (unter Angabe eines geeigneten Vertrauensintervalls) geschlossen werden, dass etwa 25 % aller  $20 + 1.000 = 1020$  Fälle der zwei Standorte fehlerhaft dokumentiert wurden. Dies lässt jedoch die fundamental unterschiedliche beobachtete Dokumentationsqualität in den Stichproben für A und B unberücksichtigt: Bei Standort A war das Datenfeld für alle 20 Fälle korrekt dokumentiert, während es bei B in der Hälfte der untersuchten 20 Fälle fehlerhaft dokumentiert war. Für Standort B ist vielmehr basierend auf den vorliegenden Informationen (unter Angabe eines Vertrauensintervalls) anzunehmen, dass das Datenfeld etwa für 500, nämlich die Hälfte seiner insgesamt 1.000 Fälle, fehlerhaft dokumentiert wurde, während für Standort A die Fehldokumentationsrate bezogen auf alle seine 20 Fälle bekannt ist. Dies ergibt eine geschätzte Fehldokumentationsrate von

$(0/20 \cdot 20 + 10/20 \cdot 1.000)/1020 = (0 + 500)/1020 \approx 49\%$  bezogen auf die insgesamt 1.020 Dokumentationen des Datenfeldes bei beiden Standorten. Die Ergebnisse für das genannte Beispiel sind übersichtshalber in Tabelle 1 dargestellt.

Tabelle 1: Schätzung der Fehldokumentationen insgesamt ohne und mit Fallzahlgewichtung

	Standort A	Standort B	Gesamt, ohne Fallzahl-Gewichtung	Gesamt, mit Fallzahl-Gewichtung
Fallzahl gesamt	20	1.000	1.020	1.020
Fehldokumentationen in der Stichprobe	0 / 20 0 %	10 / 20 50 %	10 / 40 25 %	-
Geschätzte Fehldokumentationen insgesamt	0 / 20 0 %	500 / 1.000 50 %	250 / 1.000 25 %	500 / 1.020 ca. 49 %

Den Überlegungen in diesem Beispiel folgend muss die unterschiedliche Gesamtfallzahl der betrachteten Krankenhausstandorte bei der Berechnung der Fehldokumentationsraten berücksichtigt werden. Sei für die Krankenhausstandorte  $i = 1, \dots, I$  jeweils  $J_{ik}$  die Anzahl der Fälle, in denen das Datenfeld  $k$  für die Berechnung der Fehldokumentationsrate gemäß den oben beschriebenen Regeln relevant war. Wie oben bezeichne  $J_{ik}^*$  die entsprechende Größe in der Stichprobe. Dann ergibt sich die geschätzte Fehldokumentationsrate des Datenfeldes  $k$  mit Berücksichtigung einer Fallzahl-Gewichtung als

$$F_k^{\text{Gesamt}} := \frac{1}{\sum_{i=1}^I J_i} \sum_{i=1}^I J_i \cdot F_{ik}^{\text{Stichprobe}} = \frac{1}{\sum_{i=1}^I J_i} \sum_{i=1}^I J_i \frac{\sum_{j=1}^{J_i^*} x_{ijk}}{J_i^*}. \quad (2)$$

(Levy und Lemeshow 2008). Dabei ist  $F_{ik}^{\text{Stichprobe}}$  die Fehldokumentationsrate des Datenfeldes  $k$  der Stichprobe des Krankenhausstandortes  $i$ , siehe Formel (1). Die Schätzung verwendet die Annahme, dass die Struktur der Fehldokumentationen in den in der Stichprobe untersuchten Krankenhausstandorten (insbesondere im Hinblick auf ihre Gesamtfallzahlen) repräsentativ für die Struktur der Fehldokumentationen aller Krankenhausstandorte im QS-Verfahren ist. Durch Nutzung bekannter Informationen (insbesondere über die Fallzahlen) zu den übrigen Krankenhausstandorten ist eine zukünftige Weiterentwicklung von Formel (2) denkbar.

Auch zu dieser geschätzten Fehldokumentationsrate des Datenfeldes insgesamt sollte stets ein Vertrauensintervall angegeben werden. Dieses kann ähnlich wie das in Abschnitt 2.2.2 diskutierte Vertrauensintervall aus einer geeigneten Varianzschätzung für diesen Fall einer Cluster-Stichprobe berechnet werden. Siehe dazu auch die Bemerkungen am Ende von Abschnitt 2.2.2.

### 2.1.2 Fehldokumentationsrate über alle Datenfelder hinweg

Um die Fehldokumentationsrate für alle untersuchten Datenfelder eines Falles bzw. für alle untersuchten Datenfelder aller untersuchter Fälle eines Krankenhausstandortes, oder mehrerer

Krankenhausstandorte gemeinsam, zu berechnen, kann theoretisch durch eine gewichtete Mittelung über alle Datenfelder  $k = 1, \dots, K$  auf die Formeln (1) und (2) zurückgegriffen werden.

Auf diese Weise blieben allerdings spezifikationsbedingte Abhängigkeitsstrukturen zwischen den Datenfeldern unberücksichtigt. Wurde beispielsweise ein übergeordnetes Feld „Komplikationen“ fehlerhaft dokumentiert, so sollen die Abgleichsergebnisse der davon abhängigen einzelnen Komplikationsinformationen nicht gewertet werden, da sie (falls tatsächlich Komplikationen vorlagen, diese aber nicht dokumentiert wurden) nicht ausfüllbar waren oder (falls keine Komplikationen vorlagen, aber Komplikationen dokumentiert wurden) nicht hätten ausgefüllt werden sollen. Insbesondere im Falle vieler unterschiedlicher zu dokumentierender Komplikationen ergäben sich erhebliche Verzerrungen dadurch, dass die meisten dieser abhängigen Felder trotz fehlerhafter Dokumentation des übergeordneten Feldes als korrekt dokumentiert gezählt würden (z. B. keine Sepsis, keine Knochenfraktur, keine sonstigen Komplikationen, etc.).

Daher werden abweichend von den in Abschnitt 2.1.1 definierten Regeln für die Fehldokumentationsraten von Datenfeldern die folgenden Regeln für die Fehldokumentationsraten aller untersuchter Datenfelder einer Stichprobe von Fällen festgelegt<sup>2</sup>:

1. Filterfelder und von ihnen abhängige Felder:

- Ist das Kriterium des Filterfeldes erfüllt (d. h. die abhängigen Felder sind ausfüllbar) und der Wert des Filterfeldes stimmt mit den QS-Daten überein, dann werden nur die Übereinstimmungen der abhängigen Felder (nicht das Filterfeld) bewertet.
- Ist das Kriterium des Filterfeldes nicht erfüllt (d. h. die abhängigen Felder sind nicht ausfüllbar) oder der Wert des Filterfeldes stimmt nicht mit den QS-Daten überein, dann wird nur die Übereinstimmung des Filterfeldes (nicht die abhängigen Felder) bewertet.

2. In Feldgruppen der Art, dass in einem Feld ein Wert abgefragt wird und in einem anderen „Wert nicht vorhanden“ angegeben werden kann, wird bei Abweichung stets nur *eine* Abweichung gezählt.

3. Mehrfachfelder: Statt einer dichotomen Einstufung wird hier der Anteil der übereinstimmenden Werte in der Mehrfachauswahl herangezogen.

4. Datenfelder, bei denen eine „mangelhafte Spezifikation“ vonseiten des IQTIG vorliegt, werden aus der Berechnung ausgeschlossen (siehe Abschnitt 3.3).

Die Gesamtfehldokumentationsrate für einen Krankenhausstandort – bezogen auf eine untersuchte Stichprobe, die als aus einem Krankenhausstandort stammend betrachtet wird – wird wie folgt berechnet:

$$F_{\bullet}^{\text{Stichprobe}} := \frac{\text{Anzahl Fehldokumentationen i. d. Stichprobe}}{\text{Anzahl betrachteter Ergebnisse i. d. Stichprobe}} = \frac{\sum_{j=1}^{J^*} \sum_{k=1}^{K_j} x_{jk}}{J^* \sum_{j=1}^{J^*} K_j} \quad (3)$$

Hierbei ist  $J^*$  die Anzahl der in der Stichprobe betrachteten Fälle. Der Index  $k$  läuft je Fall über diejenigen  $K_j$  Datenfelder, die gemäß der obigen Regeln für diesen Fall zu berücksichtigen sind.

<sup>2</sup> Die genannten Regeln entsprechen den in Addendum II vom 22. Oktober 2019 beschlossenen Regeln, der Vollständigkeit halber ergänzt wurde lediglich Punkt 1 (Feldgruppen).

Wie zuvor ist  $x_{jk} = 1$ , falls das Datenfeld  $k$  für Fall  $j$  fehlerhaft dokumentiert wurde und  $x_{jk} = 0$  andernfalls, mit der entsprechenden Sonderregel für Mehrfachfelder.

### 2.1.2.1 Bei einem Krankenhausstandort

Die Fehldokumentationsrate über alle betrachteten Datenfelder und alle überprüften Fälle hinweg wird gemäß Formel (3) berechnet. Die sich so ergebende Fehldokumentationsrate in der Stichprobe ist zunächst lediglich eine Aussage über die Datenqualität der (i. d. R. zwanzig) untersuchten Fälle. Sofern ausgehend hiervon auf die Fehldokumentationsrate des betreffenden Krankenhausstandortes bezogen auf alle von ihm dokumentierten Fälle im jeweiligen QS-Verfahren geschlossen werden soll, kann dieser Wert im inferenzstatistischen Sinne als Schätzer genutzt werden:

$$F_{\bullet}^{\text{Krankenhaus}} := F_{\bullet}^{\text{Stichprobe}}$$

Die Quantifizierung der Sicherheit der Schätzung kann erneut mithilfe eines Vertrauensintervalls erfolgen. Das Betrachten aller QI-relevanten Datenfelder einer zufälligen Stichprobe von 20 Fällen aus den insgesamt  $J_i$  Fällen eines Krankenhausstandortes  $i$ , wobei für jeden Fall alle gemäß der obigen Regeln zu betrachteten Datenfelder untersucht werden, entspricht einer sogenannten Cluster-Stichprobe (siehe auch Abbildung 4). Das IQTIG empfiehlt zur Berücksichtigung der statistischen Stichprobenunsicherheit für die Situation, in der die Stichprobe nur eine Teilmenge der Grundgesamtheit ist, ein logit-transformiertes Vertrauensintervall auf Basis einer Varianzschätzung für Cluster-Stichproben. Für eine detaillierte Beschreibung und Diskussion der Methodik siehe Abschnitt 2.2.2.

### 2.1.2.2 Für alle Krankenhausstandorte (bundesweit, auf Landesebene)

Ähnlich wie in Abschnitt 2.1.1.2 muss die Art der Stichprobenziehung und insbesondere die unterschiedliche Gesamtfallzahl der untersuchten Krankenhausstandorte berücksichtigt werden, wenn die Gesamtfehldokumentationsrate über alle Fallstichproben aller untersuchten Krankenhausstichproben gemeinsam berechnet werden soll.

Zur Illustration hilft wieder ein Beispiel: In einem hypothetischen QS-Verfahren mit nur zwei Datenfeldern wurde eine Stichprobe von zwei Krankenhausstandorten A und B gezogen, für die jeweils 20 Fälle untersucht werden. Standort A hat 20 Fälle und Standort B insgesamt 1.000 im betrachteten Erfassungsjahr behandelt. Feld 1 wurde bei Standort A immer korrekt dokumentiert, bei Standort B jedoch nur in 10 der 20 Fälle, die überprüft wurden. Feld 2 wurde von Standort A für 16 der 20 Fälle falsch dokumentiert, von Standort B in 2 der 20 überprüften Fälle.

Damit ergibt sich für Standort A eine Fehldokumentationsrate von  $(0 + 16)/(20 + 20) = 16/40 = 40\%$ , für Standort B  $(10 + 2)/(20 + 20) = 12/40 = 30\%$ . Unter Berücksichtigung der Fallzahlgewichtung ergibt sich für die Fehldokumentationsrate der beiden Standorte gemeinsam dann  $(16/40 \cdot 20 + 12/40 \cdot 1.000)/1020 = 308/1020 \approx 30,2\%$ . An diesem Ergebnis wird deutlich, dass die Fehldokumentationsrate von Standort B aufgrund der hohen Fallzahl mit einem viel stärkeren Gewicht in das Gesamtergebnis eingeht.

In Analogie zu Formel (2) ergibt sich die Gesamtfehldokumentationsrate im QS-Verfahren mit Berücksichtigung eines Fallzahl-Gewichts als

$$F_{\bullet}^{\text{Gesamt}} := \frac{1}{\sum_{i=1}^I J_i} \sum_{i=1}^I J_i \cdot F_i^{\text{Stichprobe}} = \frac{1}{\sum_{i=1}^I J_i} \sum_{i=1}^I J_i \cdot \frac{\sum_{j=1}^{J_i^*} \sum_{k=1}^{K_{ij}} x_{ijk}}{J_i^* \sum_{j=1}^{J_i^*} K_{ij}}. \quad (4)$$

Dabei ist  $F_i^{\text{Stichprobe}}$  die Gesamtfehldokumentationsrate des Krankenhausstandortes  $i$  in der Stichprobe, siehe Formel (3). Wieder wäre eine Erweiterung der Methodik durch Nutzung von Informationen über die Fallzahlen der übrigen Krankenhausstandorte, die nicht in der Stichprobe untersucht wurden, möglich.

Um die statistische Unsicherheit durch die mehrstufige Stichprobenziehung zu quantifizieren, sollte zu dieser geschätzten Gesamtfehldokumentationsrate stets auch ein Vertrauensintervall angegeben werden. Dies entspricht auch dem Vorgehen bei quantitativen Auswertungen und Einstufungen, wie es im Rahmenkonzept zur statistischen Auswertungsmethodik von Qualitätsindikatoren ausgeführt ist (IQTIG 2019). Dazu kann ähnlich wie in Abschnitt 2.2.2 diskutiert eine geeignete Varianzschätzung für diesen Fall einer mehrstufigen Cluster-Stichprobe genutzt werden. Siehe dazu auch die Bemerkungen am Ende von Abschnitt 2.2.2.

### 2.1.2.3 Nachträglicher Ausschluss von Datenfeldern

Aufgrund von Spezifikationsmängeln kann sich im Nachhinein die Notwendigkeit ergeben, einzelne Datenfelder aus der Übereinstimmungsrate auf Standortebene auszuschließen (siehe Abschnitt 3.3). Um eine umfangreiche und aus technischen Gründen evtl. aufwändige Neuauswertung der oben beschriebenen Regeln am IQTIG zu vermeiden, kann zum Ausschluss von Datenfeldern ersatzweise folgendes Vorgehen verwendet werden:

- Berechne Zähler und Nenner der Fehldokumentationsrate in der Stichprobe  $F = a/b$  wie oben ohne den Ausschluss der Datenfelder
- Berechne Zähler und Nenner der Fehldokumentationsrate in der Stichprobe  $F_k = a_k/b_k$  des Datenfeldes mit mangelhafter Spezifikation (kann direkt aus dem Erfassungstool übernommen werden)
- Die korrigierte Fehldokumentationsrate in der Stichprobe unter Ausschluss des Datenfeldes mit mangelhafter Spezifikation ist dann:

$$F_{\text{neu}} := \frac{a - a_k}{b - b_k}$$

Im Falle mehrerer Datenfelder mit mangelhafter Spezifikation wird das Vorgehen entsprechend wiederholt.

Das Ergebnis  $F_{\text{neu}}$  kann von der Fehldokumentationsrate abweichen, die sich aus der direkten Anwendung der Berechnungsregeln für Filterfelder, für abhängige Felder, usw. unter Ausschluss der Datenfelder ergeben würde. Dies ist dann der Fall, wenn unter den auszuschließenden Datenfeldern Filterfelder oder abhängige Felder sind, da für diese die Regeln zur Berechnung der

Übereinstimmungsrate auf Datenfeldebene von denjenigen für die Berechnung der Übereinstimmungsrate abweichen (beispielsweise werden Filterfelder bei der Bestimmung der Übereinstimmungsrate je Datenfeld immer mitgezählt).

### 2.1.3 Geschichtete Stichproben

In der Datenvalidierung wurden Fallstichproben in der Vergangenheit immer wieder statt als einfache Zufallsstichproben als geschichtete Stichproben gezogen, um sicherzustellen, dass bestimmte seltene Fallkonstellationen (z. B. Wechseleingriffe, Dekubitus Grad/Kategorie 4, sehr kleine Frühgeborene) in der Stichprobe vertreten sind. Dies ist in Abbildung 5 schematisch dargestellt. In diesem Fall ist diese Art der Stichprobenziehung bei der Berechnung der oben diskutierten Werte sowie der zugehörigen Vertrauensintervalle geeignet zu berücksichtigen. Eine detaillierte methodische Ausarbeitung hierzu ist bisher nicht erfolgt.

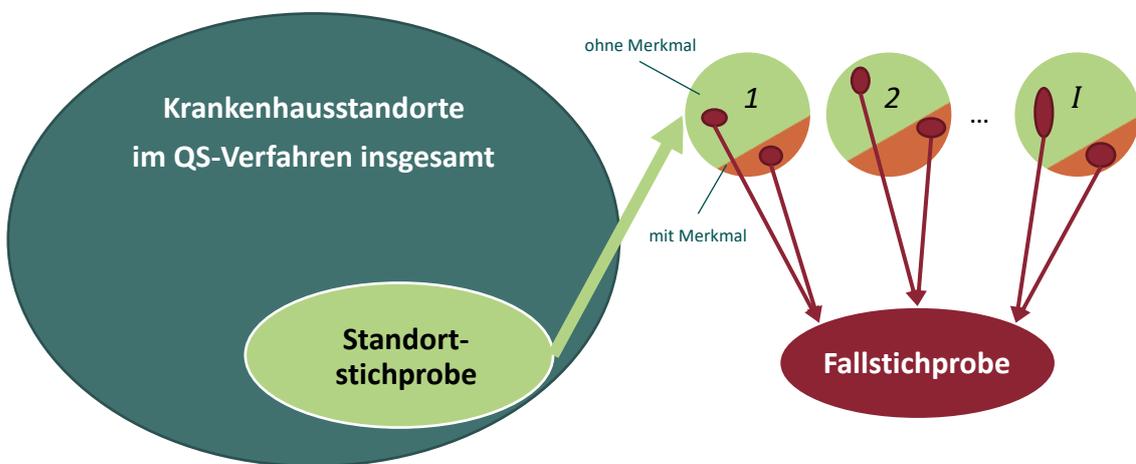


Abbildung 5: Geschichtete Stichprobe: Für jeden der I Krankenhausstandorte werden separat sowohl Fälle mit einem bestimmten Merkmal (orange) als auch Fälle ohne dieses Merkmal (grün) gezogen.

### 2.1.4 Zusammenfassung

Die in diesem Abschnitt beschriebenen Formeln sind in Tabelle 2 noch einmal zusammengefasst.

Tabelle 2: Formeln für die Berechnung der Fehldokumentationsrate eines Datenfeldes und der Gesamtfehldokumentationsrate

	Ein Krankenhausstandort	Alle Krankenhausstandorte (mit Fallzahl-Gewichtung)
Fehldokumentationsrate eines Datenfeldes	$\frac{\sum_{j=1}^{J_i^*} x_{ijk}}{J_i^*}$	$\frac{1}{\sum_{i=1}^I J_i} \sum_{i=1}^I J_i \frac{\sum_{j=1}^{J_i^*} x_{ijk}}{J_i^*}$
	Vertrauensintervall <sup>3</sup> : Hypergeometrisch exakt mit mid-p-Korrektur (Abschnitt 2.2.1)	Vertrauensintervall: Basierend auf Varianzschätzung für Cluster-Stichproben (Abschnitt 2.2.2)

<sup>3</sup> Ein Vertrauensintervall wird hier nur dann verwendet, wenn die Stichprobe nur einen Teil der Fälle des Krankenhausstandortes umfasst. Anderenfalls ist keine Stichprobenunsicherheit zu berücksichtigen.

	Ein Krankenhausstandort	Alle Krankenhausstandorte (mit Fallzahl-Gewichtung)
Fehldokumentationsrate gesamt	$\frac{\sum_{j=1}^{J^*} \sum_{k=1}^{K_j} x_{jk}}{J^* \sum_{j=1}^{J^*} K_j}$ <p>Vertrauensintervall<sup>3</sup>: Basierend auf Varianzschätzung für Cluster-Stichproben (Abschnitt 2.2.2)</p>	$\frac{1}{\sum_{i=1}^I J_i} \sum_{i=1}^I J_i \cdot \frac{\sum_{j=1}^{J_i^*} \sum_{k=1}^{K_{ij}} x_{ijk}}{J_i^* \sum_{j=1}^{J_i^*} K_{ij}}$ <p>Vertrauensintervall: Basierend auf Varianzschätzung für Cluster-Stichproben (Abschnitt 2.2.2)</p>

## 2.2 Statistische Methodik (Vertrauensintervalle)

In diesem Abschnitt wird die methodische Entwicklung der Vertrauensintervalle für die Schätzgrößen aus Abschnitt 2.1 beschrieben. Grundsätzlich wird für die statistische Datenauswertung an dieser Stelle eine *enumerative* Herangehensweise im Sinne der Diskussion in IQTIG (2019) empfohlen, da es hier primär um die Beurteilung der Korrektheit der für ein gegebenes QS-Verfahren in einem gegebenen Jahr gelieferten Datensätze geht (insgesamt oder für einen Krankenhausstandort). Damit dienen die zu definierenden Vertrauensintervalle nur der Berücksichtigung der sich aus der Stichprobenziehung ergebenden statistischen Unsicherheit. Eine weitergehende Inferenz, die Schlüsse auf den Dokumentationsprozess als solchen ermöglichen würde, erfolgt nicht. Einzelne der in den späteren Kapiteln diskutierten Maßnahmen zielen allerdings auch auf die Verbesserung der Dokumentationsqualität bei den Leistungserbringern im Sinne einer Förderung des Prozesses. Für diese Zwecke kann in Zukunft abweichend auch eine analytische Herangehensweise gewählt werden.

Für alle in diesem Bericht vorgeschlagenen Einstufungsverfahren (z. B. Häufigkeitskriterium, Relevanzkriterium, Bewertung der Dokumentationsqualität als unzureichend) wird ein Signifikanzniveau von  $\alpha = 0,05$  gewählt. Da es sich jeweils um *einseitige* Hypothesentests handelt, entspricht dies einem Abgleich der unteren Grenze eines *zweiseitigen* 90 %-Vertrauensintervalls mit dem jeweils festgelegten Referenzwert. Dieses 90 %-Vertrauensintervall soll auch regelhaft für die Berichterstattung von Ergebnissen (z. B. Fehldokumentationsraten bzw. Übereinstimmungs-raten) genutzt werden, um aufgrund der Dualität von statistischen Tests und Vertrauensintervallen eine einheitliche Darstellung zu erhalten. Abschnitt 2.2.3 ordnet die verwendete Methodik in das in IQTIG (2019) vorgeschlagene statistische Rahmenkonzept ein.

Insgesamt ist der vorliegende Abschnitt anspruchsvoller statistischer Natur und setzt fortgeschrittene Kenntnisse inferenzstatistischer Methoden und statistischer Stichprobentheorie voraus. Die genaue Vorgehensweise ist bei der Berechnung wichtig, da auf Basis der Grenzen dieser Vertrauensintervalle bei der Bewertung der Ergebnisse des Stichprobenverfahrens mit Datenabgleich die Entscheidung über eine quantitative Auffälligkeit getroffen wird (vgl. Kapitel 3, 4 und 7 in diesem Bericht sowie die Abschnitte 5.1.4 bzw. 5.3 ff. in IQTIG (2019)). Der Abschnitt ist somit wichtiger Bestandteil der fachlich fundierten Bearbeitung und transparenten Darstellung der statistischen Auswertungsmethodik bei den komplexen Datenerhebungsmechanismen der Datenvalidierung, kann jedoch von statistischen Laien übergangen werden. Die sich

aus diesem Abschnitt ergebenden methodischen Empfehlungen für die zu verwendenden Vertrauensintervalle wurden oben bereits an den betreffenden Stellen erwähnt und in Tabelle 2 zusammengefasst.

### 2.2.1 Vertrauensintervalle für einfache Zufallsstichproben ohne Zurücklegen

Die Betrachtung der Fehldokumentationsrate eines einzelnen Datenfeldes  $k$  (Notation wie in Abschnitt 2.1) bei einem einzelnen Krankenhausstandort  $i$  durch Untersuchung einer Fallstichprobe von  $n := J_i^*$  Fällen der insgesamt  $N := J_i$  Fälle des Standortes im betrachteten QS-Verfahren insgesamt kann als einfache Zufallsstichprobe aufgefasst werden. Im Stichprobenverfahren mit Datenabgleich wird für das Datenfeld  $k$  die Anzahl  $m \leq n$  der Fälle in der Stichprobe, für die es fehlerhaft dokumentiert wurde, ermittelt.

In der Praxis sind dabei die in Abschnitt 2.1.1 beschriebenen Zusatzregeln für abhängige Felder und Mehrfachfelder zu beachten. Letztere führen dazu, dass die Zahl  $m$ , wenn  $k$  ein Mehrfachfeld ist, u. U. nicht ganzzahlig ist. Für Mehrfachfelder ist an dieser Stelle also eine gesonderte Betrachtung notwendig und die hier vorgeschlagene Methodik ist entsprechend zu erweitern.

In der vereinfachten Notation ist  $m/n$  die Fehldokumentationsrate des Datenfeldes *in der Stichprobe*. Aus  $N$ ,  $n$  und  $m$  soll auf die Anzahl  $M$  aller Fehldokumentationen des Datenfeldes bezogen auf alle Fälle des Krankenhausstandortes im QS-Verfahren und damit auf die *Fehldokumentationsrate*  $M/N$  des Datenfeldes insgesamt geschlossen werden, siehe auch Abschnitt 2.1.1.1.

Dies geschieht, indem basierend auf  $N$ ,  $n$  und  $m$  ein Vertrauensintervall  $[\widetilde{U}_{N,n,m}, \widetilde{O}_{N,n,m}]$  für  $M$  bestimmt wird. Daraus ergibt sich durch Division durch den bekannten Wert  $N$  ein Vertrauensintervall  $[U_{N,n,m}, O_{N,n,m}]$  für die Gesamtfehldokumentationsrate  $M/N$ .

Sei  $X = m$  die Zufallsvariable, die die Anzahl der Fehldokumentationen  $m$  des Datenfeldes in der Stichprobe beschreibt. Da die  $n$  zu überprüfenden Fälle als einfache Zufallsstichprobe ohne Zurücklegen aus den  $N$  dokumentierten Fällen des Krankenhausstandortes ausgewählt werden, ist  $X \sim \text{Hyp}(N, M, n)$  hypergeometrisch verteilt. Für den Fall von geschichteten Stichproben, wie sie teilweise in der Datenvalidierung üblich sind, siehe die Bemerkung in Abschnitt 2.1.2.3.

Für festes  $N$  und  $n$  und ein gegebenes Ergebnis  $X = m$  der Überprüfung im Datenabgleich soll ein Vertrauensintervall mit Konfidenzniveau  $\alpha$  für  $M$  bzw.  $M/N$  zu definiert werden. Folgende Varianten wurden untersucht:

1. Das exakte 95 %-Vertrauensintervall, das durch

$$\begin{aligned} \widetilde{U}_{N,n,m} &:= \underset{M}{\operatorname{argmax}} P(X < m; N, M - 1, n) \geq 1 - \alpha/2, \\ \widetilde{O}_{N,n,m} &:= N - \widetilde{U}_{N,n-m,m} \end{aligned}$$

(für  $m = 0$  setze  $\widetilde{U}_{N,n,0} := 0$ ) gegeben ist (Wang 2015). Dabei bezeichnet  $P$  die Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung. Das so definierte Vertrauensintervall ist nicht notwendiger Weise minimal, die beiden einseitigen Intervalle  $[\widetilde{U}_{N,n,m}, 1]$  und  $[0, \widetilde{O}_{N,n,m}]$  sind jedoch per Konstruktion die kleinsten exakten einseitigen Intervalle mit Konfidenzniveau  $\alpha/2$ . Eine Division durch  $N$  liefert wie oben beschrieben ein Vertrauensintervall  $[U_{N,n,m}, O_{N,n,m}]$  für  $M/N$ .

2. Variante 2 mit mid-p-Korrektur:

$$\widetilde{U}_{N,n,m} := \operatorname{argmax}_M P(X < m; N, M - 1, n) + \frac{1}{2} P(X = m; N, M - 1, n) \geq 1 - \alpha/2,$$

$$\widetilde{O}_{N,n,m} := N - U_{N,n-m,m}$$

(für  $m = 0$  setze  $U_{N,n,m} := 0$ ). Durch den Korrekturterm kann der Problematik vorgebeugt werden, dass exakte Vertrauensintervalle oft sehr konservativ sind.

3. Die folgende Approximation mithilfe der  $t$ -Verteilung unter Verwendung einer Endlichkeitskorrektur (Kauermann und Küchenhoff 2011):

$$\frac{U_{N,n,m}}{N} := \frac{m}{n} - t_{1-\alpha/2, n-1} \cdot \sqrt{\frac{\frac{m}{n} \cdot \left(1 - \frac{m}{n}\right)}{n-1} \cdot \frac{N-n}{N}} \quad (3)$$

Hierbei ist  $t_{1-\alpha/2, n-1}$  das  $(1-\alpha/2)$ -Quantil der  $t$ -Verteilung mit  $n - 1$  Freiheitsgraden.

Abbildung 6 zeigt die Überdeckungswahrscheinlichkeiten der drei verschiedenen Vertrauensintervalle unter Annahme einer hypergeometrischen Verteilung für  $\alpha = 0,05$  (Wang 2015). Die Vertrauensintervalle aus Variante 1 weisen per Konstruktion immer eine Überdeckungswahrscheinlichkeit von  $\geq 95\%$  auf und sind damit wie erwartet sehr konservativ, d. h. sie sind sehr breit. Die approximativen Vertrauensintervalle aus Variante 3 sind dagegen insbesondere für sehr kleine und sehr große tatsächliche Fehldokumentationsraten in der Grundgesamtheit deutlich zu schmal. Dieser Effekt ist für  $N = 100$  noch einmal größer als für  $N = 25$ . Für  $N = 100$  weisen die mid-p-korrigierten Intervalle aus Variante 2 Überdeckungswahrscheinlichkeiten im Bereich von  $95\%$  auf, wobei auch diese für kleine und große Werte von  $M/N$  noch recht konservativ sind. Für  $N = 25$  ist diese Variante zwar konservativ, die Überdeckungswahrscheinlichkeiten sind aber im Vergleich mit den anderen Varianten akzeptabel. Die hohe Überdeckungswahrscheinlichkeit bei den beiden verteilungsbasierten Varianten ist teilweise auf die diskrete Natur des Problems zurückzuführen. Weiterhin spielt eine Rolle, dass diese Varianten per Konstruktion nur als einseitige Vertrauensintervalle minimal sind. Auf Grundlage dieser Ergebnisse empfiehlt das IQTIG für einfache Zufallsstichproben die Verwendung der exakten Vertrauensintervalle mit mid-p-Korrektur aus Variante 2.

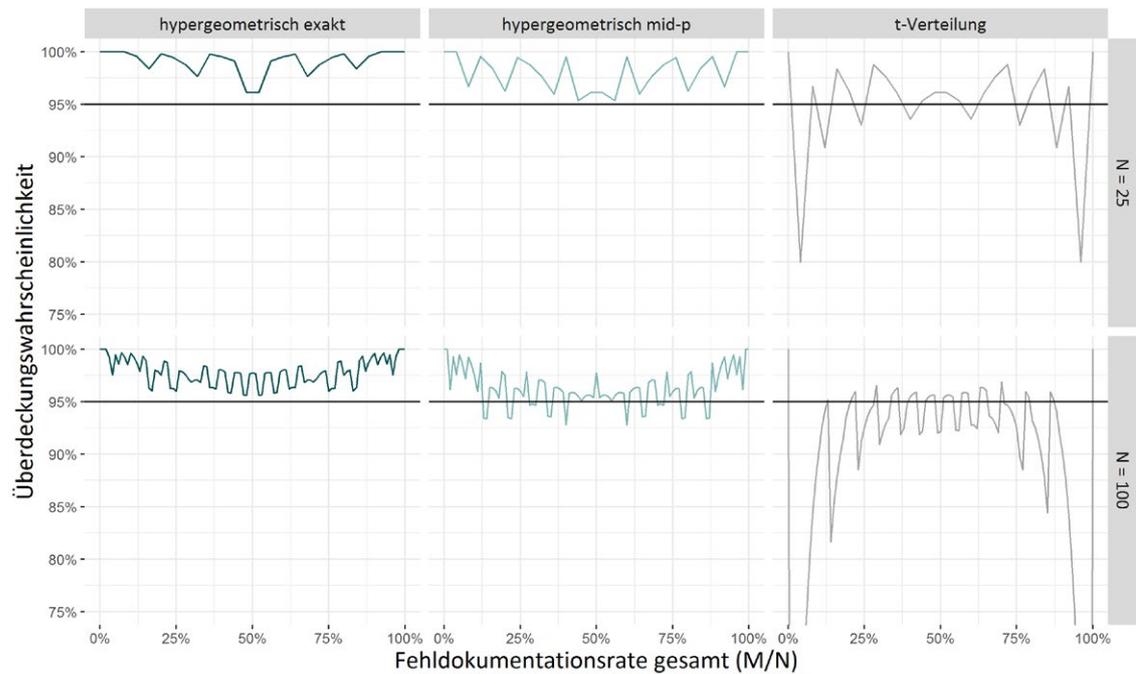


Abbildung 6: Überdeckungswahrscheinlichkeiten für die drei diskutierten Vertrauensintervalle für Fehldokumentationsraten eines Datenfeldes bei einem Krankenhausstandort für  $n = 20$  und  $N = 25$  sowie  $N = 100$ .

### 2.2.2 Vertrauensintervalle für Cluster-Stichproben

Wird bei einer Stichprobenziehung aus einer festgelegten Grundgesamtheit zunächst eine Gruppe von Einheiten (Cluster) als einfache Zufallsstichprobe (ohne Zurücklegen) gezogen, um aus diesen Einheiten erneut eine Stichprobe von Unter-Einheiten zu ziehen (oder eine Vollerhebung für diese durchzuführen), so spricht man von einer Cluster-Stichprobe.

#### Verschiedene Typen von Cluster-Stichproben in der Datenvalidierung

Eine solche Cluster-Stichprobe von Datenfeldern liegt beispielsweise bei der Ziehung von Fällen aus der Grundgesamtheit aller Fälle eines Krankenhausstandortes vor: Die Fälle werden mit gleichen Wahrscheinlichkeiten als einfache Zufallsstichprobe gezogen und für jeden ausgewählten Fall werden entsprechend einer Vollerhebung alle Datenfelder im Datenabgleich überprüft. In einer Analogie zu einem klassischen Beispiel aus der Stichprobentheorie entspricht dies der Situation, dass verschiedene Schulklassen einer Schule ausgewählt werden und für diese dann jeweils alle Schülerinnen und Schüler befragt werden.

Sollen Aussagen über die Gesamtfehldokumentationsrate nicht nur für einen einzelnen Krankenhausstandort gemacht werden, sondern für ein QS-Verfahren insgesamt, wird die Situation komplexer, da in den Datenvalidierungsverfahren Überprüfungen in der Regel nur bei einer Stichprobe von Krankenhausstandorten stattfinden. Dies entspricht der zufälligen Auswahl von Schulen (Krankenhausstandorten), in denen dann jeweils Klassen (Fälle) zufällig ausgewählt werden, für die dann alle Schülerinnen und Schüler (alle Datenfelder) befragt (abgeglichen) werden.

In diesem Abschnitt werden verschiedene Berechnungsmethodiken für ein Vertrauensintervall für die Gesamtfehldokumentationsrate eines einzelnen Krankenhausstandortes diskutiert, die

sich aus einer solchen Cluster-Stichprobe ergibt. Analoge Überlegungen können für Vertrauensintervalle für die Gesamtfehldokumentationsrate in einem QS-Verfahren insgesamt sowie für die Fehldokumentationsraten einzelner Datenfelder in einem QS-Verfahren angestellt werden, indem die hier verwendete Varianzschätzung durch eine für die jeweilige Situation adäquate Formulierung ersetzt wird (siehe z. B. Levy und Lemeshow 2008). In komplexeren Situationen kann durch Berechnung der jeweiligen (paarweisen) Auswahlwahrscheinlichkeiten eine Varianzschätzung mit dem allgemeinen Horvitz–Thompson-Schätzer vorgenommen werden (Särndal et al. 2003, Kapitel 5). Für die praktische Umsetzung in R bietet sich das R-Paket *survey* (Lumley 2019) an, welches es erlaubt, die komplexe Auswahlhierarchie mittels Designobjekten abzubilden und dann generisch die entsprechende statistische Inferenz implementiert. Als Weiterentwicklung könnten alternativ Domänen-Methoden (Särndal et al. 2003, Kapitel 10) verwendet werden, um zu berücksichtigen, dass gemäß der oben beschriebenen Auswertungsregeln nicht alle Datenfelder für alle Fälle in die Berechnung der Fehldokumentationsraten eingehen.

### Untersuchte Varianten

Für die methodische Festlegung eines Vertrauensintervalls für die Gesamtfehldokumentationsrate eines Krankenhausstandortes aus Formel (3) wurden verschiedene Möglichkeiten der Berechnung näher untersucht.

Notation: Im Folgenden wird die bereits in Abschnitt 2.1 eingeführte Notation verwendet. Zusätzlich bezeichnet jeweils  $U$  die untere und  $O$  die obere Grenze des Vertrauensintervalls  $[U, O]$  zum Konfidenzniveau  $\alpha$ . Mit  $Y = F^{\text{Stichprobe}}$  wird die Fehldokumentationsrate in der Stichprobe (des betrachteten Krankenhausstandortes) abgekürzt. Ausgehend von  $Y$  soll eine Aussage über die Fehldokumentationsrate  $\hat{Y}$  bezogen auf alle dokumentierten Fälle des Krankenhausstandortes im jeweiligen QS-Verfahren im jeweiligen Erfassungsjahr gemacht werden. Dabei gilt für die Schätzung des Mittelwertes gemäß der Konstruktion in Abschnitt 2.1.2 grundsätzlich  $\hat{Y} = Y$ .

#### 1. Aus Varianzschätzung für Cluster-Stichproben

Basierend auf einer Varianzschätzung für Cluster-Stichproben

$$\widehat{\text{Var}}(\hat{Y}) = \frac{1}{\bar{N}} \frac{J - J^*}{J} \frac{1}{J^*(J^* - 1)} \sum_{j=1}^{J^*} \left( \sum_{k=1}^{K_j} x_{jk} - K_j \hat{Y} \right)^2$$

(Kauermann und Küchenhoff 2011) – dabei kann die mittlere Anzahl Felder pro Fall  $\bar{N} := \sum_{j=1}^J K_j / J$  durch die entsprechende Größe  $\sum_{j=1}^{J^*} K_j / J^*$  in der Stichprobe approximiert werden – wird das folgende Vertrauensintervall definiert:

$$U = \hat{Y} - t_{1-\frac{\alpha}{2}, J^*-1} \widehat{\text{Var}}(\hat{Y}), \quad O = \hat{Y} + t_{1-\frac{\alpha}{2}, J^*-1} \widehat{\text{Var}}(\hat{Y}).$$

Dabei ist  $t_{1-\frac{\alpha}{2}, J^*-1}$  das  $(1-\alpha/2)$ -Quantil der  $t$ -Verteilung mit  $J^* - 1$  Freiheitsgraden.

## 2. Aus Varianzschätzung für Cluster-Stichproben, logit-transformiert

Die direkte Definition eines Vertrauensintervalls für einen Anteilswert mittels Quantilen aus der  $t$ -Verteilung (oder der Normalverteilung) hat den Nachteil, dass das resultierende Intervall  $[U, O]$  ggf. nicht vollständig im Wertebereich  $[0,1]$  enthalten ist. Dieser Problematik kann durch eine geeignete Transformation begegnet werden (Korn und Graubard 1998). Eine Möglichkeit dazu ist die folgende Konstruktion: Durch

$$U = \frac{1}{1 + \exp(-U_{\text{logit}})}, \quad O = \frac{1}{1 + \exp(-O_{\text{logit}})},$$

wobei

$$U_{\text{logit}} = \log\left(\frac{\hat{Y}}{1 - \hat{Y}}\right) - t_{1-\frac{\alpha}{2}, J^*-1} \frac{\sqrt{\text{Var}(\hat{Y})}}{\hat{Y}(1 - \hat{Y})},$$

$$O_{\text{logit}} = \log\left(\frac{\hat{Y}}{1 - \hat{Y}}\right) + t_{1-\frac{\alpha}{2}, J^*-1} \frac{\sqrt{\text{Var}(\hat{Y})}}{\hat{Y}(1 - \hat{Y})},$$

wird eine logit-transformierte Variante von Variante 1 definiert (Korn und Graubard 1998). Falls  $\hat{Y} = 0$  oder  $\hat{Y} = 1$  ist dieses Intervall nicht definiert. Ersatzweise kann in dieser Situation das Vertrauensintervall aus Variante 3 verwendet werden. Alternativ wäre es auch möglich, den Schätzer künstlich anzupassen und die Grundgesamtheit um ein korrektes und ein fehldokumentiertes Ergebnis zu ergänzen.

## 3. Exaktes Intervall aus der hypergeometrischen Verteilung mit mid-p-Korrektur

Wenn die Tatsache, dass es sich bei der Stichprobe um eine Cluster-Stichprobe handelt, vernachlässigt wird, kann die Untersuchung der  $n := J^* \sum_{j=1}^{J^*} K_j$  Datenfelder aller  $J^*$  Fälle in der Stichprobe als einfache Zufallsstichprobe (Ziehen ohne Zurücklegen) aus den insgesamt  $N := J \sum_{j=1}^J K_j$  Fällen des Krankenhausstandortes aufgefasst werden. Sei weiter  $m$  die Anzahl der Fehldokumentationen unter allen  $n$  Datenfeldern in der Stichprobe. Dann kann das exakte Vertrauensintervall mit mid-p-Korrektur aus Variante 2 in Abschnitt 2.2.1 auch in der hier vorliegenden Situation verwendet werden:

$$\tilde{U} := \tilde{U}(m) = \underset{M}{\operatorname{argmax}} P(X < m; N, M - 1, n) + \frac{1}{2} P(X = m; N, M - 1, n) \geq 1 - \alpha/2,$$

$$\tilde{O} := N - \tilde{U}(n - m)$$

Für  $m = 0$  setze wieder  $\tilde{U} := 0$  und definiere  $U := \tilde{U}/N$  und  $O := \tilde{O}/N$ . Die Betrachtung als hypergeometrische Verteilung bringt noch eine weitere Einschränkung mit sich: Sie erfordert die Formulierung der Fehldokumentationsraten als diskretes Problem, in dem ein Datenfeld entweder richtig oder falsch dokumentiert ist. Durch die in Abschnitt 2.1 festgelegten Regeln für Felder mit Mehrfachauswahl kommen aber auch nicht ganzzahlige Werte für die Zahl der Fehldokumentationen  $m$  vor. Da in allen QS-Verfahren der Anteil an solchen Mehrfachfeldern gering ist, erscheint es im Zweifel akzeptabel, diesen Aspekt an dieser Stelle zunächst zu vernachlässigen.

## Vergleich der untersuchten Varianten mittels Simulation

Aus den vergangenen Jahren stehen dem IQTIG die Ergebnisse des Datenabgleichs für diejenigen Verfahren, für die eine umfassende Datenvalidierung vorgenommen wurde, zur Verfügung. Diese enthalten je Krankenhausstandort jedoch nur eine Stichprobe von 20 Fällen. Um die beschriebenen Varianten zur Berechnung eines Vertrauensintervalls anhand ihrer Überdeckungswahrscheinlichkeiten zu vergleichen, sind jedoch auch Informationen über die übrigen, nicht in der Stichprobe untersuchten Fälle nötig.

Da diese Daten natürlicherweise nicht vorliegen, werden im Folgenden mithilfe einer Simulation Übereinstimmungsergebnisse für verschiedene fiktive Krankenhausstandorte in einem (hypothetischen) QS-Verfahren simuliert. Vereinfachend sollen die Daten so simuliert werden, dass für alle simulierten Fälle die gleichen Datenfelder überprüft wurden, abhängige Felder oder mehrfache Prozedurbögen sollen so also nicht abgebildet werden. Ergebnis der Simulation ist damit für die  $J_i$  Fälle eines Krankenhausstandortes  $i$  und  $K$  Datenfelder eine Matrix mit  $J_i \times K$  Einträgen mit den Werten 1 (bei Fehldokumentation) oder 0 (bei Übereinstimmung). Der Vorteil eines solchen Simulationssetups ist, dass – trotz der vereinfachenden Annahmen – die unterschiedlichen Größenordnungen der Einflüsse auf Fall- sowie auf Feldebene praxisnah abgebildet werden können.

Zur Modellierung wird angenommen, dass es für jedes Datenfeld jedes Falls eine individuelle Wahrscheinlichkeit der Fehldokumentation gibt. Diese wird im  $i$ -ten Krankenhaus ( $i = 1, \dots, I$ ) für den  $j$ -ten ( $j = 1, \dots, J_i$ ) Fall und das  $k$ -te Datenfeld ( $k = 1, \dots, K$ ) mit  $\pi_{ijk}$  bezeichnet. Sei  $X_{ijk}$  die Zufallsvariable, dass in Krankenhaus  $i$  bei Fall  $j$  das Feld  $k$  fehldokumentiert ist. Zu Simulationszwecken nehmen wir an, dass

$$X \sim \text{Ber}(\pi_{ijk}).$$

Zur Bestimmung der  $\pi_{ijk}$  wird ein verallgemeinertes additives gemischtes Modell (*Generalized Additive Mixed Model, GAMM*) auf Basis von tatsächlichen Daten aus dem EJ 2017 für alle Leistungserbringer geschätzt (Wood 2017). Mithilfe dieses Modells können dann neue Daten simuliert werden.

### Modell

Für das Modell wurde ein „gekreuztes Design“ gewählt, sodass sich  $\pi_{ijk}$  im Wesentlichen aus einem Datenfeld-Effekt  $\gamma_k$  und einem Fall-Effekt  $b_{ij}$  ergibt. Zusätzlich wird ein Krankenhauseffekt  $a_i$  einbezogen, um die Unterschiedlichkeit der Dokumentationsqualität zwischen den Krankenhäusern in den Ursprungsdaten abzubilden. Der Datenfeld-Effekt wurde dabei als feste Einflussgröße modelliert. Fall- und Krankenhaus-Effekt wurden aufgrund der Vielzahl der betrachteten Fälle und Krankenhäuser und weil mit dem Modell im nächsten Schritt Daten für neue, fiktive, Krankenhausstandorte und ihre Fälle erzeugt werden sollen, als zufällige Einflussgrößen (*random effects*) modelliert:

$$\text{logit}(\pi_{ijk}) = \beta_0 + a_i + b_{ij} + \gamma_k \quad (5)$$

Hierbei bezeichnet  $a_i \sim \mathcal{N}(0, \sigma^2)$  den Krankenhaus-Effekt,  $b_{ij} \sim \mathcal{N}(0, \tau^2)$  den Fall-Effekt,  $\gamma_k$  den Datenfeld-Effekt und  $\beta_0$  einen Intercept. Eine solche Modellierung erlaubt es im Vergleich zu einem logistischen Regressionsmodell auf einfache Weise Korrelationen innerhalb der Krankenhausstandorte bzw. innerhalb der Fälle abzubilden, ohne eine (unpraktikabel) hohe Anzahl an festen Effekten für jede einzelne Korrelation schätzen zu müssen.

### Datengrundlage für die Modellschätzung

Für die Modellschätzung wurden die Ergebnisse des Datenabgleichs aus dem Erfassungsjahr 2017 zum QS-Verfahren 17/1 verwendet. Der zugrunde liegende Datensatz enthält Informationen zu den überprüften Datenfeldern für eine Stichprobe von 1.179 Fällen an insgesamt 62 Krankenhausstandorten. In die Schätzung wurden dabei nur diejenigen Abgleichsergebnisse, die tatsächlich entsprechend der in Abschnitt 2.1.2 angegebenen Regeln (z. B. für Filterfelder und abhängige Felder) bei der Berechnung der Übereinstimmungsrate relevant waren, einbezogen. Um sicherzustellen, dass nach diesem Ausschluss für jedes Datenfeld hinreichend Information zur Schätzung des Modells vorhanden ist, wurden Datenfelder, deren Übereinstimmungsergebnis für weniger als 80 % der bundesweit untersuchten Fälle in die Berechnung eingegangen sind (z. B. selten ausgefüllte abhängige Datenfelder) ausgeschlossen.

### Modellschätzung

Das oben beschriebene gemischte Modell wurde mit dem R-Paket `gamm4` (Wood und Scheipl 2017) geschätzt. Der größte Datenfeldeffekt ergab sich für das Feld „Gehstrecke (bei Aufnahme oder vor der Fraktur)“ mit 1,05 (95 %-Vertrauensintervall: 0,83–1,26), den kleinsten Datenfeldeffekt hatte das Feld „Entlassungsgrund“ (das nur in den Kategorien „Tod“ und „anderer“ erhoben wird) mit -4,16 (95 %-Vertrauensintervall: -3,14–5,19). Die Varianz des Falleffekts betrug 0,30, die Varianz des Krankenhauseffekts dagegen 1,02.

### Simulation neuer Daten

Nach Schätzung der Modell-Parameter  $\beta_0$ ,  $a_i$ ,  $b_{ij}$  und  $\gamma_k$  aus Formel (5) bzw. deren Varianzen können ausgehend von diesem Modell Abgleichsergebnisse für Krankenhausstandorte simuliert werden. Zur Simulation des systematischen Einflusses des jeweiligen Krankenhausstandortes wird ein Simulationsparameter  $\kappa$  genutzt:

$$\text{logit}(\pi_{ijk}) = \beta_0 + b_{ij} + \gamma_k + \kappa \quad (6)$$

Durch Variation des Parameters  $\kappa$  zwischen  $-\infty$  und  $+\infty$  können Daten für Leistungserbringer mit unterschiedlichen Gesamtfehdokumentationsraten simuliert werden. Gegenüber der Modellgleichung in Formel (5) wurde für die Simulation neuer Daten in Formel (6) auf den Krankenhauseffekt  $a_i$  verzichtet, da die Dokumentationsqualität des Krankenhausstandortes bereits durch die Wahl von  $\kappa$  bestimmt werden kann.

Das Modell (6) dient hier der Simulation vollständig neuer Daten. Prinzipiell wäre es möglich, ausgehend von den 20 Fällen, die jeweils für die im Datenabgleich untersuchten Krankenhausstandorte vorlagen, nur die restlichen  $J_i - 20$  Fälle mithilfe von (6) zu generieren. Dies könnte jedoch eine zusätzliche, nicht gewünschte Variabilität in den generierten Daten erzeugen, wenn

das Modell nicht alle relevanten Einflüsse vollständig abbildet: Zum Beispiel wurden Abhängigkeiten zwischen den Übereinstimmungen verschiedener Datenfelder oder die Zusatzregeln aus Abschnitt 2.1.2 in der Modellierung nicht berücksichtigt. Somit ist die Simulation neuer Daten die methodisch sauberere Lösung.

### Überdeckungswahrscheinlichkeiten

Für 2.000 unterschiedliche Werte von  $\kappa$  wurden durch Ziehung von  $b_{ij} \sim \mathcal{N}(0, \tau^2)$  jeweils für fiktive Krankenhausstandorte mit 25 bzw. 100 Fällen Wahrscheinlichkeiten  $\pi_{ijk}$  bestimmt, mithilfe derer Ergebnisse des Datenabgleichs simuliert werden konnten.

Für jeden einzelnen dieser  $2 \cdot 2.000 = 4.000$  Krankenhausstandorte wurde anschließend wiederholt eine Stichprobe von 20 Fällen gezogen, für die jeweils nach den diskutierten Varianten das 95 %-Vertrauensintervall (d. h.  $\alpha = 0,05$ ) berechnet wurde. In jeder Wiederholung dieser Ziehung wurde überprüft, ob das berechnete Vertrauensintervall die tatsächliche (bekannte) Gesamtfehldokumentationsrate in der Grundgesamtheit aller Fälle des jeweiligen Krankenhausstandortes enthält. Die Umsetzung der Vertrauensintervalle erfolgte zur Erzielung geringer Laufzeiten manuell, die Implementierung der ersten beiden Varianten konnte mit dem R-Paket survey (Lumley 2019) gegengerechnet werden.

Die auf diese Weise berechneten Überdeckungswahrscheinlichkeiten durch 1.000-maliges Wiederholen dieser Stichprobenziehung sind in Abbildung 7 veranschaulicht. Die Überdeckungswahrscheinlichkeiten werden dabei nicht direkt über den (artificialen) Simulationsparameter  $\kappa$ , sondern über die Gesamtfehldokumentationsrate des simulierten Krankenhausstandortes dargestellt. Dies erleichtert die Interpretation des Ergebnisses, da das Vertrauensintervall aus den Ergebnissen in einer Stichprobe eine Aussage über diese Größe macht.

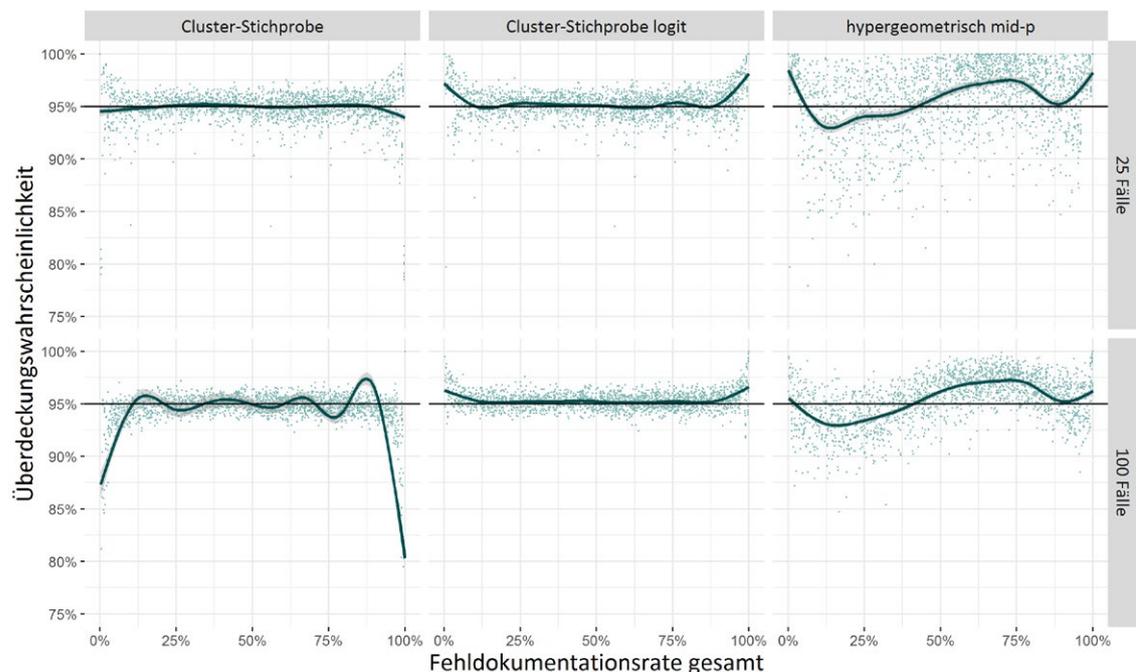


Abbildung 7: Überdeckungswahrscheinlichkeiten für verschiedene Varianten von Vertrauensintervallen für die Gesamtfehldokumentationsrate eines Krankenhausstandortes gemäß der durchgeführten Simulation.

Ergänzend zu den Datenpunkten, die die Überdeckung für jede der 4.000 Simulationen angeben, wurde eine geglättete Kurve auf Basis dieser Punkte geschätzt (mittels eines verallgemeinerten additiven Modells (GAM)), um den mittleren Verlauf trotz der hohen Dichte an Datenpunkten zu verdeutlichen. In der Grafik für Variante 1 mit Fallzahl 100 (unten links) liegen dabei 24 Datenpunkte außerhalb des dargestellten Wertebereiches.

Für die Vertrauensintervalle aus Variante 1 (Cluster-Stichprobe) liegen die Überdeckungswahrscheinlichkeiten demnach für moderate Gesamtfehldokumentationsraten um 95 % mit relativ geringer Streuung, während sie insbesondere bei einer Fallzahl von 100 für sehr kleine und sehr große Fehldokumentationsraten deutlich zu niedrig sind. Die logit-transformierten Vertrauensintervalle aus Variante 2 haben für moderate Gesamtfehldokumentationsraten sogar bessere Eigenschaften als diejenigen aus Variante 1: Die Kurve des GAM-Modells verläuft fast horizontal auf der Höhe der 95 %. Für sehr kleine und sehr große Fehldokumentationsraten ist diese Variante allerdings eher zu konservativ. Im Vergleich mit Variante 1 sind die Überdeckungen aber auch für Ergebnisse nahe bei 0 % oder 100 % noch akzeptabel, insbesondere für die Simulation auf Basis von 100 Fällen.

Die Intervalle aus der hypergeometrischen Verteilung sind zwar für sehr niedrige und sehr hohe Gesamtfehldokumentationsraten recht stabil, weichen allerdings in ihren Überdeckungswahrscheinlichkeiten im Vergleich am stärksten von den erwünschten 95 % ab, sodass sie für Gesamtfehldokumentationsraten unter 40 % eher zu schmal und für Werte darüber eher zu breit sind. Zusätzlich streuen sie stärker, sodass in einigen Fällen sehr große Abweichungen von der erwünschten Überdeckung von 95 % auftreten. Dies ist vermutlich darin begründet, dass diese Berechnungsvariante die Tatsache, dass die Stichprobe als Cluster-Stichprobe gezogen wird, nicht beachtet. Setzt man eine höhere Varianz für den Fall-Effekt fest als diejenige aus der hier verwendeten Modellschätzung, so verstärkt sich diese Beobachtung noch einmal.

Auf Basis dieser Ergebnisse ergibt sich Variante 2, d. h. das logit-transformierte Konfidenzintervall auf Basis der Varianzschätzung für Cluster-Stichproben, als eindeutig zu präferierende Variante zur Berechnung von Vertrauensintervallen für die Gesamtfehldokumentationsrate eines Krankenhausstandortes.

### 2.2.3 Einordnung in das statistische Rahmenkonzept für Auswertungen

Im Folgenden wird die in diesem Kapitel beschriebene Methodik zur Berechnung von Punktschätzern und zugehörigen Unsicherheitsintervallen für die Kenngrößen der Datenvalidierung in das statistische Rahmenkonzept für Auswertungen des IQTIG eingeordnet, um die Kohärenz der methodischen Entwicklungen zu verdeutlichen. Die in IQTIG (2019, Kapitel 5.1) dargelegten Elemente können wie folgt auf die beschriebene Methodik angewandt werden.

Die Herangehensweise bei der Datenvalidierung ist als enumerativ zu verstehen. Das heißt, es sollen Aussagen über die vorliegende Dokumentation gemacht werden. Die **Stichprobenart** für die vorgegebene Auswertungsebene hängt davon ab, ob es sich um eine Auswertung für *einen* oder *für eine Menge* an Leistungserbringern (z. B. alle Leistungserbringer in einem Bundesland)

handelt. Auf Leistungserbringerebene ist immer von einer einfachen Zufallsstichprobe (ohne Zurücklegen) von Fällen auszugehen. Pro ausgewähltem Fall handelt es sich um eine Vollerhebung aller Datenfelder. Zusammen stellen die beiden Ebenen eine Cluster-Stichprobe dar.

**Berechnungsart:** Bei allen vorgestellten Kenngrößen handelt es sich um Anteilswerte. Ein Beispiel ist der Anteil fehldokumentierter Datenfelder über alle Felder der Fälle eines Leistungserbringers hinweg (Fehldokumentationsrate gesamt). Bei der Bestimmung dieser Anteilswerte ist der Stichprobencharakter der Datengrundlage zu berücksichtigen, d. h. anhand einer Stichprobe soll auf eine klar definierte, aber in Teilen unbeobachtete Grundgesamtheit geschlossen werden. Dieser Sachverhalt bei der Berechnung soll als Berechnungsart „Rate (Cluster-Stichprobe)“ vermerkt werden.

**Bewertungsart:** Für die Klassifikation der Datenvalidierungskenngrößen eines Leistungserbringers werden, ähnlich wie bei der Bewertung von Qualitätsindikatoren, Referenzwerte definiert. Die statistische Methodik zur Klassifikation bei Datenvalidierungskenngrößen kann auch im bayesianischen Rahmen des SD-Weiterentwicklungsberichts interpretiert werden. Dabei ist bei der Datenvalidierung – im Gegensatz zur analytischen Herangehensweise – nicht die hinreichende statistische Evidenz für eine Abweichung des (unbeobachteten) Kompetenzparameters vom Referenzbereich das Ziel der Klassifikation. Stattdessen ist das Ziel, die hinreichende statistische Evidenz für eine Abweichung des (unbekannten<sup>4</sup>) Anteils in der Grundgesamtheit vom Referenzbereich festzulegen.

Anhand der A-posteriori-Verteilung für den Anteilswert kann diese Klassifikation bei vorgegebenen Schwellenwert  $\alpha$ ,  $0 \leq \alpha \leq 1$ , entweder

- über den Abgleich der entsprechenden Grenze eines zweiseitigen equi-tailed  $(1-2\alpha) \cdot 100\%$ -Unsicherheitsintervalls geschehen, oder
- durch einen Abgleich der A-posteriori-Wahrscheinlichkeit für das Nicht-Einhalten des Referenzbereiches. Das heißt, man geht von einer quantitativen Auffälligkeit aus, falls

$$P(X \leq R \cdot N | \text{Stichprobe}) \leq \alpha$$

wobei  $X$  z. B. die Anzahl an fehldokumentierten Datenfeldern ist,  $N$  die Anzahl an relevanten Feldern insgesamt (Größe der Grundgesamtheit) und  $R$  der festgelegte Referenzwert ist. Im Allgemeinen wird davon ausgegangen, dass  $X$  als Zufallsvariable betrachtet werden kann und  $N$  bekannt ist.

Wird z. B.  $\alpha = 0,05$  gewählt, und ist der Referenzbereich für den Anteil bei  $\leq 10\%$ , dann ist ein Leistungserbringer bezüglich eines Referenzwertes von  $R = 10\%$  quantitativ auffällig, falls  $P(X \leq 0,1 \cdot N | \text{Stichprobe}) \leq 0,05$ . Dies ist äquivalent zum Vorgehen einen Leistungserbringer als quantitativ auffällig einzustufen, falls die untere Grenze eines zweiseitigen  $90\%$ -Unsicherheitsintervalls für  $X/N$  über  $0,1$  liegt.

Die in Abschnitt 2.2.1 vorgestellten Methoden zur Berechnung der Konfidenzintervalle basieren alle auf einer zugrunde liegenden Verteilung für die Anzahl an fehldokumentierten Feldern  $X$ . Für Methode 1 und 2 ist das z. B. die hypergeometrische Verteilung, für Methode 3 die t-Verteilung

<sup>4</sup> Unbekannt, soweit nur durch eine Stichprobe beobachtet.

mit vorgegebenen Lokationsparameter, Skalenparameter und  $n - 1$  Freiheitsgraden. Unter der Annahme einer Gleichverteilung auf  $1, \dots, N$  als Priori für  $X$  (d. h. eine nicht informative A-priori-Verteilung) entspricht die A-posteriori-Verteilung eben dieser Verteilung. Somit kann die oben definierte A-posteriori-Wahrscheinlichkeit für das Einhalten des Referenzwertes über die Verteilungsfunktion der Posteriori-Verteilung berechnet werden.

### 3 Addendum I zum Häufigkeitskriterium

#### Teilaspekte des Gegenstands der Beauftragung Nr. I.1. und Gegenstand Nr. I.4

Gegenstand der Beauftragung war es ein Kriterium zu entwickeln, das einen gezielten Datenabgleich auslöst, wenn die Ergebnisse des Stichprobenverfahrens mit Datenabgleich im Vorjahr besonders häufige Dokumentationsfehler aufwiesen. Ebenso sollte ein Kriterium zu erheblichen Dokumentationsmängeln entwickelt werden. Hierzu konnte gemäß Beauftragung eine Bezugnahme auf die Kriterien zu „besonderer Häufigkeit“ und „Relevanz“ erfolgen (vgl. Abschnitt 1.2.).

Die am 6. März 2019 beschlossenen Kapitel zum Häufigkeitskriterium des Addendum I zum überarbeiteten Abschlussbericht vom 19. Dezember 2018 sind in dieses Kapitel übertragen worden. Nachfolgend werden die Kriterien für besonders/erheblich häufige Dokumentationsfehler erläutert. Die Krankenhausstandorte für den gezielten Datenabgleich auf der Grundlage der Kriterien zur besonderen und erheblichen Häufigkeit wurden zum 15. April 2019 erstmals ermittelt und für das EJ 2018 vor Ort erneut im Rahmen einer Zweiterfassung datenvalidiert.

#### 3.1 Dokumentationsfehler von besonderer Häufigkeit

##### Teilaspekt des Gegenstands der Beauftragung Nr. I.1

Auf Basis der Ergebnisse des Stichprobenverfahrens mit Datenabgleich (Zufallsstichprobe) des Vorjahres wird folgendes Kriterium für die Auslösung eines erneuten Datenabgleichs bei einem Krankenhausstandort vorgeschlagen:

Zur Beurteilung, ob im betrachteten Erfassungsjahr bei einem Krankenhausstandort Dokumentationsfehler von besonderer Häufigkeit vorliegen, wird die Übereinstimmungsrate über alle Datenfelder und Fälle hinweg herangezogen. Die Übereinstimmungsrate der Stichprobe im Stichprobenverfahren mit Datenabgleich berechnet sich als Anteil der Datenfelder unter allen überprüften Datenfeldern, für die im Stichprobenverfahren mit Datenabgleich QS-Daten und Zweiterfassungsdaten übereinstimmen. Umgekehrt bezeichnet die Fehldokumentationsrate der Stichprobe den Anteil der Datenfelder, für die keine Übereinstimmung festgestellt werden konnte:

$$\frac{\text{Anzahl Fehldokumentationen i. d. Stichprobe}}{\text{Anzahl betrachteter Datenfelder aller Fälle i. d. Stichprobe}}$$

Die genaue Berechnungsvorschrift ist in Abschnitt 2.1.2 wiedergegeben. In die Berechnung der Fehldokumentationsrate fließen alle Datenfelder ein, für die ein Datenabgleich stattgefunden hat, wobei die in Abschnitt 2.1.2 aufgeführten zusätzlichen Regeln (z. B. für Filterfelder und abhängige Felder) gelten.<sup>5</sup>

Bei dem vorgeschlagenen statistischen Verfahren wird berücksichtigt, dass im Stichprobenverfahren mit Datenabgleich in Abhängigkeit von der Zahl der dort behandelten Fälle u. U. nur ein

<sup>5</sup> Die dort genannten Regeln entsprechen den in Addendum II vom 22. Oktober 2019 beschlossenen Regeln, der Vollständigkeit halber ergänzt wurde lediglich eine Regel zu Feldgruppen.

kleiner Teil aller tatsächlich in dem jeweiligen Krankenhausstandort für das jeweilige QS-Verfahren dokumentierten Fälle überprüft wird.

Nach Beschluss vom 6. März 2019 liegen besonders häufige Dokumentationsfehler vor, wenn für einen Krankenhausstandort auch unter Berücksichtigung der sich hieraus ergebenden Stichprobenunsicherheit davon auszugehen ist, dass bezogen auf alle in dem jeweiligen QS-Verfahren dokumentierten Fälle eine Gesamtfehldokumentationsrate von  $> 10\%$  über alle betrachteten Felder und alle Fälle vorliegt. Zur Entscheidung, ob das Kriterium erfüllt ist, wird die für den betrachteten Standort gezogene Zufallsstichprobe von 20 Fällen gemäß QSKH-RL herangezogen.

Für Krankenhausstandorte mit vier bis 20 Fällen im betrachteten QS-Verfahren entspricht die Stichprobe der Grundgesamtheit aller dokumentierten Fälle. Für diese Standorte ist das Kriterium erfüllt, sobald die Fehldokumentationsrate in der Stichprobe  $> 10\%$  beträgt.

Für Krankenhausstandorte mit mehr als 20 dokumentierten Fällen im betrachteten QS-Verfahren wird ein statistisches Verfahren angewendet, um zu bewerten, ob genügend Evidenz gegen die Hypothese vorliegt, dass die Fehldokumentationsrate des Krankenhausstandortes  $\leq 10\%$  beträgt. Hierzu wird ein einseitiges 95 %-Vertrauensintervall für die Gesamtfehldokumentationsrate berechnet. Liegt die untere Grenze dieses Vertrauensintervalls über  $10\%$ , so liegen besonders häufige Dokumentationsfehler vor. Wie in Abschnitt 2.1.3 beschrieben kann äquivalent dazu ein bayesianisches Framework unter Verwendung eines p-Wertes genutzt werden.

Infolgedessen muss hier ggf. eine höhere Fehldokumentationsrate als  $10\%$  in der untersuchten Stichprobe vorliegen, damit das Kriterium erfüllt ist. Die jeweilige minimale Fehldokumentationsrate in der Stichprobe, ab der im Sinne der hier vorgeschlagenen Definition besonders häufige Dokumentationsfehler vorliegen, hängt dabei sowohl von der Varianz der Fehldokumentationen zwischen den in der Stichprobe untersuchten Fällen als auch von der Gesamtfallzahl des jeweiligen Krankenhausstandortes ab. In Tabelle 3 ist die Abhängigkeit von der Gesamtfallzahl auf Basis einer Simulation exemplarisch dargestellt. Das genaue statistische Verfahren ist in Abschnitt 2.2.2 beschrieben.

*Tabelle 3: Fehldokumentationsrate für besonders häufige Dokumentationsfehler in Abhängigkeit von der Gesamtfallzahl (Simulation)*

Anzahl Fälle des KH	„Kleines“ QS-Verfahren (8 Datenfelder pro Fall)	„Großes“ QS-Verfahren (40 Datenfelder pro Fall)
10 (alle Fälle in Stichprobe)	$F > 10,0\%$	$F > 10,0\%$
20 (alle Fälle in Stichprobe)	$F > 10,0\%$	$F > 10,0\%$
30 (Stichprobe 20 Fälle)	$F > 12,7\%$	$F > 11,1\%$
50 (Stichprobe 20 Fälle)	$F > 13,4\%$	$F > 11,4\%$
100 (Stichprobe 20 Fälle)	$F > 13,9\%$	$F > 11,7\%$
1.000 (Stichprobe 20 Fälle)	$F > 14,3\%$	$F > 11,9\%$

In der Tabelle ist für zwei vereinfachte Beispielfälle dargestellt, ab welcher Fehldokumentationsrate  $F$  in der Stichprobe gemäß dem verwendeten statistischen Verfahren – in Abhängigkeit der Gesamtfallzahl – besonders häufige Dokumentationsfehler vorliegen. Tatsächlich hängen diese Werte auch von der Verteilung der Fehldokumentationen auf die einzelnen Fälle ab (diese wurde hier mittels einfacher Zufallsauswahl wiederholt simuliert, dargestellt sind die Mittelwerte aus den Simulationen), sodass Abweichungen möglich sind.

### 3.2 Dokumentationsfehler von erheblicher Häufigkeit

#### Gegenstand der Beauftragung Nr. I.4

Die QSKH-RL sieht seit 2018 vor, dass Krankenhausstandorte, bei denen erhebliche Dokumentationsmängel vorliegen, gegenüber dem Unterausschuss Qualitätssicherung des G-BA (direkte Verfahren) bzw. dem zuständigen Lenkungsgremium auf Landesebene (indirekte Verfahren) zu benennen sind (§ 13 Abs. 3 Satz 2 QSKH-RL). Für die Festlegung der Definition von **erheblichen Dokumentationsmängeln** wurde vom IQTIG zunächst ausschließlich Bezug auf das Kriterium der Häufigkeit genommen, da das Kriterium zur Relevanz zu diesem Zeitpunkt noch nicht entwickelt war. Weist ein Krankenhausstandort mindestens 15 % Dokumentationsfehler im Stichprobenverfahren mit Datenabgleich auf, kann davon ausgegangen werden, dass die Defizite im Bereich der Dokumentationsqualität so gravierend sind, dass eine sofortige Benennung an das zuständige Lenkungsgremium auf Landesebene bzw. den Unterausschuss Qualitätssicherung des G-BA nötig ist.

Nach Beschluss vom 6. März 2019 werden Krankenhausstandorte somit gegenüber dem Lenkungsgremium auf Landesebene / dem Unterausschuss Qualitätssicherung benannt, bei denen auch unter Berücksichtigung der sich aus der Stichprobenziehung ergebenden statistischen Unsicherheit davon auszugehen ist, dass sie bezogen auf alle in dem jeweiligen QS-Verfahren dokumentierten Fälle eine Fehldokumentationsrate von  $> 15\%$  aufweisen. Somit bestehen erhebliche Dokumentationsmängel, wenn erheblich häufige Dokumentationsfehler vorliegen.

Die Berechnung dieses Kriteriums erfolgt analog zu der Darstellung zum Kriterium für besonders häufige Dokumentationsfehler (mit entsprechend höherem Grenzwert von  $> 15\%$ ); vgl. Tabelle 4 für die zwei Beispielsituationen analog zu Tabelle 3. Das genaue statistische Verfahren ist in Abschnitt 2.2.2 beschrieben.

Tabelle 4: Fehldokumentationsrate für erheblich häufige Dokumentationsfehler in Abhängigkeit von der Gesamtfallzahl (Simulation)

Anzahl Fälle des KH	„Kleines“ QS-Verfahren (8 Datenfelder pro Fall)	„Großes“ QS-Verfahren (40 Datenfelder pro Fall)
10 (alle Fälle in Stichprobe)	FS $> 15,0\%$	FS $> 15,0\%$
20 (alle Fälle in Stichprobe)	FS $> 15,0\%$	FS $> 15,0\%$
30 (Stichprobe 20 Fälle)	FS $> 18,0\%$	FS $> 16,3\%$
50 (Stichprobe 20 Fälle)	FS $> 19,0\%$	FS $> 16,7\%$

Anzahl Fälle des KH	„Kleines“ QS-Verfahren (8 Datenfelder pro Fall)	„Großes“ QS-Verfahren (40 Datenfelder pro Fall)
100 (Stichprobe 20 Fälle)	FS > 19,6 %	FS > 17,0 %
1.000 (Stichprobe 20 Fälle)	FS > 20,0 %	FS > 17,2 %

In der Tabelle ist für zwei vereinfachte Beispielfälle dargestellt, ab welcher Fehldokumentationsrate  $F$  in der Stichprobe gemäß dem verwendeten statistischen Verfahren – in Abhängigkeit von der Gesamtfallzahl – erheblich häufige Dokumentationsfehler vorliegen. Tatsächlich hängen diese Werte auch von der Verteilung der Fehldokumentationen auf die einzelnen Fälle ab (diese wurde hier mittels einfacher Zufallsauswahl wiederholt simuliert, dargestellt sind die Mittelwerte aus den Simulationen), sodass Abweichungen möglich sind.

### 3.3 Weitere Aspekte zu Kriterien und Prüfmethodik von Dokumentationsfehlern mit besonderer und erheblicher Häufigkeit

#### Ausschluss von Datenfeldern mit Spezifikationsmängeln

Als Grundlage für die Anwendung der Kriterien sollen nur Dokumentationsfehler dienen, die Dokumentationsdefizite eines Krankenhausstandortes darstellen und von diesem auch behoben werden können. Sind Dokumentationsfehler mit hoher Wahrscheinlichkeit auf Mängel in der Spezifikation (etwa unzureichend beschriebene oder fehlende Datenfelder, unzureichende Ausfüllhinweise oder Plausibilitätsregeln) zurückzuführen, so müssen diese Probleme vom IQTIG für künftige Erfassungsjahre gelöst werden. Der Hinweis bzw. die Feststellung einer unzureichenden Spezifikation eines Datenfeldes kann sich erst aufgrund der Auswertung der Ergebnisse des Datenabgleichs vor Ort und somit retrospektiv ergeben. Folglich müssen diese Datenfelder aus der Berechnung der Ergebnisse ausgeschlossen werden. Für Details zur Berechnung des Ausschlusses siehe Abschnitt 2.1.2.3. Um Datenfelder zu identifizieren, bei denen die Dokumentationsfehler aufgrund von Spezifikationsmängeln durch das IQTIG vorliegen, wird Folgendes definiert: Datenfelder, bei denen über alle überprüften Krankenhausstandorte hinweg im Vorjahr > 15 % Dokumentationsfehler auftreten, werden im Hinblick auf mögliche Spezifikationsmängel begutachtet (zur Berechnung des Wertes und zugehöriger Vertrauensintervalle siehe Abschnitt 2.1.1.2). Diese Begutachtung erfolgt durch das IQTIG in Absprache mit Expertinnen und Experten. Auf Grundlage dieser Begutachtung wird entschieden, ob ein Datenfeld mangelhaft spezifiziert ist und aus der Berechnung der Kriterien ausgeschlossen wird.

#### Standorte des erstmalig gezielten Datenabgleichs 2019 (EJ 2018)

Für den gezielten Datenabgleich im Jahr 2019 kamen die drei Verfahren, welche im Jahr 2018 (EJ 2017) im Stichprobenverfahren mit Datenabgleich waren, potentiell in Betracht:

- 17/1: Hüftgelenknahe Femurfraktur mit osteosynthetischer Versorgung
- 09/2: Herzschrittmacher-Aggregatwechsel
- HCH-KCH: Koronarchirurgie, isoliert

In Tabelle 5 werden die Anzahl der Standorte der o. g. Verfahren berichtet, welche die Kriterien zur besonderen und erheblichen Häufigkeit erfüllten. Dabei wird deutlich, dass nur Standorte der direkten Verfahren eines der beiden Kriterien erfüllten und kein überprüfter Standort des direkten QS-Verfahrens HCH-KCH, weshalb für dieses QS-Verfahren auch kein gezielter Datenabgleich im Folgejahr oder eine Benennung gegenüber dem Unterausschuss Qualitätssicherung des G-BA vorgenommen werden musste. In den beiden indirekten QS-Verfahren erfüllten mehrere Standorte eines der beiden Kriterien, weshalb 18 Standorte im QS-Verfahren *Hüftgelenknahe Femurfraktur mit osteosynthetischer Versorgung* und 3 Standorte im Modul Herzschrillmacher-Aggregatwechsel gezielt im Folgejahr überprüft wurden. Gemäß QSKH-RL § 9 Abs. 10 Satz 3 sollten für die Benennung gegenüber dem Lenkungsgrremium bei erheblichen Dokumentationsmängeln keine Informationen des EJ 2017 zugrunde gelegt werden. Deshalb wurden die entsprechenden Standorte in Tabelle 5 mit erheblichen Dokumentationsmängeln in einer Übergangslösung für das EJ 2017 ebenfalls im Folgejahr zum EJ 2018 gezielt überprüft. Hierdurch ist es einmalig zu einem erhöhten Aufwand der gezielten Überprüfung im Jahr 2019 für die auf Landesebene beauftragten Stellen gekommen.

Tabelle 5: QS-Verfahren/QS-Modul und Anzahl der Standorte für den gezielten Datenabgleich 2019

QS-Verfahren	Standorte		Fälle die max. gezogen werden	Anzahl des max. Datenfeldabgleichs
	Besondere Häufigkeit > 10 %	Erhebliche Häufigkeit > 15 %		
<b>17/1</b> (von n = 62 im Stichprobenverfahren mit Datenabgleich)	18	14	n = 20	n = 34
<b>09/2</b> (von n = 48 im Stichprobenverfahren mit Datenabgleich)	3	7	n = 20	n = 17
<b>HCH-KCH</b> (von n = 5 im Stichprobenverfahren mit Datenabgleich)	0	0	n = 20	n = 40

### 3.4 Prozessablauf bei Dokumentationsfehlern von besonderer oder erheblicher Häufigkeit

Wenn bei einem Krankenhausstandort Dokumentationsfehler von besonderer Häufigkeit festgestellt werden, wird die Dokumentationsqualität dieses Krankenhausstandortes im Folgejahr erneut anhand eines Datenabgleichs der QS-Dokumentation mit der Patientenakte überprüft (vgl. Abbildung 8). Dabei wird eine Stichprobe in gleicher Größe wie für das Stichprobenverfahren mit Datenabgleich gemäß § 9 Abs. 4 QSKH-RL (20 Fälle) herangezogen. Die Stichprobenziehung der Fälle für jeden zu überprüfenden Standort soll sich dabei nach Möglichkeit auf Fälle aus dem dritten und vierten Quartal des Vorjahres beziehen, da in der dazwischenliegenden Zeit geeignete Verbesserungsmaßnahmen zur Dokumentationsqualität erfolgt sein sollten. Ebenfalls werden nach Möglichkeit die gleichen Datenfelder erneut überprüft. Das verantwortliche Gremium nach § 14 QSKH-RL soll im Nachhinein unabhängig von den Auswertungsergebnissen über diese informiert werden (Übersendung des bundeseinheitlichen Prüfberichts). Da eine Veröffentlichung der Ergebnisse in den Qualitätsberichten der Krankenhäuser nach § 136b Abs. 1 Satz

1 Nr. 3 SGB V vorgesehen ist, geschieht die Informationsweitergabe unabhängig davon, ob in der erneuten Überprüfung besonders häufige oder erheblich häufige Dokumentationsfehler vorlagen. Die Einordnung der Ergebnisse des gezielten Datenabgleichs obliegt damit dem Gremium nach § 14 QSKH-RL, welches sich an den vom IQTIG vorgeschlagenem Häufigkeitskriterium orientieren kann.

Eine sofortige Entanonymisierung gegenüber dem Lenkungsgremium auf Landesebene bzw. dem Unterausschuss Qualitätssicherung des G-BA erfolgt, wenn die im Rahmen des Stichprobenverfahrens mit Datenabgleich durch das IQTIG bzw. den auf Landesebene beauftragten Stellen festgestellten Fehldokumentationen so erheblich sind, dass sofort Maßnahmen angeraten sind und eine erneute Überprüfung als nicht ausreichend erscheint. Der Unterausschuss Qualitätssicherung bzw. das zuständige Lenkungsgremium auf Landesebene können sich bei der Einordnung der Ergebnisse an den vom IQTIG vorgeschlagenen Kriterien zu besonders häufigen Dokumentationsfehlern orientieren.

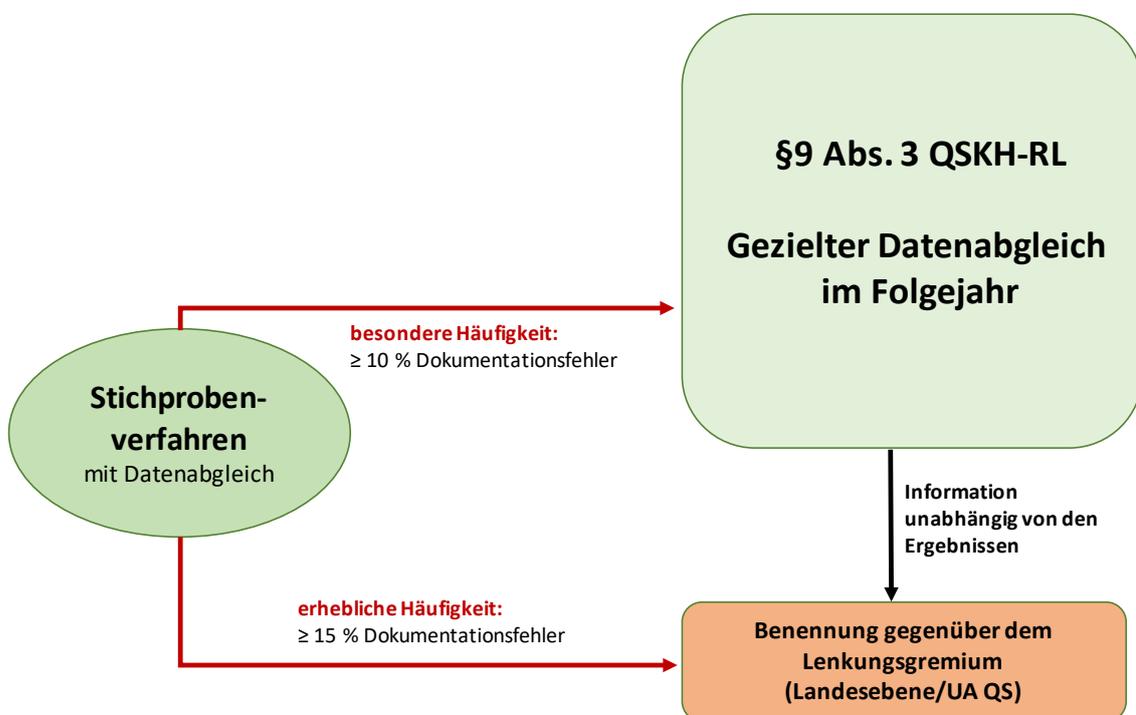


Abbildung 8: Gezielter Datenabgleich nach besonders häufigen Dokumentationsfehlern und sofortige Benennung gegenüber dem Lenkungsgremium bei erheblichen Dokumentationsfehlern

### 3.5 Limitationen

Das IQTIG weist auf mögliche Limitationen für die Definition und Auslösung des Häufigkeitskriteriums hin. Als mögliche Limitation könnte die Festlegung der Definition von erheblichen Dokumentationsmängeln betrachtet werden, da diese ausschließlich auf dem Kriterium zur Häufigkeit von Dokumentationsfehlern rekurriert. Das Kriterium zur Relevanz war zu diesem Zeitpunkt noch nicht entwickelt und wird auch nachfolgend nicht in die Festlegung erheblicher Dokumentationsmängel einbezogen (vgl. Begründungen in Abschnitt 4.6)

Grundsätzlich ist für die Ergebnisse des gezielten Datenabgleichs zu beachten, dass aufgrund der geringen Stichprobengröße von 20 Fällen ein Vergleich der Ergebnisse mit den Ergebnissen des Stichprobenverfahrens mit Datenabgleich im Vorjahr im Sinne einer **Trendaussage nur eingeschränkt möglich** ist. Daher empfiehlt das IQTIG dem jeweils zuständigen Gremium nach §14 QSKH-RL, die zeitliche Entwicklung der Fehldokumentationsraten in den beiden Stichproben nicht in die Bewertung der Dokumentationsfehler einzubeziehen. Perspektivisch ist hier auch eine gemeinsame Bewertung der Ergebnisse der beiden betrachteten Jahre unter Verwendung der im Vorbericht zur Weiterentwicklung des Strukturierten Dialogs (IQTIG 2019) beschriebenen methodischen Ansätze denkbar.

## 4 Addendum II zum Relevanzkriterium

### Teilaspekte des Gegenstands der Beauftragung Nr. I.1.

Gegenstand der Beauftragung war es, die Kriterien für Relevanz in Bezug auf die Ergebnisse des Stichprobenverfahrens mit Datenabgleich im Vorjahr zu definieren. Das Kriterium soll berücksichtigen, ob beispielsweise Datenfelder von Fehldokumentationen betroffen sind, die besonders relevante Ereignisse erfassen (z. B. Todesfälle oder gravierende Komplikationen) oder die in besonderem Maße Einfluss auf das Ergebnis eines Indikators haben. Grundsätzlich wird das Relevanzkriterium als Ergänzung zum Häufigkeitskriterium betrachtet, um damit insbesondere andere Standorte zu erkennen, an denen Fehldokumentationen auftreten, die durch die Identifikation von besonders und erheblich häufigen Fehldokumentationen nicht abgedeckt werden.

Das am 4. Dezember 2019 beschlossene Kapitel zum Relevanzkriterium des Addendum II ist in dieses Kapitel übertragen worden. Nachfolgend wird das Kriterium für relevante Dokumentationsfehler sowie die statistische Auswertungsmethodik erläutert. Die Krankenhausstandorte für den gezielten Datenabgleich auf der Grundlage des Kriteriums zu relevanten Dokumentationsfehlern sollen zum 15. April 2020 erstmals für das EJ 2019 ermittelt werden und vor Ort erneut im Rahmen eines gezielten Datenabgleichs überprüft werden.

### 4.1 Methodisches Vorgehen

Neben einer orientierenden Literaturrecherche zum Relevanzkriterium wurden Datenfelder systematisch und manuell ausgewählt und neben Proberechnungen inhaltlich in einem Expertengremium diskutiert. Auf Grundlage dieser Ergebnisse hat das IQTIG erneute Analysen durchgeführt.

#### Literaturrecherche

Zur Erarbeitung eines Kriteriums zur Identifizierung von Dokumentationsfehlern mit Relevanz in der Gesundheitsversorgung wurde eine Literaturrecherche in den Datenbanken *Cochrane Methodology Register* und *PubMed* sowie in *Google Scholar* durchgeführt<sup>6</sup>. Durch die manuelle Suche wurden fünf relevante Publikationen identifiziert. Das systematische Review von Kramer und Drews (2017) über Checklisten in der Gesundheitsversorgung hat keine Hinweise zur Identifizierung von Dokumentationsfehlern dargestellt. Gemäß den Ergebnissen des systematischen Reviews zu Fehlern in der Gesundheitsversorgung werden Dokumentationsfehler insgesamt unzureichend untersucht (Alsulami et al. 2013). Andere Studien berichten eher über Kriterien zur Erstellung valider und reliabler Erfassungsbögen (Pesudovs et al. 2007, Terwee et al. 2007). Inhaltliche Kriterien zur Identifizierung relevanter Datenfelder wurden in der Literaturrecherche nicht gefunden.

---

<sup>6</sup> Suchbegriffe: clinical documentation quality; clinical documentation error; significant documentation errors; medical documentation error.

Im Zusammenhang zu medizinischen Fehlern fokussieren (Murff et al. 2003) unerwünschte Ereignisse im Behandlungsverlauf einer Patientin oder eines Patienten. Ein unerwünschtes Ereignis wird in diesem Review nach der Definition von Kohn et al. (2000) aus dem *Institute of Medicine* (IOM) herangezogen und als „eine Verletzung, die durch das medizinische Management und nicht durch den zugrundeliegenden Zustand des Patienten verursacht wurde“ (Murff et al. 2003: S. 132) beschrieben. Zur Identifizierung unerwünschter Ereignisse fasst das Review von Murff et al. (2003) verschiedene Methoden zusammen. Auf die Erfassungsbögen bzw. Datensätze der Qualitätssicherung ließen sich folgende Methoden anwenden:

1. Kriterien nach Harvard Medical Practice Study z. B. Neurologisches Defizit bei der Entlassung und Tod
2. ICD-Codes von (potentiellen) Komplikationen
3. Stichwörter (inklusive Testlogik zur Reduzierung falsch positiver Ergebnisse), z. B.: Für welche Zielgruppe ist ein Laborparameter relevant? Welche Bedingungen erklären Abweichungen eines Parameters und unter welchen Bedingungen wird im Gegensatz ein Dokumentationsfehler vermutet?
4. Theoretisches Modell: Welche Fehler können auftreten? Welche Mechanismen vermitteln Fehler? Wann, wie und warum treten Fehler auf?

Aus der Literaturrecherche geht zusammenfassend hervor, dass es bislang keine standardisierten Kriterien zur Identifizierung relevanter Dokumentationsfehler gibt. Murff et al. (2003) bieten eine Übersicht zur Identifizierung von unerwünschten Ereignissen mithilfe verschiedener Methoden an. Auf dieser Grundlage werden (potentiell) unerwünschte Ereignisse für die Identifizierung relevanter Dokumentationsfehler fokussiert. Fehler innerhalb der Dokumentation dieser Ereignisse können somit als relevant bewertet werden. Um möglichst alle unerwünschten Ereignisse zu identifizieren, sollten Kombinationen der obigen Testmethoden angewendet werden (Murff et al. 2003).

### **Erste Analysen nach den Methodiken I bis III (werden im Folgenden beschrieben)**

Für die Datenfelder des Stichprobenverfahrens aus dem EJ 2017 wurden auf Basis einer exemplarischen Auswahl von Datenfeldern Proberechnungen durchgeführt und anschließend in einem Expertentreffen diskutiert.

Dies betrifft die QS-Verfahren: 09/2, 17/1 und HCH-KCH. Weitere Analysen zu früheren Verfahren der Zweiterfassung sind aufgrund der neuen Auswertungslogik und dahinterstehenden Aufbereitung des Datensatzes aus Aufwandsgründen nicht erfolgt. Ferner erscheinen solche Analysen nicht zielführend, da es letztlich um die Festlegung eines normativen Kriteriums geht, für die empirische Überlegungen höchstens im Hinblick auf die Frage nach der prinzipiellen Praktikabilität von Bedeutung sind.

Ergebnis dieser Untersuchungen war, dass eine Betrachtung der Übereinstimmungsraten einer Auswahl von Datenfeldern in diesem Sinne zwar möglich ist, aber im Wesentlichen zu keinen gravierend anderen Schlussfolgerungen als das bereits bestehende Häufigkeitskriterium führt. Zusätzlich stellte die sehr unterschiedliche Dokumentationsqualität der als inhaltlich relevant

eingestuften Datenfelder eine große Herausforderung dar: Während im QS-Verfahren 17/1 beispielsweise für das Feld „Entlassungsgrund“ eine sehr hohe Übereinstimmungsrate vorlag, war die Übereinstimmung im Feld „ASA“ wesentlich niedriger. Gleichzeitig kann allein durch die Betrachtung jedweder Abweichung bei den ausgewählten Datenfeldern nicht zwischen tatsächlich für die Indikatorberechnung relevanten und nicht relevanten Fehldokumentationen unterschieden werden (wenn z. B. den ASA-Klassifikationen 4 und 5 in den Risikomodellen das gleiche Risiko zugeordnet ist, spielt eine Fehldokumentation von ASA 4 statt ASA 5 keine Rolle für die QI-Berechnung). Zu den Verfahren im Einzelnen:

### **I. Automatisierte Auswahl von Datenfeldern nach Einfluss auf die Rechenregel**

Auf der Grundlage, dass bestimmte Datenfelder für die Berechnung von Qualitätsindikatoren benötigt werden, wurde die folgende Möglichkeit für die Entwicklung von Kriterien für relevante Dokumentationsfehler geprüft. Zunächst wurde geprüft, inwieweit durch allgemeine Überlegungen eine Regel gefunden werden kann, aus den Rechenregeln der Qualitätsindikatoren eines Verfahrens eine Gruppe von für das Verfahren besonders relevanten Datenfeldern zu ermitteln, ohne eine verfahrensspezifische manuelle Auswahl zu treffen. Eine Option wäre die Auswahl aller Datenfelder, die in den Zähler-Formeln der Indikatoren verwendet werden. Dies erschien allerdings wenig zielführend, da gerade auch Fehler bei Datenfeldern, für die Dokumentationsfehler zum Ein- und Ausschluss aus der Grundgesamtheit mehrerer Indikatoren führen, besonders gravierend erscheinen (Beispiel: Dokumentierter Therapieverzicht im QS-Verfahren *Ambulant erworbene Pneumonie* führt zu Ausschluss aus mehreren QIs). Ferner können einzelne Datenfelder einen hohen Einfluss auf das E eines O/E-Indikators haben. Andererseits ist auch die schlichte Gewichtung der Datenfelder danach, in wie vielen Indikatoren sie verwendet werden, nicht zielführend. Eine solche Regelung ergäbe für das Verfahren *Hüftgelenknahe Femurfraktur mit osteosynthetischer Versorgung* (17/1) eine hohe Relevanz des Alters, obwohl dieses nur in sehr groben Kategorien ( $\geq 18$  Jahre,  $\geq 65$  Jahre) einen Einfluss auf Indikatorergebnisse hat.

### **II. Neuberechnung von QI-Ergebnissen**

Eine vollständige Abbildung des Effektes der tatsächlichen Fehldokumentationen in der Stichprobe ließe sich durch eine erneute Auswertung der Qualitätsindikatoren auf Basis der nach dem Datenabgleich korrigierten QS-Daten erreichen. Dies erscheint jedoch aus verschiedenen Gesichtspunkten nicht zielführend: Falls ein Krankenhausstandort in einem solchen Verfahren relevante Dokumentationsfehler attestiert bekäme, wäre nicht einfach kommunizierbar, welche konkreten Fehldokumentationen dazu geführt haben. Gleichzeitig können sich, wenn nur die Abweichung des QI-Ergebnisses insgesamt betrachtet wird, Unter- und Überdokumentationen gegenseitig ausgleichen. Eine Extrapolation des Ergebnisses von der Stichprobe auf alle Fälle des Krankenhausstandortes ist auf Basis von lediglich 20 untersuchten Fällen nicht ohne weiteres möglich. Zudem ist unklar, ab welcher Schwelle für die Abweichung im QI-Ergebnis Fehldokumentationen als relevant erachtet werden (eine prozentuale Abweichung funktioniert beispielsweise nicht in Fällen, in denen eine Veränderung von 0 % auf einen anderen Wert stattgefunden hat). Zu diesen Aspekten kommt hinzu, dass nicht alle Qualitätsindikatoren fallweise ausgewertet werden: Für die Verfahren *Herzschrittmacherversorgung* und *Implantierbare Defibrillatoren*

sowie für das Verfahren *Pflege: Dekubitusprophylaxe* werden Qualitätsindikatoren nicht fallweise auf einem einzelnen Datensatz sondern durch Aggregation unterschiedlicher Informationen auf Standortebene ausgewertet. An dieser Stelle würden sich viele Herausforderungen in der Umsetzung des Kriteriums stellen.

### III. Manuelle verfahrensspezifische Auswahl von Datenfeldern

Ergänzend wurde der Ansatz verfolgt, verfahrensspezifisch Datenfelder auf Basis von fachlichen Überlegungen auszuwählen. Dazu wurden die potentiell relevanten Datenfelder aller Verfahren betrachtet und in einer Übersicht zusammengetragen. Insbesondere wurden einer näheren Betrachtung unterzogen:

- a) Datenfelder, die für die Berechnung von Qualitätsindikatoren verwendet werden, welche im Zeitraum von EJ 2016 bis EJ 2018 einen besonderen Handlungsbedarf aufwiesen
- b) Datenfelder, welche die Verfahrensmanager aufgrund von Rückmeldungen und Erfahrungen in der Leitung des Verfahrens für besonders relevant erachtet haben

Die entsprechenden Datenfelder wurden zunächst in einer Übersicht erfasst und Übereinstimmungen markiert, auch im Hinblick auf oben erwähnte Auswahlstrategien wie das Vorkommen in einer Zählerrechenregel. Ebenfalls hervorgehoben wurden Datenfelder des Zählers für „Sentinel-Event-Indikatoren“, auf die in Abschnitt 4.4 genauer eingegangen wird. Grundsätzlich gibt es nur wenige Datenfelder, die in allen QS-Verfahren gleich sind. Hieraus ergibt sich konsequenterweise, dass eine verfahrensspezifische Auswahl relevanter Datenfelder erfolgen müsste.

Gerade im Hinblick auf die Tatsache, dass alle im Datenabgleich überprüften Datenfelder in der einen oder anderen Form QI-relevant sind, führt dieses Vorgehen leicht dazu, dass eine große Anzahl von Datenfeldern ausgewählt würde. Betrachtet man dann von diesen Datenfeldern jeweils die Übereinstimmungsraten, hat dies zur Konsequenz, dass das sich ergebende Kriterium – anders als intendiert – eine große Nähe zum bereits bestehenden Häufigkeitskriterium aufweist.

#### Betrachtung unerwünschter Ereignisse

Anknüpfend an die Ergebnisse der Literaturrecherche und die Diskussionsergebnisse des Expertengremiums wurden *unerwünschte Ereignisse* als mögliches Relevanzkriterium neu betrachtet und analysiert. Aufgrund des feststehenden Begriffs „unerwünschte Ereignisse“ in Arzneimittelstudien werden nachfolgend „unerwünschte Ereignisse“ als *für das Indikatorergebnis relevante Ereignisse* bezeichnet.

Demnach wurden Datenfelder und Datenfeldkombinationen, die in die Berechnung von Qualitätsindikatoren eingehen, auf Anhaltspunkte, die für das *Indikatorergebnis relevante Ereignisse* darstellen, überprüft. Diese können sowohl *unerwünscht* und damit negativ formuliert sein, *als auch erwünscht* und damit positiv formuliert sein. Beispiele für negativ formulierte Ereignisse sind Komplikationen oder Todesfälle, Beispiele für positiv formulierte Ereignisse sind die Durchführung einer Antibiotikaprophylaxe oder das Überleben.

Positiv oder negativ formulierte Ereignisse können sich direkt aus den Zähler-Rechenregeln der Qualitätsindikatoren eines Verfahrens ergeben. Sie können auch nur einen Teil der Zähler-Bedingung abbilden oder aus einer Verknüpfung einer Zähler-Bedingung mit der Nenner-Bedingung (z. B. „Notfallkaiserschnitt und E-E-Zeit größer 20 Minuten“) oder einer Bedingung, die einen hohen Einfluss auf die Risikoadjustierung hat (z. B. „ASA > = 4“), bestehen. Aus Praktikabilitätsgründen werden nur dichotome Ereignisse betrachtet.

Grundsätzlich kann statt positiv formulierten Ereignissen durch logische Negation das entsprechende negativ formulierte Ereignis, dass das Gegenteil eingetreten ist, betrachtet werden (z. B. „Tod“ statt „Überleben“, „Keine Antibiotikaphylaxe durchgeführt“ statt „Antibiotikaphylaxe durchgeführt“).

In Tabelle 6 sind beispielhaft für 20 Fälle vier Ereignisse dargestellt (z. B. Komplikationen: ja/nein). Die roten Felder stellen die für das Indikatorergebnis relevanten Ereignisse dar (z. B. Komplikation = ja) und die Kreuze (X) inwiefern diese Ereignisse fehdokumentiert wurden, d. h. in der QS-Dokumentation nicht angegeben waren und damit eine relevante Fehdokumentation bedeuten. Durch die Betrachtung von Ereignissen wird klar, dass es sich nicht um die korrekte Dokumentation einzelner Datenfelder handeln soll, da diese schwer als „relevant“ definiert werden können (siehe Argumentation oben), sondern um für das Indikatorergebnis relevante Ereignisse, die sich aus einer Konstellation von verschiedenen Datenfeldern ergeben können (z. B. Alter größer 65 und keine Sturzprophylaxe durchgeführt).

Tabelle 6: Beispielhafte Darstellung von unterdokumentierten relevanten Ereignissen

Fälle	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Σ
Ereignis 1				X									X								2/3
Ereignis 2								X													1/3
Ereignis 3											X										1/2
Ereignis 4									X												1/4
Σ																					5/12 41,7%

Grundsätzlich können entsprechend Tabelle 6 für das Indikatorergebnis relevante Ereignisse sowohl über- als auch unterdokumentiert werden. Bei der Zweiterfassung vor Ort werden die Angaben aus der QS-Dokumentation mit denen der Primärdokumentation verglichen. Daraus wird ersichtlich, wie viele Ereignisse korrekt für die Qualitätssicherung dokumentiert wurden oder fälschlich bzw. fälschlich nicht für die Qualitätssicherung dokumentiert wurden.

Differenziert nach positiv und negativ formulierten Ereignissen ergeben sich für die Unter- oder Überdokumentation folgende Möglichkeiten:

- Eine **Unterdokumentation** eines positiv formulierten Ereignisses (erwünscht) bedeutet, dass ein Krankenhausstandort im Qualitätsergebnis schlechter abgebildet wird als es ist.
- Eine **Unterdokumentation** eines negativ formulierten Ereignisses (unerwünscht) bedeutet, dass ein Krankenhausstandort im Qualitätsergebnis besser abgebildet wird als es ist.

- c) Eine **Überdokumentation** eines positiv formulierten Ereignisses (erwünscht) bedeutet, dass ein Krankenhausstandort im Qualitätsergebnis besser abgebildet wird als es ist.
- d) Eine **Überdokumentation** eines negativ formulierten Ereignisses (unerwünscht) bedeutet, dass ein Krankenhausstandort im Qualitätsergebnis schlechter abgebildet wird, als es ist.

Grundsätzlich ist die Überdokumentation eines positiv formulierten bzw. die Unterdokumentation eines negativ formulierten Ereignisses in gleichem Maße schwerwiegend im Hinblick auf die Berechnung von Indikatorergebnissen. Dennoch schlägt das IQTIG vor, zunächst nur die Unterdokumentation negativ formulierter Ereignisse (Kategorie b)) zu betrachten, um das Relevanzkriterium einfach und kommunizierbar zu halten. Da allerdings positiv formulierte Ereignisse durch logische Negation in negativ formulierte Ereignisse umgeformt werden können, können durch diese Fokussierung auch Überdokumentationen positiv formulierter Ereignisse (Kategorie c)) betrachtet werden (siehe auch Abschnitt 4.2 zur genauen Definition des Kriteriums). Nachfolgend werden negativ formulierte Ereignisse als Ereignisse betrachtet, die auf mögliche Unterdokumentationen zu untersuchen sind. Dies entspricht der Betrachtung der Aspekte b) und c) aus obiger Auflistung.

Prinzipiell wäre es möglich, zusätzlich zu dem im Folgenden vorgeschlagenen Kriterium auch einen normativen Grenzwert für die Überdokumentation negativ formulierter Ereignisse festzulegen (für a) und b)). Für negativ formulierte Ereignisse aus den Zähler-Rechenregeln gilt, dass eine Unterdokumentation in der Qualitätssicherung einen Einfluss zugunsten des Leistungserbringers hat, während eine Überdokumentation sich zuungunsten des Leistungserbringers auswirkt.

Eine Betrachtung des Anteils der Über- und Unterdokumentationen von negativ formulierten Ereignissen erscheint deutlich zielführender als die Analyse der Übereinstimmungsraten: Speziell im Falle von Ereignissen mit niedriger Prävalenz, wie beispielsweise Wundinfektionen, ist die bloße Übereinstimmungsrate von Datenfeldern oft sehr hoch, da in den allermeisten Fällen sowohl in der QS-Dokumentation als auch in der Patientenakte übereinstimmend angegeben ist, dass keine Wundinfektion vorlag. Gleichzeitig ist es für einen Indikator, der den Anteil an Wundinfektionen nach einer Operation misst, von erheblicher Bedeutung, dass nicht die wenigen tatsächlich vorliegenden Infektionen unterdokumentiert wurden oder gar Wundinfektionen dokumentiert wurden, die tatsächlich gar nicht aufgetreten sind.

Dies wird im Folgenden durch zwei Beispiele mit Ergebnissen aus dem QS-Verfahren 17/1 im Erfassungsjahr 2017 genauer verdeutlicht. In Tabelle 7 ist die Über- und Unterdokumentation des für das Indikatorergebnis relevanten Ereignisses „Tod“ über alle überprüften Fälle aller Krankenhausstandorte hinweg dargestellt. Die Fehldokumentationsrate für dieses Ereignis beträgt insgesamt 0,34 % (4/1.179).

Tabelle 7: Ergebnisse zur Dokumentation des Ereignisses „Tod“ in 17/1

		Ereignis in Patientenakte dokumentiert		Gesamt	
		tot	lebend		
Ereignis für die QS dokumentiert	tot	48	2 (fälschlicherweise in der QS dokumentiert)	50	Überdokumentation 4,00 % (2/50)
	lebend	2 (fälschlicherweise nicht in der QS dokumentiert)	1.127	1.131	
	Gesamt	50	1.129	1.179	
Fehldokumentationsrate 0,34 % (4/1.179)		Unterdokumentation 4,00 % (2/50)			

Von 1.179 Fällen waren zwei Fälle gemäß den Angaben in der Patientenakte verstorben, die in der QS-Dokumentation allerdings als lebend dokumentiert wurden. Es ist zu erkennen, dass der prozentuale Anteil der Unterdokumentation des für das Indikatorergebnis relevanten Ereignisses mit 4,0 % gering ausfällt.

In einem weiteren Beispiel zum selbigen QS-Verfahrens 17/1 wird für das Ereignis „Spezifische Komplikationen“ in Tabelle 8 ebenfalls die Über- und Unterdokumentation dargestellt (zu beachten ist, dass nur eine Auswahl der spezifischen Komplikationen im Datenabgleich überprüft wurde). Die Fehldokumentationsrate für dieses Datenfeld beträgt 0,68 % (8/1.179). Gleichzeitig wurden in 5 von 12 Fällen Komplikationen in der Patientenakte dokumentiert, die allerdings nicht in der QS-Dokumentation angegeben wurden. Die Unterdokumentation des für das Indikatorergebnis relevanten Ereignisses liegt damit in diesem Beispiel bei 42 %, obgleich die Fehldokumentationsrate insgesamt mit unter einem Prozent sehr niedrig ist.

Tabelle 8: : Ergebnisse zur Dokumentation spezifischer Komplikationen (nur im Datenabgleich überprüfte Komplikationen) in 17/1

		Ereignis in Patientenakte dokumentiert		Gesamt	
		Komplikation	Keine Komplikation		
Ereignis für die QS dokumentiert	Komplikation	7	3 (fälschlicherweise in der QS dokumentiert)	10	Überdokumentation 30,00 % (3/10)
	Keine Komplikation	5 (fälschlicherweise nicht in der QS dokumentiert)	1.164	1.169	
	<b>Gesamt</b>	<b>12</b>	<b>1.167</b>	<b>1.179</b>	
<b>Fehldokumentationsrate</b> 0,68 % (8/1.179)		<b>Unterdokumentation</b> 41,67 % (5/12)			

Ermittelt wird also jeweils, wie bereits oben beschrieben, der Anteil der Unterdokumentationen an allen tatsächlich laut Patientenakte vorliegenden für das Indikatorergebnis relevanten Ereignissen. Zusammenfassend liegen relevante Dokumentationsfehler also dann vor, wenn eine festgelegte Grenze für den Anteil der Unterdokumentationen an allen tatsächlich laut Patientenakte vorliegenden Ereignissen überschritten wird.

Aufgrund der Tatsache, dass die Prävalenz derartiger für das Indikatorergebnis relevanter Ereignisse in der Regel niedrig ist, und weil die Festsetzung von individuellen Schwellenwerten je Ereignis nicht praktikabel erscheint, empfiehlt das IQTIG die Festlegung eines Schwellenwertes für den Anteil der Unterdokumentationen *über alle definierten für das Indikatorergebnis relevanten Ereignisse des betrachteten QS-Verfahrens hinweg* (und nicht etwa für jedes Ereignis separat).

## 4.2 Definition des Relevanzkriteriums

Das Relevanzkriterium überprüft die Unterdokumentation von für das Indikatorergebnis relevanten Ereignissen, die über die Rechenregeln Einfluss auf das Ergebnis eines oder mehrerer Qualitätsindikatoren des Verfahrens haben.

### Für das Indikatorergebnis relevante Ereignisse

Dazu werden für die Verfahren im Stichprobenverfahren mit Datenabgleich verfahrensspezifisch wesentliche Ereignisse ausgewählt, die Einfluss auf die QI-Ergebnisse haben. Damit ist jedes Ereignis was betrachtet wird für ein jeweiliges Indikatorergebnis relevant. Die verschiedenen Ereignisse werden anschließend gemeinsam betrachtet (siehe unten). Aus Praktikabilitätsgründen werden ausschließlich dichotome Ereignisse (ja/nein) berücksichtigt. Durch diese Einschränkung sind zwar ggf. relevante Sachverhalte nicht abbildbar, sie ist jedoch notwendig, um zu einem möglichst einfachen und verständlichen Kriterium zu gelangen.

Für das Indikatorergebnis relevante Ereignisse in diesem Sinne können sein:

- Ereignisse, die direkt über binäre Datenfelder abgebildet sind (z. B. „Komplikation = ja“)
- Bedingungen an den Wertebereich von Datenfeldern (z. B. „ENTLGRUND = ‚07‘“; „Körpergewicht > 1500 g“)
- Komplexere Bedingungen wie Und- und Oder-Verknüpfungen (z. B. „alter >= 65 UND ERFASS-GSTURZRISIKO = 1 UND PRAEVMASSNAHMEN = 1“)

Bei der Auswahl werden *dichotome Ereignisse, die den Zähler eines oder mehrerer Qualitätsindikatoren bestimmen*, berücksichtigt. Dabei kommen alle Qualitätsindikatoren, für die solche Ereignisse als einfache Bedingungen formulierbar sind, gleichermaßen in Betracht. In Anbetracht des unten beschriebenen Sentinel-Event-Kriteriums erscheint es zweckmäßig, Sentinel-Event-Indikatoren bei der Ereignis-Definition nicht zu berücksichtigen (siehe Abschnitt 4.4).

Grundsätzlich werden die für das Indikatorergebnis relevanten Ereignisse als *negativ formulierte* Ereignisse verfasst (z. B. Tod, Auftreten von Komplikationen, keine Antibiotikaphylaxe durchgeführt), unabhängig von der Ausrichtung des zugehörigen Indikators. Dies erfordert ggf. die logische Negation des Ausdrucks aus der Rechenregel („Keine Antibiotikaphylaxe durchgeführt“ statt „Antibiotikaphylaxe durchgeführt“).

Für das Indikatorergebnis relevante Ereignisse können sich ergeben:

1. Aus Bedingungen aus der Zähler-Rechenregel (z. B. „Komplikation aufgetreten“)
2. Aus Bedingungen aus der Zähler-Rechenregel in Verbindung mit der Nenner-Rechenregel (z. B. „Keine antimikrobielle Therapie innerhalb von acht Stunden, obwohl kein dokumentierter Therapieverzicht und keine Aufnahme aus anderem Krankenhaus“)
3. Aus Bedingungen aus der Zähler-Rechenregel in Verbindung mit Merkmalen, deren Vorhandensein eine deutliche Erhöhung des Risikos bei risikoadjustierten Indikatoren bewirkt (z. B. binäre Risikofaktoren mit hohem Odds-Ratio)

Falls ein für das Indikatorergebnis relevantes Ereignis nur über ein einzelnes Datenfeld definiert ist entspricht die Unterdokumentationsrate dieses Ereignisses der Unterdokumentation dieses Datenfeldes (1 - Sensitivität).

Bei der Auswahl der entsprechenden Ereignisse ist neben der einfachen Abbildbarkeit der logischen Bedingung zu beachten, dass möglichst kein Datenfeld in zwei oder mehr Ereignisse gleichzeitig eingeht, damit nicht die Fehldokumentation eines einzelnen Feldes dazu führen kann, dass mehrere Ereignisse als unterdokumentiert gewertet werden. Ferner werden so statistische Abhängigkeiten vermieden, die bei der Berechnung von Vertrauensintervallen für die Stichprobe methodische Schwierigkeiten aufwerfen können.

Zur Beurteilung, ob im betrachteten Erfassungsjahr bei einem Krankenhausstandort relevante Dokumentationsfehler vorliegen, wird je Krankenhausstandort der Anteil an Unterdokumentationen in der QS-Dokumentation an allen für das Verfahren ausgewählten Ereignissen, die laut den Patientenakten in der untersuchten Stichprobe vorlagen, ermittelt.

Dieser Wert entspricht dem Wert unten rechts in Tabelle 6 (dort  $5/12 \approx 41,7\%$ ). Um die unterschiedliche Bedeutung der Unterdokumentationen der für die jeweiligen Indikatorergebnisse

relevanten Ereignisse zu berücksichtigen, wird eine Gewichtung der einzelnen Ereignisse vorgeschlagen. Prinzipiell könnte diese Gewichtung rein inhaltlich erfolgen. Einen guten Anhaltspunkt für den Einfluss der Fehldokumentationen auf die QI-Ergebnisse eines Krankenhausstandortes liefert die Prävalenz eines Ereignisses: Falls ein (negativ formuliertes) Ereignis exakt der Rechenregel eines ratenbasierten QI entspricht, ist die durch eine einzelne Unterdokumentation verursachte relative Abweichung im Ergebnis des entsprechenden QI umso höher, je seltener das Ereignis ist.

Aus diesem Grund und um zu verhindern, dass beispielsweise ein einzelnes der definierten Ereignisse aufgrund einer sehr hohen Prävalenz das Gesamtergebnis für den Anteil aller Unterdokumentationen bestimmt, wird eine Gewichtung mit dem Kehrwert der geschätzten Prävalenz des Ereignisses insgesamt vorgenommen (gewichtetes arithmetisches Mittel). Seltene Ereignisse werden damit höher gewichtet als häufigere Ereignisse.

Beispiel für eine Gewichtung mittels inverser Prävalenz der Ereignisse: Für ein QS-Verfahren werden die beiden für die Indikatorergebnisse relevanten Ereignisse „Komplikationen“ (Prävalenz 1 %) und „Keine Antibiotikatherapie“ (Prävalenz 10 %) definiert. In einer Stichprobe von 20 Fällen eines Krankenhausstandortes wird festgestellt, dass ein Fall eine Komplikation aufwies und in drei Fällen keine Antibiotikatherapie stattfand. Dabei waren die Komplikation sowie eine der drei fehlenden Antibiotikatherapien unterdokumentiert. Dann ergibt sich ohne Gewichtung eine Unterdokumentation von  $2/4 = 50\%$  der Ereignisse. Mit Gewichtung beträgt die Unterdokumentation aufgrund der Höhergewichtung der Komplikation sogar ca. 84,6 % (siehe Tabelle 9). Wäre die Komplikation nicht unterdokumentiert gewesen, betrüge die gewichtete Unterdokumentationsrate lediglich  $10/130 \approx 7,7\%$ .

Tabelle 9: Gewichtung der für das Indikatorergebnis relevanten Ereignisse invers proportional zu ihrer Prävalenz

	Unterdokumentation	Prävalenz	Gewicht	Unterdokumentation (gewichtet)
<b>Komplikationen</b>	1 / 1	1 %	$1/0,01 = 100$	100 / 100
<b>Keine Antibiotikatherapie</b>	1 / 3	10 %	$1/0,10 = 10$	10 / 30
<b>Gesamt</b>	<b>2 / 4</b>			<b>110 / 130</b> <b>(<math>\approx 84,6\%</math>)</b>

Falls jedes definierte für das Indikatorergebnis relevante Ereignis die Rechenregel eines Qualitätsindikators berücksichtigt, sorgt die vorgeschlagene Gewichtung approximativ für eine Gleichgewichtung der Qualitätsindikatoren bezüglich der Einflüsse der Fehldokumentationen auf ihre Ergebnisse: In der Regel hat die Unterdokumentation von Ereignissen, die selten auftreten, einen größeren relativen Einfluss auf das Ergebnis eines zugehörigen ratenbasierten QI als die Unterdokumentation von häufigeren Ereignissen.

Die zur Gewichtung notwendigen Prävalenzen der Ereignisse können (mit Methoden ähnlich wie in Abschnitt 2.1 für die Fehldokumentationsraten beschrieben) aus den Ergebnissen des Datenabgleiches für alle untersuchten Standorte bundesweit geschätzt werden. Alternativ wäre eine Ermittlung der Prävalenzen aus den QS-Daten des jeweiligen Verfahrensjahres denkbar, diese können jedoch durch Unter- oder Überdokumentationen von den tatsächlichen Prävalenzen laut Patientenakte abweichen, weshalb das IQTIG zur Umsetzung den ersten Vorschlag bevorzugt.

### Definition

Insgesamt basiert das Relevanzkriterium auf folgendem Wert:

$$X = \frac{\text{Gewichtete Anzahl Unterdokumentationen in der Stichprobe}}{\text{Gewichtete Anzahl aller für das Indikatorergebnis relevanten Ereignisse laut Patientenakten in der Stichprobe}}$$

Demnach ist das Relevanzkriterium definiert als Unterdokumentation spezifisch festgelegter negativ formulierter Ereignisse die für das Indikatorergebnis relevant sind. Der Quotient bezieht sich dabei auf alle ausgewählten Ereignisse sowie Merkmale gemeinsam. Der Anteil an Unterdokumentationen für jedes Einzelereignis wird nicht betrachtet.

Es wird also nicht wie im Häufigkeitskriterium eine Übereinstimmungsrate über alle Fälle berechnet, sondern gezielt der Anteil an Unterdokumentationen unter den (wenigen) tatsächlich aufgetretenen für das Indikatorergebnis relevanten Ereignissen untersucht. So können z. B. auch Fehldokumentationen bei seltenen Komplikationen oder nicht dokumentierte Sterbefälle erkannt werden.

### Auslösung des Relevanzkriteriums

Auf Basis der Ergebnisse des Stichprobenverfahrens mit Datenabgleich (Zufallsstichprobe) des Vorjahres wird folgendes Kriterium für die Auslösung eines erneuten Datenabgleichs bei einem Krankenhausstandort vorgeschlagen:

Für den oben definierten Anteil X wird analog zum Vorgehen beim Häufigkeitskriterium die untere Grenze eines einseitigen 95 %-Vertrauensintervalls ermittelt. Ist diese Grenze > 10 %, so liegen an dem betrachteten Krankenhausstandort relevante Dokumentationsfehler vor. Prinzipiell ist jede Unterdokumentation eines für das Kriterium ausgewählten Ereignisses maßgebend. Der normative Grenzwert von 10 % wurde rein unter Praktikabilitätsaspekten gewählt und kann zu einem späteren Zeitpunkt auf Basis von Erfahrungen mit dem Verfahren in der Praxis angepasst werden.

Durch die Verwendung des Vertrauensintervalls wird die statistische Unsicherheit aus der Stichprobenziehung, die von der Gesamtfallzahl des Standortes abhängt, berücksichtigt.

### Ausschluss von Datenfeldern mit Spezifikationsmängeln

Analog zum Vorgehen beim Häufigkeitskriterium werden die für das Indikatorergebnis relevanten Ereignisse genau dann retrospektiv ausgeschlossen, wenn deren Berechnung durch Fehldo-

kumentationen von Datenfeldern beeinflusst ist, für die Spezifikationsmängel festgestellt wurden. Um Datenfelder zu identifizieren, bei denen die Dokumentationsfehler in der Verantwortung des IQTIG liegen könnten, wird auf die für das Häufigkeitskriterium gewählte Definition zurückgegriffen: Datenfelder, bei denen über alle überprüften Krankenhausstandorte hinweg im Vorjahr > 15 % Dokumentationsfehler auftreten, werden in Hinblick auf mögliche Spezifikationsmängel begutachtet (zur Berechnung des Wertes und zugehöriger Vertrauensintervalle siehe Abschnitt 2.1.1.2).

### Modellrechnung für 17/1 im Erfassungsjahr 2017

Beispielhaft wurden für die Ergebnisse des Stichprobenverfahrens mit Datenabgleich im Erfassungsjahr 2017 für das QS-Verfahren *Hüftgelenknahe Femurfraktur mit osteosynthetischer Versorgung* (17/1) eine Modellrechnung mit dem hier vorgeschlagenen Relevanzkriterium durchgeführt.

Folgende dichotome, für das Indikatorergebnis relevante Ereignisse wurden auf Basis der Rechenregeln der Qualitätsindikatoren aus der QIDB 2017 identifiziert (Tabelle 10):

Tabelle 10: Exemplarisch für die Modellrechnung ausgewählte Ereignisse zum Relevanzkriterium in 17/1

Ereignis	Rechenregel
Präoperative Verweildauer über 24 Stunden (QI 54030)	praeopminutenMin1 % > % 1440
Keine Sturzprophylaxe bei Patientinnen/Patienten über 65 (QI 54050)	alter > = 65 & !(ERFASSGSTURZRISIKO % = % 1 & PRAEVMASSNAHMEN % = % 1)
Implantatfehlagen, Dislokationen, Frakturen (QI 54036)	IMPLANTATFEHLLAGE % = % 1   IMPLANTATDSLOKATION % = % 1   FRAKTUR % = % 1
Spezifische Komplikationen (QI 54029)	NERVENSCHADEN % = % 1   GEFAESSLAESION % = % 1   IMPLANTATFEHLLAGE % = % 1   IMPLANTATDSLOKATION % = % 1   FRAKTUR % = % 1   POSTOPCDC %in % c(2,3)   WUNDDEHISZE % = % 1   NEKROSEWUND % = % 1
Wundhämatome/Nachblutungen (QI 54042)	HAEMATBLUTUN % = % 1
Allgemeine Komplikationen (QI 54042)	PNEUMONIE % = % 1   KARDVASKKOMP % = % 1   THROMBOSEN % = % 1   LUNGEMBOLIE % = % 1   HARNWEGSINF % = % 1   TIA % = % 1   GASTROBLUTUNG % = % 1   NIERENINSUFFIZIENZJL % = % 1
Todesfälle (QI 54046)	ENTLGRUND % = % "07"

<sup>7</sup> Die folgende Modellrechnung basiert auf dem Stand zum Zeitpunkt der Ausarbeitung von Addendum II. Damit sind hier weder die oben beschriebene Gewichtung der Ereignisse umgekehrt proportional zu ihrer Prävalenz, noch neuere Entwicklungen bezüglich der Methodik der Vertrauensintervalle berücksichtigt. Dennoch kann sie als Orientierung sowie zur Verdeutlichung der Definition des Kriteriums dienen.

Tabelle 11 zeigt über alle Krankenhausstandorte hinweg den Anteil der Unterdokumentationen an allen für die jeweiligen Indikatorergebnisse relevanten Ereignisse, die laut Patientenakte für die im Stichprobenverfahren untersuchten Fälle vorlagen. Für „Todesfälle und spezifische Komplikationen“ wurden die Werte bereits in Tabelle 7 bzw. Tabelle 8 angegeben.

Tabelle 11: Unterdokumentation definierter Ereignisse über alle Krankenhausstandorte hinweg

Ereignis	Unterdokumentation
Präoperative Verweildauer über 24 Stunden (QI 54030)	5,31 % (12/226)
Keine Sturzprophylaxe bei Patientinnen/Patienten über 65 (QI 54050)	52,38 % (33/63)
Implantatfehlagen, Dislokationen, Frakturen (QI 54036)	50,00 % (3/6)
Spezifische Komplikationen (QI 54029)	41,67 % (5/12)
Wundhämatome/Nachblutungen (QI 54042)	28,57 % (2/7)
Allgemeine Komplikationen (QI 54042)	34,13 % (43/126)
Todesfälle (QI 54046)	4,00 % (2/50)
<b>Insgesamt</b>	<b>31,85 % (179 / 562)</b>

Das Ergebnis der Auswertung auf Ebene der Krankenhausstandorte ist in

Abbildung 9 dargestellt. Dazu wurde der in Tabelle 11 unter „Insgesamt“ angegebene Anteilswert für jeden einzelnen der 62 untersuchten Krankenhausstandorte berechnet. Anschließend wurde ein einseitiges 95 %-Vertrauensintervall unter Berücksichtigung der Gesamtfallzahl des jeweiligen Standortes über die Stichprobe hinaus ermittelt.<sup>8</sup> Gemäß der obigen Definition ist die untere Grenze dieses Vertrauensintervalls ausschlaggebend für die Einstufung im Relevanzkriterium: Liegt sie oberhalb von 10 %, so liegen relevante Dokumentationsfehler vor. Dies ist in Abbildung 9 für alle Krankenhausstandorte rechts der eingezeichneten senkrechten Linie der Fall.

<sup>8</sup> Die Methodik für das hier verwendete Vertrauensintervall lehnt sich an diejenige des Häufigkeitskriteriums (Stand Ausarbeitung Addendum II) an. Gegebenenfalls werden hier noch Weiterentwicklungen vorgenommen, die zu einer (geringfügigen) Abweichung führen können.

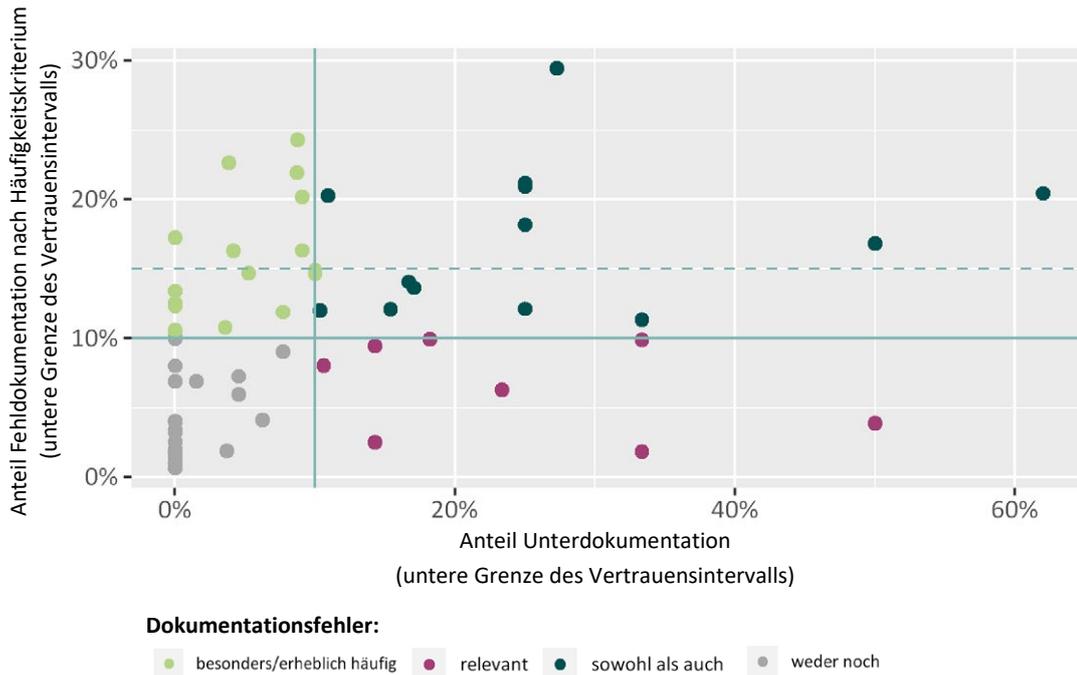


Abbildung 9: Ergebnisse des Häufigkeits- und Relevanzkriteriums im Vergleich

Die Abbildung zeigt die Ergebnisse der 62 untersuchten Standorte für das Häufigkeits- und das Relevanzkriterium im Vergleich. Zur Orientierung sind die Grenzen für besondere Häufigkeit und Relevanz (10 %) sowie für erhebliche Häufigkeit (15 %) dargestellt. Deutlich erkennbar ist, dass Relevanz- und Häufigkeitskriterium durchaus unterschiedliche Sachverhalte abbilden und einige Standorte nur relevante, aber keine häufigen bzw. nur häufige, aber keine relevanten Dokumentationsfehler aufweisen. Insbesondere ergibt sich, dass zusätzlich zum Häufigkeitskriterium für acht weitere Krankenhausstandorte (dunkelrote Punkte) ein wiederholter Datenabgleich aufgrund des Relevanzkriteriums durchgeführt würde.

### 4.3 Statistisches Verfahren (Vertrauensintervall)

In mathematischer Notation kann das Relevanzkriterium wie folgt ausgedrückt werden: Sei  $l_{jk} = 1$ , falls Ereignis  $k$  für Fall  $j$  in der Stichprobe ( $j = 1, \dots, J^*$ ) laut Patientenakte eingetreten ist und  $l_{jk} = 0$  sonst. Sei ferner  $x_{jk} = 1$ , falls das Eintreten nicht in der QS-Dokumentation angegeben wurde (d. h. das Ereignis unterdokumentiert wurde) und  $x_{jk} = 0$  sonst. Dann ist

$$\frac{\sum_{j=1}^{J^*} x_{jk} \cdot l_{jk}}{\sum_{j=1}^{J^*} l_{jk}}$$

der Anteil der Unterdokumentationen von Ereignis  $k$  eines Krankenhausstandortes mit  $J^*$  Fällen in der Stichprobe. Die grundlegende Notation lehnt sich dabei an diejenige aus Kapitel 2 an. Für das Relevanzkriterium wird für jeden Krankenhausstandort das folgende gewichtete arithmetische Mittel berechnet:

$$X = \frac{\sum_{j=1}^{J^*} \sum_{k=1}^K w_k x_{jk} l_{jk}}{\sum_{j=1}^{J^*} \sum_{k=1}^K w_k l_{jk}}$$

Das Stichprobendesign entspricht dem einer Domänen-Stichprobe (Särndal et al. 2003 Kapitel 10), in der für eine Grundgesamtheit von untersuchten Elementen (hier: alle  $K \cdot J$  Ereignisse aller  $J$  Fälle) eine Stichprobe (hier: alle  $K \cdot J^*$  Ereignisse der  $J^*$  Fälle in der Stichprobe) gezogen wird und eine Aussage über eine Teilpopulation (hier: alle Ereignisse, die laut Patientenakten eingetreten sind) gemacht werden soll. Zu ermitteln ist der Domänen-Mittelwert der stetigen Zufallsvariablen

$$X_{jk} := \frac{w_k}{\sum_{k'=1}^K w_{k'}} x_{jk}$$

für  $j = 1, \dots, J^*$  und  $k = 1, \dots, K$  auf der Domänen  $\{(j, k) \mid l_{jk} = 1\}$  derjenigen Kombinationen aus Fällen  $j$  und Ereignissen  $k$ , für die Ereignis  $k$  bei Fall  $j$  eingetreten ist.

Zur Berechnung einer Varianzschätzung können die Methoden aus Särndal et al. (2003) sowie die Funktionen des R-Paketes survey (Lumley 2019) herangezogen werden. Anschließend kann wie in Abschnitt 2.2.2 beschrieben ein logit-transformiertes Vertrauensintervall auf Basis der  $t$ -Verteilung berechnet werden.

#### 4.4 Sentinel-Event-Kriterium

Der gezielte Datenabgleich soll gem. § 9 Abs. 3 Satz 1 Punkt 3 QSHK-RL auch bei festgestellter Nicht-Dokumentation eines Sentinel Events im Vorjahr erfolgen. Insgesamt gibt es derzeit 34 Sentinel-Event-Indikatoren über alle QS-Verfahren. Im Stichprobenverfahren mit Datenabgleich zum EJ 2017 waren keine Sentinel-Event-Indikatoren in den QS-Verfahren enthalten. Für das Stichprobenverfahren mit Datenabgleich zum EJ 2018 gibt es im QS-Verfahren NEO einen Sentinel-Event-Indikator.

Gemäß § 9 QSKH-RL erfolgt für die Unterdokumentation eines Sentinel Event im Vorjahr keine Beschränkung auf das Stichprobenverfahren mit Datenabgleich. Dies bedeutet, dass auch die Ergebnisse aus dem Strukturierten Dialog zu den Qualitätsindikatoren herangezogen werden müssten und es damit zu einer Ausweitung auf alle QS-Verfahren kommt. Somit müssten Sentinel Events

- a) aus dem Stichprobenverfahren mit Datenabgleich sowie
- b) aus dem Strukturierten Dialog zu den Qualitätsindikatoren

betrachtet werden.

Eine Unterdokumentation von Sentinel Events im Strukturierten Dialog zu den QI kann aber allenfalls zufällig erkannt und nicht systematisch ermittelt werden. Dies bedeutet, dass eine Umfrage unter den Bundesländern erfolgen müsste um zu erfahren, inwiefern zufällig beim Strukturierten Dialog von anderen Qualitätsindikatoren auch eine weitere Unterdokumentation eines Sentinel Events entdeckt wurde. Die Erfolgsquote wird vom IQTIG zunächst als sehr gering eingeschätzt. Somit bleibt zu hinterfragen, inwiefern der Aufwand einer solchen Umfrage gerechtfertigt ist. Trotz der nicht in § 9 QSKH-RL erfolgten Beschränkung für die Unterdokumentation

eines Sentinel Event auf das Stichprobenverfahren mit Datenabgleich empfiehlt das IQTIG deshalb, diese nicht auf den Strukturierten Dialog der Qualitätsindikatoren auszuweiten, weshalb nur Punkt a) weiter betrachtet werden soll.

Ein Sentinel Event kann entsprechend der Definition des Relevanzkriteriums als dichotomes relevantes Ereignis aufgefasst werden. Sentinel-Event-Indikatoren werden jedoch nicht in das oben beschriebene Kriterium einbezogen, sondern gesondert betrachtet. Hier gilt, dass bereits die festgestellte Unterdokumentation eines einzigen Sentinel Events zu einem gezielten Datenabgleich führt.

#### **Zusammenhang zum Relevanzkriterium**

Bezüglich der im Rahmen des Stichprobenverfahrens mit Datenabgleich festgestellten Unterdokumentation von Sentinel Events (Punkt a)) besteht folgender Zusammenhang zum oben definierten Relevanzkriterium: Sentinel Events können als relevante Ereignisse aufgefasst werden (Zähler in Verbindung mit Nenner). Die Feststellung der Nicht-Dokumentation eines Sentinel Events entspricht also der Unterdokumentation eines negativ formulierten Ereignisses.

Im Unterschied zum Relevanzkriterium wird die Unterdokumentation von Sentinel Events als relevante Ereignisse aber einzeln statt aggregiert betrachtet und es löst bereits eine einzige Unterdokumentation einen gezielten Datenabgleich aus und nicht erst eine Unterdokumentation über 10 %.

#### **Ausschluss von Datenfeldern mit Spezifikationsmängeln**

Analog zum Vorgehen beim Häufigkeitskriterium werden Sentinel Events retrospektiv ausgeschlossen, deren Berechnung durch Fehldokumentationen von Datenfeldern beeinflusst ist, für die Spezifikationsmängel festgestellt wurden. Um Datenfelder zu identifizieren, bei denen die Dokumentationsfehler in der Verantwortung des IQTIG liegen könnten, wird auf die für das Häufigkeitskriterium gewählte Definition zurückgegriffen: Datenfelder, bei denen über alle überprüften Krankenhausstandorte hinweg im Vorjahr > 15 % Dokumentationsfehler auftreten, werden in Hinblick auf mögliche Spezifikationsmängel begutachtet (zur Berechnung des Wertes und zugehöriger Vertrauensintervalle siehe Abschnitt 2.1.1.2).

### **4.5 Prozessablauf bei Unterdokumentation relevanter Ereignisse und Nicht-Dokumentation eines Sentinel Event**

Die Auswahl relevanter Ereignisse wird jährlich durch das IQTIG im Rahmen des Routineverfahrens der Datenvalidierung vorgenommen. Dabei werden aus Praktikabilitätsgründen grundsätzlich nur dichotome Ereignisse (ja/nein) berücksichtigt. Somit erfolgt eine jährliche Festlegung entsprechend der ausgewählten QS-Verfahren für das Stichprobenverfahren mit Datenabgleich.

Für den Beschluss der Ereignisse (bezogen auf die Leistungsbereiche des Stichprobenverfahrens mit Datenabgleich eines Erfassungsjahres) und die damit einhergehend abgefragten Datenfelder und Rechenregeln stellt das IQTIG dem Unterausschuss Qualitätssicherung des G-BA eine beschlussfähige Liste bereit.

Die Ermittlung einer Nicht-Dokumentation eines Sentinel Event wird erstmals im QS-Verfahren Neonatologie zum 15. April 2020 auf Basis der Ergebnisse des Stichprobenverfahrens mit Datenabgleich zum EJ 2018 des Sentinel-Event-Indikators mit der QI-ID 51070 „Sterblichkeit bei Risiko-Lebendgeborenen (ohne zuverlegte Kinder)“ herangezogen werden.

Liegen relevante Fehldokumentationen der ausgewählten Ereignisse aus dem Stichprobenverfahren mit Datenabgleich vor, sollen diese dazu führen, dass ein gezielter Datenabgleich im Folgejahr erfolgt. In Abbildung 10 ist das Verfahren des gezielten Datenabgleichs für das Relevanzkriterium dargestellt.

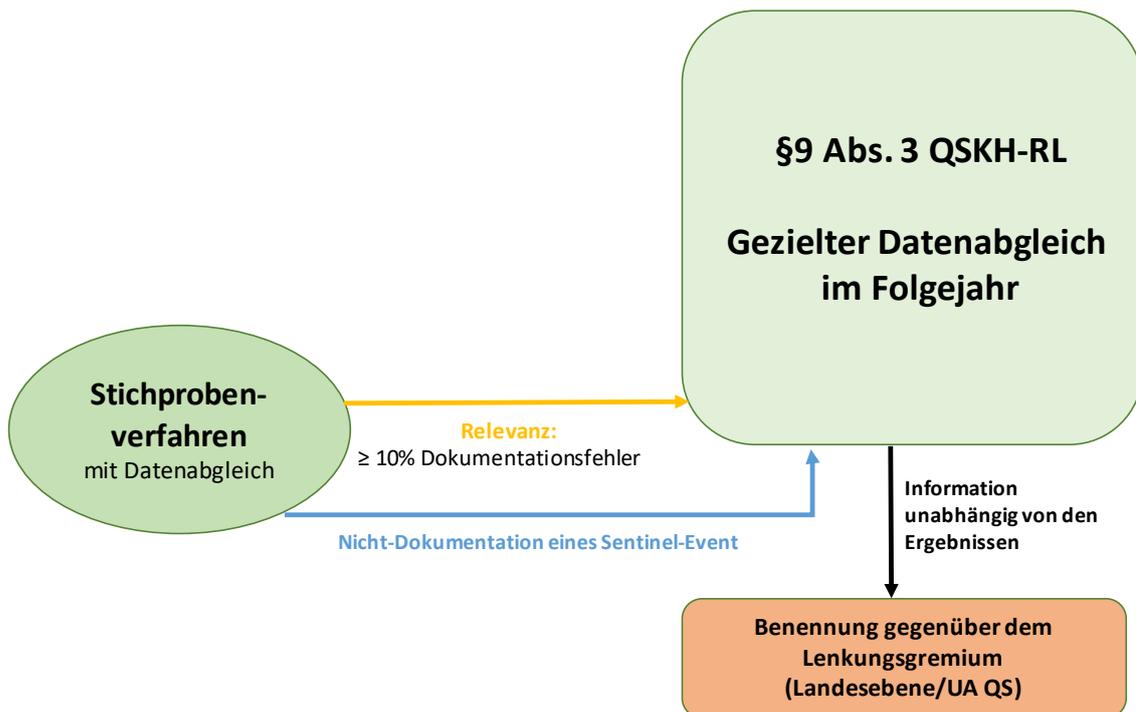


Abbildung 10: Gezielter Datenabgleich nach relevanten Dokumentationsfehlern und der Nicht-Dokumentation eines Sentinel Event.

Für die erneute Überprüfung eines Standortes wird eine Stichprobe der Fälle in gleicher Größe wie für das Stichprobenverfahren mit Datenabgleich gemäß § 9 Abs. 4 QSKH-RL (20 Fälle) herangezogen. Die Stichprobenziehung der Fälle für jeden zu überprüfenden Standort soll sich dabei nach Möglichkeit auf Fälle aus dem dritten und vierten Quartal des Vorjahres beziehen, da in der dazwischenliegenden Zeit geeignete Verbesserungsmaßnahmen zur Dokumentationsqualität erfolgt sein sollten. Ebenfalls werden nach Möglichkeit die gleichen Datenfelder erneut überprüft.

Unter folgenden Konstellationen soll ein gezielter Datenabgleich im Folgejahr ausgelöst werden, oder eine Benennung gegenüber dem entsprechend zuständigen Lenkungsgremium erfolgen:

- Bei erheblichen Fehldokumentationen erfolgt die Benennung gegenüber dem entsprechenden Lenkungsgremium (Beschluss vom 6. März 2019)
- Bei besonders häufigen Fehldokumentationen erfolgt ein gezielter Datenabgleich (Beschluss vom 6. März 2019)

- Bei Vorliegen von Fehldokumentationen, die alleinig das Relevanzkriterium betreffen, wird ein gezielter Datenabgleich im Folgejahr ausgelöst
- Bei gemeinsamem Vorliegen von besonders häufigen und relevanten Fehldokumentationen wird ein gezielter Datenabgleich im Folgejahr ausgelöst
- Bei Vorliegen von Nicht-Dokumentationen eines Sentinel Events wird ein gezielter Datenabgleich im Folgejahr ausgelöst
- Das Vorliegen von erheblich häufigen und relevanten Fehldokumentationen löst bereits durch das Häufigkeitskriterium die Entanonymisierung gegenüber dem entsprechenden Lenkungsgremium aus. (s. o.)

Das verantwortliche Gremium nach § 14 QSKH-RL soll im Nachhinein unabhängig von den Auswertungsergebnissen über diese informiert werden (Übersendung des bundeseinheitlichen Prüfberichts). Da eine Veröffentlichung der Ergebnisse in den Qualitätsberichten der Krankenhäuser nach § 136b Abs. 1 Satz 1 Nr. 3 SGB V vorgesehen ist, geschieht die Informationsweitergabe unabhängig davon, ob in der erneuten Überprüfung besonders häufige oder erheblich häufige oder relevante Dokumentationsfehler vorlagen. Die Einordnung der Ergebnisse des gezielten Datenabgleichs obliegt damit dem Gremium nach § 14 QSKH-RL, welches sich an den vom IQTIG vorgeschlagenen Kriterien zu besonders häufigen und relevanten Dokumentationsfehlern orientieren kann.

### **Perspektivische Überlegungen**

Bei der Festlegung der Ereignisse zur Ermittlung relevanter Fehldokumentationen werden wie oben beschrieben aus Praktikabilitätsgründen ausschließlich dichotome Ereignisse (ja/nein) berücksichtigt. Hier zeichnet sich ggf. zukünftig Weiterentwicklungspotenzial ab. Grundsätzlich ist vorgesehen, die vom IQTIG vorgeschlagenen Kriterien zum gezielten Datenabgleich, insbesondere die Ereignisse zum Relevanzkriterium, im Rahmen des Routineverfahrens unter Einbindung der Beratung durch externe Dritte (z. B. der Projektgruppe Datenvalidierung) zu begutachten und ggf. zu anpassen. Ein umfassendes Evaluationskonzept ist bisher noch nicht beauftragt.

## **4.6 Limitationen**

Das IQTIG weist auf mögliche Limitationen für die Definition und Auslösung des Relevanzkriteriums hin. Aufgrund einer nur orientierenden Literaturrecherche können Publikationen zu relevanten Datenfeldern oder relevanten Fehldokumentationen unentdeckt geblieben sein. Des Weiteren impliziert Relevanz bereits eine Wertung aus einer bestimmten Perspektive, weshalb auch keine weitere Literatur bezüglich möglicher Grenzwerte für die Auslösung des Relevanzkriteriums hinzugezogen wurde. Vielmehr erfolgte die Festlegung normativ unter Berücksichtigung von statistischen Probeanalysen.

Grundsätzlich ist anzumerken, dass Fälle mit seltenen oder schweren unerwünschten bzw. negativ formulierten Ereignissen trotz der methodischen Annäherung weiterhin unentdeckt bleiben können, wie die Beispiele für die Unterdokumentation von „Tod“ und „Komplikationen“ des QS-Verfahrens 17/1 anhand der Daten zum EJ 2017 zeigen. Die zwei unterdokumentierten Todesfälle werden durch die normative Festlegung eines maximal tolerierten Anteils von > 10 %

Unterdokumentationen über alle definierten Ereignisse hinweg auch zukünftig nicht entdeckt werden. Ferner können durch den Ansatz, die Unterdokumentation über alle Ereignisse hinweg und nicht pro Ereignis zu betrachten, spezifische Fehldokumentationen eines Standortes für ein bestimmtes Ereignis ggf. nicht von dem hier vorgestellten Kriterium detektiert werden. Angesichts der Tatsache, dass ansonsten aufwändig ereignisindividuelle Schwellenwerte definiert werden müssten, erscheint dies jedoch vertretbar.

Aus Praktikabilitätsgründen soll nur die Unterdokumentation der definierten jeweils für die Indikatorergebnisse relevanten Ereignisse betrachtet werden. Fehldokumentationen, die zur Überdokumentation führen, werden derzeit nicht in das Verfahren eingeschlossen.

Als mögliche weitere Limitation könnte angemerkt werden, dass das Relevanzkriterium nicht bei der Festlegung der Definition von erheblichen Dokumentationsmängeln berücksichtigt wird und es folglich bei Vorliegen relevanter Fehldokumentationen nicht zu einer Benennung gegenüber dem entsprechenden Lenkungsgremium kommen soll, sondern zunächst zu einer erneuten Überprüfung durch einen gezielten Datenabgleich. Da das Relevanzkriterium ein gänzlich neues Konstrukt ist, empfiehlt das IQTIG zunächst dessen Einführung und Praxistauglichkeit abzuwarten. Im Rahmen einer möglichen Evaluation des gezielten Datenabgleichs könnte bei hoher Treffsicherheit des Relevanzkriteriums über eine Integration in die Definition erheblicher Dokumentationsmängel nachgedacht werden.

Die kleine Stichprobengröße von nur 20 Fällen pro Krankenhausstandort führt in Kombination mit der in der Regel sehr niedrigen Prävalenz der betrachteten Ereignisse dazu, dass Entscheidungen ggf. aufgrund von sehr wenigen Beobachtungen getroffen werden müssen. Hier macht sich bemerkbar, dass die verwendete Zufallsstichprobe nicht gezielt zur Betrachtung der Ereignisse des Relevanzkriteriums gezogen wurde. Prinzipiell können auf diese Weise Konstellationen auftreten, in denen die Fehldokumentation eines einzelnen Datenfeldes zur Auslösung eines gezielten Datenabgleichs führt. Durch die Aggregation der Ereignisse in der Berechnung des Kriteriums wird diesen Umständen bestmöglich Rechnung getragen. Ferner kann letztlich jede Unterdokumentation eines für das Indikatorergebnis relevanten Ereignisses eine wesentliche Auswirkung auf ein QI-Ergebnis des Standortes haben und ist dementsprechend zu problematisieren. Eine zukünftige Erhöhung des Grenzwertes (beispielsweise auf 20 %) ist möglich.

Perspektivisch ist für das Relevanzkriterium genau wie schon für das Häufigkeitskriterium auch eine gemeinsame Bewertung der Ergebnisse der beiden betrachteten Jahre (Stichprobenverfahren mit Datenabgleich und gezielter Datenabgleich) unter Verwendung der im Vorbericht zur Weiterentwicklung des Strukturierten Dialogs beschriebenen methodischen Ansätze denkbar.

Je nach Dokumentationsqualität der Standorte kann sich der Aufwand für die Bundesländer durch die Anzahl der nochmals zu überprüfenden Standorte im gezielten Datenabgleich unterscheiden. Ein Vorgehen zur Aufwandsbegrenzung wird nach Betrachtung der aktuellen Zahlen zum gezielten Datenabgleich 2019 jedoch zunächst nicht empfohlen.

Fraglich bleibt allerdings wie oben bereits angemerkt, inwiefern weitere Unterdokumentationen von Sentinel Events im Strukturierten Dialog zu den Qualitätsindikatoren entdeckt werden können. Die Erfolgsquote einer möglichen Umfrage hierzu unter den Bundesländern wird vom IQTIG sowohl als gering als auch als zu aufwändig im Verhältnis zum Nutzen eingeschätzt. Für das Stichprobenverfahren mit Datenabgleich können die Unterdokumentationen von Sentinel Events am IQTIG berechnet werden, wodurch eine verlässlichere Basis zur Erkennung dieses Kriterium vorhanden ist.

Auch an dieser Stelle soll noch einmal darauf hingewiesen werden, dass grundsätzlich für die Ergebnisse des gezielten Datenabgleichs zu beachten ist, dass aufgrund der geringen Stichprobengröße von 20 Fällen ein Vergleich der Ergebnisse mit den Ergebnissen des Stichprobenverfahrens mit Datenabgleich im Vorjahr im Sinne einer Trendaussage nur eingeschränkt möglich ist. Daher empfiehlt das IQTIG dem jeweils zuständigen Gremium nach § 14 QSKH-RL, die zeitliche Entwicklung der Fehldokumentationsraten in den beiden Stichproben nicht in die Bewertung der Dokumentationsfehler einzubeziehen. Perspektivisch ist hier auch eine gemeinsame Bewertung der Ergebnisse der beiden betrachteten Jahre unter Verwendung der im Vorbericht zur Weiterentwicklung des Strukturierten Dialogs beschriebenen methodischen Ansätze denkbar.

## 5 Addendum III zum Kriterium der wiederholten rechnerischen Auffälligkeit

### Gegenstand der Beauftragung Nr. I.2

Gegenstand der Beauftragung ist es, die Kriterien für wiederholte rechnerische Auffälligkeiten in Bezug auf die Ergebnisse der Statistischen Basisprüfung zu definieren.

Die Kriterien sollen berücksichtigen, inwiefern Krankenhausstandorte zum wiederholten Mal in einem Auffälligkeitskriterium rechnerisch auffällig werden. In diesem Kapitel wird das Kriterium für wiederholte rechnerische Auffälligkeiten definiert. Die Krankenhausstandorte für den gezielten Datenabgleich auf der Grundlage des Kriteriums zur wiederholten rechnerischen Auffälligkeit könnten zum 15. April 2024 erstmals ermittelt werden und vor Ort erneut im Rahmen eines gezielten Datenabgleichs überprüft werden.

### 5.1 Methodisches Vorgehen

Bezogen auf wiederholte rechnerische Auffälligkeiten werden die Ergebnisse aus dem Strukturierten Dialog zu den Auffälligkeiten betrachtet. In den Auffälligkeitskriterien zur Plausibilität und Vollständigkeit sowie zur Vollzähligkeit lagen in den letzten Jahren unterschiedlich viele wiederholte rechnerische Auffälligkeiten vor. Diese kann zum einen auf unterschiedliche Berechnungsarten in den Berichten zurückzuführen sein oder auch auf eine variierende Anzahl von Auffälligkeitskriterien (AK) in den Verfahren.

#### Ergebnisse der SD-Berichte zu den Erfassungsjahren 2015, 2016 und 2017:

Aus dem Datenvalidierungsbericht für das EJ 2015 wurde die durchschnittliche Auffälligkeit pro AK und pro QS-Verfahren für zwei aufeinanderfolgende Jahre ermittelt. Bei den Auffälligkeitskriterien zur Plausibilität und Vollständigkeit waren durchschnittlich 2,3 Standorte pro AK wiederholt rechnerisch auffällig im EJ 2015 (d. h. auf Basis des EJ 2014 und 2015). Dabei lagen Schwankungen zwischen 0 und 16 Standorten vor. Pro AK wiesen durchschnittlich 0,7 Standorte eine im SD festgestellte wiederholte Fehldokumentation auf (Minimum: 0; Maximum: 5).

Für die Auffälligkeitskriterien zur Vollzähligkeit lagen zum EJ 2015 durchschnittlich pro QS-Verfahren für die Unterdokumentation 1,6 (min: 0; max: 10), für die Überdokumentation 0,4 (min: 0; max 3) und für Minimaldatensätze 1,4 (min: 0, max: 13) rechnerisch auffällige Standorte vor, die ebenso bereits im Vorjahr rechnerisch auffällig waren.

Aus dem Datenvalidierungsbericht zum EJ 2016 geht hervor, dass über alle AKs zur Plausibilität und Vollständigkeit von den Standorten, die die 1.406 rechnerischen Auffälligkeiten des EJ 2016 aufwiesen, 142 schon im EJ 2015 und 16 bereits sowohl im EJ 2015 als auch im EJ 2014 rechnerisch auffällig waren. Über alle AKs zur Vollzähligkeit waren von den Standorten, die 1.194 rechnerische Auffälligkeiten des EJ 2016 aufwiesen, 314 schon im EJ 2015 und 19 auch sowohl im EJ 2015 als auch im EJ 2014 rechnerisch auffällig. Auf eine zusammenfassende tabellarische Darstellung wurde in diesem Bericht verzichtet.

Im Datenvalidierungsbericht für das EJ 2017 wurden nicht die durchschnittlich auffälligen Standorte pro AK ermittelt, sondern die absolute Zahl aller auffälligen Standorte für Auffälligkeitskriterien zur Plausibilität und Vollständigkeit, sowie zur Vollzähligkeit. Für die Auffälligkeiten zur Plausibilität und Vollständigkeit lagen 153 wiederholt rechnerische Auffälligkeiten im EJ 2017 auf Grundlage der Auffälligkeiten für EJ 2016 vor. Bezogen auf wiederholte Auffälligkeiten der letzten drei Erfassungsjahre waren 32 Standorte wiederholt auffällig. Der Anteil der wiederholten fehlerhaften Dokumentation zum Vorjahr lag bei 14,2 % (73 Standorte) und bezogen auf drei Jahre bei 2,5 % (13 Standorten). Für die AKs zur Vollzähligkeit konnte eine wiederholte rechnerische Auffälligkeit für 259 Standorte in Bezug auf das Vorjahr (EJ 2016) festgestellt werden. In Bezug auf drei aufeinanderfolgende Jahre waren es 78 Standorte (EJ 2017 und auffällig in EJ 2016 und 2015). Die wiederholte fehlerhafte Dokumentation zum Vorjahr lag mit 25,8 % (140 Standorten) höher als bei den Plausibilitäts-AKs. Die wiederholt fehlerhafte Dokumentation bezogen auf drei Jahre in Folge lag bei 2,6 % (14 Standorte).

### Betrachtung von wiederholt auffälligen AKs vs. wiederholt auffälligen QS-Verfahren

Das IQTIG empfiehlt die getrennte Betrachtung der wiederholten rechnerischen Auffälligkeit für jedes Auffälligkeitskriterium. Alternativ wäre auch eine Betrachtung pro QS-Verfahren, d. h. Auslösung eines gezielten Datenabgleiches bei wiederholter rechnerischer Auffälligkeit in mindestens einem AK eines QS-Verfahrens, denkbar. Dies erscheint aufgrund der unterschiedlichen Anzahl und Art von Auffälligkeitskriterien je QS-Verfahren jedoch nicht als sinnvoll. In Tabelle 1 ist die Anzahl der Auffälligkeitskriterien je QS-Verfahren für das EJ 2018 dargestellt.

Die Anzahl der Auffälligkeitskriterien zur Plausibilität und Vollständigkeit schwankte zwischen 0 bis 6 AKs in den indirekten Verfahren und zwischen 0 bis 4 AKs in den direkten Verfahren. Die Auffälligkeitskriterien zur Vollzähligkeit sind etwas konstanter verteilt, da in jedem QS-Verfahren AKs zur Unter- und Überdokumentation vorhanden sind. Bei alleiniger Betrachtung von Plausibilitäts-AKs je Verfahren wäre es in einem Verfahren mit mehreren AKs, wie z. B. *Ambulant erworbene Pneumonie*, eher wahrscheinlich innerhalb von z. B. drei Jahren wiederholt rechnerisch auffällig zu werden, als bei den Verfahren die z. B. nur ein Auffälligkeitskriterium zur Plausibilität und Vollständigkeit aufweisen. Hinzu kommt: auch wenn es mehrere AKs pro Verfahren gibt, können die Auffälligkeitskriterien aufgrund ihrer Unterschiedlichkeit inhaltlich häufig nicht zusammengefasst werden.

Tabelle 12: Anzahl der Auffälligkeitskriterien je QS-Verfahren für das EJ 2018

QS-Verfahren	AKs zur Plausibilität und Vollständigkeit	AKs zur Vollzähligkeit				
		Unter- und Überdokumentation	MDS	GKV-Unterdokumentation	Doku-Raten FU*	Zählleistungsbereiche
09n1-HSM-IMPL	2	2	1	1		
09n2-HSM-AGGW	1	2	1	1		
09n3-HSM-REV	4	2	1	1		
09n4-DEFI-IMPL		2	1	1		

QS-Verfahren	AKs zur Plausibilität und Vollständigkeit	AKs zur Vollständigkeit				
		Unter- und Überdokumentation	MDS	GKV-Unterdokumentation	Doku-Raten FU*	Zähleleistungsbereiche
09n5-DEFI-AGGW		2	1	1		
09n6-DEFI-REV		2	1	1		
10n2-KAROT	1	2	1			
15n1-GYN-OP	1	2	1			
16n1-GEBH	2	2	1			
17n1-HÜFT-FRAK	3	2	1			
HEP	3	2	1	1		4
KEP	4	2	1	1		4
18n1-MAMMA	4	2	1			
DEK	3	2 + 3	1			
NEO	6	2	1			
PNEU	5	2	1			
HCH-AORT-CHIR	2	2				4
HCH-AORT-KATH	2					
HCH-KCH	2					
HCH-KOMB	2					
HCH-D			1			
HTXM-D	3	2			3	2
HTXM-TX	4	2			3	2
HTXM-MKU	3	2			3	
LTX	3	2			3	
LUTX	3	2			3	
LLS		2			3	
NLS						
PNTX-D						

\* Die Dokumentationsrate wird auf Standortebene betrachtet und greift nicht auf die Soll-Statistik zurück.

### Problematik des Einbezugs von Auffälligkeitskriterien zur Vollständigkeit

Das IQTIG empfiehlt, die Betrachtung der wiederholten rechnerischen Auffälligkeit auf die AKs zu Vollständigkeit und Plausibilität aus zwei Gründen zu beschränken.

1. Da die Soll-Statistik ab 2020 auf IKNR-Ebene basiert, kann keine standortbezogene Auswertung für die Auffälligkeitskriterien zur Vollständigkeit vorgenommen werden, welche auf der

Soll-Statistik basieren. Dies bedeutet, dass wiederholt rechnerische Auffälligkeiten zwar festgestellt werden können, allerdings nicht auf Standortebene, sondern auf Krankensebene. Bei Einbezug in den gezielten Datenabgleich bliebe bei dem Krankenträger zu ermitteln, für welchen Standort Fehldokumentationen vorlagen und der gezielte Datenabgleich erfolgen sollte, falls das überhaupt möglich ist. Dies ist in dem Sinne ungünstig, da nicht genau nachvollzogen werden kann, ob derjenige Standort überprüft wird, der auch für die rechnerischen Auffälligkeiten verantwortlich ist. Gegebenenfalls kann die rechnerische Auffälligkeit bei einem Vollzähligkeits-AK auch erst durch die gemeinsame Betrachtung aller Standorte entstehen. Die Überprüfung der Gründe für die rechnerischen Auffälligkeiten sowie weitere Maßnahmen sind an dieser Stelle besser durch den Strukturierten Dialog zu adressieren. Möglich wäre in diesen Fällen nicht nur schriftliche Stellungnahmen einzuholen, sondern weitere verpflichtende Maßnahmen folgen zu lassen, wie z. B. ein kollegiales Gespräch o. ä. Hiervon ausgenommen werden könnten Auffälligkeiten zur Unterdokumentation, für die bereits Sanktionen nach §24 QSKH-RL bestehen.

2. Grundsätzlich handelt es sich bei den AKs zur Unter- und Überdokumentation sowie zu den Minimaldatensätzen um Datenfehler, die dadurch entstehen, dass beispielsweise nicht alle Fälle übermittelt wurden (wie z. B. bei der Unterdokumentation). Somit ist bei diesen Auffälligkeiten nicht von einem Dokumentationsfehler bestimmter Datenfelder auszugehen, sondern von Datenfehlern bzgl. der Übertragung, die z. B. auch auf technische o. a. Ursachen zurückgehen können. Eben weil die Ursachen nicht bezogen auf Fehldokumentationen in Datenfeldern zurückzuführen sind, würde ein gezielter Datenabgleich mit der Überprüfung von QI-relevanten Datenfeldern für einen Leistungsbereich sein Ziel verfehlen. Aus diesem Grund werden AKs zur Unterdokumentation und Überdokumentation und zu den Minimaldatensätzen von den nachfolgenden methodischen Überlegungen ausgeschlossen. Dies gilt auch für die Zählleistungsbereiche und die Unterdokumentation von GKV-Patientinnen und GKV-Patienten. Bzgl. der AKs zum Follow-up (Dokumentationsraten) gilt zwar, dass diese nicht auf die Soll-Statistik zurückgreifen und auf Standortebene betrachtet werden können (keine Nutzung der IKNR) und im weiteren Sinn zur Vollzähligkeit gehören, dennoch eine Überprüfung im gezielten Datenabgleich ihr Ziel verfehlen würde, da diese AKs nicht die konkrete Dokumentation in der Patientenakte, sondern das Ausfüllen des Follow-up-Bogens der QS an sich überprüfen.

## 5.2 Modellrechnungen

Berücksichtigt sind alle Auffälligkeitskriterien zur Vollständigkeit und Plausibilität, nicht jedoch die Auffälligkeitskriterien zur Vollzähligkeit.

Datengrundlage sind die am IQTIG berechneten Ergebnisse zu rechnerischen Auffälligkeiten in den AKs für die Auswertungen zu den Jahren 2016, 2017 und 2018 – also nicht die von den Bundesländern an das IQTIG zurückgemeldeten SD-Ergebnisse. Hierdurch können sich Abweichungen von den im Bericht zur Datenvalidierung veröffentlichten Zahlen ergeben.

In Abbildung 11 sind alle Auffälligkeitskriterien aufgeführt, die über einen Zeitraum von drei Jahren konstant geblieben sind. Bei Betrachtung wiederholter rechnerischer Auffälligkeiten über diesen Zeitraum konnten in neun QS-Verfahren Standorte im jeweils gleichen AK bezogen auf die Erfassungsjahre EJ 2016, 2017 und 2018 festgestellt werden (vgl. Tabelle 13). Insgesamt gab es 44 wiederholt rechnerisch auffällige Standorte, die in einem gezielten Datenabgleich überprüft werden müssten.

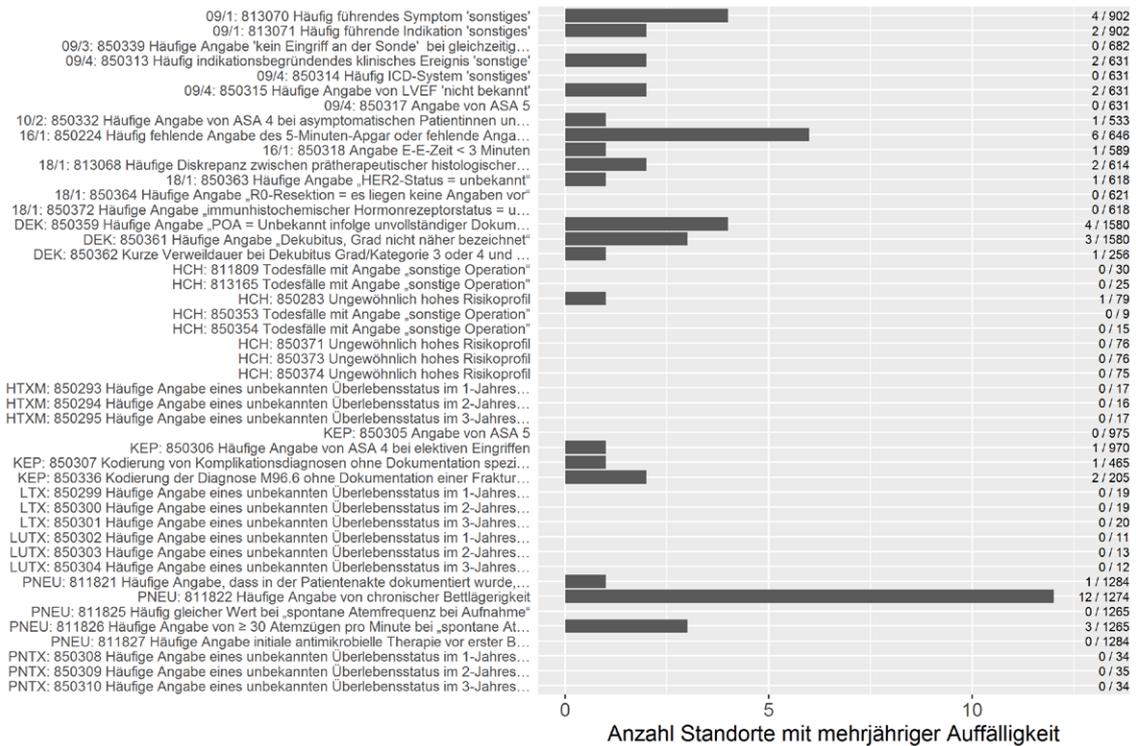


Abbildung 11: Anzahl der Standorte mit mehrjähriger Auffälligkeit je Auffälligkeitskriterium für die Erfassungsjahre 2016, 2017 und 2018 (Grundgesamtheit: Standorte, die in allen drei Jahren für das jeweilige AK relevante (Nenner-)Fälle hatten).

Tabelle 13: Wiederholt rechnerisch auffällige Standorte pro QS-Verfahren bezogen auf EJ 2016, EJ 2017 und EJ 2018.

09/1	09/4	10/2	17/1	18/1	DEK	HCH	KEP	PNEU	Gesamt
5	4	1	7	3	8	1	4	16	44

Würden ausschließlich Standorte einbezogen werden, die in zwei aufeinanderfolgenden Jahren in demselben AK rechnerisch auffällig wären, läge die Anzahl deutlich höher. Beispielrechnungen für EJ 2017 und EJ 2018 zeigen, dass insgesamt 158 Standorte in 14 QS-Verfahren anhand der momentan vorliegenden Anzahl an AKs überprüft werden müssten (vgl. Tabelle 14). Hierdurch würde der Aufwand zur Überprüfung der Standorte in einem gezielten Datenabgleich insbesondere auf Landesebene enorm erhöht, sodass das IQTIG von einer Überprüfung aufgrund von wiederholter rechnerischer Auffälligkeit innerhalb eines Zwei-Jahres-Zeitraums absehen

würde. Diese Empfehlung beruht einzig auf Praktikabilitätsgründen und müsste zukünftig ggf. angepasst werden, wenn sich die Anzahl der AKs ändern sollte.

Tabelle 14: Wiederholt rechnerisch auffällige Standorte pro QS-Verfahren bezogen auf EJ 2017 und EJ 2018.

09/2	09/3	09/4	10/2	15/1	16/1	17/1	18/1	DEK	HCH	HEP	KEP	NEO	PNEU	Gesamt
12	2	8	1	3	13	11	11	35	3	10	10	9	42	<b>158</b>

Für das Kriterium zu wiederholten rechnerischen Auffälligkeiten gelten ein paar Besonderheiten, die im Nachhinein geschildert werden:

### 5.3 Herausforderungen zum Kriterium zur wiederholten rechnerischen Auffälligkeit

Änderungen bezüglich der Standorte (Schließung, Standortverlagerung von Leistungsbereichen, Übernahme von anderem Krankenhaus oder andere Gründe für einen Wechsel des Instituts-kennzeichens) innerhalb und zwischen den Erfassungsjahren müssen der BAS durch die DAS bis zum jeweiligen Datenannahmeschluss als sogenanntes Standortmapping bereitgestellt werden. Mithilfe dieser Information kann die Bundesauswertungsstelle die Daten eines Leistungserbringerstandortes, die aufgrund der Standortkennungsänderungen mit unterschiedlichen Leistungserbringerpseudonymen geliefert werden, zusammenführen. Da zum 1. Januar 2020 die neue Standortdefinition gemäß § 293 SGB V verpflichtend für die Krankenhäuser gilt, führt dies im Vergleich zu aktuell geführten Standorten eines Krankenhauses ggf. zu einer geänderten Standortaufteilung (insbesondere unter dem Gesichtspunkt der räumlichen Ausdehnung, die die definitorische Grundlage der neuen Standortdefinition bildet). Ein offizielles Mapping der Standorte aus den vorherigen Erfassungsjahren mit denen zum Erfassungsjahr 2020 und folgende ist aktuell nicht beauftragt. Darüber hinaus wird dies für einen gewissen Teil der Standorte nicht möglich sein, wenn bei neuen Standortstrukturen eine in-Beziehung-Setzung von neuen und alten Standorten unmöglich ist (z. B. bei Aufteilung eines bisher geführten Standortes auf ein bis mehrere Standorte).

Ab dem Jahr 2020 werden die standortbezogenen QS-Daten (QSKH) anstelle der bisher genutzten Kombination aus Institutionskennzeichen (IKNR) und 2-stelligen Standortkennzeichen mit einem neuen 6-stelligen Standortkennzeichen an das IQTIG übermittelt. Durch die möglicherweise geänderte Zuordnung von Standorten und das fehlende Mapping können bei zusammenfassender Betrachtung von Daten über mehrere Jahre keine verlässlichen Aussagen zu den Standorten vorgenommen werden. Erhält die Bundesauswertungsstelle die Mappinginformationen nicht, kann sie keine Vorjahresdaten einbeziehen.

Hinzu kommt ein weiterer Bruch in der Auswertung, da mit Überführung der QSKH-Verfahren in die DeQS-RL ab 2020 bzw. spätestens 2021 die standortbezogenen QS-Daten (die bis dato die IKNR und Standortkennung enthalten) mit Umstellung auf das dezentrale Pseudonymisierungsverfahren mit verfahrensspezifischen Standortpseudonymen an das IQTIG übermittelt werden. Diese verfahrensspezifische Pseudonymisierung bedeutet, dass ein Standort für jedes von ihm

gelieferte QS-Verfahren ein Pseudonym erhält. Nach aktuellem Diskussionsstand werden nach der Überführung der QSKH-Verfahren in die DeQS-RL 19 separate Verfahren definiert und damit bis zu 19 Standortpseudonyme für einen Standort geführt (vgl. Anlage 3 zu TOP 7, AG DeQS vom 13.08.2019). Das IQITG als Bundeauswertestelle kann somit keinen Vorjahresbezug der nach der DeQS-RL-Überführung übermittelten QS-Daten herstellen.

Letztlich kann die Kontinuität der jahresübergreifenden Auswertungen für die in die DeQS-RL überführten Verfahren nur dadurch erreicht werden, dass der BAS bei einer dezentralen Pseudonymisierung von der DAS entsprechende Mappingtabellen (Standort-ID, Pseudonym) zu Verfügung gestellt werden. Ist diese Offenlegung nicht gewünscht, ist mit der Überführung entsprechend ein Bruch der Auswertungen die Folge (vgl. Anlage 3 zu TOP 7, AG DeQS vom 13.08.2019).

Weiterhin besteht aufgrund der Überführung der Verfahren in die DeQS-RL und momentan noch nicht festgelegter Regelungen für die Datenvalidierung die Problematik, dass das hier vorgestellte Konzept nur auf Grundlage der QSKH-RL erfolgen konnte. Anders als die QSKH-RL spricht die DeQS-RL nicht von „rechnerischen Auffälligkeiten“, sondern schlicht von „Auffälligkeiten“, und lässt mehr methodischen Freiraum für die Ausgestaltung dieses Begriffs. Im Falle der Anwendung einer anderen Methodik wäre separat zu prüfen, inwieweit sich diese für die mehrjährige Betrachtung von Auffälligkeiten eignet bzw. ob diese Methodik bereits Vorschläge enthält, wie sich mehrjährige Betrachtungen realisieren lassen. Ebenfalls unklar bleibt, ob die Regelung zu wiederholten rechnerischen Auffälligkeiten dann ebenfalls für die Vorjahre Gültigkeit hätten. Plante man, wie oben vorgeschlagen, einen ersten gezielten Datenabgleich auf Grundlage dieses Kriteriums im Jahr 2024, würde dies voraussetzen, dass die Verfahren im Jahr 2021 unter den Regelungen der DeQS-RL Bestand haben. Für die indirekten Verfahren ist eine Überführung in die DeQS-RL zum Jahr 2021 geplant, wodurch eine Überprüfung ab 2024 möglich erscheint.

Unter Aufwandsaspekten wäre denkbar, nur dasjenige QS-Verfahren für den gezielten Datenabgleich aufgrund von wiederholten Auffälligkeiten in der statistischen Basisprüfung auszuwählen, in welchem es zuzüglich zur Erfüllung des Kriteriums einen hohen Anteil an im SD bestätigter wiederholter Fehldokumentationen gab. Die Überprüfung von nur einem QS-Verfahren bei Vorliegen mehrerer wiederholter rechnerischer Auffälligkeiten in verschiedenen QS-Verfahren an einem Standort ist aufgrund der Standortpseudonymisierung pro QS-Verfahren ab 2021 jedoch nicht mehr möglich, da dem IQITG unbekannt sein wird, ob es sich um den gleichen Standort handeln wird.

### 5.3.1 Schlussfolgerung des IQTIG

Da der Einbezug des Kriteriums *wiederholter rechnerischer Auffälligkeiten* unabhängig davon ist, ob im Strukturierten Dialog Fehldokumentationen festgestellt werden konnten, empfiehlt das IQTIG dieses Kriterium nicht anzuwenden. Rechnerische Auffälligkeiten ohne Klärung durch den Strukturierten Dialog weisen damit nicht eindeutig auf zugrundeliegende Dokumentationsfehler hin, sondern können lediglich ein Aufgreifkriterium darstellen. Insbesondere kann ein Krankenhausstandort über mehrere Jahre hinweg nachvollziehbar dargelegt haben, warum die rechnerische Auffälligkeit keinen Mangel in der Dokumentationsqualität darstellt.

Sollte sich der G-BA dennoch für die Einführung des Kriteriums der wiederholten rechnerischen Auffälligkeit entscheiden, schlägt das IQTIG folgende Definition und Prozessabläufe vor:

#### **5.4 Definition und Prozessablauf zum Kriterium zur wiederholten rechnerischen Auffälligkeit**

Das Kriterium zur wiederholten rechnerischen Auffälligkeit wird ausgelöst, wenn ein Krankenhausstandort in Folge von drei Jahren in einem Auffälligkeitskriterium zur Plausibilität und Vollständigkeit rechnerisch auffällig war [ $EJ = X$ ,  $EJ = X-1$ ,  $EJ = X-2$ ]. Hierfür wird eine Vergleichbarkeit zum Vorjahr vorausgesetzt, sodass bei maßgeblichen Veränderungen der Rechenregeln (Einstufung als „nicht vergleichbar“) innerhalb dieses Zeitraums keine Auslösung stattfindet.

Bei wiederholt rechnerisch auffälligen Krankenhausstandorten wird aufgrund der Informationen aus der Statistischen Basisprüfung demnach vermutet, dass die QS-Daten in den Vorjahren fehlerhaft oder unvollständig waren. Für jede wiederholte rechnerische Auffälligkeit innerhalb von drei Jahren soll ein gezielter Datenabgleich erfolgen, auch wenn dies bedeutet, dass in einem Jahr an einem Standort mehrere Überprüfungen aufgrund dieses Kriteriums stattfinden können.

Der Einbezug von wiederholten rechnerischen Auffälligkeiten ist unabhängig davon, ob im Strukturierten Dialog Fehldokumentationen festgestellt werden konnten. Da Krankenhausstandorte Leistungen in unterschiedlich vielen Versorgungsbereichen erbringen und eine verfahrensübergreifende Auswertung technisch nicht möglich ist, soll die wiederholte rechnerische Auffälligkeit pro Auffälligkeitskriterium je QS-Verfahren betrachtet werden.

Weil die Überprüfung von nur einem QS-Verfahren bei Vorliegen mehrerer wiederholter rechnerischer Auffälligkeiten in verschiedenen QS-Verfahren an einem Standort aufgrund der Standortpseudonymisierung pro QS-Verfahren ab 2021 nicht möglich sein wird (s. Abschnitt 5.3), kann es sein, dass ein Standort in mehreren QS-Verfahren (je nach Zuordnung des AK) überprüft werden muss. Dies bedeutet, dass für jedes QS-Verfahren die QI-relevanten Datenfelder in einer Zweiterfassung bei einer zufälligen Stichprobe von maximal 20 Fällen pro Standort überprüft werden.

Um zu vermeiden, dass die Krankenhausstandorte in den Folgejahren im selben QS-Verfahren erneut überprüft werden müssen, wird der Standort für die Überprüfung aufgrund der teilweise gleichen Datengrundlage gesperrt: Ein Krankenhausstandort, der über drei Jahre wiederholt rechnerisch auffällig ist [ $EJ = X$ ,  $EJ = X - 1$ ,  $EJ = X-2$ ], wird in  $EJ = X + 1$  gezielt überprüft. Gleichzeitig würden in diesem Jahr für die Feststellung einer wiederholten rechnerischen Auffälligkeit die Jahre  $EJ = X + 1$ ,  $EJ = X$ ,  $EJ = X - 1$  herangezogen werden. Um dies auszuschließen, soll ein gezielter Datenabgleich erst wieder ab  $EJ = X + 3$  für die wiederholte rechnerische Auffälligkeit im selben AK der drei aufeinanderfolgenden Jahre  $EJ = X + 1$ ,  $EJ = X + 2$ ,  $EJ = X + 3$  erfolgen. Krankenhausstandorte, die jährlich alternierend außerhalb und innerhalb des Referenzbereichs der Auffälligkeitskriterien liegen, werden mit diesem Kriterium nicht wiederholt auffällig.

In Bezug auf die Vergleichbarkeit zum Vorjahr ist zu erwähnen, dass die Kriterien für den Zähler- oder Nenner-Einschluss eines Qualitätsindikators durch neue medizinisch-wissenschaftliche Er-

kenntnisse an den aktuellen Forschungsstand angepasst werden und folglich auch eine Anpassung der Auffälligkeitskriterien, die auf diesen Qualitätsindikatoren beruhen, erfolgen. Inhaltliche Änderungen der Auffälligkeitskriterien des jeweils letzten Jahres werden in der Qualitätsindikatorendatenbank (Historie der Auffälligkeitskriterien) dokumentiert. Zudem ist dort vermerkt, ob die Neuanpassung des Auffälligkeitskriteriums mit der vorherigen Version des Auffälligkeitskriteriums vergleichbar, eingeschränkt vergleichbar oder nicht vergleichbar ist. Um zu gewährleisten, dass Krankenhausstandorte nur bei wiederholten rechnerischen Auffälligkeitskriterien in den gezielten Datenabgleich gelangen, wenn diese mit den Auffälligkeitskriterien des Vorjahres vergleichbar sind, dürften nur Auffälligkeitskriterien beachtet werden, die im ersten und darauffolgenden Jahr eingeschränkt vergleichbar oder vergleichbar sind.

Die Einführung des gezielten Datenabgleichs für dieses Kriterium sollte aufgrund der o. g. Herausforderungen für wiederholt rechnerische Auffälligkeiten über drei Jahre hinweg in der Statistischen Basisprüfung für jeden Standort erst ab 2024 für die Erfassungsjahre 2021, 2022 und 2023 durchgeführt werden. Ab diesem Zeitpunkt könnte für jede wiederholte rechnerische Auffälligkeit innerhalb von drei Jahren ein gezielter Datenabgleich erfolgen, auch wenn dies bedeutet, dass in einem Jahr an einem Standort mehrere Überprüfungen aufgrund dieses Kriteriums stattfinden können. Diese Empfehlung gilt auch für die direkten Verfahren, für welche ein Mapping prinzipiell möglich, aber aufwendig und ggf. fehleranfällig sein wird. Nach den oben gezeigten Modellrechnungen (vgl. Tabelle 13) wäre ein einziger Standort in HCH betroffen, was den Aufwand einer teilweisen Einführung nur für die direkten Verfahren vor dem Jahr 2024 nicht rechtfertigen würde.

In Abbildung 10 ist der Vorschlag für das Verfahren des gezielten Datenabgleichs für Adendum III dargestellt, wenn das Kriterium umgesetzt werden sollte. Bei wiederholter rechnerischer Auffälligkeit könnte ein gezielter Datenabgleich im Folgejahr durchgeführt werden, wenn rechnerische Auffälligkeiten in drei aufeinanderfolgenden Jahren in einem AK aufgetreten sind. Je nachdem, aus welchem QS-Verfahren das AK in Folge rechnerisch auffällig war, könnte der gezielte Datenabgleich für das QS-Auswertungsmodul erfolgen, welches dem AK zugeordnet werden kann. Bei Vorliegen von wiederholt rechnerisch auffälligen AKs in mehreren QS-Verfahren an einem Standort kann es jedoch, wie oben beschrieben, auch zu mehreren unterschiedlichen Überprüfungen kommen.

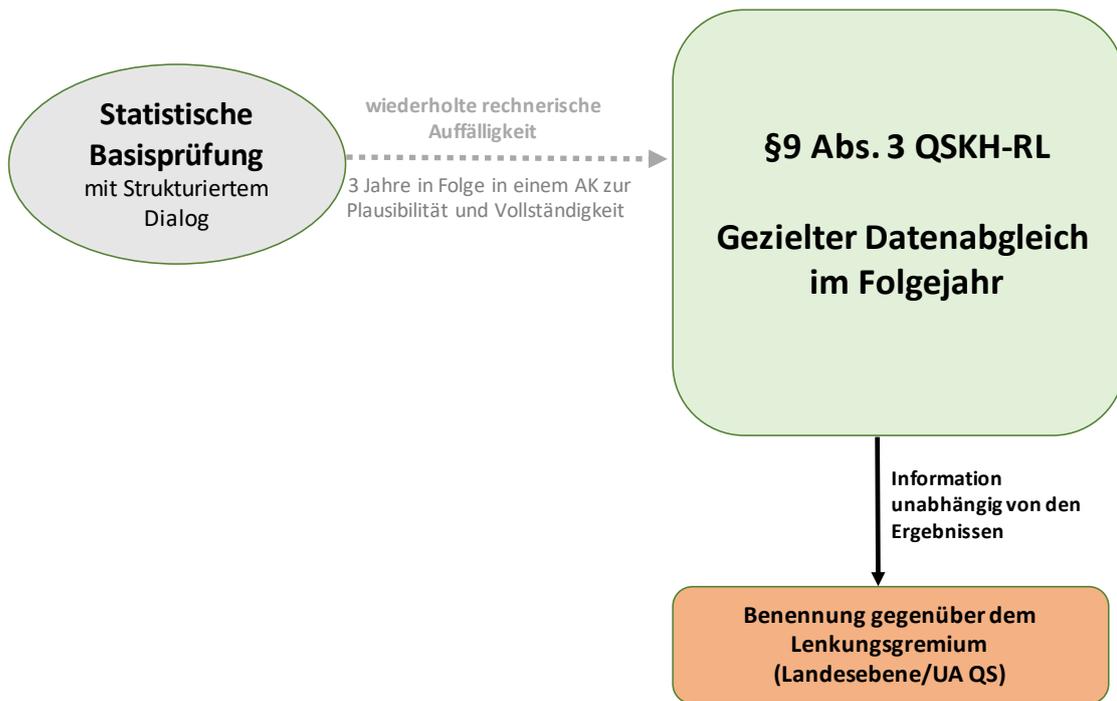


Abbildung 12: Gezielter Datenabgleich nach wiederholt rechnerischer Auffälligkeit in der Statistischen Basisprüfung.

## 5.5 Limitationen

Das IQTIG weist auf mögliche methodische Limitationen für die Definition und Auslösung des o. g. Kriteriums neben den oben geschilderten verfahrensbezogenen Herausforderungen unter Abschnitt 5.3 hin. Die Anzahl der AKs unterliegt grundlegenden Schwankungen, wodurch sich auch Änderungen hinsichtlich auffälliger Standorte und Aufwände für die erneute Überprüfung über die Jahre ergeben können. Ebenso kann die Qualität der AKs in Bezug auf die Aussagekraft von rechnerisch auffälligen Ergebnissen Schwankungen unterliegen, da bei ihrer Entwicklung bisher keine einheitliche Methodik verwendet wurde, die im Methodenpapier des IQTIG festgelegt ist. Hinzu kommt, dass die Vergleichbarkeit zu den Vorjahren eingeschränkt sein kann, wenn es zu Rechenregel Anpassungen kam. Rechnerische Auffälligkeiten ohne Klärung durch den Strukturierten Dialog weisen damit nicht eindeutig auf zugrunde liegende Dokumentationsfehler hin, sondern können lediglich ein Aufgreifkriterium darstellen, weshalb das IQTIG die Einführung nicht empfiehlt. Durch das hier beschriebene Kriterium würde ein Standort, der plausibel dargelegt hat, wieso die rechnerische Auffälligkeit kein Dokumentationsdefizit darstellt, trotzdem für den gezielten Datenabgleich ausgewählt werden. Wurden hingegen im SD wiederholte Fehldokumentationen festgestellt, kann tatsächlich von Dokumentationsfehlern ausgegangen werden. Insofern erscheint es sinnvoller, statt der wiederholten rechnerischen Auffälligkeit, wiederholte und im SD bestätigte Fehldokumentation zu betrachten.

## 6 Prozessablauf des gezielten Datenabgleichs insgesamt

Das Gesamtkonzept (Abbildung 13) sieht demnach vor, bei Vorliegen von Dokumentationsfehlern im Stichprobenverfahren mit Datenabgleich als auch in der Statistischen Basisprüfung für alle ausgearbeiteten Kriterien einen gezielten Datenabgleich im Folgejahr auszulösen.

Als Besonderheit erfolgt eine sofortige Benennung gegenüber dem entsprechend zuständigen Lenkungsgremium, wenn bereits im Stichprobenverfahren mit Datenabgleich erhebliche Fehldokumentationen vorlagen. Für alle anderen Konstellationen gilt, dass im Folgejahr am entsprechenden Krankenhausstandort die Dokumentationsqualität anhand eines Datenabgleichs der QS-Dokumentation mit der Patientenakte erneut überprüft wird. Für die wiederholten rechnerischen Auffälligkeiten gilt (nur bei Einführung), dass der gezielte Datenabgleich für das QS-Auswertungsmodul erfolgt, welches dem AK zugeordnet werden kann das wiederholt auffällig war. Bei Vorliegen von wiederholt rechnerisch auffälligen AKs in mehreren QS-Verfahren an einem Standort kann es demnach auch zu mehreren unterschiedlichen Überprüfungen an einem Standort kommen. Dies trifft ebenfalls auf die Feststellung bei Unterdokumentationen eines Sentinel Events zu.

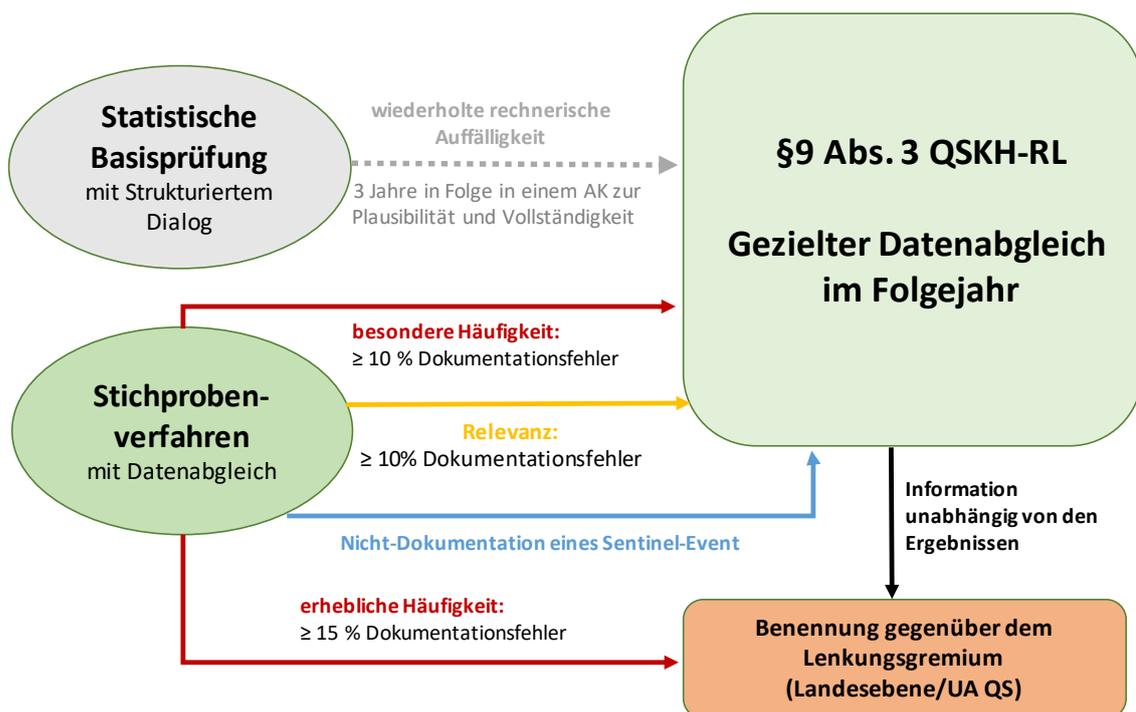


Abbildung 13: Darstellung der Datenvalidierung nach § 9 QSKH-RL: Wege der Krankenhausstandorte in die gezielte Stichprobe über Dokumentationsfehler in den verschiedenen Kriterien.

Nach erfolgtem gezieltem Datenabgleich wird das verantwortliche Gremium nach § 14 QSKH-RL im Nachhinein unabhängig von den Auswertungsergebnissen über diese informiert (Übersendung des bundeseinheitlichen Prüfberichts). Die Einordnung der Ergebnisse des gezielten Datenabgleichs obliegt damit dem Gremium nach §14 QSKH-RL, welches sich an den vom IQTIG

vorgeschlagenen Kriterien zu besonders häufigen oder relevanten Dokumentationsfehlern orientieren kann.

Der zeitliche Prozess ist in Abbildung 14 dargestellt und sieht wie folgt aus: Nach Besprechung in EJ = X zu den festzulegenden Leistungsbereichen für das Stichprobenverfahren mit Datenabgleich in EJ = X + 1 (zu EJ = X) werden für die vorgeschlagenen Leistungsbereiche alle Datenfelder, die QI-relevant sind und ggf. auch für die Risikoadjustierung, dem G-BA vorgelegt. Ebenfalls werden bereits die Ereignisse zum Relevanzkriterium für die vorgeschlagenen Leistungsbereiche definiert. Anschließend erfolgen die Beschlüsse zu den Leistungsbereichen, zu den Datenfeldern und der Liste der Ereignisse des Relevanzkriteriums. Anzumerken ist, dass die Zeitschiene sehr eng ist und Datenfelder sowie die Ereignisse bereits festgelegt werden müssen, obwohl die Leistungsbereiche, die im Stichprobenverfahren mit Datenabgleich überprüft werden sollen, zu dem Zeitpunkt noch nicht beschlossen sind. Anschließend erfolgen Anfang des Jahres EJ = X + 1 die Beschlüsse zu den Datenfeldern und relevanten Ereignissen. Die Datenfelder, die zum Stichprobenverfahren mit Datenabgleich in EJ = X + 1 (zu EJ = X) beschlossen werden, sollen auch für eine mögliche Überprüfung im gezielten Datenabgleich zu EJ = X + 2 (zu EJ + 1) gelten. Die Integration der Verfahren und Datenfelder sowie der Ereignisse in das Erfassungstool geschieht unmittelbar im Anschluss. Es wird ein Handbuch zur Datenvalidierung erstellt, welches zu jedem überprüften Datenfeld Hinweise zum Vorgehen vor Ort enthält, um eine einheitliche Erfassung zu gewährleisten. Zu Anfang April findet die Stichprobenziehung durch das IQTIG statt und die entsprechenden QS-Daten der zufällig gezogenen Krankenhausstandorte werden in das Tool integriert.

Am 15. April des EJ = X + 1 wird das Erfassungstool mit den Standorten sowie das Datenvalidierungshandbuch an die Bundesländer übermittelt. Das Stichprobenverfahren mit Datenabgleich findet ab diesem Tag bis zum 31. Oktober des EJ = X + 1 statt. Die Standorte werden ab dem 15. April über die Zweiterfassung informiert und gebeten, eine Rückmeldung zwecks Akteneinsicht zurückzumelden. Wenn die Patientenakten nicht durch das IQTIG/die auf Landesebene beauftragten Stellen eingesehen werden können, muss der MDK beauftragt werden. Dies bedeutet ggf., dass bei den direkten Verfahren die bundesweit zufällig gezogen werden, ein neues Tool für den zu überprüfenden Standort durch den MDK generiert werden muss. Im Anschluss an die Übermittlung der Daten durch die Länder für die indirekten Verfahren und die Übermittlung der Daten für die direkten Verfahren zum 31. Oktober werden diejenigen Standorte dem jeweiligen Lenkungsgremium benannt, welche im Stichprobenverfahren mit Datenabgleich erheblich häufige Dokumentationsfehler gemacht haben. Durch das IQTIG erfolgen ab Ende des Stichprobenverfahrens mit Datenabgleich die Berechnungen zu den in Abbildung 13 dargestellten Kriterien. Zu den Standorten mit erheblich häufigen Dokumentationsfehlern müssen Rückmeldungen an die auf Landesebene beauftragten Stellen gegeben werden.

Zum 31. Januar im EJ = X + 2 werden die Ergebnisse des Stichprobenverfahrens mit Datenabgleich zu EJ = X veröffentlicht. Ab dem 15. März könnten die wiederholten rechnerischen Auffälligkeiten aus der Statistischen Basisprüfung berechnet werden. Die Standorte und Verfahren (hier wird es ggf. mehr Verfahren geben als bei den für das Stichprobenverfahren mit Datenabgleich ausgewählten) müssten daraufhin in einem sehr kurzen Zeitraum in das Tool integriert

werden. Am 15. April des EJ = X + 2 wird das Erfassungstool mit den Standorten und das Datenvalidierungshandbuch an die Bundesländer übermittelt und es startet der gezielte Datenabgleich. Nach Beendigung am 31. Oktober werden die Ergebnisse dem jeweiligen Lenkungsgremium mitgeteilt.

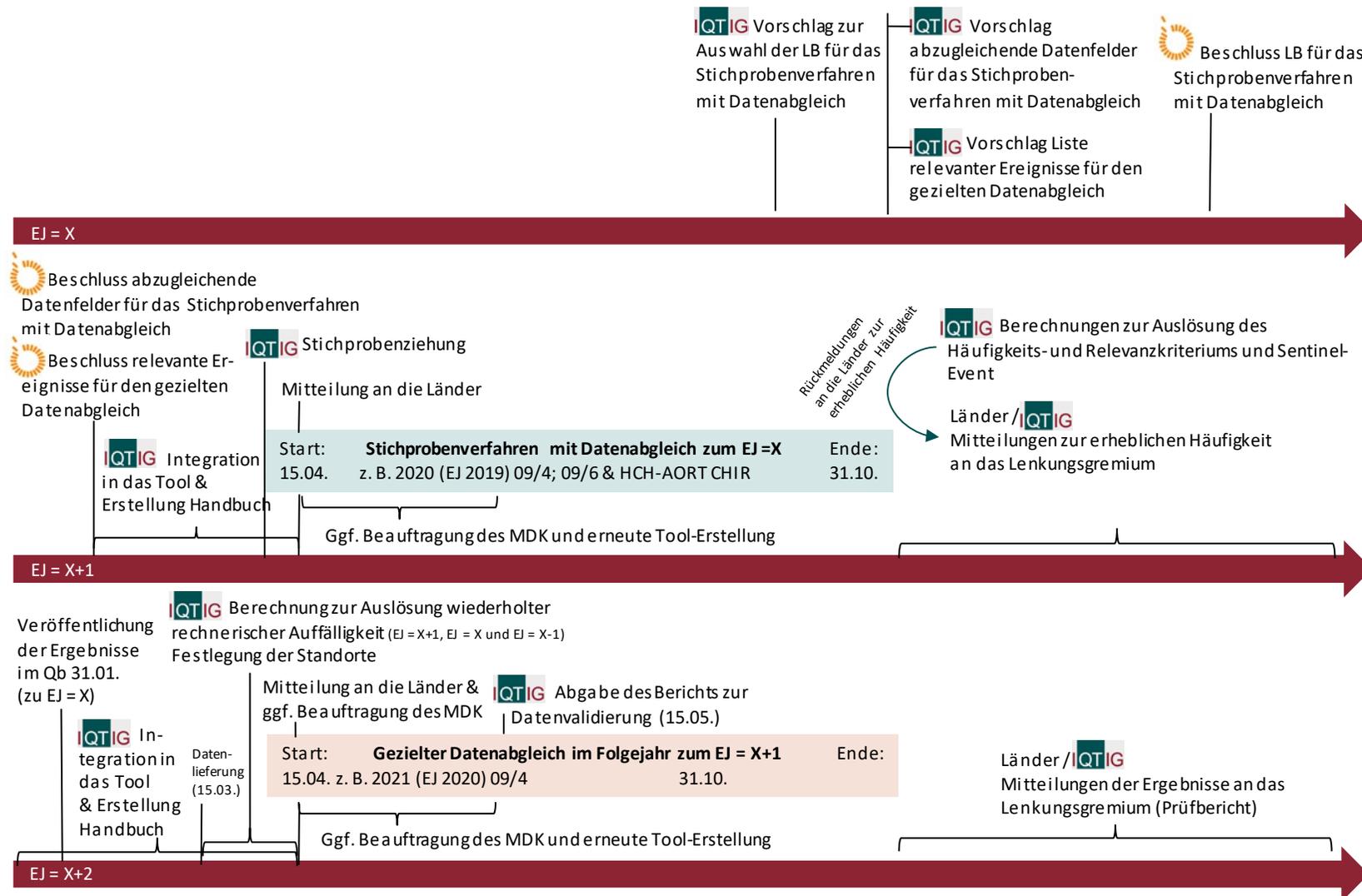


Abbildung 14: Zeitablauf des Stichprobenverfahrens und gezielten Datenabgleichs

## 7 Bewertung der Dokumentationsqualität

### Gegenstand der Beauftragung Nr. I.5

Mit Beschluss vom 17. November 2017 wurde das IQTIG außerdem vom G-BA beauftragt, Bewertungskriterien für die Ergebnisse des Dokumentationsqualitätsabgleichs für den Bericht zur Datenvalidierung zu entwickeln und die bisher verwendeten Kategorien „verbesserungswürdig“, „gut“ und „hervorragend“ zu überarbeiten.

Im Bericht zur Datenvalidierung des IQTIG nach § 15 Abs. 2 QSKH-RL wird die Dokumentationsqualität auf der Ebene des jeweiligen Datenfeldes über alle Standorte hinweg dargestellt. In vorherigen Berichten wurde nicht nur über die Dokumentationsqualität berichtet, sondern auch eine anschließende Bewertung der Dokumentationsqualität auf Datenfeldebene vorgenommen und in die Berichterstattung eingeschlossen (AQUA 2014).

Der Vorschlag zur Bewertung der Dokumentationsqualität wird im vorliegenden Abschlussbericht deshalb für die Ergebnisse des Stichprobenverfahrens mit Datenabgleich des Datenvalidierungsverfahrens nach § 9 QSKH-RL gemäß Tabelle 15 vorgeschlagen. Die hier vorgeschlagene Methodik zur Bewertung der Dokumentationsqualität kann auf Standortebene auch zur Bewertung der Ergebnisse des gezielten Datenabgleichs verwendet werden. Aufgrund der evtl. geringen Anzahl der Standorte für die ein gezielter Datenabgleich durchgeführt wird, kann hier auf Bundesebene keine Bewertung erfolgen.

Zum einen fließen die Ergebnisse des Datenabgleichs rein quantitativ in das Kriterium der Dokumentationsfehler von besonderer und erheblicher Häufigkeit ein, das die Ergebnisse der Übereinstimmungsrate betrachtet. Zum anderen werden qualitative Abweichungen über das Kriterium der Relevanz und der Nicht-Dokumentation eines Sentinel Events betrachtet. Hier werden indikatorrelevante Abweichungen bzw. Abweichungen in einem Sentinel-Event-Indikator quantitativ betrachtet. Beide Aspekte sollen in den vorgeschlagenen Bewertungskategorien zur Dokumentationsqualität berücksichtigt werden.

### 7.1 Methodisches Vorgehen

Zur Erarbeitung einer Bewertung von Dokumentationsfehlern in der Gesundheitsversorgung wurde eine manuelle Literaturrecherche in den Datenbanken *PubMed* und *Google Scholar* durchgeführt. In der überarbeiteten Leitlinie zur Datenqualität in der medizinischen Forschung von Nonnemacher et al. (2014) wird ein adaptives Management von Datenqualität in Kohortenstudien und Registern beschrieben. Hierfür werden 51 verschiedene Indikatoren zur Messung von Datenqualität angewendet. Die Bewertung der Datenqualität erfolgt anschließend durch Anwendung eines Qualitätsscores mit bis zu 100 Punkten in sehr schlecht, schlecht, moderat, gut und sehr gut. Eine Übernahme zur Anwendung auf den Abgleich zwischen QS-Dokumentation und Patientenakte ist in dem Sinne nicht möglich, weil es sich hier um Registerdaten und Kohortenstudien handelt, und das IQTIG zudem keine Indikatoren im Originaldatenabgleich zur Beurteilung der Datenqualität im Stichprobenverfahren an einem Standort einsetzt. Das beschriebene Vorgehen kommt eher den Auffälligkeitskriterien in der Statistischen Basisprüfung

nahe. Ein (zusätzlicher) Originaldatenabgleich im Rahmen von Stichproben wird in der Leitlinie aber explizit empfohlen.

In die Leitlinie von Nonnemacher et al. (2014) wurden 39 Literaturstudien einbezogen, die z. B. eine Analyse von Datenqualität behandeln. Manche der einbezogenen Studien beschäftigen sich eher mit den Gründen für Datenfehler und dem Aufbau eines einheitlichen Verfahrens zur Definition und Erhebung von Daten (Kahn et al. 2016, Nahm et al. 2008, Arts et al. 2002). Andere Studien sind auch für die Bewertung von Datenqualität bei Originaldatenabgleich für die hier vorzuschlagende Bewertung der Dokumentationsqualität gesichtet worden. So haben Duda et al. (2012) beispielsweise die Vollständigkeit und Korrektheit von Routinedokumentationen durch Vor-Ort-Audits untersucht und dabei die Fehler in „Geringfügiger Fehler („minor error“), „Bedeutsamer Fehler (major error“), fehlender Wert in der Studiendatenbank („Missing data“), Quellenloser Wert („sourceless“) und korrekte Werte („correct“) eingeteilt. Aus den Ergebnissen wurden Fehlerraten berechnet, die sich bei 5 von 7 untersuchten Kliniken auf > 10 % beliefen. Es ist allerdings keine Beurteilung der Dokumentationsqualität vorgenommen worden. Ebenso ist die Aufteilung in minor und major errors nach klinischer Bedeutsamkeit erfolgt, die im Zweifel aus Sicht des IQTIG eher schwer festzulegen ist.

Tuble (2011) unternahm die Aufteilung zur Datenqualität in korrekte und komplette Daten, inkorrekte Daten, und fehlende Daten (entweder nicht eingegeben oder nicht in der Akte gefunden) vor. Die Variablen mit Fehldokumentationen wurden basierend auf der Fehldokumentation in Kategorien eingeteilt (0-25 %, 25-50 %, 51-75 %, 75-100 %). Fehlende, ungenaue, nicht eingegebene oder nicht validierbare Werte wurden als Abweichung gewertet. Dies entspricht der Ansicht und Umsetzung durch das IQTIG beim Häufigkeitskriterium. Die Einteilung der Fehldokumentation in Kategorien wurde jedoch nicht begründet und ist deshalb schwer nachvollziehbar. Ebenso wurde keine Bewertung der Dokumentationsqualität für eine der Kategorien vorgenommen. Insgesamt konnte gezeigt werden, dass die Daten der überprüften 60 Fälle in 7 % inkorrekt gewesen sind und in 3 % fehlten. Höhere Übereinstimmungsraten konnten von Datta et al. (2007) für ein Register von Traumapatientinnen und -patienten gezeigt werden. Die Übereinstimmungsrate von 1.400 Datenfeldern (100 Patientinnen und Patienten mal 14 überprüfte Felder) lag bei 98 % in Bezug auf Korrektheit, was darauf hindeutet, dass auch bei vielen überprüften Datenfeldern eine hohe Übereinstimmung erreicht werden kann.

Die gesichtete Literatur weißt insgesamt sehr heterogene Fehldokumentationsraten/Übereinstimmungsraten auf, die auf unterschiedlichen Studienpopulationen und Datenbankabgleichen basieren. Insgesamt fanden sich in der wissenschaftlichen Literatur keine Hinweise auf einen Standard zur Bewertung der Dokumentationsqualität in Bezug auf eine Quellenverifizierung (Abgleich mit den Originaldaten). Vielmehr ist die Bewertung der Dokumentationsqualität vom Verwendungszweck der Daten abhängig (Houston et al. 2015). Somit konnte bei der Bildung von Bewertungskriterien kein Konzept direkt übernommen werden, sondern war spezifisch zu entwickeln.

In der Literatur werden allerdings beim Abgleich von Registerdaten mit Patientenakten die Daten mit einer Übereinstimmungsrate von  $\geq 95$  % als konkordant angenommen (Jacke et al. 2012,

Thoburn et al. 2007). Daraus folgt, dass bei einer Richtigkeit von unter 95 % bereits Optimierungsbedarf für die Dokumentation zu bestehen scheint. Vor dem Hintergrund wichtiger Datenfelder aus Patientensicht, wie einer richtigen Indikationsstellung oder komplikationslosen Operationen, erscheint dies von grundlegender Bedeutung. Für andere Datenfelder könnte die Richtigkeit aus Patientensicht ggf. nicht so maßgebend sein. Aus Sicht des IQTIG sind jedoch alle QI-relevanten Datenfelder von gleichem Gewicht, weshalb von einer je Datenfeld unterschiedlichen Übereinstimmungsempfehlung abgesehen wurde, zumal diese auch impraktikabel erscheint. Deshalb soll eine Lösung angestrebt werden, die gleichermaßen für alle Datenfelder gilt.

Grundsätzlich muss betont werden, dass alle Fehldokumentationen relevant für die Richtigkeit der Ergebnisse der Qualitätsindikatoren sein können. Dies unterstreicht auch die Wichtigkeit einer korrekten Dokumentation (Richtigkeit). In Bezug auf die korrekte Anzahl an zu dokumentierenden Fällen (Vollzähligkeit) ist bereits das Maß schrittweise auf 100 % erhöht worden und geht ggf. mit Abschlagszahlungen gem. §24 QSKH-RL einher. In Bezug auf die Richtigkeit der übermittelten Datensätze gibt es weder einen Standard, ab welchem Grad für die Richtigkeit immer noch eine gute Datenqualität angenommen werden kann, noch eine normative Regelung. Deshalb kann an dieser Stelle lediglich eine Annäherung vorgenommen werden, die sich derjenigen für die Konkordanz der Dokumentationsqualität in Registern annähern soll.

Die Beauftragung des G-BA vom 17. November 2017 sieht vor, dass bei der Entwicklung von Bewertungskriterien eine Kategorisierung der Ergebnisse nicht nur in Bezug auf die Übereinstimmungsrate, sondern auch in Bezug auf Sensitivität und Spezifität erfolgen soll. Da Sensitivität und Spezifität einen Goldstandard voraussetzt, wird die Primärdokumentation weiterhin als Goldstandard betrachtet. Es besteht jedoch die grundlegende Problematik, dass nicht alle im Stichprobenverfahren oder im gezielten Datenabgleich zu überprüfenden indikatorrelevanten Datenfelder binär kodiert sind. Nur für binär kodierte Datenfelder kann die Sensitivität und Spezifität berechnet werden, kategorial und metrisch kodierte Datenfelder müssten geeignet dichotomisiert werden. Beispielsweise waren für das Stichprobenverfahren mit Datenabgleich 2019 (EJ 2018) im QS-Verfahren Neonatologie 13 von 40 Datenfelder binär kodiert (32,5 %). Für das QS-Verfahren Herzunterstützungssysteme/Kunsth Herzen waren von 27 zu überprüfenden Datenfeldern 5 binär kodiert (18,5 %). Dies bedeutet, dass entweder nur für diese Datenfelder in der Bewertung der Dokumentationsqualität der zusätzliche Aspekt von Sensitivität und Spezifität berücksichtigt werden kann, oder eine Dichotomisierung der übrigen kategorial und metrisch ausgeprägten Datenfelder erfolgen müsste. Eine Dichotomisierung der übrigen Datenfelder kann sich ggf. als schwierig erweisen und die Ergebnisse würden je nach Gruppenbildung ggf. unterschiedlich ausfallen. Bei metrischen Datenfeldern müsste ein idealer Cut-off-Point gefunden werden. Insgesamt wäre dieses Verfahren sehr aufwändig, weshalb diese Möglichkeit vorerst ausgeschlossen werden soll. Der Vorschlag, für binär kodierte Datenfelder zusätzlich Sensitivität und Spezifität zu betrachten, bedeutet, dass die Datenfelder insgesamt sehr heterogen bewertet würden. Entweder würde nur die Übereinstimmungsrate bewertet, oder zusätzlich die Sensitivität und Spezifität herangezogen, was dem Vorgehen der damaligen Institution nach § 137a SGB V gleichkäme. Somit wären Verfahren, die wenige binäre Datenfelder verwenden ggf. bessergestellt als Verfahren mit vielen binären Datenfeldern. Weiterhin besteht die Problematik einer wissenschaftlich fundierten Synthese beider Bewertungsaspekte: beispielsweise

dann, wenn eine gute Übereinstimmungsrate eines Datenfeldes insgesamt vorliegt, aber einer niedrigen Sensitivität, die auf wenigen Fällen beruht. Hierfür kann das Beispiel aus Addendum II Tabelle 8 betrachtet werden. Die Übereinstimmungsrate zur Dokumentation spezifischer Komplikationen liegt mit 99,3 % sehr hoch. Die Sensitivität beträgt allerdings nur 58,3 %, was bedeutet, dass im interessierenden Ereignis „Komplikationen“ eine Unterdokumentation von knapp 42 % vorlag. Eine gute Sensitivität und Spezifität kann angenommen werden, wenn diese nahe bei 100 % liegt (Raslich et al. 2007). Ansonsten gibt es in der Literatur keine allgemeingültigen Angaben, ab welchem Anteil die Sensitivität und Spezifität als „ausreichend“ angenommen werden kann. Angaben zur einem optimalen Cut-off-Point beziehen sich immer ganz konkret auf den untersuchten Aspekt (Schwarzer et al. 2001). Die Fehldokumentationsrate des genannten Beispiels bezieht sich auf 8 von 1.179 Fälle, die Unterdokumentation auf lediglich 5 von 8 Fälle. Sensitivität und Spezifität sind grundsätzlich von der Prävalenz unabhängig (Allouche et al. 2006). Eine Synthese beider Aspekte (Übereinstimmungsrate und Sensitivität/Spezifität) in der Bewertung der Dokumentation des Datenfeldes erscheint vor diesem Hintergrund schwierig. In diesem Aspekt besteht auch der Unterschied zum Relevanzkriterium, welches eindeutig interessierende QI-relevante Ereignisse betrachtet und aufgrund seiner Datenfeldkombinationen sowie dem Einbezug mehrerer Ereignisse allerdings nicht auf einem einzelnen Datenfeld rekurriert. Mit den nachfolgend aufgeführten Betrachtungsebenen und einer anschließenden Synthese ist versucht worden, sich der geschilderten Problematik in der Bewertung der Dokumentationsqualität anzunähern.

## 7.2 Bewertungsebenen

Die Bewertung der Dokumentationsqualität kann grundsätzlich auf mehreren Ebenen erfolgen: zum Beispiel auf Ebene des Datenfeldes, des QS-Verfahrens, des Leistungserbringers oder Varianten hiervon (z. B. auf Datenfeldebene pro Leistungserbringer oder auf Datenfeldebene pro QS-Verfahren oder auch QS-verfahrensübergreifend auf Leistungserbringerebene). Da ein Krankenhausstandort unterschiedliche Arten von Leistungen erbringt und somit zur Dokumentation und Lieferung von Daten für unterschiedlich viele QS-Verfahren verpflichtet sein kann und in den QS-Verfahren unterschiedlich viele Datenfelder befüllt werden, schließt das IQTIG eine verfahrensübergreifende Bewertung der Dokumentationsqualität aus.

Der hier vorliegende Vorschlag umfasst eine Bewertung der Dokumentationsqualität pro Standort und ist standortübergreifend (Tabelle 15).

Ziel der vorgeschlagenen Aspekte zur Bewertung auf zwei Ebenen ist es, eine Aussage über die Dokumentationsqualität im jeweiligen Erfassungsjahr für das jeweils betrachtete QS-Verfahren zu erhalten, und zwar einerseits für jeden überprüften Krankenhausstandort (1) und andererseits auf standortübergreifender Ebene (2). Für letzteren Zweck wird sowohl die Verteilung der Ergebnisse für die einzelnen Standorte ( $2_1$ ) als auch die auf Basis der Stichprobe aller untersuchter Fälle geschätzte bundesweite Dokumentationsqualität ( $2_2$ ) herangezogen. Die Unterteilung in ( $2_1$ ) und ( $2_2$ ) erfolgt, da sich die Situation ergeben kann, dass beispielsweise ein großer Standort sehr schlecht dokumentiert, aber alle kleinen Standorte sehr gut bzw. umgekehrt ein großer Standort sehr gut, dafür aber alle kleinen eher schlecht dokumentieren. Somit wäre im ersten

Fall fraglich, ob das QS-Verfahren standortübergreifend als schlecht dokumentiert gelten kann, wenn theoretisch nur ein Standort sehr schlecht dokumentiert hat. Auf Bundesebene könnte dies aber bedeuten, dass viele Fälle fehldokumentiert wurden. Generell kann eine hohe Fehldokumentation im Sinne von (2<sub>1</sub>) die Standortergebnisse und deren Verteilung beeinflussen, während eine hohe Fehldokumentation gemäß (2<sub>2</sub>) zu Abweichungen im Bundesergebnis von Qualitätsindikatoren führen kann.

Sowohl auf Standortebene als auch auf der standortübergreifenden Ebene können verschiedene Aspekte (a–d in Tabelle 15) betrachtet werden, die jeweils eine eigene Aussage zur Dokumentationsqualität mit sich bringen. Deshalb sollen die nachfolgenden Aspekte soweit wie möglich auch im Bericht zur Datenvalidierung aufgeführt werden. Es werden jedoch nur die Aspekte zur Bewertung der Dokumentationsqualität herangezogen, die mit einem ✓ in Tabelle 15 versehen sind. Aspekte, die mit (✓) versehen sind, werden als ergänzende Informationen betrachtet.

Tabelle 15: Betrachtungsebenen der Dokumentationsqualität inklusive Beispiele

	<b>1) Standortebene</b> (pro individuelm Standort)	<b>2) Standortübergreifende Ebene</b> (für alle Standorte oder alle Fälle aller Standorte)	
		2 <sub>1</sub> ) Verteilung der Standort- ergebnisse	2 <sub>2</sub> ) Bundesergebnis über alle Fälle
	Statistische Inferenz: 20 Fälle → gesamter Standort (Einbezug stat. Unsicherheit)	Statistische Inferenz: Alle geprüften KH (5 %) → Alle datenliefernden KH (Einbezug stat. Unsicherheit)	
<b>a) Pro Datenfeld</b>			
<b>Übereinstimmungsrate</b>	z. B. <b>85 %</b> (17/20) für Komplikationen <u>Aussage</u> zur Übereinstimmungsrate des einzelnen Standortes je Datenfeld. <u>Darstellung</u> : Im Prüfbericht für die Standorte aber <u>ohne</u> Inferenzstatistik <u>Formeln und statistische Methodik</u> : Abschnitt 2.1.1.1	✓ z. B. 10 % der Standorte haben Dokumentationsfehler von mehr als 5 % <u>Aussage</u> zur Verteilung der Übereinstimmungsrate der Standorte je Datenfeld. <u>Darstellung</u> : Benchmark-Diagramm im Bericht zur Datenvalidierung	✓ Bundesergebnis für alle Fälle: z. B. <b>99,66 %</b> (auf Basis von 1.179 untersuchten Werten für das Datenfeld Komplikationen) <u>Aussage</u> zur Übereinstimmung jedes Datenfeldes in Bezug auf alle Fälle aller Standorte <u>Darstellung</u> : Im Bericht zur Datenvalidierung <u>Formeln und statistische Methodik</u> : Abschnitt 2.1.1.2
<b>Sensitivität</b> (nur bei binär kodierten DF möglich)	z. B. für das DF „Inotrope“ im QS-Verfahren HCH: 0 % (0/2) → 1-Sensitivität = 100 % Unterdokumentation <u>Aussage</u> : Das DF geht in die Risikoadjustierung ein und wurde von dem Standort 2 Mal zuungunsten dokumentiert	(✓) z. B. 10 % der Standorte haben weniger als 90 % Sensitivität im betrachteten Datenfeld erreicht (d. h. mehr als 10 % ihrer Fälle unterdokumentiert) <u>Aussage</u> wie hoch der Anteil an Standorten ist, die mindestens 10 % in der QS unterdokumentiert haben.	(✓) z. B. <b>60 %</b> (auf Basis von 3 interessierenden Ereignissen aus 5 dokumentierten) <u>Aussage</u> wie viele der in der Patientenakte dokumentierten Werte auch in der QS dokumentiert wurden. Unterdokumentation = 1-Sensitivität

	1) Standortebene (pro individuelm Standort)	2) Standortübergreifende Ebene (für alle Standorte oder alle Fälle aller Standorte)	
		2 <sub>1</sub> ) Verteilung der Standort- ergebnisse	2 <sub>2</sub> ) Bundesergebnis über alle Fälle
	<u>Darstellung:</u> Die Sensitivität pro DF wird nicht für die Standorte im Prüfbericht ausgegeben. Für die direkten Verfahren erfolgt bei Relevanz eine Rückmeldung des IQTIG mit entsprechendem Hinweis (inkl. der Übermittlung des Prüfberichts).	<u>Darstellung:</u> Im Bericht zur Datenvalidierung	<u>Darstellung:</u> Im Bericht zur Datenvalidierung <u>Ergänzende Darstellung:</u> Im Anhang des Berichts zur Datenvalidierung wird für jedes DF (alle Kategorien) eine Kreuztabelle erstellt.
<b>Spezifität</b> (nur bei binär kodierten DF möglich)	z. B. für das DF „Inotrope“ im QS-Verfahren HCH: <b>100 % (18/18)</b> → 1-Spezifität = 0 % Überdokumentation <u>Aussage:</u> Das DF geht in die Risikoadjustierung ein und wurde nicht fälschlicherweise in der QS überdokumentiert. <u>Darstellung:</u> Die Sensitivität pro DF wird nicht für die Standorte im Prüfbericht ausgegeben. Für die direkten Verfahren erfolgt bei Relevanz eine Rückmeldung dem Hinweis (inkl. der Übermittlung des Prüfberichts).	(✓) z. B. 10 % der Standorte haben weniger als 90 % Spezifität im betrachteten Datenfeld erreicht (d. h. mehr als 10 % ihrer Fälle überdokumentiert). <u>Aussage</u> wie hoch der Anteil an Standorten ist, die mindestens 10 % in der QS unterdokumentiert haben. <u>Darstellung:</u> Im Bericht zur Datenvalidierung	(✓) z. B. <b>99,8 %</b> (auf Basis von 1.172 interessierenden aus 1.174 dokumentierten Ereignissen) <u>Aussage</u> dazu, wie viele Werte nicht in der Patientenakte dokumentiert waren, die aber in der QS dokumentiert wurden. Überdokumentation: 1-Spezifität. <u>Darstellung:</u> Im Bericht zur Datenvalidierung <u>Ergänzende Darstellung:</u> Im Anhang des Berichts zur Datenvalidierung wird für jedes DF (alle Kategorien) eine Kreuztabelle erstellt.
<b>b) alle Datenfelder</b>			
<b>Übereinstimmungsrate</b> (Häufigkeitskriterium)	✓ z. B. <b>95,07 % (347/365)</b> für das QS-Verfahren HTXM-MKU. Betrachtung aller geprüften Datenfelder unter Berücksichtigung statistischer Unsicherheit. <u>Aussage</u> über die allgemeine Dokumentationsqualität des Krankenhausstandortes <u>Darstellung:</u> Im Prüfbericht für die Standorte und ggf. in den Qualitätsberichten nach § 136b Abs. 1 Satz 1 Nr. 3 SGB V. <u>Formeln und statistische Methodik:</u> Abschnitt 2.1.2.1	z. B. X % der Standorte haben Dokumentationsfehler von mehr als 10 % <u>Aussage</u> zur Verteilung der jeweiligen Übereinstimmungsraten der überprüften Standorte. <u>Darstellung:</u> Benchmark-Diagramm im Bericht zur Datenvalidierung.	z. B. <b>87,8 %</b> (auf Basis von 19.508 interessierenden Ereignissen für das QS-Verfahren 17/1) <u>Aussage</u> wie korrekt das QS-Verfahren von allen Standorten (bundesweit) dokumentiert wird. <u>Darstellung:</u> Nennung im Bericht zur Datenvalidierung pro überprüfem QS-Verfahren. <u>Formeln und statistische Methodik:</u> Abschnitt 2.1.2.2

	<b>1) Standortebe- ne</b> (pro einzel- nem Standort)	<b>2) Standortübergreifende Ebene</b> (für alle Standorte oder alle Fälle aller Standorte)	
		2 <sub>1</sub> ) Verteilung der Standort- ergebnisse	2 <sub>2</sub> ) Bundesergebnis über alle Fälle
<b>c) Relevanzkriterium</b>			
<b>Unterdokumen- tation relevan- ter Ereignisse</b>	✓ z. B. <b>12,5 %</b> (1/8) für das QS- Verfahren HTXM-MKU <u>Aussage</u> , wie häufig am überprüften Standort eine Unterdokumentation rele- vanter Ereignisse eingetre- ten ist. <u>Darstellung</u> : Im Prüfbericht für die Standorte und ggf. in den Qualitätsberichten nach § 136b Abs. 1 Satz 1 Nr. 3 SGB V.	z. B. X % der Standorte ha- ben relevante Ereignisse zu mehr als 10 % unterdoku- mentiert <u>Aussage</u> , bei welchem Anteil der überprüften Standorte eine Unterdokumentation relevanter Ereignisse festge- stellt wurde. <u>Darstellung</u> : Benchmark-Dia- gramm im Bericht zur Da- tenvalidierung.	z. B. <b>31,85 %</b> (auf Basis von 179 interessierenden aus 562 dokumentierten Ereig- nissen für das QS-Verfahren 17/1) <u>Aussage</u> , wie häufig eine Un- terdokumentation relevanter Ereignisse in QS-Verfahren XY eingetreten ist. <u>Darstellung</u> : Nennung im Be- richt zur Datenvalidierung pro überprüfem QS-Verfah- ren.
<b>d) Sentinel Event</b>			
<b>Nicht dokumen- tiertes Sentinel Event</b>	✓ <u>Aussage</u> ob am überprüften Standort eine Nicht-Doku- mentation eines Sentinel Events eingetreten ist. <u>Darstellung</u> : Ggf. in den Qualitätsberichten nach § 136b Abs. 1 Satz 1 Nr. 3 SGB V.	z. B. im QS-Verfahren NEO zum QI: 51070: Sterblichkeit bei Risiko-Lebendgeborenen (ohne zuverlegte Kinder)“ <u>Darstellung</u> : Nennung der Anzahl der Standorte, für die dieses Kriterium einge- troffen ist im Bericht zur Da- tenvalidierung.	<u>Aussage</u> über die Anzahl der festgestellten unterdoku- mentierten Sentinel Events bundesweit. <u>Darstellung</u> : keine

Auf Standortebe-  
ne wird dabei ausgehend von den i. d. R. 20 untersuchten Fällen eine Aussage  
über die Dokumentationsqualität bei der Gesamtheit der Fälle des Krankenhausstandortes ge-  
macht. Auf der standortübergreifenden Ebene wird ausgehend von der Fallstichprobe von 20  
Fällen bei der Krankenhausstichprobe in der Größenordnung von 5 % aller Standorte eine Aus-  
sage über die Dokumentation des QS-Verfahrens insgesamt gemacht. Hierzu werden jeweils ge-  
eignete statistische Verfahren verwendet. Näheres ist in Kapitel 2 erläutert.

### 7.3 Synthese

Für die verschiedenen Ebenen und betrachteten Aspekte aus Tabelle 15 wurden in einem ersten  
Schritt einzelne Grenzwerte normativ festgelegt (Tabelle 16; Tabelle 17). In einem zweiten  
Schritt erfolgt die Synthese der festgelegten Grenzwerte zu den unterschiedlichen Aspekten für  
die Standortebe-  
ne und die standortübergreifende Ebene (Systemebene), in welcher die eigent-  
liche Bewertung der Dokumentationsqualität festgelegt wurde (Tabelle 18).

Es wird ein binäres Bewertungsschema vorgeschlagen:

- *Kein Hinweis auf unzureichende Dokumentationsqualität*
- *Unzureichende Dokumentationsqualität*

Diese Art der binären Kategorisierung wird vorgenommen, weil methodisch ein einseitiger statistischer Test verwendet wird, um festzustellen, ob genügend Evidenz gegen die Hypothese besteht, dass (gemäß dem weiter unten definierten Kriterium) *keine* unzureichende Dokumentationsqualität vorliegt (bei einem Signifikanzniveau von  $\alpha = 0,05$ ). Ist dies nicht der Fall, ist die Dokumentationsqualität keinesfalls als „gut“ o. ä. zu bewerten, da es nicht genug Anhaltspunkte dafür gibt, dass sie tatsächlich unzureichend ist.

Wichtig ist, dass die in Tabelle 16 für die Standorte ebene vorgenommenen Grenzwertfestlegungen gemeinsam betrachtet werden sollten, so wie in der in Tabelle 18 vorgeschlagenen Synthese. Ansonsten würde beispielsweise eine Übereinstimmungsrate von 91 % auf Standorte ebene bereits bedeuten, dass kein Hinweis auf unzureichende Dokumentationsqualität vorliegt. Dies kann aber nur in Verbindung mit dem Relevanzkriterium wirklich beurteilt werden. Der genannte Aspekt wird in Abbildung 9 in Abschnitt 4.2 zum Addendum II (Seite 62) sehr deutlich. In genannter Abbildung werden die Ergebnisse des Häufigkeits- und Relevanzkriteriums im Vergleich beispielhaft für 62 Standorte dargestellt. Es ist erkennbar, dass bei alleiniger Betrachtung des Häufigkeitskriteriums mit Dokumentationsfehlern  $\leq 10\%$  insgesamt 8 Standorte relevante Ereignisse unterdokumentiert haben. Für diese Standorte kann allerdings nicht ausgesagt werden, dass kein Hinweis auf unzureichende Dokumentationsqualität vorlag.

Anders verhält sich die Beurteilung der Dokumentationsqualität für die Grenzwertfestlegung auf standortübergreifender Ebene (Tabelle 17), für welche die Übereinstimmung für jedes einzelne Datenfeld anstelle der Übereinstimmungsrate über alle Datenfelder und Fälle herangezogen wird. Sensitivität und Spezifität sollen als ergänzende Kennzahlen hinzugezogen werden.

Die jeweiligen Grenzwerte in Tabelle 16, Tabelle 17 und Tabelle 18 sind im Hinblick auf die Diskussion in Kapitel 2 unter Berücksichtigung der sich aus der Stichprobenziehung ergebenden statistischen Unsicherheit zu verstehen. Beispielsweise bedeutet die Klassifikation „unzureichende Dokumentationsqualität“, dass mithilfe eines statistischen Tests zu überprüfen ist, ob genügend Evidenz gegen die Hypothese vorliegt, dass der entsprechende Wert über 90 % liegt. Ansonsten ist „kein Hinweis auf unzureichende Dokumentationsqualität“ anzunehmen. Hierzu kann ein entsprechender p-Wert mit dem Signifikanzniveau  $\alpha = 0,05$  verglichen werden. Alternativ kann – aufgrund der Dualität zwischen statistischen Tests und Vertrauensintervallen – die untere Grenze eines geeigneten einseitigen 95 %-Vertrauensintervalls genutzt werden (entspricht einem zweiseitigen 90 %-Vertrauensintervall). Für methodische Ansätze zur Bestimmung der Vertrauensintervalle siehe Abschnitt 2.2. Dieses Vorgehen entspricht dem statistischen Verfahren, das zur Auslösung des Häufigkeits- und Relevanzkriteriums genutzt wird.

Eine ausführliche Begründung für die jeweiligen Grenzwerte und die Synthese erfolgt auf Seite 82).

Tabelle 16: Grenzwerte für die Bewertung der Dokumentationsqualität auf Standortebene

<b>1) Standortebene (pro individuelm Standort)</b>		
<b>b) alle Datenfelder: Festlegung der Grenzwerte</b>		
<b>Übereinstimmungsrate (Häufigkeitskriterium)</b>	✓ > 90 %: „Kein Hinweis auf unzureichende Dokumentationsqualität“ ≤ 90 %: „Unzureichende Dokumentationsqualität“ (aufgrund des Vorliegens von besonders häufigen Dokumentationsfehlern)	
<b>c/d) Relevanzkriterium und Sentinel Event: Festlegung der Grenzwerte</b>		
<b>Unterdokumentation relevanter Ereignisse oder Unterdokumentation eines Sentinel Events</b>	✓ Unterdokumentation relevanter Ereignisse: > 90 %: „Kein Hinweis auf unzureichende Dokumentationsqualität“ ≤ 90 %: „Unzureichende Dokumentationsqualität“	✓ Unterdokumentation eines Sentinel Events = „Unzureichende Dokumentationsqualität“

Tabelle 17: Grenzwerte für die Bewertung der Dokumentationsqualität auf standortübergreifender Ebene

<b>2) Standortübergreifende Ebene (für alle Standorte oder alle Fälle aller Standorte)</b>		
	2 <sub>1</sub> )Verteilung der Standortergebnisse	2 <sub>2</sub> )Verteilung der Fälle (Bundesergebnis)
<b>a) Pro Datenfeld: Festlegung der Grenzwerte</b>		
<b>Übereinstimmungsrate</b>	✓ ≤ 10 % der Standorte haben mehr als 5 % Fehldokumentationen = Kein Hinweis auf unzureichende Dokumentationsqualität > 10 % der Standorte haben mehr als 5 % Fehldokumentationen = Unzureichende Dokumentationsqualität	✓ ≥ 95 % = Kein Hinweis auf unzureichende Dokumentationsqualität < 95 % = Unzureichende Dokumentationsqualität
<b>Sensitivität (nur bei binär kodierten DF möglich)</b>	✓ ≤ 10 % der Standorte haben weniger als 90 % Sensitivität erreicht = Kein Hinweis auf unzureichende Dokumentationsqualität > 10 % der Standorte haben weniger als 90 % Sensitivität erreicht = Unzureichende Dokumentationsqualität	✓ ≥ 90 % = Kein Hinweis auf unzureichende Dokumentationsqualität < 90 % = Unzureichende Dokumentationsqualität
<b>Spezifität (nur bei binär kodierten DF möglich)</b>	✓ ≤ 10 % der Standorte haben weniger als 90 % Spezifität erreicht = kein Hinweis auf unzureichende Dokumentationsqualität > 10 % der Standorte haben weniger als 90 % Spezifität erreicht = Unzureichende Dokumentationsqualität	✓ ≥ 90 % = Kein Hinweis auf unzureichende Dokumentationsqualität < 90 % = Unzureichende Dokumentationsqualität

Tabelle 18: Synthese der Grenzwerte aus Tabelle 16 und Tabelle 17 und Bewertungsschema für die Standorte Ebene und standortübergreifende Ebene

1) Standorte Ebene (pro individuellem Standort)		2) Standortübergreifende Ebene (für alle Standorte oder alle Fälle aller Standorte)		
		Es kann in 2 <sub>1</sub> ) der Grenzwert überschritten werden, dann sind viele Standorte betroffen oder es kann in 2 <sub>2</sub> ) der Grenzwert überschritten werden, dann sind viele Patientinnen und Patienten betroffen (die z. B. in größeren Einrichtungen behandelt wurden).		
<ul style="list-style-type: none"> <li>▪ Häufigkeitskriterium</li> <li>▪ Relevanzkriterium</li> <li>▪ Unterdokumentation Sentinel Event</li> </ul>	Bewertung	<ul style="list-style-type: none"> <li>▪ Übereinstimmungsrate pro DF</li> </ul>	<ul style="list-style-type: none"> <li>▪ Sensitivität/Spezifität</li> </ul>	Bewertung und Zusatzinformation
		Entweder die Aspekte in Ebene 2 <sub>1</sub> ) Verteilung der Standortergebnisse ODER 2 <sub>2</sub> ) Verteilung der Fälle (Bundesergebnis) treffen zu		
Kein Kriterium ist erfüllt.	<b>Kein Hinweis auf unzureichende Dokumentationsqualität.</b>	<ul style="list-style-type: none"> <li>≤ 10 % der Standorte haben mehr als 5 % Fehldokumentationen</li> <li>UND</li> <li>≥ 95 % Übereinstimmungsrate pro DF bundesweit</li> </ul>	<ul style="list-style-type: none"> <li>≤ 10 % der Standorte haben weniger als 90 % von 1- Sensitivität/Spezifität erreicht</li> <li>UND</li> <li>1-Sensitivität/Spezifität &gt; 90 % bundesweit</li> </ul>	<p><b>Kein Hinweis auf unzureichende Dokumentationsqualität.</b></p> <p><u>Keine ergänzende Information.</u></p>
			<ul style="list-style-type: none"> <li>&gt; 10 % der Standorte haben weniger als 90 % 1- Sensitivität/Spezifität erreicht</li> <li>ODER</li> <li>1-Sensitivität/Spezifität &lt; 90 % bundesweit</li> </ul>	<p><b>Kein Hinweis auf unzureichende Dokumentationsqualität.</b></p> <p>ABER <u>ergänzende Information:</u> Bei diesem Datenfeld wurde</p> <ul style="list-style-type: none"> <li>▪ bei mehr als 10 % der Standorte mehr als 10 % fälschlicherweise nicht in der QS/fälschlicherweise für die QS dokumentiert</li> <li>▪ bundesweit mehr als 10 % fälschlicherweise nicht in der QS/fälschlicherweise für die QS dokumentiert.</li> </ul>
Eines der drei Kriterien ist erfüllt.	<b>Unzureichende Dokumentationsqualität.</b>	<ul style="list-style-type: none"> <li>&gt; 10 % der Standorte haben mehr als 5 % Fehldokumentationen</li> <li>ODER</li> <li>&lt; 95 % Übereinstimmungsrate pro DF bundesweit</li> </ul>	<ul style="list-style-type: none"> <li>≤ 10 % der Standorte haben weniger als 90 % 1-Sensitivität/Spezifität erreicht</li> <li>UND</li> <li>1-Sensitivität/Spezifität &gt; 90 % bundesweit</li> </ul>	<p><b>Unzureichende Dokumentationsqualität.</b></p> <p>UND <u>ergänzende Information:</u> Kein Hinweis auf erhöhten Anteil an fälschlicherweise nicht in der QS oder fälschlicherweise für die QS dokumentierten Fällen für dieses Datenfeld.</p>
			<ul style="list-style-type: none"> <li>&gt; 10 % der Standorte haben weniger als 90 % 1-Sensitivität/Spezifität erreicht</li> <li>ODER</li> <li>1-Sensitivität/Spezifität &lt; 90 % bundesweit</li> </ul>	<p><b>Unzureichende Dokumentationsqualität</b></p> <p>UND <u>ergänzende Information:</u> Bei diesem Datenfeld wurde</p> <ul style="list-style-type: none"> <li>▪ bei mehr als 10 % der Standorte mehr als 10 % fälschlicherweise nicht in der QS/fälschlicherweise für die QS dokumentiert</li> <li>▪ bundesweit mehr als 10 % fälschlicherweise nicht in der QS/fälschlicherweise für die QS dokumentiert.</li> </ul>

**Begründung zu den festgelegten Grenzwerten und die Bewertung der Dokumentationsqualität für die Standortebene:**

Die festgelegten Grenzwerte sind wie oben erläutert nicht für den einzelnen betrachteten Aspekt allein zu interpretieren, sondern nur in der vorgeschlagenen Synthese gemeinsam. Deshalb wird vorgeschlagen, nicht nur die Übereinstimmungsrate über alle Datenfelder, sondern auch die Relevanz der Dokumentationsfehler in die Bewertung der Dokumentationsqualität einzubeziehen. Eine Unterdokumentation relevanter Ereignisse ab > 10 % wird als unzureichend angenommen (ggf. mit Berücksichtigung von Unsicherheit). Das Relevanzkriterium umfasst wichtige QI-relevante Datenfeldkombinationen, für welche eine Unterdokumentation betrachtet wird. Ähnliches erfolgt durch Betrachtung der Sensitivität pro Datenfeld. Beide Kriterien erfassen damit, wie viele in der Patientenakte dokumentierte Ereignisse nicht für die QS dokumentiert wurden. Durch die Schnittmenge des Relevanzkriteriums zur Sensitivität sowie zur Spezifität durch die im Relevanzkriterium vorgenommene logische Negation negativ formulierter Ereignisse werden die Sensitivität und Spezifität pro Datenfeld nicht in die Bewertung auf Standortebene einbezogen. Der Vorteil durch den Einbezug des Relevanzkriteriums anstelle der Sensitivität und Spezifität je Datenfeld ist, erstens, dass sie sich eindeutig an den Rechenregeln der Qualitätsindikatoren orientieren und zweitens, dass nicht nur das Ausmaß der Fehldokumentation in die eine oder andere Richtung betrachtet wird, sondern durch die Zusammensetzung der Ereignisse des Relevanzkriteriums klar eine Unterdokumentation zugunsten des Krankenhausstandortes betrachtet wird. Ebenso wird nicht die Übereinstimmungsrate jedes Datenfeldes einzeln bewertet, da bereits eine einzelne Fehldokumentation pro Datenfeld ausreichen würde um die Dokumentationsqualität dieses Datenfeldes als unzureichend zu bewerten. Eine Übereinstimmungsrate pro Datenfeld von > 95 % sollte dennoch grundlegend auch von jedem Standort angestrebt werden. Einzig für die Bewertung auf Standortebene hält das IQTIG die vorgeschlagene Grundlage, nämlich das Häufigkeits- und Relevanzkriterium, aufgrund der kleinen Stichprobengröße von 20 Fällen für angemessen.

Das vorgeschlagene Bewertungsschema ermöglicht damit eine datenfeldübergreifende Bewertung der Dokumentationsqualität auf Krankenhausstandortebene. Eine striktere Bewertung als für die Auslösung des gezielten Datenabgleichs wird an dieser Stelle nicht vorgenommen, da die Anwendung der Kriterien zur Häufigkeit und Relevanz bereits als eine Art Bewertung der Dokumentationsqualität betrachtet werden können. Der Vorteil dieser Festlegung ist, dass es für die Standorte keine diskrepante Aussage zur Auslösung eines gezielten Datenabgleichs gibt. In der Betrachtung der Übereinstimmungsrate des Standortes werden wie in Abschnitt 2.1.2 beschrieben nicht alle Datenfelder einbezogen, weshalb eine Unterscheidung zur Betrachtung auf standortübergreifender Ebene, die sich auf jedes Datenfeld bezieht, durchaus sinnvoll erscheint.

**Synthese**

Eine unzureichende Dokumentationsqualität im betrachteten Erfassungsjahr liegt auf Standortebene jeweils dann vor, wenn eins der Kriterien gemäß Tabelle 16 erfüllt ist. Treten somit Do-

kumentationsdefizite in mindestens einem der einzelnen Aspekte auf, dann ist von unzureichender Dokumentationsqualität auszugehen. Insgesamt ergeben sich in der Gesamtbeurteilung die Kombinationsmöglichkeiten mit folgender Bewertung:

1. Keines der Kriterien ist erfüllt, d. h. der Standort hat in seiner Übereinstimmungsrate für alle Datenfelder mehr als 90 % richtig dokumentiert und es lagen weder relevante Unterdokumentationen noch die Unterdokumentation eines Sentinel Events vor.

→ **Kein Hinweis auf unzureichende Dokumentationsqualität**

2. Mindestens eins der Kriterien ist erfüllt, d. h. der Standort hat in seiner Übereinstimmungsrate für alle Datenfelder weniger als 90 % richtig dokumentiert oder es lagen relevante Unterdokumentationen vor oder die Unterdokumentation eines Sentinel Events.

→ **Unzureichende Dokumentationsqualität**

#### **Begründung zu den festgelegten Grenzwerten und die Bewertung der Dokumentationsqualität für die standortübergreifende Ebene:**

Für die Bewertung auf Systemebene wird vorgeschlagen, diese pro Datenfeld vorzunehmen und im Ergebnis einerseits die Verteilung der Krankenhausstandorte (2<sub>1</sub>) zu betrachten und andererseits die Behandlungsfälle (2<sub>2</sub>) bundesweit.

Einer der beiden Faktoren muss den Grenzwert überschreiten, um einen Hinweis auf eine unzureichende Dokumentationsqualität zu erhalten.

Die Festlegung des Anteils auf 10 % der Standorte für eine Betrachtung von weniger oder mehr als 5 % Fehldokumentationen orientiert sich an der Betrachtungsweise von besonderem Handlungsbedarf bei QI-Ergebnissen. Anders als im Konzept für den besonderen Handlungsbedarf ist hier die Stichprobenunsicherheit entsprechend durch einen statistischen Test bzw. die Betrachtung eines geeigneten Konfidenzintervalls zu berücksichtigen, da nur für eine Stichprobe der Standorte ein Ergebnis vorliegt. Dies führt aufgrund der Tatsache, dass lediglich 5 % der Standorte untersucht werden dazu, dass deutlich mehr als 10 % der Standorte in der Standortstichprobe entsprechende Fehldokumentationen aufweisen müssen, um dieses Kriterium auszulösen.

Dies bedeutet, dass unter Betrachtung statistischer Unsicherheit davon ausgegangen werden kann, dass in mehr als jeder zehnten Einrichtung die Übereinstimmungsrate des entsprechend betrachteten Datenfeldes niedriger als 95 % ist.

Wie in Abschnitt 7.1 bereits beschrieben wird für eine Übereinstimmung der Daten mit  $\geq 95\%$  eine angemessene Dokumentationsqualität angenommen. Die für die standortübergreifende Ebene normativ festgelegten Grenzwerte orientieren sich damit an denen in der Literatur für die Konkordanz der Dokumentationsqualität in Registern. Es kann daher davon ausgegangen werden, dass eine Übereinstimmungsrate bei Einbezug aller Fälle pro Datenfeld von unter 95 % als unzureichende Dokumentationsqualität angenommen werden kann. In diesem Fall ist eine Verbesserung der Dokumentation des Datenfeldes erforderlich.

Die Ergebnisse der Sensitivität und Spezifität werden nur für binär kodierte Datenfelder ergänzend betrachtet. Es soll lediglich ein Hinweis auf eine vorliegende Unter- oder Überdokumentation gegeben werden und kein expliziter Einbezug in die Bewertung der Datenfelddokumentation aufgrund der ebenfalls in Abschnitt 7.1 erläuterten Problematik einer heterogenen Bewertung von Datenfeldern und eines fehlenden wissenschaftlich fundierten allgemeingültigen Grenzwertes für die Sensitivität und Spezifität.

An dieser Stelle unterscheidet sich das vorliegende Konzept von der durch AQUA (2014) vorgenommenen Bewertung der Dokumentationsqualität auf Datenfeldebene dahingehend, dass nicht zwischen dem Einbezug von entweder der Übereinstimmungsrate oder der Sensitivität und Spezifität unterschieden wird. Deshalb erfolgt mit dem Vorschlag des IQTIG unter allen erläuterten Schwierigkeiten der Kompromiss der zusätzlichen Nennung von Unter- und Überdokumentationen (fälschlicherweise nicht in der QS oder fälschlicherweise für die QS) für binär kodierte Datenfelder, da diese Information bei interessierenden Ereignissen von besonderer Wichtigkeit ist. Aufgrund einer fehlenden Querverbindung zu den Auswirkungen auf die Ergebnisse der Qualitätsindikatoren bei vorliegender Unter- oder Überdokumentation für die Rechenregeln interessierender Ereignisse ist die Bewertung der Dokumentationsqualität auf Datenfeldebene limitiert (siehe Limitationen in Abschnitt 7.4).

Die Festlegung eines Grenzwertes von  $< 90\%$  Sensitivität und  $< 90\%$  Spezifität ist ebenfalls normativ festgelegt. Aufgrund dessen, dass der Einbezug nur für binär kodierte Datenfelder erfolgt und interessierende Ereignisse für die QI-Berechnung aufgrund sich daraus ggf. ergebener Fehlberechnungen nicht unter- oder überdokumentiert werden sollten, erscheinen die festgelegten Grenzwerte durchaus praktikabel. Eine Unter- oder Überdokumentation von  $10\%$  soll dennoch keineswegs als vertretbar fehlinterpretiert werden. Dennoch ist, wie am Beispiel in Abschnitt 7.1 verdeutlicht, insbesondere der Anteil der Sensitivität nicht als stabil in Bezug auf die Fallzahl anzusehen. So lange kein Konzept für die Übertragung von Fehldokumentationen auf QI-Ergebnisse vorliegt, erscheint die Lösung der Betrachtung der Übereinstimmungsrate von Datenfeldern und einem Hinweis von Unter-/Überdokumentationen bei binär kodierten Datenfeldern als geeigneter Kompromiss.

### **Zusammenfassung**

Unzureichende Dokumentationsqualität im betrachteten Erfassungsjahr liegt auf Standortebene jeweils dann vor, wenn eines der Kriterien für einen Hinweis auf Mängel in der Dokumentationsqualität in Tabelle 17 erfüllt ist. Treten somit Dokumentationsdefizite der einzelnen Aspekte auf, besteht in der Synthese ein hinreichender Hinweis auf unzureichende Dokumentationsqualität. Insgesamt ergeben sich in der Gesamtbeurteilung zwei Kombinationsmöglichkeiten mit folgender Bewertung und der ergänzenden Zusatzinformation:

2<sub>1</sub>) Die Übereinstimmungsrate pro Datenfeld liegt bundesweit über  $95\%$ .

2<sub>2</sub>) Die Übereinstimmungsrate pro Datenfeld ist bei weniger als  $10\%$  der Standorte mit mehr als  $5\%$  fehdokumentiert.

→ **Kein Hinweis auf unzureichende Dokumentationsqualität**

2<sub>1</sub>) Die Übereinstimmungsrate pro Datenfeld liegt bundesweit unter 95 %.

2<sub>2</sub>) Die Übereinstimmungsrate pro Datenfeld ist bei mehr als 10 % der Standorte mit mehr als 5 % fehdokumentiert.

→ **Unzureichende Dokumentationsqualität**

#### Ergänzende Information:

Wenn das Datenfeld binär kodiert ist und **zusätzlich** 1 - Sensitivität und/oder 1 - Spezifität bundesweit unter 90 % liegt oder mehr als 10 % der Standorte im Ergebnis 1 - Sensitivität und/oder 1 - Spezifität weniger als 90 % erreicht haben, dann:

→ Bei diesem Datenfeld waren bundesweit mehr als 10 % fälschlicherweise nicht in der QS/fälschlicherweise für die QS dokumentiert, oder

→ Bei diesem Datenfeld wurde bei mehr als 10 % der Standorte mehr als 10 % fälschlicherweise nicht in der QS/fälschlicherweise für die QS dokumentiert.

## 7.4 Limitationen

Die standortbezogene Bewertung lässt Ergebnisse der Statistischen Basisprüfung mit Strukturiertem Dialog für den jeweiligen Standort bisher außen vor. Ebenso werden die Ergebnisse bisher nicht in den Qualitätsberichten der Krankenhäuser veröffentlicht. An dieser Stelle wird grundlegendes Weiterentwicklungspotenzial gesehen, zunächst für die Veröffentlichung der Ergebnisse der Statistischen Basisprüfung mit Strukturiertem Dialog und weiterhin für einen Einbezug dieser Ergebnisse in die Gesamtbeurteilung der Dokumentationsqualität eines Standortes.

In Bezug auf das vorgeschlagene Bewertungssystem werden für die Standortebebene und die standortübergreifende Ebene unterschiedliche Kriterien herangezogen: Einmal die Übereinstimmungsrate über alle Datenfelder (Häufigkeitskriterium) und das Relevanzkriterium sowie für die standortübergreifende Ebene die Übereinstimmungsrate je Datenfeld und eine Zusatzinformation zur Sensitivität und Spezifität. Vor dem Hintergrund des Einbezugs von Sensitivität und Spezifität wurde die Problematik bereits erläutert, dass nicht alle Datenfelder für diese Aspekte betrachtet werden können. Eine Dichotomisierung der kategorial und metrisch ausgeprägten Datenfelder wird aus Aufwandsgründen nicht vorgenommen. Um die unterschiedlichen Arten von Datenfeldern nicht heterogen zu bewerten, wird nur die Übereinstimmungsrate je Datenfeld gezählt. Für binäre Datenfelder erfolgt – sofern zutreffend – die oben beschriebene „ergänzende Information“.

Dadurch soll die Information, dass bei bestimmten Datenfeldern ggf. Problematiken in Bezug auf eine Über- oder Unterdokumentation vorliegen, erhalten bleiben und sich ggf. Handlungsbedarf ableiten lassen. Diese Informationen fehlen allerdings bei anders kategorisierten Datenfeldern.

Aufgrund der oben geschilderten Herausforderungen für die Beurteilung der Dokumentationsqualität auf Standortebebene sowie auf standortübergreifender Ebene erscheint der vorliegende Vorschlag als ein solider Kompromiss.

Für den Bewertungsaspekt (2<sub>1</sub>) wird gefordert, dass 10 % aller Standorte ein bestimmtes Maß an Fehldokumentationen aufweisen. Auch nach Berücksichtigung der Tatsache, dass es sich um eine (kleine) Zufallsstichprobe von Standorten handelt (Vertrauensintervall), kann dies insbesondere in den direkten Verfahren bedeuten, dass sehr wenige Standorte mit Dokumentationsfehlern ausreichen, um das Kriterium auszulösen. Hier kann nach Erfahrungen in der Praxis ggf. eine Anpassung des Grenzwertes vorgenommen werden.

Grundlegendes Weiterentwicklungspotenzial wird an dieser Stelle in Bezug auf die Querverbindung zu QI-Ergebnissen gesehen. Auswirkungen festgestellter Fehldokumentationen auf QI-Ergebnisse werden bisher im Regelbetrieb der Datenvalidierung nicht benannt und bewertet. Die Entwicklung eines Konzepts für die Übertragung der Erkenntnisse aus der Datenvalidierung auf die Qualitätsindikatoren könnte daher hilfreich sein. Das vorliegende Bewertungsschema könnte zunächst als Aufgreifkriterium für die Überprüfung bestimmter Qualitätsindikatoren dienen. Die Bewertung der Dokumentationsqualität bliebe damit zunächst von den Auswirkungen unabhängig. Allerdings könnte in einer Weiterentwicklung auch ein Zusammenhang hergestellt werden.

Es wird vorgeschlagen, die Ergebnisse der Betrachtungsebenen der Dokumentationsqualität aus Tabelle 15 im Bericht zur Datenvalidierung nach §15 Abs. 2 QSKH-RL aufzuführen. Dies bedeutet jedoch umfassendere Auswertungen (z. B. Benchmarks), die nicht alle bis zum Datenvalidierungsbericht 2020 umgesetzt werden können. Langfristig ist die Berichterstattung dieser Aspekte für ein vollumfängliches Bild zur Dokumentationsqualität allerdings unerlässlich.

## 8 Berichterstattung nach § 136b SGB V

### Gegenstand der Beauftragung Nr. 1.7

Die Krankenhäuser sind nach § 136b Abs. 1 Satz 1 Nr. 3 SGB V verpflichtet, strukturierte Qualitätsberichte auf ihrer eigenen Webseite zu veröffentlichen. Nach § 136b Abs. 6 und 7 sind zudem der G-BA sowie die Landesverbände der Krankenkassen und die Ersatzkassen verpflichtet, die Ergebnisse im Internet zu veröffentlichen. Für die Veröffentlichung der Ergebnisse des Datenvalidierungsverfahrens in den Qualitätsberichten nach § 136b Abs. 1 Satz 1 Nr. 3 SGB V werden die im Folgenden genannten Formulierungen vorgeschlagen. In den Vorschlag zur Berichterstattung zunächst nicht eingeschlossen wurde das Kriterium zur „wiederholten rechnerischen Auffälligkeit“ aufgrund der vom IQTIG benannten Limitationen (vgl. Abschnitt 5.5).

### 8.1 Berichterstattung nach Durchführung des Stichprobenverfahrens mit Datenabgleich

Wenn die Dokumentationsqualität eines Krankenhausstandortes im Stichprobenverfahren mit Datenabgleich überprüft wurde:

*Der Krankenhausstandort wurde zufällig für eine Überprüfung der Dokumentationsqualität im QS-Verfahren [jeweiliges QS-Verfahren der Zufallsstichprobe] im Erfassungsjahr [jeweiliges Erfassungsjahr] ausgewählt.*

Falls der Krankenhausstandort keine Auffälligkeiten im Häufigkeits- und Relevanzkriterium einschließlich der Unterdokumentation eines Sentinel Events gemäß Tabelle 18 aufweist, wird folgende Formulierung in die Qualitätsberichte aufgenommen:

**→ Kein Hinweis auf unzureichende Dokumentationsqualität.**

*Die Ergebnisse des Stichprobenverfahrens mit Datenabgleich sind folgende:*

- *Die Fehldokumentationsrate des Krankenhausstandorts lag bei: X % (90 %-Vertrauensintervall: Y % – Z %)*
- *Die Unterdokumentationsrate der untersuchten relevanten Ereignisse lag bei X % (90 %-Vertrauensintervall: Y % – Z %)*
- *Unterdokumentation eines Sentinel Events: [lag nicht vor]*

Falls der Krankenhausstandort Auffälligkeiten im Häufigkeits- oder Relevanzkriterium oder die Unterdokumentation eines Sentinel Events gemäß Tabelle 18 aufweist wird folgende Formulierung in die Qualitätsberichte aufgenommen:

**→ Unzureichende Dokumentationsqualität.**

*Die Ergebnisse des Stichprobenverfahrens mit Datenabgleich sind folgende:*

- *Die Fehldokumentationsrate des Krankenhausstandorts lag bei: X % (90 %-Vertrauensintervall: Y % – Z %)*

- Die Unterdokumentationsrate der untersuchten relevanten Ereignisse lag bei X % (90 %-Vertrauensintervall: Y % – Z %)
- Unterdokumentation eines Sentinel Events: [lag vor/ lag nicht vor]

Abschließender Satz:

*Die Dokumentation wird zum Folgejahr erneut überprüft.*

Falls der Krankenhausstandort bei Fehldokumentationen von erheblicher Häufigkeit auffällig wurde, wird folgendes berichtet:

**→ Unzureichende Dokumentationsqualität.**

*Die Ergebnisse des Stichprobenverfahrens mit Datenabgleich sind folgende:*

- *Es wurden Fehldokumentationen von erheblicher Häufigkeit festgestellt.*
- *Die Fehldokumentationsrate des Krankenhausstandortes lag bei: X % (90 %-Vertrauensintervall: Y % – Z %)*
- *Die Unterdokumentationsrate der untersuchten relevanten Ereignisse lag bei X % (90 %-Vertrauensintervall: Y % – Z %)*
- *Unterdokumentation eines Sentinel-Events: [lag vor/ lag nicht vor]*
- *Das verantwortliche Gremium auf Landes- oder Bundesebene (§ 13 Abs. 3 Satz 2 QSKH-RL) wurde über diese Dokumentationsdefizite informiert und entscheidet über weitere Maßnahmen.*

## **8.2 Berichterstattung nach Durchführung des gezielten Datenabgleichs**

Falls der Krankenhausstandort Auffälligkeiten im Häufigkeits- und Relevanzkriterium einschließlich der Unterdokumentation eines Sentinel Events im Stichprobenverfahren mit Datenabgleich des Vorjahres gemäß Tabelle 18 aufgewiesen hat, wird folgende Formulierung verwendet:

*Bei diesem Krankenhausstandort wurden im vorherigen Jahr im QS-Verfahren [QS-Verfahren] im Stichprobenverfahren mit Datenabgleich Fehldokumentationen festgestellt, die einen gezielten Datenabgleich ausgelöst haben. Daher wurde eine erneute Überprüfung der Dokumentation auf Basis der Daten des Erfassungsjahres [Erfassungsjahr der Zufallsstichprobe +1] vorgenommen.*

Falls der Krankenhausstandort keine Auffälligkeiten im Häufigkeits- und Relevanzkriterium einschließlich der Unterdokumentation eines Sentinel Events gemäß Tabelle 18 aufweist, wird folgende Formulierung in die Qualitätsberichte aufgenommen:

**→ Kein Hinweis auf unzureichende Dokumentationsqualität.**

*Die Ergebnisse des gezielten Datenabgleichs sind folgende:*

- *Die Fehldokumentationsrate des Krankenhausstandorts lag bei: X % (90 %-Vertrauensintervall: Y % – Z %)*

- Die Unterdokumentationsrate der untersuchten relevanten Ereignisse lag bei X % (90 %-Vertrauensintervall: Y % – Z %)
- Unterdokumentation eines Sentinel Events: [lag nicht vor]
- Das Ergebnis der erneuten Überprüfung wurde dem nach §14 QSKH-RL auf Landes- oder Bundesebene verantwortlichen Gremium zur Information mitgeteilt.

Falls der Krankenhausstandort nach gezieltem Datenabgleich Auffälligkeiten im Häufigkeits- oder Relevanzkriterium oder die Unterdokumentation eines Sentinel Events gemäß Tabelle 18 aufweist, wird folgende Formulierung in die Qualitätsberichte aufgenommen:

**→ Wiederholt unzureichende Dokumentationsqualität.**

*Die Ergebnisse des gezielten Datenabgleichs sind folgende:*

- Die Fehldokumentationsrate des Krankenhausstandortes lag bei: X % (90 %-Vertrauensintervall: Y % – Z %)
- Die Unterdokumentationsrate der untersuchten relevanten Ereignisse lag bei X % (90 %-Vertrauensintervall: Y % – Z %)
- Unterdokumentation eines Sentinel Events: [lag vor/ lag nicht vor]

Abschließender Satz:

*Das Ergebnis der erneuten Überprüfung wurde dem nach § 14 QSKH-RL auf Landes- oder Bundesebene verantwortlichen Gremium zur Information mitgeteilt. Dieses entscheidet über weitere Maßnahmen.*

### 8.3 Limitationen

Durch die Berichterstattung in den Qualitätsberichten nach § 136b Abs. 1 Satz 1 Nr. 3 SGB V über eine zufällige Teilnahme am Stichprobenverfahren mit Datenabgleich erfolgt eine Deanonymisierung des Standorts gegenüber dem UA QS/Lenkungsgremium auf Landesebene. Dies ist an anderer Stelle eine mögliche Konsequenz dauerhaft schlechter Versorgungsqualität oder mangelnder Kooperationsbereitschaft. Beim Stichprobenverfahren mit Datenabgleich besteht – anders als beim gezielten Datenabgleich – eine reine Zufallsauswahl für die 5 %-Stichprobe der zu validierenden Standorte. Eine Deanonymisierung des Standorts gegenüber dem UA QS/Lenkungsgremium auf Landesebene könnte daher fehlinterpretiert werden.

Zur Berichterstattung der Ergebnisse aus der Statistischen Basisprüfung ist anzumerken, dass rechnerische Ergebnisse der Auffälligkeitskriterien keine Aussage zur Dokumentationsqualität eines Krankenhausstandortes zulassen, weshalb dieses Kriterium nicht im vorliegendem Vorschlag abgebildet wird. Krankenhausstandorte, die in den Auffälligkeitskriterien zur Plausibilität und Vollständigkeit rechnerisch auffällig sind, können trotzdem korrekte Daten geliefert haben, wenn sie z. B. eine besondere Versorgungsstruktur oder eine besondere Patientenklientel auf-

weisen. Die rechnerische Auffälligkeit sagt nur aus, dass die übermittelten Werte unwahrscheinlich sind. Werte, die sicher unplausibel sind, werden bereits über die Plausibilitätsregeln der Spezifikation abgefangen (vgl. Limitationen zu Kapitel 5).

## 9 Prüfung nicht übermittelter Fälle

### Gegenstand der Beauftragung Nr. I.3

Es sollte laut Beauftragung eine Methodik zur Prüfung nicht übermittelter Fälle entwickelt werden. Die Krankenhausstandorte sind nach § 23 QSKH-RL verpflichtet, die QS-Dokumentation zu allen Fällen, die durch den QS-Filter des jeweiligen QS-Verfahrens ausgelöst wurden, an die auf der Landesebene beauftragten Stellen zu übermitteln (Ist-Fälle) und zugleich eine Konformitätserklärung über die Richtigkeit dieser übermittelten Daten abzugeben. Wenn Krankenhausstandorte nicht alle zu dokumentierenden Fälle an die auf der Landesebene beauftragten Stellen übermitteln, verletzen sie das Vollständigkeitskriterium. Nicht übermittelte Fälle können zudem die Indikatorergebnisse eines Krankenhausstandortes verändern. Für einen fairen Leistungserbringervergleich ist es somit essenziell, dass alle Fälle übermittelt werden. Um die Dokumentationsqualität eines Falles zu überprüfen wird normalerweise ein Abgleich zwischen den QS-Daten und der Patientenakte vorgenommen. Da bei den nicht übermittelten Fällen keine QS-Daten übermittelt wurden, kann kein Abgleich der QS-Daten mit der Patientenakte erfolgen. Somit ist auch eine zufällige Identifizierung eines nicht übermittelten Falles im Rahmen eines Stichprobenverfahrens mit Datenabgleich ausgeschlossen.

Sowohl durch das Verfahren nach § 23 QSKH-RL als auch durch die Auffälligkeitskriterien zur Unterdokumentation der Statistischen Basisprüfung kann auf Krankenhaus- (und Standortebene) die Anzahl nicht übermittelter Fälle festgestellt werden. Eine Sanktionierung durch Qualitätssicherungsabschläge auf Krankenhausebene erfolgt durch § 24 QSKH-RL. Gründen für eine rechnerisch auffällige Unterdokumentation kann auf Standortebene durch einen Strukturierten Dialog gemäß § 11–14 QSKH-RL nachgegangen werden. Gemäß § 9 Abs. 9 QSKH-RL ist beabsichtigt, die Ergebnisse des Datenvalidierungsverfahrens (inkl. der Berichterstattung des gezielten Datenabgleichs, vgl. Kapitel 8) in den Qualitätsberichten der Krankenhäuser zu veröffentlichen. Durch die bereits jetzt in den Qualitätsberichten nach § 136b Abs. 1 Satz 1 Nr. 3 SGB V Abschnitt C1–1 veröffentlichten Dokumentationsraten (IST/Soll-Abgleich) wird deutlich, wenn der Standort Fälle nicht übermittelt hat.

### Perspektivische Überlegungen

Grundsätzlich kann überlegt werden, anhand der Soll-Statistik bei Unter- und Überdokumentationen ab einer gewissen Schwelle der z. B. im Stichprobenverfahren mit Datenabgleich gezogenen Standorte diese Standorte anzuhalten, die nicht übermittelten Fälle nachzuliefern und gezielt zu überprüfen. Ebenso könnten die zu viel übermittelten Fälle dahingehend überprüft werden, ob diese richtig dokumentiert wurden, oder die Unterdokumentationen Eigenschaften aufweisen, die auf systematische Unterdokumentation bestimmter Fallkonstellationen hindeuten. Allerdings bezöge sich die Divergenz zwischen Soll und Ist ggf. nicht auf das überprüfte Verfahren. Demnach wäre auch zu überlegen, ob der Soll-Ist-Abgleich nicht bundesweit betrachtet werden sollte und Standorte mit einem gewissen Prozentsatz an Divergenz in den Pool für eine mögliche Überprüfung mitaufgenommen werden sollten. Es könnten dann X-Prozent herange-

zogen werden, die die Daten für diese Fälle nachliefern müssten. Dann bliebe zu überlegen, inwiefern ein gezielter Fallabgleich erfolgen kann, da wahrscheinlich mehrere QS-Verfahren betroffen sein werden. Demnach wäre ggf. der erste Schritt, die QS-Daten für eine neue QI-Berechnung heranzuziehen und im zweiten Schritt eine weitere Auswahl zu überprüfender Standorte zu wählen. Wie eingangs erwähnt ist eine Querverbindung zwischen Dokumentationsqualität und QS-Auswertungen sinnvoll, bedarf jedoch eines gesonderten Konzepts, in dem z. B. auch die Ergebnisse von im SD festgestellten Fehldokumentationen in den AKs und/oder QIs mitbedacht werden.

## 10 Spezifikation des bundeseinheitlichen Prüfberichts

### Gegenstands der Beauftragung Nr. I.6

Für die Auswahl der im bundeseinheitlichen Prüfbericht darzustellenden Datenfelder gilt, dass mindestens alle Datenfelder abzugleichen sind, die in die Berechnung der Qualitätsindikatoren des jeweiligen Verfahrens eingehen (§ 9 Abs. 4 Satz 4 QSKH-RL).

Der bundeseinheitliche Prüfbericht enthält je Krankenhausstandort die Ergebnisse des Abgleichs für alle zweiterfassten Datenfelder aller überprüften Fälle. Folgende Informationen können dem bundeseinheitlichen Prüfbericht entnommen werden:

- Aus Patientenakte nicht ersichtlich: Anzahl der Fälle pro Datenfeld, bei denen die Angabe aus der Patientenakte nicht ersichtlich war.
- Anzahl Abweichungen: Anzahl der Fälle pro Datenfeld, bei denen eine Abweichung zwischen der Patientenakte und der QS-Dokumentation vorlag.
- Anzahl Übereinstimmungen: Anzahl der Fälle pro Datenfeld, bei denen eine Übereinstimmung zwischen der Patientenakte und der QS-Dokumentation vorlag.
- Anzahl Datensätze pro Feld: Anzahl gezählter Teildatensätze (Bögen) für das jeweilige Feld
- Übereinstimmungsrate pro Feld (%): Anteil derjenigen Fälle pro Datenfeld, bei denen die Angabe aus der Patientenakte mit der Angabe aus der QS-Dokumentation übereinstimmt, an allen für dieses Datenfeld zweiterfassten Fällen. Diese Übereinstimmungsrate ist zusammen mit dem beschriebenen statistischen Verfahren Grundlage für die Auslösung des gezielten Datenabgleichs aufgrund von besonders häufigen Dokumentationsfehlern.
- Übereinstimmungsrate gesamt: Anzahl und Anteil (%) der Übereinstimmungen aller gezählten zweiterfassten Datenfelder für alle geprüften und abgeschlossenen Fälle (zweiterfassten Vorgangsnummern) insgesamt.
- Abgeschlossene Fälle: Anzahl der Fälle, die abgeschlossen wurden.
- Anzahl nicht zweiterfasster Fälle: Anzahl der Fälle, die nicht datenvalidiert wurden.
- Anzahl Reservefälle: Anzahl der abgeschlossenen Reservefälle.
- Fehldokumentationsrate relevanter Ereignisse: Je Ereignis die Angabe, dass X von Y laut Patientenakte eingetretenen Ereignissen unterdokumentiert wurden.

Neben der Auswertung aller geprüften Fälle (bundeseinheitlicher Prüfbericht), werden je geprüftem Fall sogenannte Fallabgleich erstellt. Diese enthalten Informationen, welche Werte je Datenfeld für die externe stationäre Qualitätssicherung dokumentiert wurden und welche Werte je Datenfeld aus der Patientenakte im Krankenhaus entnommen wurden. Es wird je Feld farblich dargestellt ob die Werte kongruent waren und somit eine Übereinstimmung festgestellt wurde (Fallabgleich).

- Ergänzende Kommentare der überprüfenden Stelle zu jedem Fall

Die Krankenhäuser können unmittelbar im Anschluss an die Dokumentationsüberprüfung sehen, bei welchen Datenfeldern mögliche Defizite in der Dokumentationsqualität vorliegen bzw. wo es Verbesserungspotenzial gibt. Das unter Gegenstand der Beauftragung I.8 entwickelte Erfassungstool erstellt den bundeseinheitlichen Prüfbericht automatisch.

**Hinweis**

Die im Prüfbericht vorgelegten Übereinstimmungsraten zum Häufigkeits- und Relevanzkriterium werden vom IQTIG nach Abschluss der Überprüfungen zum 31. Oktober eines jeden Jahres im Anschluss für die Auslösung der Kriterien erneut berechnet. Erst dabei werden einerseits mögliche Ausschlüsse auf Basis von Mängeln in der Spezifikation (siehe Abschnitte 3.3, 4.2 bzw. 0) durchgeführt sowie die sich aus der Stichprobenziehung ergebende statistische Unsicherheit (Abschnitt 2.1) berücksichtigt. Daher sind die abschließenden Bewertungsergebnisse, die zur Auslösung eines gezielten Datenabgleichs führen, für die jeweiligen Kriterien noch nicht unmittelbar für die Standorte im Prüfbericht darstellbar.

# 11 Entwicklung eines Erfassungstools für die prüfenden Stellen für den Datenabgleich vor Ort

## Gegenstands der Beauftragung Nr. I.8

Um einer einheitlichen Datenvalidierung zu entsprechen hat das IQTIG beschlossen, ein Erfassungstool zu entwickeln, das nicht nur ausschließlich für die Datenvalidierung nach § 9 QSKH-RL genutzt werden kann, sondern eine Software umzusetzen, die allen Bedürfnissen und Aufgabenstellungen der Datenvalidierung nach QSKH-RL, der Richtlinie zu planungsrelevanten Qualitätsindikatoren (plan. QI-RL) und der DeQS-RL entspricht.

Das neue Erfassungstool für die Zweiterfassung mit Datenabgleich erfüllt folgende Kriterien:

- Schnelle Bereitstellung für ein Erfassungsjahr
- Flexible Auswahl von Erfassungsfeldern
- Wartbarkeit
- Leichte Testbarkeit
- Stabilität

Eine schnelle Bereitstellung des Erfassungstools wird dadurch erreicht, dass die Datenfelder der QS-Verfahren automatisch aus der Spezifikation des zugrunde liegenden Erfassungsjahres gelesen werden.

Trotz gleicher Menüpunkte wurde die Stabilität des neuen Zweiterfassungstools deutlich erhöht, indem bekannte potentielle Fehlerquellen durch entsprechende Prüffunktionen oder Absicherungen ausgeschlossen wurden. Zudem kommen Komponenten zum Einsatz, mit denen die Anforderungen besser erfüllt werden, als es mit MS Access möglich war.

Das neue Tool wurde auf Basis der Programmiersprache Java erstellt, da der Verbreitungsgrad und die Plattformunabhängigkeit von Java wichtige Vorteile dieser Programmiersprache darstellen. Die Plattformunabhängigkeit stellt sicher, dass das Erfassungstool für die Anwenderinnen und Anwender in verschiedensten Umgebungen eingesetzt werden kann.

Detaillierte Beschreibungen und Begründungen zum Gegenstand der Beauftragung Nr. I.8. sind dem Abschlussbericht „Erfassungstool für die Zweiterfassung mit Datenabgleich“ vom 31. Oktober 2018 zu entnehmen.

## 12 Abschließende Empfehlungen

Grundlegend gehen aus den in den einzelnen Kapiteln genannten Limitationen für die Umsetzung verschiedene Verfahrensempfehlungen hervor. Auf Grundlage dieses Weiterentwicklungsprojekts lassen sich folgende generelle Empfehlungen für die zukünftige Weiterentwicklung der Datenvalidierung ableiten:

### **Qualitätsberichte der Krankenhäuser**

Die Ergebnisse des Strukturierten Dialogs zu den Auffälligkeitskriterien werden bisher (noch) nicht in den Qualitätsberichten der Krankenhäuser veröffentlicht. Allerdings werden momentan noch ca. 1/3 Hinweise versendet, sodass eine Veröffentlichung ggf. nur eingeschränkt sinnvoll ist. Im Rahmen der Weiterentwicklung zum Strukturierten Dialog wurden Auffälligkeitskriterien bisher nicht betrachtet. Eine Abschaffung von Hinweisen wäre jedoch auch für die Auffälligkeitskriterien ggf. geeignet. Zukünftig sollte über eine Veröffentlichung der Ergebnisse zu den Auffälligkeitskriterien in den Qualitätsberichten der Krankenhäuser beraten werden.

### **Strukturierter Dialog zu den Auffälligkeitskriterien**

Das Konzept zur Weiterentwicklung des Strukturierten Dialogs sollte um die Betrachtung von Auffälligkeitskriterien erweitert werden.

### **Bericht zur Datenvalidierung**

Der Bericht zur Datenvalidierung muss durch die Einführung des gezielten Datenabgleichs und den vorgeschlagenen Ebenen zur Berichterstattung in erheblichem Maße angepasst werden. Der Aufbau des Routinebetriebs sollte deshalb mit Ressourcen versehen werden – analog zur Überführung neuer QS-Verfahren in den Routinebetrieb. Eine Beratung hierüber erscheint zweckmäßig.

### **Auswirkungen der Dokumentation auf die Ergebnisse der Qualitätsindikatoren**

Modellhaft könnten die Auswirkungen von Fehldokumentationen (Statistische Basisprüfung und Stichprobenverfahren mit Datenabgleich) auf die Ergebnisse der Qualitätsindikatoren untersucht werden. Ausgehend davon könnte in einem nächsten Schritt ein Konzept zur Ermittlung/Berechnung dieser Auswirkungen und zu ihren Konsequenzen erstellt werden. Gegebenenfalls wäre dann auch eine Anpassung der Berichterstattung notwendig.

### **Evaluation des gezielten Datenabgleichs sowie Aufwandsabschätzung**

Ein Beschlussentwurf liegt vor, wurde aber bis zur Abgabe dieses Berichts noch nicht beraten und beschlossen. Es wird jedoch empfohlen, zunächst die Ergebnisse des gezielten Datenabgleichs abzuwarten, bevor mit einer Evaluation begonnen wird. Grundlegend sollte nicht nur der Aufwand abgeschätzt werden, sondern auch eine inhaltliche Überprüfung der Grenzwerte für die ausgearbeiteten Kriterien auf Basis der Erfahrungen im Regelbetrieb vorgenommen werden.

## Literatur

- Allouche, O; Tsoar, A; Kadmon, R (2006): Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43(6): 1223-1232. DOI: 10.1111/j.1365-2664.2006.01214.x.
- Alsulami, Z; Conroy, S; Choonara, I (2013): Medication errors in the Middle East countries: a systematic review of the literature. *European Journal of Clinical Pharmacology* 69(4): 995-1008. DOI: 10.1007/s00228-012-1435-y.
- AQUA [Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen] (2014): Qualitätsreport 2013. [Stand:] 01.08.2014. Göttingen: AQUA. ISBN: 978-3-9809434-8-2. URL: <http://www.sgg.de/sqg/upload/CONTENT/Qualitaetsberichte/2013/AQUA-Qualitaetsreport-2013.pdf> (abgerufen am: 22.01.2020).
- Arts, DGT; de Keizer, NF; Scheffer, GJ (2002): Defining and Improving Data Quality in Medical Registries: A Literature Review, Case Study, and Generic Framework. *Journal of the American Medical Informatics Association* 9(6): 600-611. DOI: 10.1197/jamia.m1087.
- Datta, I; Findlay, C; Kortbeek, JB; Hameed, SM (2007): Evaluation of a regional trauma registry. *Canadian Journal of Surgery* 50(3): 210-203. URL: <http://canjsurg.ca/wp-content/uploads/2014/03/50-3-210.pdf> (abgerufen am: 17.12.2019).
- Duda, SN; Shepherd, BE; Gadd, CS; Masys, DR; McGowan, CC (2012): Measuring the Quality of Observational Study Data in an International HIV Research Network. *PLoS One* 7(4): e33908. DOI: 10.1371/journal.pone.0033908.
- G-BA [Gemeinsamer Bundesausschuss] (2017): Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des IQTIG: Entwicklung von Kriterien zu Anhaltspunkten für den gezielten Datenabgleich der Datenvalidierung nach § 9 QSKH-RL (n. F.) und zur Bewertung der Datenvalidität – einschließlich Erfassungstool und einheitlichem Berichtsformat. [Stand:] 17.11.2017. Berlin: G-BA. URL: [https://www.g-ba.de/downloads/39-261-3135/2017-11-17\\_IQTIG-Beauftragung-Datenvalidierung.pdf](https://www.g-ba.de/downloads/39-261-3135/2017-11-17_IQTIG-Beauftragung-Datenvalidierung.pdf) (abgerufen am: 22.01.2020).
- Houston, L; Probst, Y; Humphries, A (2015): Measuring Data Quality Through a Source Data Verification Audit in a Clinical Research Setting. In: Georgiou, A; Grain, H; Schaper, LK: *Driving Reform: Digital Health is Everyone's Business. Selected Papers from the 23rd Australian National Health Informatics Conference (HIC 2015)*. (Studies in Health Technology and Informatics, Vol. 214). Amsterdam, NL-NH: IOS Press, 107-113. DOI: 10.3233/978-1-61499-558-6-107.
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018a): Bericht zur Datenvalidierung 2017 (nach QSKH-RL). Erfassungsjahr 2016. Stand: 12.07.2018. Berlin: IQTIG. URL: [https://iqtig.org/downloads/berichte/2016/IQTIG\\_Bericht-zur-Datenvalidierung-2017\\_2018-07\\_12.pdf](https://iqtig.org/downloads/berichte/2016/IQTIG_Bericht-zur-Datenvalidierung-2017_2018-07_12.pdf) (abgerufen am: 11.12.2019).

- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018b): Methodische Grundlagen V1.1s. Entwurf für das Stellungnahmeverfahren. Stand: 15.11.2018. Berlin: IQTIG. URL: [https://iqtig.org/dateien/dasiqtig/grundlagen/IQTIG\\_Methodische-Grundlagen-V1.1s-2018-11-15.pdf](https://iqtig.org/dateien/dasiqtig/grundlagen/IQTIG_Methodische-Grundlagen-V1.1s-2018-11-15.pdf) (abgerufen am: 28.01.2020).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2019): Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern. Vorbericht zu Stufe 1 und Stufe 2. Stand: 10.10.2019. Berlin: IQTIG. [unveröffentlicht].
- Jacke, CO; Kalder, M; Koller, M; Wagner, U; Albert, US (2012): Systematische Bewertung und Steigerung der Qualität medizinischer Daten. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 55(11-12): 1495-503. DOI: 10.1007/s00103-012-1536-x.
- Kahn, MG; Callahan, TJ; Barnard, J; Bauck, AE; Brown, J; Davidson, BN; et al. (2016): A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *eGEMs (Generating Evidence & Methods to improve patient outcomes)* 4(1): 18. DOI: 10.13063/2327-9214.1244.
- Kauermann, G; Küchenhoff, H (2011): Stichproben. Methoden und praktische Umsetzung mit R. Heidelberg [u. a.]: Springer. ISBN: 978-3-642-12317-7.
- Kohn, LT; Corrigan, JM; Donaldson, MS; Hrsg. (2000): To Err Is Human. Building a Safer Health System. Washington, US-DC: National Academy Press. ISBN: 978-0-309-26174-6.
- Korn, EL; Graubard, BI (1998): Confidence Intervals for Proportions With Small Expected Number of Positive Counts Estimated From Survey Data. *Survey Methodology* 24(2): 193-201. URL: <https://www150.statcan.gc.ca/n1/pub/12-001-x/1998002/article/4356-eng.pdf> (abgerufen am: 25.11.2019).
- Kramer, HS; Drews, FA (2017): Checking the lists: A systematic review of electronic checklist use in health care. *Journal of Biomedical Informatics* 71(Suppl.): S6-S12. DOI: 10.1016/j.jbi.2016.09.006.
- Levy, PS; Lemeshow, S (2008): Sampling of Populations. Methods and Applications. Fourth Edition. (Wiley Series in Survey Methodology). Hoboken, US-NJ [u. a.]: Wiley. ISBN: 978-0-470-04007-2.
- Lumley, T (2019): survey: Analysis of Complex Survey Samples [*Open Source Software*]. Version 5.36. Published: 27.04.2019. Vienna, AT: R Foundation for Statistical Computing. URL: <https://CRAN.R-project.org/package=survey> [Package source > survey\_3.36.tar.gz] (abgerufen am: 25.11.2019).
- Murff, HJ; Patel, VL; Hripcsak, G; Bates, DW (2003): Detecting adverse events for patient safety research: a review of current methodologies. *Journal of Biomedical Informatics* 36(1–2): 131-143. DOI: 10.1016/j.jbi.2003.08.003.
- Nahm, ML; Pieper, CF; Cunningham, MM (2008): Quantifying Data Quality for Clinical Trials Using Electronic Data Capture. *PLoS One* 3(8): e3049. DOI: 10.1371/journal.pone.0003049.

- Nonnemacher, M; Nasseh, D; Stausberg, J (2014): Datenqualität in der medizinischen Forschung. Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern. 2., aktualisierte und erweiterte Auflage. (Schriftenreihe der TMF [Technologie- und Methodenplattform für die Vernetzte Medizinische Forschung], Band 4, 2. Auflage). Berlin: MWV [Medizinisch Wissenschaftliche Verlagsgesellschaft]. ISBN: 978-3-95466-121-3.
- Pesudovs, K; Burr, JM; Harley, C; Elliott, DB (2007): The Development, Assessment, and Selection of Questionnaires. *Optometry and Vision Science* 84(8): 663-674. DOI: 10.1097/OPX.0b013e318141fe75.
- Raslich, MA; Markert, RJ; Stutes, SA (2007): Selecting and interpreting diagnostic tests. *Biochemistry Medica* 17(2): 151-161. DOI: 10.11613/bm.2007.014.
- Särndal, CE; Swensson, B; Wretman, J (2003): Model Assisted Survey Sampling. (Springer Series in Statistics). New York, US-NY [u. a.]: Springer-Verlag. ISBN: 978-0-387-40620-6.
- Schwarzer, G; Türp, JC; Antes, G (2001): Sensitivität und Spezifität: Auswirkung der Wahl des Trennpunktes. *Deutsche Zahnärztliche Zeitschrift* 56(8): 446-447.
- Terwee, CB; Bot, SDM; de Boer, MR; van der Windt, DAWM; Knol, DL; Dekker, J; et al. (2007): Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 60(1): 34-42. DOI: 10.1016/j.jclinepi.2006.03.012.
- Thoburn, KK; German, RR; Lewis, M; Nichols, PJ; Ahmed, F; Jackson-Thompson, J (2007): Case Completeness and Data Accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries. *Cancer* 109(8): 1607-1616. DOI: 10.1002/cncr.22566.
- Tuble, SC (2011): Perfusion Downunder Collaboration Database – Data Quality Assurance: Towards a High Quality Clinical Database. Perfusion Downunder Winter Meeting, Queens-town, NZ-OTA, 05.-08.10.2010. *The Journal of Extra-Corporeal Technology* 43: P44-P51. URL: [https://amsect.smithbucklin.com/JECT/PDFs/2011\\_volume43/issue1/Tuble.pdf](https://amsect.smithbucklin.com/JECT/PDFs/2011_volume43/issue1/Tuble.pdf) (abgerufen am: 17.12.2019).
- Wang, W (2015): Exact Optimal Confidence Intervals for Hypergeometric Parameters. *Journal of the American Statistical Association* 110(512): 1491-1499. DOI: 10.1080/01621459.2014.966191.
- Wood, S; Scheipl, F (2017): gamm4: Generalized Additive Mixed Models using 'mgcv' and 'lme4' [Open Source Software]. Version 0.2-5. Published: 25.07.2017. Vienna, AT: R Foundation for Statistical Computing. URL: <https://CRAN.R-project.org/package=gamm4> [Package source > gamm4\_0.2-5.tar.gz] (abgerufen am: 25.11.2019).
- Wood, SN (2017): Generalized Additive Models. An Introduction with R. Second Edition. (Texts in Statistical Science). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC. ISBN: 978-1-4987-2833-1.