



Institut für Qualitätssicherung und
Transparenz im Gesundheitswesen

Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern

Abschlussbericht zu Stufe 1 und Stufe 2
Anhang

Erstellt im Auftrag des
Gemeinsamen Bundesausschusses

Stand: 31. Januar 2020

Impressum

Thema:

Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern. Abschlussbericht zu Stufe 1 und Stufe 2. Anhang

Ansprechpartnerin:

Andrea Wolf

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

18. Januar 2018

Datum der Abgabe:

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

<https://www.iqtig.org>

Inhaltsverzeichnis

Tabellenverzeichnis.....	5
Abbildungsverzeichnis.....	6
1 Protokoll zum Workshop zur Weiterentwicklung des Strukturierten Dialogs am 27. Mai 2019.....	8
2 Auszug des Protokolls zum LQS-IQTIG-Treffen am 18. Juni 2019	17
3 Ergänzende Informationen zu Kapitel 2 – Analyse der Ausgangssituation.....	19
4 Biometrische Analyse: Heterogenität der Vorgehensweise und Bewertung im Strukturierten Dialog – Exemplarische Analyse anhand von Daten aus dem Erfassungsjahr 2016	22
4.1 Methodik.....	22
4.2 Ergebnisse	26
4.3 Zusammenfassung und Diskussion	32
4.4 Details zur statistischen Modellierung.....	33
4.4.1 Fixed- vs. Random-Effects-Modelle	33
4.4.2 Operationalisierung des Begriffes vergleichbarer Leistungserbringer	34
4.4.3 Modelle ohne Berücksichtigung der Leistungserbringer-Ergebnisse.....	37
4.4.4 Implementation.....	39
4.5 Deskriptive Auswertungen zu den Ergebnissen des Strukturierten Dialogs zum Erfassungsjahr 2016 für ausgewählte Indikatoren	40
4.5.1 HEP: QI 54003 Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur.....	41
4.5.2 09/1: QI 54139 Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen	44
4.5.3 DEK: QI 52010 Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4	47
5 Beispielfälle für die Qualitätsbewertung im Stellungnahmeverfahren.....	50
5.1 Fachlich-inhaltliche Einflussfaktoren	50
5.2 Datenfehler	54
6 Textvorschläge Einholung Stellungnahmen	56
7 Beispiel zur Übermittlung von Vorgangsnummern in Stellungnahmen.....	60
8 Beispiel zur Berücksichtigung mehrerer Einflussfaktoren bei der partiellen Nachberechnung	62

9	Berechnungsvorschrift für die partielle Nachberechnung	64
9.1	Berechnung des informativen Indikatorwerts	64
9.2	Berechnung des bereinigten Referenzbereiches	65
9.3	Retrospektive Schätzung der in Referenzbereichen enthaltenen Toleranz für Qualitätsdefizite	66
10	Gesamtalgorithmus für die fachliche Bewertung	69
11	Zielvereinbarungen	70
11.1	Formular zum Sachstandsbericht einzelner Zielvereinbarungen	70
11.2	Erläuterung zum Formular	72
	Literatur.....	75

Tabellenverzeichnis

Tabelle 1: Anzahl von Fachgruppen und Mitgliederzahl pro Fachgruppe in der externen stationären Qualitätssicherung (Winkler-Komp et al. 2014).....	19
Tabelle 2: Einstufungsschema für rechnerische Auffälligkeiten bei Qualitätsindikatoren nach Abschluss des Strukturierten Dialogs (bisheriges Vorgehen)	19
Tabelle 3: Anteil von Hinweisen an allen rechnerisch auffälligen Ergebnissen nach Bundesland zum Erfassungsjahr 2015 (modifizierte Tabelle aus dem Bericht zum Strukturierten Dialog 2016; IQTIG 2017: 52)	21
Tabelle 4: Entscheidungsverfahren in den Bundesländern, erfasst anhand der Frage: „Wie wird festgelegt, ob ein Ergebnis als qualitativ auffällig oder unauffällig beurteilt wird?“ (Winkler-Komp et al. 2014).....	21
Tabelle 5: Parameterschätzungen der drei Modelle zu Fragestellung 1: Heterogenität bezüglich des Anforderns einer Stellungnahme	26
Tabelle 6: Parameterschätzungen der drei Modelle zu Fragestellung 2: Heterogenität bezüglich der Bewertung als qualitativ auffällig.....	27
Tabelle 7: (Konditionaler) AIC der Modellvarianten M_1 - M_4 zu Frage 1	38
Tabelle 8: (Konditionaler) AIC der Modellvarianten M_1 - M_4 zu Frage 2	38
Tabelle 9: Parameterschätzungen Modellvariante M_1 der drei Modelle zu Fragestellung 1: Heterogenität bezüglich des Anforderns einer Stellungnahme	39
Tabelle 10: Parameterschätzungen Modellvariante M_1 der drei Modelle zu Fragestellung 2: Heterogenität bezüglich der Bewertung als qualitativ auffällig	39
Tabelle 11: Fiktives Beispiel für die Angabe von Vorgangsnummern zur Stellungnahme zu einem Indikator.....	60
Tabelle 12: Fiktives Beispiel für die fallbezogene Berücksichtigung von Einflussfaktoren bei der partiellen Neuberechnung.....	62

Abbildungsverzeichnis

Abbildung 1: Ergebnisse des Strukturierten Dialoges zum Qualitätsindikator HEP: QI 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ für das Erfassungsjahr 2016 aus zwei exemplarisch ausgewählten Bewertungsstellen.	23
Abbildung 2: HEP QI 54003 (ratenbasierter Indikator). Streuung der geschätzten Bewertungswahrscheinlichkeit in Quartilen der Leistungserbringerergebnisse	30
Abbildung 3: DEK QI 52010 (Sentinel-Event-Indikator). Streuung der geschätzten Bewertungswahrscheinlichkeit in Quartilen der Leistungserbringerergebnisse	31
Abbildung 4: 09/1 QI 54139 (ratenbasierter Indikator). Streuung der geschätzten Bewertungswahrscheinlichkeit in Quartilen der Leistungserbringerergebnisse	31
Abbildung 5: Illustration zu alternativen Definitionen von <i>rlb</i> anhand einer nach Fallzahl differenzierten Darstellung der rechnerischen Ergebnisse. Segmente gleicher Farbe markieren Gruppen von Leistungserbringern deren Ergebnisse als vergleichbar eingestuft würden. Links die p-Wert-basierten Definition; rechts die ratenbasierte Definition.	35
Abbildung 6: Qualitätsindikator HEP: 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“. Anfordern einer Stellungnahme nach rechnerischer Auffälligkeit	41
Abbildung 7: Qualitätsindikator HEP: 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“. Bewertungen nach Einleitung des Stellungnahmeverfahrens	42
Abbildung 8: Ergebnisse des strukturierten Dialoges zum Qualitätsindikator HEP: 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ für das Erfassungsjahr 2016 differenziert nach Fallzahl und Bewertungsstelle. Diese Daten sind die Grundlage für die modellbasierten Auswertungen in Abschnitt 2.2.....	43
Abbildung 9: Qualitätsindikator 09/1: 54134 „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“. Anfordern einer Stellungnahme nach rechnerischer Auffälligkeit.....	44
Abbildung 10: Qualitätsindikator 09/1: 54134 „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“. Bewertungen nach Einleitung des Stellungnahmeverfahrens.....	45
Abbildung 11: Ergebnisse des Strukturierten Dialogs zum Qualitätsindikator 09/1: 54134 „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“ für das Erfassungsjahr 2016 differenziert nach Fallzahl und Bewertungsstelle. Diese Daten sind die Grundlage für die modellbasierten Auswertungen in Abschnitt 2.2.	46
Abbildung 12: Qualitätsindikator DEK: 52010 „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“. Anfordern einer Stellungnahme nach rechnerischer Auffälligkeit	47

Abbildung 13: Qualitätsindikator DEK: 52010 „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“. Bewertungen nach Einleitung des Stellungnahmeverfahrens	48
Abbildung 14: Ergebnisse des Strukturierten Dialogs zum Qualitätsindikator DEK: 52010 „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“ für das Erfassungsjahr 2016 differenziert nach Fallzahl und Bewertungsstelle. Diese Daten sind die Grundlage für die modellbasierten Auswertungen in Abschnitt 2.2.....	49

1 Protokoll zum Workshop zur Weiterentwicklung des Strukturierten Dialogs am 27. Mai 2019

TOP	Thema
TOP 1	<p>Begrüßung</p> <p>Das IQTIG begrüßt die Teilnehmerinnen und Teilnehmer des Workshops und stellt die Beauftragung durch den G-BA vor. Die Ziele des Workshops sind die Vorstellung des methodischen Hintergrunds und des Rahmenkonzepts sowie die Einholung der Expertise seitens der mit der Durchführung des Strukturierten Dialogs Betrauten, um anschließend mögliche Herausforderungen zu identifizieren und ggf. Lösungsansätze zu entwickeln.</p> <p>Zunächst werden die Erwartungen/Wünsche der Teilnehmerinnen und Teilnehmer an den Workshop eingeholt und besprochen. Die Teilnehmerinnen und Teilnehmer wünschen, dass ihr Input aufgenommen wird und in die Weiterentwicklung einfließt. Ebenso wird ein transparenterer Umgang hinsichtlich der Entscheidung darüber, welche Aspekte aufgenommen/nicht aufgenommen werden, gewünscht. Hinsichtlich des Konzepts wird sich Praxistauglichkeit und Entbürokratisierung gewünscht. Das IQTIG versichert, dass die Anmerkungen der Landesvertreterinnen und -vertreter eingehend geprüft werden und nach Möglichkeit für die Stufe 2 der „Weiterentwicklung des Strukturierten Dialogs“ berücksichtigt werden.</p>
TOP 2	<p>Präsentation methodischer Hintergrund und Rahmenkonzept</p> <p>Das IQTIG erläutert den methodischen Hintergrund und das Rahmenkonzept (siehe Präsentationsfolien 6–17). Vor dem Hintergrund des richtlinienunabhängigen Konzepts wird erläutert, dass die Begriffe Strukturiertes Dialog und Stellungnahmeverfahren synonym verwendet werden. Da die QSKH-RL bald ausläuft, sieht das IQTIG vor, den Schwerpunkt auf Empfehlungen für die DeQS-Richtlinie zu legen.</p> <p>Im methodischen Hintergrund werden die Funktionen des bisherigen Strukturierten Dialogs vorgestellt. Das IQTIG hat die Methodik des bisherigen Verfahrens analysiert. Hierzu wurde u. a. eine biometrische Analyse zur Heterogenität (bewertungsstellenabhängig bezüglich der Einholung der Stellungnahme und bezüglich der Bewertung als qualitativ auffällig) im Strukturierten Dialog durchgeführt.</p> <p>Ziele der Weiterentwicklung sind einerseits eine aussagekräftige Qualitätsbewertung für alle Indikatorergebnisse zu erlangen und andererseits durch einheitliche Verfahrensregeln das Stellungnahmeverfahren mit abschließender Qualitätsbewertung effizienter zu gestalten, um eine Ressourcenverschiebung von Qualitätsbewertung hin zur Qualitätsförderung zu erreichen.</p> <p>Im methodischen Hintergrund wurden Gütekriterien für die Bewertungsprozesse zur Minimierung der Heterogenität eingeführt. Vor diesem Hintergrund empfiehlt das IQTIG zukünftig eine Trennung von Qualitätsbewertung und -förderung.</p> <p>Einige Teilnehmerinnen und Teilnehmer weisen darauf hin, dass die Richtlinien (plan. QI-RL und DeQS-RL) untereinander abgestimmt werden müssen.</p> <p>Das IQTIG erläutert, dass diese Abstimmung laut Beauftragung vorgesehen ist.</p>

TOP	Thema
TOP 3	<p data-bbox="389 255 1374 322">Analyse Herausforderungen und Diskussion Lösungen – „Vor der Bewertung der Stellungnahme“</p> <p data-bbox="389 338 1374 405">I. Operationalisierung des „Hinreichenden Verdachts auf ein Qualitätsdefizit“</p> <p data-bbox="389 421 1386 629">Das IQTIG präsentiert Möglichkeiten zur Effizienzsteigerung hinsichtlich der Auslösung des Stellungnahmeverfahrens anhand der Qualitätsindikatoren (siehe Präsentationsfolien 19–26). Am Beispiel der HEP-Indikatoren wird der große Arbeitsaufwand auf Landesebene im Sinne der Anzahl an rechnerisch auffälligen Ergebnissen dargestellt, sowie die Diskrepanz zwischen der Anzahl an rechnerisch auffälligen Ergebnissen und der Anzahl an qualitativ auffälligen Ergebnissen verdeutlicht.</p> <p data-bbox="389 645 1386 786">Die statistische Operationalisierung des Kriteriums „Hinreichender Verdacht auf ein Qualitätsdefizit“ erfolgt nach einer Taxonomie bestehend aus Studienart, Stichprobenart und Berechnungsart und führt immer zu einem Indikatorwert für den Standort, ggf. einem Unsicherheitsintervall, und einer binären Klassifikation.</p> <p data-bbox="389 801 879 831">Herausforderungen und Lösungsansätze:</p> <ol data-bbox="389 846 1386 1946" style="list-style-type: none"> <li data-bbox="389 846 1386 1245">1. Bei einigen Teilnehmerinnen und Teilnehmern besteht die Sorge, dass der Begriff „hinreichender Verdacht auf ein Qualitätsdefizit“ bei den Leistungserbringern als Generalverdacht, welcher entkräftet werden müsse, interpretiert werden könnte. <ul data-bbox="421 1003 1386 1245" style="list-style-type: none"> <li data-bbox="421 1003 1386 1245">▫ Das IQTIG erläutert, dass es sich dabei eben nicht um einen Generalverdacht handelt, sondern im Gegenteil zunächst von guter Versorgungsqualität ausgegangen wird. Das Ziel ist, eine spezifischere Auslösung anhand der Qualitätsindikatoren zu erreichen, um eine Verschlinkung des Stellungnahmeverfahrens zu erzielen. Das bedeutet, dass nur bei denjenigen Leistungserbringern ein Stellungnahmeverfahren eingeleitet werden soll, bei denen ein Qualitätsdefizit anhand der Daten begründet zu vermuten ist. <p data-bbox="421 1261 1386 1357">Einige Teilnehmerinnen und Teilnehmer weisen darauf hin, dass unter dem Aspekt der Akzeptanzförderung Begriffe, die scheinbar aus dem Strafrecht entlehnt sind, vermieden werden sollten.</p> <ul data-bbox="421 1373 1386 1440" style="list-style-type: none"> <li data-bbox="421 1373 1386 1440">▫ Das IQTIG wird den Begriff „Hinreichender Verdacht auf ein Qualitätsdefizit“ beraten. <li data-bbox="389 1462 1386 1648">2. Die aktuelle Anzahl an Qualitätsindikatoren wird von einigen Teilnehmerinnen und Teilnehmern als zu groß eingeschätzt. <ul data-bbox="421 1541 1386 1648" style="list-style-type: none"> <li data-bbox="421 1541 1386 1648">▫ Das IQTIG erläutert, dass im Rahmen der beauftragten methodischen Überarbeitung der Qualitätsindikatoren ggf. auch Indikatoren, die die Eignungskriterien nicht im ausreichenden Maß erfüllen, zur Streichung empfohlen werden. <li data-bbox="389 1664 1386 1946">3. Einige Teilnehmerinnen und Teilnehmer merken an, dass mit den bestehenden Qualitätsindikatoren Qualität nicht abgebildet werden könne. Zudem wird von einigen Teilnehmerinnen und Teilnehmern angemerkt, dass eine Vielzahl der aktuellen Qualitätsindikatoren nicht geeignet sei, um zwischen guter und schlechter Qualität zu unterscheiden und daher eine Anwendung des Konzepts anhand der aktuellen Indikatoren kritisch gesehen wird. <ul data-bbox="421 1888 1386 1946" style="list-style-type: none"> <li data-bbox="421 1888 1386 1946">▫ Das IQTIG erläutert, dass Qualitätsindikatoren gemäß der Methodischen Grundlagen des Instituts geeignet sein müssen, um Qualität abzubilden. Vor

TOP	Thema
	<p>dem Hintergrund, der unterschiedlichen Verwendungszwecke von Qualitätsmessungen in der gesetzlichen Qualitätssicherung ist es erforderlich, zu jedem Qualitätsindikatorergebnis eine abschließende Qualitätsaussage zu erhalten. Das IQTIG bestätigt, dass viele der aktuellen Qualitätsindikatoren gemäß den Kriterien der Methodischen Grundlagen zu überarbeiten sind und hierzu bereits ein Konzept erarbeitet wird. Es wird zudem erläutert, dass es notwendig ist, ein Konzept zur Weiterentwicklung des Strukturierten Dialogs auf der Basis geeigneter Qualitätsindikatoren (siehe Eignungskriterien von Qualitätsindikatoren, Methodische Grundlagen des IQTIG 2019) zu erstellen und nicht auf der Basis ungeeigneter Indikatoren.</p> <p>4. Die Risikoadjustierung erfolgt anhand der Ergebnisse der Vorjahre. Im Stimmnahmeverfahren wird nach patientenbezogenen Besonderheiten gefragt. Es wird die Frage gestellt, inwiefern das einbezogen wird.</p> <ul style="list-style-type: none"> ▫ Das IQTIG erläutert, dass bei der Indikatorergebnisberechnung nicht-berücksichtigte Risikofaktoren können im Stimmnahmeverfahren als Ausnahmetatbestände geltend gemacht werden. <p>5. Eine Teilnehmerin fragt, ob die, in den Qualitätssicherungsergebnisberichten (DeQS) ausgewiesenen Dokumentationsfehler in der vorgestellten statistischen Auswertungsmethodik berücksichtigt werden. Zudem wird berichtet, dass es bei dem PCI-Verfahren bspw. einen hohen Anteil an Dokumentationsproblemen gegeben habe. Aus Sicht der Teilnehmerin wäre es hilfreich, darauf zurückzugreifen, um abzuleiten, bei welchen Indikatoren Nachbesserungsbedarf besteht.</p> <ul style="list-style-type: none"> ▫ Das IQTIG erläutert, dass die Problematik zu vieler Dokumentationsfehler nicht mittels statistischer Verfahren im Rahmen der QI-Berechnung gelöst werden könne. <p>6. Hinsichtlich der gesundheitsökonomischen Betrachtungsweise und der Frage: „Ist ein entdecktes Qualitätsdefizit bei einem kleinen Leistungserbringer gleichermaßen ‚teuer‘ wie ein Qualitätsdefizit bei einem großen Leistungserbringer (da mehr Patientinnen und Patienten gefährdet werden)?“ (siehe Präsentationsfolie 24) wurde angemerkt, dass die Erfahrungen gezeigt haben, dass Häuser mit geringen Fallzahlen eher Unterstützung benötigten und eher in die Qualitätsförderung einbezogen werden sollten.</p> <ul style="list-style-type: none"> ▫ Das IQTIG wird diesen Hinweis beraten. <p>7. Eine Teilnehmerin merkt an, dass 1/3 der Abweichungen vom Referenzwert auf 1 oder 2 Fällen beruhten. Der Effekt verstärkte sich bei perzentilbasierten Referenzbereichen. Dies sei ein Problem, dass sich mathematisch nicht lösen lasse. Es wird weiterhin berichtet, dass bei einem Fall im Zähler die Kommunikation mit den Leistungserbringern schwierig sei.</p> <ul style="list-style-type: none"> ▫ Das IQTIG weist darauf hin, dass dies bei der zu entwickelnden statistische Auswertungsmethodik berücksichtigt wird, speziell bei der Methodik zur Bestimmung von Perzentilreferenzbereichen. Bei den HEP-Indikatoren sind ca. 13 % der rechnerischen auffälligen Ergebnisse bei Standorten mit 1–4 Fällen festgestellt worden. <p>Ein Teilnehmer berichtet, dass bspw. die Akzeptanz für Qualitätsindikatorergebnisse zur Cholezystektomie (Landesverfahren in Hessen), die mittels statistischer Auswertungsmethodik (Signifikanztest) berechnet werden größer sei.</p>

TOP	Thema
	<p data-bbox="389 248 995 277">II. Formale Kriterien für die Stellungnahmen</p> <p data-bbox="389 293 1386 360">Das IQTIG präsentiert die Empfehlungen zu den formalen Anforderungskriterien an Stellungnahmen (siehe Präsentationsfolie 28).</p> <p data-bbox="389 376 879 405">Herausforderungen und Lösungsansätze:</p> <ol data-bbox="389 421 1386 1986" style="list-style-type: none"> <li data-bbox="389 421 1386 568">1. Ein Teilnehmer fordert, dass die Angaben der Leistungserbringer (insbesondere bei plan. QI), die in den Stellungnahmen gemacht wurden, grundsätzlich als korrekt erachtet werden sollten. <ul data-bbox="421 539 879 568" style="list-style-type: none"> <li data-bbox="421 539 879 568">▫ Das IQTIG wird den Hinweis beraten. <li data-bbox="389 584 1386 1570">2. Einige Teilnehmerinnen und Teilnehmer stellen die Frage, ob die Nicht-Einhaltung formaler Vorgaben (z. B. des Datenschutzes, digitale Stellungnahme) dazu führen dürfe, dass das Ergebnis des Leistungserbringers als qualitativ auffällig bewertet wird. Im Nachhinein könne nicht mehr differenziert werden, ob ein Leistungserbringer aufgrund nicht-ordnungsgemäßer Stellungnahmen oder aufgrund unzureichender Versorgungsqualität als qualitativ auffällig bewertet wurde. Es wird die Sorge geäußert, dass dies zu einer „Beschädigung“ des Verfahrens führen könnte. Zudem wird in Frage gestellt, inwieweit die vorgestellten formalen Kriterien (z. B. digitale Stellungnahme, Unterschrift durch juristisch verantwortliche Instanz) für die Akzeptanz des Gesamtverfahrens förderlich sind. Die Teilnehmerinnen und Teilnehmer haben hierzu unterschiedliche Ansichten. Eine Teilnehmerin macht den Vorschlag die Kategorie A, Ziffer 4 (qualitativ auffällig aufgrund nicht-Einhaltung der formalen Kriterien) aus der Kategorie A herauszulösen und gesondert (nicht unter Kategorie A) auszugeben. Wenn an dieser Bewertung festgehalten werde, wird vermutet, dass die Akzeptanz der QS-Verfahren darunter leiden würde und das aufgebaute Vertrauen der Leistungserbringer Schaden nehme. Zudem wird vermutet, dass damit auch keine Transparenzerhöhung zu erreichen sei. <ul data-bbox="421 1240 1386 1570" style="list-style-type: none"> <li data-bbox="421 1240 1386 1375">▫ Das IQTIG wird die einzelnen Empfehlungen zu den formalen Kriterien prüfen. Die Wahrung der Anonymität der Patientinnen und Patienten ist jedoch auch im Rahmen des Stellungnahmeverfahrens (siehe Datenschutzgesetz) sicherzustellen. <li data-bbox="421 1391 1386 1570">▫ Das IQTIG erläutert, dass die Konsequenzen bei Nicht-Einhaltung der formalen Kriterien, wie sie auch in anderen Kontexten umgesetzt werden, dazu führen, dass die Kriterien eingehalten werden. Zudem differenziert das Bewertungsschema bereits, ob eine qualitative auffällige Bewertung aufgrund der Nicht-Einhaltung der formalen Kriterien vergeben wurde. <li data-bbox="389 1585 1386 1800">3. Die Teilnehmerinnen und Teilnehmer diskutieren, ob im Rahmen der Qualitätsförderung die juristisch verantwortliche Instanz die Stellungnahme autorisieren muss. <ul data-bbox="421 1697 1386 1800" style="list-style-type: none"> <li data-bbox="421 1697 1386 1800">▫ Das IQTIG wird diesen Hinweis beraten. Im Rahmen der Qualitätsförderung ist ggf. auch eine Autorisierung durch die fachlich verantwortliche Person ausreichend. <li data-bbox="389 1816 1386 1986">4. Seitens der Teilnehmerinnen und Teilnehmer wird die im Konzept empfohlene 4-Wochen-Frist für die Abgabe einer Stellungnahme kontrovers diskutiert. Einige Teilnehmerinnen und Teilnehmer geben zu bedenken, dass der Strukturierte Dialog oder das Stellungnahmeverfahren genau über die Haupturlaubszeit abzuwickeln ist, in der die Krankenhäuser mit einer deutlich reduzierten Personaldecke

TOP	Thema
	<p>die Patientenversorgung aufrechterhalten müssen und deshalb solche administrativen Aufgaben nicht erste Priorität erhalten könnten.</p> <ul style="list-style-type: none"> ▫ Der Vorschlag seitens einiger Teilnehmerinnen und Teilnehmer einer 3-Wochen-Frist plus 2-Wochen-Nachrichtfrist und weitere Alternativen werden durch das IQTIG geprüft werden. <p>5. Inwieweit die Anforderung, die Stellungnahme digital einzureichen, ein sinnvolles formales Kriterium sei, wird unter den Teilnehmerinnen und Teilnehmern kontrovers diskutiert. Eine solche Anforderung könne für einzelne Leistungserbringer ein Problem sein.</p> <ul style="list-style-type: none"> ▫ Das IQTIG wird diesen Hinweis beraten. <p>6. Einige Teilnehmerinnen und Teilnehmer berichten, dass es Fachgruppen gebe, die einen strukturierten Fragebogen (teilweise zwischen Geschäftsstellen abgestimmt) entwickelt und Anfragetexte zusammengestellt haben, welche dann scheinbar nicht mehr genutzt werden können, wenn die Texte zur Einholung von Stellungnahmen vorgeschrieben seien. Dies könne für die Akzeptanz des Verfahrens abträglich sein.</p> <ul style="list-style-type: none"> ▫ Das IQTIG wird diesen Hinweis beraten. <p>7. Einige Teilnehmerinnen und Teilnehmer äußern die Sorge, dass durch eine Formalisierung der Druck auf die Leistungserbringer steigen könne. Zudem wird die Sorge geäußert, dass durch die Empfehlungen des IQTIG das „Formelle“ stärker als bislang in die Bewertung mit einbezogen werde.</p> <ul style="list-style-type: none"> ▫ Das IQTIG wird den Hinweis beraten und erläutert, dass die formalen Kriterien zu einer Effizienzsteigerung bei der Qualitätsbewertung beitragen können, um mehr Ressourcen für die Qualitätsförderung zur Verfügung zu haben. <p>Das IQTIG erläutert abschließend, dass auch formale Kriterien für eine Vereinheitlichung des Vorgehens beauftragt wurden. Um eine Vereinheitlichung zu erreichen, seien einheitliche Verfahrensregeln dazu, wie sich Leistungserbringer zu auffälligen Indikatorergebnissen äußern können, unabdingbar.</p>
TOP 4	<p>Analyse Herausforderungen und Diskussion Lösungsansätze – „Bewertung der Stellungnahme“ und „Einleitung Maßnahmen zur Qualitätsförderung“</p> <p>I. Fachliche Bewertung</p> <p>Das IQTIG präsentiert das Ablaufschema zum Bewertungsprozess, die Definition des Ausnahmetatbestands und das Bewertungsschema (vgl. Präsentationsfolien 33–36).</p> <p>Herausforderungen und Lösungsansätze:</p> <p>1. Hinsichtlich der Definition des Ausnahmetatbestandes wird seitens einer Teilnehmerin angemerkt, dass die Qualitätsanforderung, die zu erfüllen ist, die Erfüllung des Indikatorziels sei. Da die Indikatoren teilweise veraltet und ungeeignet seien, könne die Erfüllung eines Indikatorziels nicht mit guter Qualität gleichgesetzt werden.</p> <ul style="list-style-type: none"> ▫ Das IQTIG weist darauf hin, dass aus methodischer Sicht bei Qualitätsindikatoren, die die Eignungskriterien erfüllen, die Erfüllung des Indikatorziels guter

TOP	Thema
	<p>Qualität entspräche. Das IQTIG erläutert, dass aktuell ein Konzept zur Weiterentwicklung der Qualitätsindikatoren anhand der Eignungskriterien erarbeitet wird.</p> <ol style="list-style-type: none"> 2. Einige Teilnehmerinnen und Teilnehmer sprechen sich dafür aus, dass auch Aspekte der Pflege bei der Definition von Ausnahmetatbeständen berücksichtigt werden sollten. <ul style="list-style-type: none"> ▫ Das IQTIG wird über die Definition des Ausnahmetatbestandes beraten und prüfen, was unter medizinisch-inhaltlich (z. B. auch pflegerisch) zu verstehen ist. 3. Es wird der Hinweise gegeben, dass der Begriff „organisatorisch-strukturelle Gründe“ sehr viel Interpretationsspielraum lässt. <ul style="list-style-type: none"> ▫ Das IQTIG wird den Hinweis beraten. 4. Einige Teilnehmerinnen und Teilnehmer kritisieren den Begriff „verschuldet“, da er an das Strafrecht erinnere. Es wird der Vorschlag gemacht, diesen durch „verantwortet“ zu ersetzen. <ul style="list-style-type: none"> ▫ Das IQTIG wird den Vorschlag und weitere Alternativen zum Begriff „Schuld“ prüfen. 5. Ein Teilnehmer merkt an, dass bei unverschuldeten Fehlern in der Datengrundlage (U1) nicht bewertet werden könne, ob ein Leistungserbringer eine gute Versorgungsqualität habe. Daher könne keine unauffällige Versorgungsqualität bescheinigt werden. <ul style="list-style-type: none"> ▫ Das IQTIG erläutert, dass in Abhängigkeit davon, wie gravierend die Fehler sind ggf. noch eine Bewertung vorgenommen werden kann (siehe gestrichelte Linie im Prozessablauf der Bewertung Folie 33). Da grundsätzlich zunächst von guter Versorgungsqualität ausgegangen werde, gelte die Unschuldsvermutung, falls unverschuldete Datenfehler belegt werden können. 6. Es wird seitens der Teilnehmerinnen und Teilnehmer gefragt, wo die Grenze sei, zwischen Dokumentationsfehlern, die der Leistungserbringer verschuldet und solchen, die er nicht verschuldet habe. <ul style="list-style-type: none"> ▫ Das IQTIG wird den Hinweis beraten und ggf. Entscheidungshilfen zur Verfügung stellen. Das IQTIG erläutert in diesem Zusammenhang, dass die Leistungserbringer verpflichtet seien, korrekt zu dokumentieren. Die derzeit bestehende Lücke, bei Dokumentationsfehlern keine Qualitätsbewertung vornehmen zu können, soll mit diesem Konzept geschlossen werden. 7. Einige Teilnehmerinnen und Teilnehmer fordern eine deutlichere Differenzierung des Bewertungsschemas in Bezug auf Fehldokumentationen der Leistungserbringer (Kategorie „D“). Patientinnen und Patienten sollten ein Recht darauf haben, zu wissen, ob eine Abteilung falsch dokumentiert hat oder ob ein medizinisch-fachlicher Qualitätsmangel bestand. Hier ist die Kategorie D „Bewertung nicht möglich wegen fehlerhafter Dokumentation“ eine sinnvolle Alternative. <ul style="list-style-type: none"> ▫ Das IQTIG wird den Hinweis beraten. 8. Leistungserbringer mit einem unauffälligen Qualitätsindikatorergebnis auch als qualitativ unauffällig zu bewerten wird von einigen Teilnehmerinnen und Teilnehmern kritisch gesehen. Nach dem Konzept wird denjenigen, die nicht im Stel-

TOP	Thema
	<p>lungnahmeverfahren geprüft werden, weil sie ein unauffälliges Indikatorergebnis aufweisen, eine unauffällige Qualität zugestanden. Dies sei nicht vertretbar, da nur nach einem Stellungnahmeverfahren ein qualitativ unauffälliges Ergebnis bescheinigt werden könne. Daher solle die Kategorie R (Ergebnis liegt innerhalb des Referenzbereichs) beibehalten werden.</p> <ul style="list-style-type: none"> ▫ Das IQTIG erläutert, dass per Definition von unauffälliger Qualität gesprochen werde, wenn ein Leistungserbringer den Referenzbereich erreiche. Gibt die Datengrundlage keinen Anlass (aufgrund ausreichender Evidenz) eine weitere Prüfung vorzunehmen, weil das „Erwartbare“ erfüllt wurde, könne daher von unauffälliger Qualität bezogen auf das jeweilige Qualitätsziel des Indikators gesprochen werden. <p>9. Es wird die Frage gestellt, wie „Mischergebnisse“ zu bewerten seien (z. B. wenn bei 5 Fällen Dokumentationsfehler angegeben und 5 Fälle als Ausnahmetatbeständen bewertet wurden).</p> <ul style="list-style-type: none"> ▫ Das IQTIG wird den Hinweis beraten. <p>10. Eine Teilnehmerin merkt an, dass die Kategorie A0 und A4 zusammengefasst werden könne. Die übrigen Kategorien seien sinnvoll gewählt.</p> <ul style="list-style-type: none"> ▫ Das IQTIG wird den Hinweis beraten. <p>II. Zusammensetzung der Expertengremien</p> <p>Das IQTIG präsentiert die Empfehlungen zur Qualifikationen Expertinnen und Experten und zur Zusammensetzung des Expertengremiums (vgl. Präsentationsfolien 38–39).</p> <p>Herausforderungen und Lösungsansätze:</p> <p>1. Eine Teilnehmerin stellt die Frage, ob diese Empfehlungen nur für neue Expertinnen und Experten gelten sollen oder ob alle bisher mitarbeitenden Expertinnen und Experten nochmal überprüft werden müssten. Die Anforderungen seien zu hoch, insbesondere in kleinen Ländern gebe es nicht genug Fachexpertinnen und Fachexperten, die diese Anforderungen erfüllen und bereit seien mitzuwirken. Zudem wird die Frage stellt, wer die fachliche Qualifikation bewerten und über die Mitwirkung der Expertinnen und Experten entscheiden solle.</p> <ul style="list-style-type: none"> ▫ Das IQTIG wird diesen Hinweis beraten. <p>2. Aus Sicht einiger Teilnehmerinnen und Teilnehmer werden Expertinnen und Experten benötigt, die bereits QS-Bögen ausgefüllt und Erfahrungen mit der esQS haben bzw. Verantwortung dafür tragen, die „Kodierer“ im Haus zu schulen etc. Es sollten, neben der Fachexpertise, umfangreiche QS-Erfahrungen vorliegen.</p> <ul style="list-style-type: none"> ▫ Das IQTIG wird diesen Hinweis beraten. <p>3. Einige Teilnehmerinnen und Teilnehmer erläutern, dass der Begriff „Zuständige Stelle“ und LAG vor dem Hintergrund der unterschiedlichen Richtlinien zu konkretisieren sei. Gegebenenfalls müssten Abläufe/Zeitpläne für die LAG geändert werden. Sinnvoll wäre eine Orientierung an der DeQS-RL und es müsse geklärt werden, welches Organ innerhalb der LAG welche Kompetenzen übertragen bekommen solle.</p> <ul style="list-style-type: none"> ▫ Das IQTIG wird den Hinweis beraten.

TOP	Thema
	<p>4. Einige Teilnehmerinnen und Teilnehmer fordern, dass die Bewertung zwingend von der Fachgruppe durchgeführt werden solle. Die LQS/LAG sollte nicht die abschließende Entscheidung treffen und diese verantworten müssen.</p> <ul style="list-style-type: none"> ▫ Das IQTIG wird den Hinweis beraten, weist jedoch darauf hin, dass formal die LQS/LAG die Verantwortung für den Prozess trägt. <p>III. Kriterien für das Einleiten von Maßnahmen zur Qualitätsförderung</p> <p>Das IQTIG präsentiert die Empfehlungen hinsichtlich der Kriterien für die Einleitung von Maßnahmen zur Qualitätsförderung (vgl. Präsentationsfolien 42–43).</p> <p>Herausforderungen und Lösungsansätze:</p> <ol style="list-style-type: none"> 1. Einige Teilnehmerinnen und Teilnehmer empfehlen, dass es sich bei der zwingenden Einleitung von Fördermaßnahmen bei einem wiederholt auffälligen Ergebnis um denselben Qualitätsindikator oder einen thematisch ähnlichen Indikator handeln sollte. <ul style="list-style-type: none"> ▫ Das IQTIG wird diesen Hinweis beraten. 2. Ein Teilnehmer weist darauf hin, dass die Chance, dass ein Leistungserbringer, der z. B. im EJ 2018 auffällig war und diese Information Mitte des Folgejahres (EJ 2019) erhält, bereits im EJ 2019 ein unauffälliges Ergebnis erzielen zu können, gering sei. <ul style="list-style-type: none"> ▫ Das IQTIG nimmt den Hinweis auf und erläutert, dass in der DeQS-RL quartalsweise Auswertungen vorgesehen sind, sodass das interne QM frühzeitig intervenieren könne. Es gibt weiterhin zu bedenken, dass es sich bei den qualitätsfördernden Maßnahmen auch um sehr niedrigschwellige Maßnahmen handeln könne. 3. Einige Teilnehmerinnen und Teilnehmer erachten einen 5-Jahres-Zeitraum bis zwingend qualitätsfördernde Maßnahmen eingeleitet werden müssen als zu lang. <ul style="list-style-type: none"> ▫ Der Vorschlag „innerhalb von 3 Jahren bei thematisch ähnlichen Indikatoren“ wird durch das IQTIG geprüft. 4. Einige Teilnehmerinnen und Teilnehmer wünschen Kriterien, die vorgeben, wann welche Fördermaßnahmen einzuleiten sind. <ul style="list-style-type: none"> ▫ Das IQTIG wird den Hinweis beraten.
TOP 5	<p>Analyse Herausforderungen und Diskussion Lösungsansätze – „Verkürzung Gesamtverfahrens (I.) und Strukturierter Dialog und Auffälligkeitskriterien (AK) (II.)“</p> <p>I. Verkürzung des Gesamtverfahrens</p> <p>Das IQTIG stellt die Frage, welche weiteren Maßnahmen zu einer „Verschlankung“ des Verfahrens im Sinne einer Verkürzung führen würden.</p> <p>Die Teilnehmerinnen und Teilnehmer äußern, dass die Auswertungen früher kommen müssten. Außerdem sei der in der DeQS-RL vorgesehene Prozess bis zur Anforderung einer Stellungnahme zu aufwendig.</p> <p>II. Der Strukturierte Dialog und Auffälligkeitskriterien</p> <ol style="list-style-type: none"> 1. Einige Teilnehmerinnen und Teilnehmer erklären, dass die Bewertung der Stellungnahmen zu den AK und Qualitätsindikatoren bisher in einem Prozess/Vorgang abgehandelt wurden und dies auch zukünftig beibehalten werden solle.

TOP	Thema
	<p>Das Verfahren bei den AK solle sich nicht von dem für Qualitätsindikatoren unterscheiden.</p> <ul style="list-style-type: none"> ▫ Das IQTIG erklärt, dass das Konzept für den Strukturierten Dialog zu den AK in der Entwicklung ist. Die bisherige Bewertung der AK wurden in korrekte Dokumentation, fehlerhafte Dokumentation und „Sonstiges“ eingeteilt. <p>2. Das IQIG bittet die Teilnehmerinnen und Teilnehmer, ihre Vorschläge zum Verfahren und zur Klassifikation/Bewertung der AK im Nachgang zum Workshop dem IQTIG schriftlich mitzuteilen.</p> <p>3. Bezogen auf die Frage des IQTIG, inwiefern qualitätsfördernde Maßnahmen bezüglich der AK eingeleitet wurden, nennen einzelne Teilnehmerinnen und Teilnehmer Maßnahmen zur Förderung der Dokumentationsqualität. Aus Sicht der Teilnehmerinnen und Teilnehmer wäre eine mögliche „dokumentationsqualitätsfördernde Maßnahme“ zu überlegen, inwiefern die Akten anzufordern seien, wenn es wiederholt nicht korrekte Dokumentationen gab. Auch der Aspekt zur Qualitätsförderung bzgl. der Dokumentationsqualität soll im Konzeptvorschlag bedacht werden.</p> <ul style="list-style-type: none"> ▫ Das IQTIG wird den Hinweis beraten.
TOP 6	<p>Weiteres Vorgehen</p> <p>Das IQTIG teilt mit, dass das Protokoll den Workshop-Teilnehmerinnen und -Teilnehmern zeitnah zugesandt wird. Das Konzept wird am 28. Mai im G-BA und beim LQS-Treffen am 18. Juni 2019 vorgestellt.</p> <p>Das IQTIG teilt mit, dass im Nachgang eine E-Mail zur Erinnerungen zu den Detailfragen an die Teilnehmerinnen und Teilnehmer versendet wird.</p>
TOP 7	<p>Sonstiges</p> <p>Ein Teilnehmer äußert, dass das Konzept bislang keine Empfehlungen zum Umgang mit Best-Practice-Leistungserbringern beinhaltet. Das IQTIG erläutert, dass das Teil von Stufe 2 der Beauftragung sei.</p>

2 Auszug des Protokolls zum LQS-IQTIG-Treffen am 18. Juni 2019

TOP	Thema
TOP 3	<p>Weiterentwicklung Strukturiertes Dialog</p> <p>Präsentation der Ergebnisse des Workshops</p> <p>Das IQTIG präsentiert das Rahmenkonzept zur Weiterentwicklung des Strukturierten Dialogs (analog zur Präsentation des entsprechenden Workshops vom 27. Mai 2019) sowie die Workshop-Ergebnisse (Präsentationsfolien 5–30).</p> <p>Das IQTIG erläutert, dass das Protokoll des Workshops dazu dient, die Workshop-Ergebnisse zu sichern. Die seitens der Workshop-Teilnehmerinnen und -Teilnehmer eingegangenen Anmerkungen zum Protokoll werden geprüft und das Protokoll ggf. überarbeitet. Darüber hinaus wird das Workshop-Protokoll voraussichtlich Bestandteil des Abschlussberichts sein. Das IQTIG erläutert, dass das Protokoll und die eingehenden schriftlichen Hinweise der LAG- und LQS-Vertreterinnen und -Vertreter wichtige Dokumente für die weitere Entwicklung des Konzepts sind und das IQTIG nach eingehender Prüfung die Anmerkungen ggf. entsprechend berücksichtigen wird. Des Weiteren wird erläutert, dass es ein Stellungnahmeverfahren zum Abschlussbericht geben wird. Das IQTIG wird das finale Protokoll gemeinsam mit den Workshop-Folien zum Rahmenkonzept an alle LAG- und LQS-Vertreterinnen und -Vertreter senden.</p> <p>Einige Teilnehmerinnen und Teilnehmer merken an, dass ein rechnerisch unauffälliges Ergebnis nicht automatisch als qualitativ unauffälliges Ergebnis bewertet und dass die Nicht-Einhaltung formaler Kriterien nicht mit unzureichender medizinischer Versorgungsqualität gleichgesetzt werden dürfe. Auch im Hinblick auf das Qualitätsportal wird angemerkt, dass Patientinnen und Patienten anhand der Bewertung erkennen sollten, ob das qualitativ auffällige Ergebnis aufgrund formaler Fehler oder aufgrund unzureichender medizinischer Qualität zustande gekommen ist.</p> <p>Die Teilnehmerinnen und Teilnehmer äußern sich kontrovers über die Abgabefrist für Stellungnahmen.</p> <p>Es wird der Hinweis gegeben, dass sich das Rahmenkonzept an der DeQS-RL orientieren sollte und das Konzept die zeitlichen Vorgaben der Richtlinie berücksichtigen müsse. Das IQTIG bestätigt, dass sich die Empfehlungen im Abschlussbericht auf die DeQS-RL beziehen werden.</p> <p>Ein Teilnehmer macht folgende Anmerkungen:</p> <ol style="list-style-type: none"> 1. Ein Konzept zu „Lernen von den Besten“ ist bislang nicht im Konzept enthalten. 2. Die Möglichkeit Hinweise versenden zu können wird vermisst. 3. Auch die Sozialdaten sollten im Konzept berücksichtigt werden. 4. Die QIDB sollte möglichst frühestmöglich geliefert werden, um den Strukturierten Dialog rechtzeitig zu starten. Ideal wäre z. B. Anfang Februar.

5. Für die Anforderung der Stellungnahmen sollten konkrete und einheitliche Fragen formuliert werden.

Das IQTIG nimmt die Hinweise auf und wird diese prüfen.

Seitens einiger Teilnehmerinnen und Teilnehmer wird angemerkt, dass eine einheitliche Aufwandentschädigung aller Mitglieder der Expertengremien geleistet werden sollte. Des Weiteren wird die Begrifflichkeit „hinreichender Verdacht“ hinterfragt und gebeten diese zu prüfen.

Das Konzept müsse zudem berücksichtigen, dass Fördermaßnahmen, die eingeleitet wurden sich oftmals nicht direkt im Folgejahr in einem unauffälligem QI-Ergebnis niederschlagen.

Das IQTIG führt aus, dass die Operationalisierung des hinreichenden Verdachts auf ein Qualitätsdefizit im Abschlussbericht vorgenommen wird.

Einige Teilnehmerinnen und Teilnehmer merken an, dass ein Vorgehen im Umgang mit „Hybridstellungennahmen“ (ATB, Dokumentationsfehlern, nicht akzeptierte Gründe) gefunden werden müsse.

Das IQTIG nimmt die Hinweise auf und wird diese prüfen.

Das IQTIG bittet alle LAG- und LQS-Vertreterinnen und -Vertreter ihre Vorschläge zum Rahmenkonzept und zu Optionen zur Verkürzung des Verfahrens bis zum **26. Juni 2019** einzureichen. Das IQTIG wird hierzu eine Erinnerungs-E-Mail versenden.

3 Ergänzende Informationen zu Kapitel 2 – Analyse der Ausgangssituation

Tabelle 1: Anzahl von Fachgruppen und Mitgliederzahl pro Fachgruppe in der externen stationären Qualitätssicherung (Winkler-Komp et al. 2014)

Bundesland	Anzahl der Fachgruppen	Mitgliederzahl / Fachgruppe	
		Durchschnitt	Spannweite
Baden-Württemberg	9	6,7	8–10
Bayern	5	9,4	6–8
Berlin	7	8	8
Brandenburg	8	4	3–6
Bremen	7	9,9	6–13
Hamburg	7	11	8–15
Hessen	7	7,1	4–13
Mecklenburg-Vorpommern	7	7,1	6–11
Niedersachsen	6	5,8	4–8
Nordrhein-Westfalen	7	9,4	5–18
Saarland	7	7,7	7–8
Sachsen	7	7,3	5–10
Sachsen-Anhalt	7	5,1	3–6
Schleswig-Holstein	9	5,8	4–8
Thüringen	8	7,5	6–8
Gesamt	114	7	3–18

Tabelle 2: Einstufungsschema für rechnerische Auffälligkeiten bei Qualitätsindikatoren nach Abschluss des Strukturierten Dialogs (bisheriges Vorgehen)

Kategorie	Einstufung	Ziffer	Begründung
N	Bewertung nicht vorgesehen	01	Qualitätsindikator ohne Ergebnis, da entsprechende Fälle nicht aufgetreten sind
		02	Referenzbereich ist für diesen Indikator nicht definiert
		99	Sonstiges (im Kommentar erläutert)

Kategorie	Einstufung	Ziffer	Begründung
R	Ergebnis liegt im Referenzbereich	10	Ergebnis rechnerisch unauffällig, daher kein Strukturierter Dialog erforderlich
H	Einrichtung auf rechnerisch auffälliges Ergebnis hingewiesen	20	Aufforderung an das einrichtungsinterne Qualitätsmanagement zur Analyse der rechnerischen Auffälligkeit
		99	Sonstiges (im Kommentar erläutert)
U	Bewertung nach Strukturiertem Dialog als qualitativ unauffällig	30	Korrekte Dokumentation wird bestätigt (Datenvalidierung)
		31	Besondere klinische Situation
		32	Das abweichende Ergebnis erklärt sich durch Einzelfälle
		33	Kein Hinweis auf Mängel der med. Qualität (vereinzelte Dokumentationsprobleme)
		99	Sonstiges (im Kommentar erläutert)
A	Bewertung nach Strukturiertem Dialog als qualitativ auffällig	40	Fehlerhafte Dokumentation wird bestätigt (Datenvalidierung)
		41	Hinweise auf Struktur- und Prozessmängel
		42	Keine (ausreichend erklärenden) Gründe für die rechnerische Auffälligkeit benannt
		99	Sonstiges (im Kommentar erläutert)
D	Bewertung nicht möglich wegen fehlerhafter Dokumentation	50	Unvollständige oder falsche Dokumentation
		51	Softwareprobleme haben eine falsche Dokumentation verursacht
		99	Sonstiges (im Kommentar erläutert)
S	Sonstiges	90	Verzicht auf Maßnahmen im Strukturierten Dialog
		91	Strukturierter Dialog noch nicht abgeschlossen
		99	Sonstiges (im Kommentar erläutert)

Tabelle 3: Anteil von Hinweisen an allen rechnerisch auffälligen Ergebnissen nach Bundesland zum Erfassungsjahr 2015 (modifizierte Tabelle aus dem Bericht zum Strukturierten Dialog 2016; IQTIG 2017: 52)

Bundesland	rechnerische Auffälligkeit n	Hinweis H20 + H99	
		n	%
Bayern	2759	1684	61,0
Baden-Württemberg	1556	901	57,9
Berlin	551	181	32,8
Brandenburg	525	71	13,5
Bremen	112	6	5,4
Hamburg	257	92	35,8
Hessen	977	461	47,2
Mecklenburg-Vorpommern	373	79	21,2
Niedersachsen	1410	165	11,7
Nordrhein-Westfalen	3357	1165	34,7
Rheinland-Pfalz	1152	491	42,6
Saarland	208	2	1,0
Sachsen	800	110	13,8
Sachsen-Anhalt	612	249	40,7
Schleswig-Holstein	439	105	23,9
Thüringen	397	177	44,6
IQTIG	371	0	0,0
Gesamt	15856	5939	37,5

Tabelle 4: Entscheidungsverfahren in den Bundesländern, erfasst anhand der Frage: „Wie wird festgelegt, ob ein Ergebnis als qualitativ auffällig oder unauffällig beurteilt wird?“ (Winkler-Komp et al. 2014)

Art der Entscheidung	Häufigkeit
im Konsens	9
teilweise durch Mehrheitsbeschluss	4
Konsens und teilweise durch Mehrheitsbeschluss	2
Konsens in LQS oder in Fachgruppe auf Landesebene	2

4 Biometrische Analyse: Heterogenität der Vorgehensweise und Bewertung im Strukturierten Dialog – Exemplarische Analyse anhand von Daten aus dem Erfassungsjahr 2016

Im Folgenden wird eine Methodik vorgestellt, welche ermöglicht die Heterogenität in den Bewertungen des Strukturierten Dialoges quantitativ zu analysieren. Auf Basis dieser Methodik wurden im Abschlussbericht in Abschnitt 2.5 Analyseergebnisse formuliert, die an dieser Stelle ausführlich hergeleitet und beschrieben werden. Angewandt wird die Methodik auf folgende Qualitätsindikatoren: „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ (QI-ID 54003), „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“ (QI-ID 54139) sowie „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“ (QI-ID 52010), vgl. Tabelle 2 im Abschlussbericht, Abschnitt 2.5.

Für die Analyse der Unterschiede zwischen den Bewertungsweisen der Bundesländer wurden folgende zwei Fragestellungen formuliert:

1. **Fragestellung 1:** Wie groß ist die landesstellenabhängige Heterogenität bezüglich der Aufnahme eines Stellungnahmeverfahrens bei rechnerischer Auffälligkeit eines Leistungserbringers?
2. **Fragestellung 2:** Wie groß ist die landesstellenabhängige Heterogenität bezüglich der Bewertung als qualitativ auffällig nach Aufnahme eines Stellungnahmeverfahrens?

Diese Fragestellungen sind jeweils unter Berücksichtigung der rechnerischen Ergebnisse der Leistungserbringer in ihrem Bundesland zu beantworten, um auszuschließen, dass sich die beobachteten Unterschiede durch Unterschiede zwischen den Leistungserbringern erklären. Es geht also darum zu charakterisieren, wie die jeweiligen Bewertungsstellen *vergleichbare rechnerische Ergebnisse* bewerten.¹

4.1 Methodik

Zur Beantwortung von Fragestellung 1 und 2 ist zu untersuchen, wie verschiedene Bewertungsstellen vergleichbare rechnerische Ergebnisse von Leistungserbringern bewerten. Dazu ist eine Definition zu treffen, welche rechnerischen Ergebnisse als vergleichbar einzustufen sind. Die QSKH-Richtlinie sieht zwar keine Differenzierung der rechnerischen Ergebnisse etwa nach Größe der Leistungserbringer oder Ausmaß der Überschreitung des Referenzwertes vor.² Für eine differenzierte Analyse der Bewertungsergebnisse des Strukturierten Dialogs ist es allerdings unabdingbar die rechnerischen Ergebnisse genauer mit einzubeziehen. Ein geeigneter Ausgangspunkt

¹ Dabei dienen die rechnerischen Ergebnisse der Leistungserbringer für die Analyse der qualitativ auffälligen Bewertungen (Fragestellung 2) lediglich als Approximation, um tatsächliche Qualitätsunterschiede zwischen den Leistungserbringern zu modellieren (vgl. Abschnitt 2.3).

² Mit Ausnahme der sog. Ein-Fall-Regel

dafür ist eine nach Bundesland und Größe (Fallzahl) der Leistungserbringer differenzierte Auswertung der Ergebnisse des strukturierten Dialoges. Derartige Auswertungen werden exemplarisch in Abbildung 1 für einen Qualitätsindikator und zwei ausgewählte Bewertungsstellen dargestellt. In Abschnitt 4.5 finden sich vollständige Auswertungen für alle drei ausgewählten Qualitätsindikatoren. Dort finden sich auch weitere Darstellungen der Ergebnisse des strukturierten Dialoges zu diesen Indikatoren.

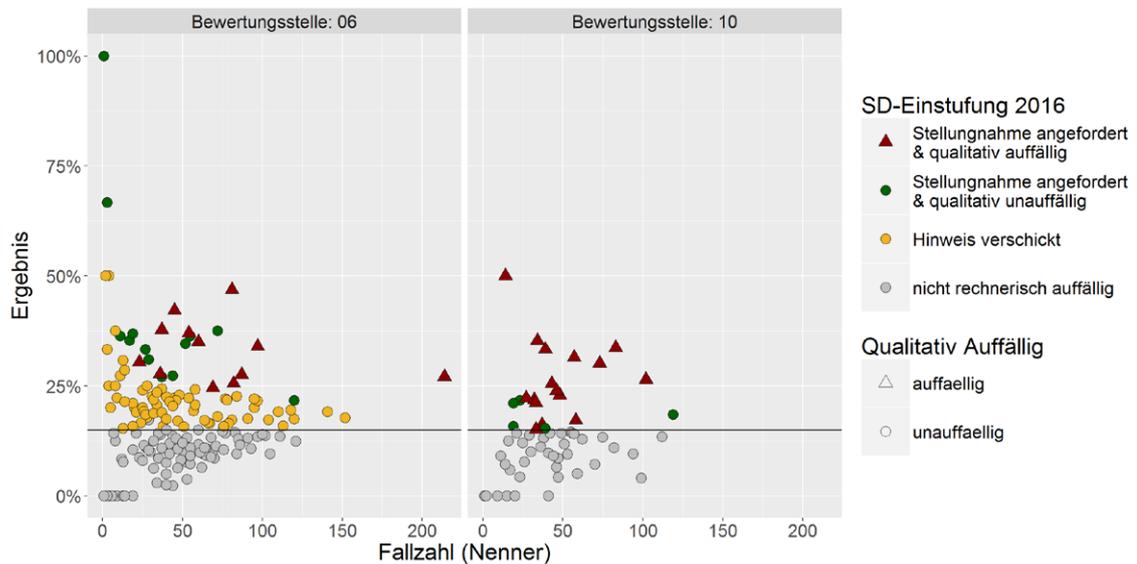


Abbildung 1: Ergebnisse des Strukturierten Dialoges zum Qualitätsindikator HEP: QI 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ für das Erfassungsjahr 2016 aus zwei exemplarisch ausgewählten Bewertungsstellen.

Es zeigt sich, dass in der Praxis mehrerer Bewertungsstellen inferenzstatistische Methoden (z. B. Konfidenzintervalle) zum Einsatz kommen, um die Entscheidung für die Aufnahme eines Stellungnahmeverfahrens zu treffen. Die Handhabung des Strukturierten Dialoges fällt somit i. d. R. bereits innerhalb einer Bewertungsstelle unterschiedlich aus, je nach Größe des Leistungserbringers und Ausmaß der Referenzbereichsüberschreitung. Die vorliegende Analyse untersucht, wie unterschiedlich die Bewertungen bei *ähnlichen* Fallzahlen und *ähnlichen* rechnerischen Ergebnissen der Leistungserbringer ausfallen.

Ein erster Ansatz die Heterogenität der Bewertungen des Strukturierten Dialogs zu quantifizieren könnte sein, Gruppen von Leistungserbringern mit vergleichbaren Ergebnissen zu bilden und diese anschließend nach Bundesland zu stratifizieren. Da es allerdings in einem einzelnen Qualitätsindikator selbst bei Indikatoren mit besonderem Handlungsbedarf aus den größten QS-Verfahren bundesweit nur wenige hundert rechnerisch auffällige Standorte gibt, kann man auf Ebene der Bundesländer keine statistisch aussagekräftigen Vergleiche durchführen, wenn man zusätzlich zu der Stratifizierung nach Bundesland weitere Untergruppen von Leistungserbringern mit ähnlichem Ergebnis betrachtet.

Analysemodell

Ein alternativer Ansatz an Stelle einer Stratifizierung der Leistungserbringerergebnisse ist ein modellbasierter Ansatz. Um die Heterogenität im Strukturierten Dialog unter Berücksichtigung von Unterschieden zwischen den Leistungserbringern in ihren rechnerischen Ergebnissen und Unterschieden in ihren Fallzahlen zu analysieren, modellieren für einen Qualitätsindikator die Wahrscheinlichkeit, dass

1. von einem Leistungserbringer eine Stellungnahme angefordert wird (Fragestellung 1),
2. ein Leistungserbringer nach Anforderung einer Stellungnahme als qualitativ auffällig bewertet wird (Fragestellung 2),

in Abhängigkeit des rechnerischen Leistungserbringerergebnisses und der Bewertungsstelle. Sei Y_{lb} das interessierende Bewertungsergebnis im Strukturierten Dialog eines Leistungserbringers l durch Bewertungsstelle b . Je nach Fragestellung 1 oder 2 beschreibt Y_{lb} das Anfordern einer Stellungnahme (ja = 1, nein = 0) oder die Bewertung als qualitativ auffällig (ja = 1, nein = 0). Wir formulieren folgendes logistisches Regressionsmodell mit bewertungsstellenspezifischen Effekten:

$$\log \left[\frac{P(Y_{lb} = 1 | r_{lb}, u_b)}{1 - P(Y_{lb} = 1 | r_{lb}, u_b)} \right] = \beta_0 + \beta_{Q_2} \cdot Q_2(r_{lb}) + \beta_{Q_3} \cdot Q_3(r_{lb}) + \beta_{Q_4} \cdot Q_4(r_{lb}) + u_b. \quad (1)$$

Hierbei sind $\beta_0, \beta_{Q_2}, \beta_{Q_3}, \beta_{Q_4}$ und u_b zu bestimmende Modellparameter. Die globale Konstante β_0 quantifiziert die Grundhäufigkeit von Stellungnahmen (Fragestellung 1) bzw. qualitativen Auffälligkeiten (Fragestellung 2) für den betrachteten Qualitätsindikator. Mit r_{lb} quantifizieren wir die statistische Evidenz, dass ein Leistungserbringer außerhalb des Referenzbereiches liegt. Je nach Bewertungsart des Indikators fließen dabei die rechnerischen Ergebnisse des Leistungserbringers unterschiedlich in r_{lb} ein. Für jeden Leistungserbringer ist dabei r_{lb} eine bekannte Zahl, welche von den beobachteten Ergebnissen des Leistungserbringers (Zähler und Nenner) abhängt. In der Art und Weise wie r_{lb} definiert wird, wird festgelegt welche Leistungserbringerergebnisse als vergleichbar eingestuft werden (siehe Anhang 4.4.2 für eine ausführliche Diskussion). Mit $Q_2(r_{lb}), Q_3(r_{lb})$ und $Q_4(r_{lb})$ bezeichnen wir das zweite bis vierte Quartil von r_{lb} . Die Koeffizienten β_{Q_2}, β_{Q_3} und β_{Q_4} quantifizieren somit, wie stark die Wahrscheinlichkeit $P(Y_{lb} = 1 | r_{lb}, u_b)$ ansteigt, wenn ein Leistungserbringer im zweiten, dritten oder vierten Quartil der rechnerischen Ergebnisse r_{lb} liegt im Vergleich mit dem ersten Quartil.

Wahl von r_{lb}

Für die beiden ratenbasierten Indikatoren wählen wir zur Definition von r_{lb} p-Werte eines einseitigen exakten Binomialtests [vgl. z. B. Spiegelhalter et al. (2012)], welcher sowohl das Ratenergebnis als auch die Fallzahl des Leistungserbringers berücksichtigt. Diese Vorgehensweise wird gewählt, da die Streuung der rechnerischen Leistungserbringerergebnisse von der Fallzahl abhängt und somit Fallzahl und Ratenergebnis keine unabhängigen Kovariablen darstellen.³ Beispielsweise ist davon auszugehen, dass bei einem Referenzbereich $\leq 15\%$ die Ergebnisse 2 von

³ In einem ratenbasierten Qualitätsindikator streuen die Ergebnisse der Leistungserbringer mit kleiner Fallzahl i. d. R. breiter. Beispielsweise kann ein Leistungserbringer mit nur fünf behandelten Fällen nur die Ergebnisse

4 und 50 von 100 jeweils zu einer anderen Bewertung durch die Fachgruppen führen, obwohl beide einem Ergebnis von 50 % entsprechen. In der Regressionsmodellierung ist daher zwingend eine Interaktion zwischen Fallzahl und Ergebnis zu berücksichtigen. Der p-Wert ist eine einfache Möglichkeit diese Interaktion abzubilden, welche zudem auch zu einer klaren Interpretation der bewertungsstellenspezifischen Effekte u_b führt: Diese Stellen die Unterschiede in der Bewertung bei gleicher statistischer Evidenz der Leistungserbringerergebnisse dar. Die Parameter β_{Q_2} , β_{Q_3} und β_{Q_4} beschreiben, wie sehr die stat. Evidenz r_{lb} in das Bewertungsergebnis Y_{lb} einfließen. Für den Sentinel-Event-Indikator ist r_{lb} definiert als die Anzahl beobachteter Sentinel-Events pro Leistungserbringer, ohne Berücksichtigung der Fallzahl, vgl. Abschnitt 4.4.2 für weitere Details.

Random-Effects-Modell

Im oben beschriebenen Ansatz ist die statistische Schätzung der Modellparameter bei geringer Anzahl von Leistungserbringern eine Herausforderung, weshalb für die Modellierung des Einflusses der Bundesländer ein sogenanntes Random-Effects-Modell [vgl. z. B. Agresti (2013), Stroup (2013)] benutzt wird, welches die Schätzung der Effekte u_b vor allem bei kleiner Anzahl von Leistungserbringern pro Bundesland stabilisiert (vgl. Ash et al. (2012) für eine Diskussion zu Random-Effects-Modellen im Kontext der Qualitätsbewertung von Krankenhäusern). Dabei wird die Annahme getroffen, dass die bewertungsstellenspezifischen Effekte u_b einer übergeordneten Normalverteilung mit Varianz τ^2 unterliegen. Die Standardabweichung τ ist dabei ein weiterer zu schätzender Modellparameter, welcher explizit die Heterogenität der Bewertungen zwischen den Bundesländern quantifiziert. Der Parameter τ ist dabei ein geeignetes Maß für die Beantwortung von Fragestellung 1 und 2, da es quantifiziert, wie groß die Unterschiede in der Wahrscheinlichkeit $P(Y_{lb} = 1 | r_{lb}, u_b)$ zwischen den Bundesländern sind, die nicht durch die rechnerischen Ergebnisse der Leistungserbringer erklärt werden können. Je größer der geschätzte Wert von τ ausfällt, desto größer ist die auf die Bewertungsstellen zurückzuführenden Heterogenität. Bei der Interpretation der Standardabweichung τ ist allerdings zu beachten, dass damit die Standardabweichung der u_b auf Ebene des linearen Prädiktors (Logit-Ebene) modelliert wird. Die Streuung der Wahrscheinlichkeiten $P(Y_{lb} = 1 | r_{lb}, u_b)$ je nach Bewertungsstelle hängt neben dem Wert von τ auch von den rechnerischen Leistungserbringerergebnissen r_{lb} ab. Beispielsweise ist zu erwarten, dass für extreme Ergebnisse r_{lb} eine geringere Streuung in den Wahrscheinlichkeiten $P(Y_{lb} = 1 | r_{lb}, u_b)$ vorherrscht als für Ergebnisse, die den Referenzbereich nur knapp überschreiten.

Eine mögliche Interpretation der Heterogenitätsparameter τ bieten sog. *Median Odds Ratios* [vgl. Larsen et al. (2000)]. Würde man zwei Leistungserbringer A und B aus dem gleichen Quartil der rechnerischen Ergebnisse von zwei zufällig ausgewählten Bewertungsstellen bewerten lassen, so wäre der Odds-Ratio zwischen Leistungserbringer A mit dem höheren Risiko und Leistungserbringer B im Median:

$$\text{MOR}(\tau) = \exp(\sqrt{2 \cdot \tau^2} \cdot \Phi^{-1}(3/4)),$$

0 %, 20 %, 40 %, 60 %, 80 %, 100 % erreichen. Ein einzelner Behandlungsfall hat somit einen größeren Einfluss auf das Leistungserbringerergebnis.

dabei ist Φ^{-1} die inverse Verteilungsfunktion der Normalverteilung.

Datengrundlage

Datengrundlage der Analyse zu Fragestellung 1 sind alle im Erfassungsjahr 2016 im jeweils betrachteten Qualitätsindikator rechnerisch auffällig gewordenen Standorte. Für die beiden ratenbasierten Qualitätsindikatoren (QI 54003 und 54139) wurden zusätzlich alle Standorte ausgeschlossen, welche nur auf Grund eines einzigen Zählerfalles rechnerisch auffällig wurden. In diesem Fall wird im strukturierten Dialog nach der sog. Ein-Fall-Regel fast immer ein Hinweis verschickt, und somit unterliegt das Anfordern einer Stellungnahme i. d. R. nicht dem Ermessen der Landesfachgruppen.

Datengrundlage der Analyse zu Fragestellung 2 sind alle Standorte, von welchen zum Erfassungsjahr 2016 eine Stellungnahme angefordert wurde und für welche keine D-Bewertung („Bewertung nicht möglich wegen fehlerhafter Dokumentation“) getroffen wurde.

4.2 Ergebnisse

Die Berechnung der Modelle zu Fragestellung 1 (Aufnahme eines Stellungnahmeverfahrens bei rechnerischer Auffälligkeit) ergab für die drei Qualitätsindikatoren die in Tabelle 5 dargestellten Parameterschätzungen.

Tabelle 5: Parameterschätzungen der drei Modelle zu Fragestellung 1: Heterogenität bezüglich des Anforderns einer Stellungnahme

Parameter	Schätzung	(Median-)Odds-Ratio	Konfidenzintervall (Median-)Odds-Ratio	
			2,5 %	97,5 %
HEP: QI 54003 (ratenbasiert)				
β_0	0,57	-	-	-
β_{Q_2}	1,28	3,6	1,1	14
β_{Q_3}	4,28	72	20	360
β_{Q_4}	7,21	$1,4 \cdot 10^3$	300	$8,4 \cdot 10^3$
τ	4,67	86	15	$5,1 \cdot 10^3$
DEK: QI 52010 (Sentinel Event)				
β_0	13,65	-	-	-
β_{Q_3}	1,37	3,9	0,33	70
β_{Q_4}	6,27	530	68	$1,1 \cdot 10^4$
τ	8,08	$2,2 \cdot 10^3$	19	$7,9 \cdot 10^{12}$
09/1: QI 54139 (ratenbasiert)				

Parameter	Schätzung	(Median-)Odds-Ratio	Konfidenzintervall (Median-)Odds-Ratio	
			2,5 %	97,5 %
β_0	0,43	-	-	-
β_{Q_2}	0,61	1,8	0,54	6,9
β_{Q_3}	4,04	57	15	270
β_{Q_4}	4,48	88	23	430
τ	5,00	120	18	$9,2 \cdot 10^3$

Die Heterogenität zwischen den Bewertungsstellen lässt sich aus den Schätzungen für den Parameter τ bzw. dem daraus resultierenden Median-Odds-Ratio ablesen. Für alle drei Indikatoren ergeben die modellbasierten Analysen einen Wert für τ , dessen Konfidenzintervall 0 nicht einschließt. Dies bedeutet, dass bei allen drei Indikatoren signifikante Heterogenität darin vorliegt, ob bei vergleichbarem rechnerischen Ergebnis eine Stellungnahme eingeholt wird oder lediglich ein Hinweis versendet wird. Diese Heterogenität lässt sich dementsprechend nicht auf Unterschiede in der Fallzahl oder im Ausmaß der rechnerischen Auffälligkeit zwischen den Leistungserbringern der Bundesländer erklären. Der Modellparameter β_0 quantifiziert die Grundwahrscheinlichkeit für den jeweiligen Indikator, dass eine Stellungnahme eingeholt wird. Die Modellparameter β_{Q_2} bis β_{Q_4} sind auch deutlich von 0 verschieden. Dies bedeutet, dass der in r_{lb} zusammengefasste Einfluss von Fallzahl und Ausmaß der rechnerischen Auffälligkeit einen messbaren Einfluss auf die Entscheidung hat, ob eine Stellungnahme eingeholt wird oder ein Hinweis versendet wird. Dieser in r_{lb} kombinierte Einfluss kann als die Evidenz für die Verfehlung des Referenzbereichs interpretiert werden. Von Quartil Q_1 bis Quartil Q_4 steigt demnach die Evidenz dafür, dass die Leistungserbringer im jeweiligen Quartil den Referenzbereich verfehlt haben. Die von Q_2 bis Q_4 größer werdenden β -Parameter zeigen, dass mit zunehmender Evidenz für die Verletzung des Referenzbereichs die Bundesländer mit zunehmender Wahrscheinlichkeit eine Stellungnahme einholen statt einen Hinweis zu versenden.

Die Ergebnisse zu Fragestellung 2 (Qualitative Auffälligkeit nach Aufnahme eines Stellungnahmeverfahrens) finden sich in Tabelle 6.

Tabelle 6: Parameterschätzungen der drei Modelle zu Fragestellung 2: Heterogenität bezüglich der Bewertung als qualitativ auffällig

Parameter	Schätzung	(Median-)Odds-Ratio	Konfidenzintervall (Median-)Odds-Ratio	
			2,5 %	97,5 %
HEP: QI 54003 (ratenbasiert)				
β_0	-2,71	-	-	-
β_{Q_2}	1,03	2,8	0,95	9,3

Parameter	Schätzung	(Median-)Odds-Ratio	Konfidenzintervall (Median-)Odds-Ratio	
			2,5 %	97,5 %
β_{Q_3}	1,91	6,8	2,5	21
β_{Q_4}	3,02	21	7,7	66
τ	1,02	2,7	1,7	5,8
DEK: QI 52010 (Sentinel Event)				
β_0	-2,18	-	-	-
β_{Q_3}	1,18	3,2	1,5	7,3
β_{Q_4}	1,93	6,9	3,5	14
τ	1,59	4,6	2,6	12
09/1: QI 54139 (ratenbasiert)				
β_0	-2,83	-	-	-
β_{Q_2}	0,93	2,5	0,64	11
β_{Q_3}	0,93	2,5	0,72	10
β_{Q_4}	2,054	7,8	2,2	33
τ	1,289	3,4	1,5	14

Auch für Fragestellung 2 ergeben die modellbasierten Analysen für alle drei Indikatoren einen Wert für τ , dessen Konfidenzintervall 0 nicht miteinschließt. Somit gibt es auch in der Bewertung der Leistungserbringer als qualitativ auffällig eine messbare Heterogenität zwischen den Bewertungsstellen, die sich nicht durch die rechnerischen Ergebnisse der Leistungserbringer erklären lässt. Mit dem Quartil Q_2 bis Q_4 größer werdende β -Parameter belegen, dass mit zunehmender Evidenz r_{lb} für die Verletzung des Referenzbereichs die Wahrscheinlichkeit steigt, dass ein Leistungserbringer mit rechnerischem Ergebnis im entsprechenden Quartil abschließend als qualitativ auffällig bewertet wird.

Zur besseren Interpretation der Ergebnisse stellen wir die Verteilung der durch das Modell geschätzten Wahrscheinlichkeiten für beide Fragestellungen und alle drei Qualitätsindikatoren in Abbildung 2 bis Abbildung 4 dar. Für jedes Quartil der Leistungserbringerergebnisse wird durch die Punkte dargestellt, was die Bewertungswahrscheinlichkeit aller 16 Bewertungsstellen wäre, wenn diese einen Leistungserbringer aus dem entsprechenden Quartil beurteilt hätten. Dies wird auf Grundlage der geschätzten Werte der Parameter $\beta_0, \beta_{Q_2}, \beta_{Q_3}, \beta_{Q_4}$ und u_b für jede Bewertungsstelle b extrapoliert. In der grafischen Darstellung entspricht in jedem Quartil der Leistungserbringerergebnisse jeder Bewertungsstelle ein Punkt.

Vergleicht man die Ergebnisse der modellbasierten Analysen zwischen den Indikatoren und Fragestellungen, so stellt man fest, dass für die beiden ratenbasierten Indikatoren aus den QS-Verfahren 09/1 und HEP die Heterogenität beim Anfordern von Stellungnahmen (Fragestellung 1)

deutlich größer ist als in der Bewertung der Leistungserbringer als qualitativ auffällig (Fragestellung 2). In den Ergebnissen zu Fragestellung 1 (vgl. Abbildung 2 bis Abbildung 4 oben) lässt sich erkennen, dass sich die Bewertungsstellen tendenziell in zwei Gruppen einteilen: Bewertungsstellen, die fast immer Stellungnahmen einholen und Bewertungsstellen, welche seltener und nur bei deutlichen Überschreitungen des Referenzbereiches eine Stellungnahme anfordern. Diese bimodale Verteilung der Bewertungsstelleneffekte deutet auf eine Verletzung der Normalverteilungsannahme des Random-Effects-Modelles hin. Da es sich immer noch um eine symmetrische Verteilung handelt, kann der Parameter τ allerdings immer noch als ein valides Heterogenitätsmaß angesehen werden.

Ein anderes Bild ergibt sich bei dem Sentinel-Event-Indikator (Dekubitus Grad 4): 15 von 16 Landesstellen nehmen bei rechnerischer Auffälligkeit eines Leistungserbringers fast immer ein Stellungnahmeverfahren auf, sodass sich hier (bis auf ein Bundesland) ein sehr homogenes Bild bzgl. Fragestellung 1 zeichnet. Die Standardabweichung τ wird in diesem Fall allerdings sehr hoch geschätzt (mit breitem Konfidenzintervall), was sich vor allem durch das stark abweichende Verhalten einer Bewertungsstelle begründet. Auch hier ist davon auszugehen, dass die Normalverteilungsannahme des Random-Effects-Modells verletzt ist. Anders als bei den beiden ratenbasierten Indikatoren, stellt der Parameter τ für diese Art von Verteilung mit einem starken Ausreißer kein verlässliches Heterogenitätsmaß dar. Während für diesen Sentinel-Event-Indikator das Anfordern von Stellungnahmen recht homogen im Vergleich zu den beiden ratenbasierten Indikatoren gehandhabt wird, ist die Heterogenität in der Bewertung der Leistungserbringer als qualitativ auffällig für den Sentinel-Event-Indikator am größten.

Zusätzlich zum Einfluss der Bewertungsstellen lässt sich übergreifend für alle drei Qualitätsindikatoren und beide Fragestellungen ein deutlicher Einfluss der rechnerischen Leistungserbringerergebnisse auf die modellierten Wahrscheinlichkeiten feststellen. Im Extremfall hat bei der Frage nach dem Anfordern einer Stellungnahme (Fragestellung 1) im HEP-Indikator „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ ein Leistungserbringer aus dem vierten Quartil der rechnerischen Ergebnisse r_{lb} gegenüber einem Leistungserbringer aus dem ersten Quartil einen Odds-Ratio von $1,4 \cdot 10^3$ (95 % KI: [300; $8,4 \cdot 10^3$]). Für den Sentinel-Event-Indikator fallen erstes und zweites Quartil der rechnerischen Leistungserbringerergebnisse (also die Anzahl der Sentinel-Event-Ereignisse) für beide Fragestellungen zusammen: Mehr als die Hälfte der Leistungserbringer sind durch ein einzelnes Sentinel-Event-Ereignis rechnerisch auffällig geworden. Gegenüber diesen Leistungserbringern haben Leistungserbringer mit 2 Sentinel-Event-Fällen (drittes Quartil Q_3) und Leistungserbringer mit mehr als drei Sentinel-Event-Fällen (viertes Quartil Q_4) bei Fragestellung 1 einen Odds-Ratio von 3,9 (95 % KI: [0,33; 70]) für das dritte Quartil bzw. 530 (95 % KI: [68; $1,1 \cdot 10^4$]) für das vierte Quartil. Der Einfluss der rechnerischen Leistungserbringerergebnisse auf die Einstufung als qualitativ auffällig (Fragestellung 2) fällt insgesamt geringer aus als für Fragestellung 1. Auch hier stellt der Indikator „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ den Extremfall dar: Leistungserbringer im vierten Quartil der rechnerischen Leistungserbringerergebnisse haben gegenüber dem ersten Quartil einen Odds-Ratio von 21 (95 % KI: [7,7; 66]). Ein geringerer Effekt ist dagegen beim Indikator „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“ zu erkennen. Gegenüber dem ersten

Quartil der rechnerischen Leistungserbringerergebnisse r_{lb} liegt der Odds-Ratio für das vierte Quartil bei 7,8 (95 % KI: [2,2: 33]).

Bei allen Ergebnissen ist durch tendenziell große Konfidenzintervalle zu erkennen, dass die Parameterschätzungen einer deutlichen Unsicherheit unterliegen. Dies erklärt sich durch eine vergleichsweise kleine Datenbasis im Vergleich zur spezifizierten Modellkomplexität: In jedem Modell werden mindestens 20 Parameter geschätzt. Die Analysen zu beiden Fragestellungen basieren jedoch auf wenigen Hundert Datenpunkten (vgl. Abschnitt 2.5, Tabelle 2 im Abschlussbericht)

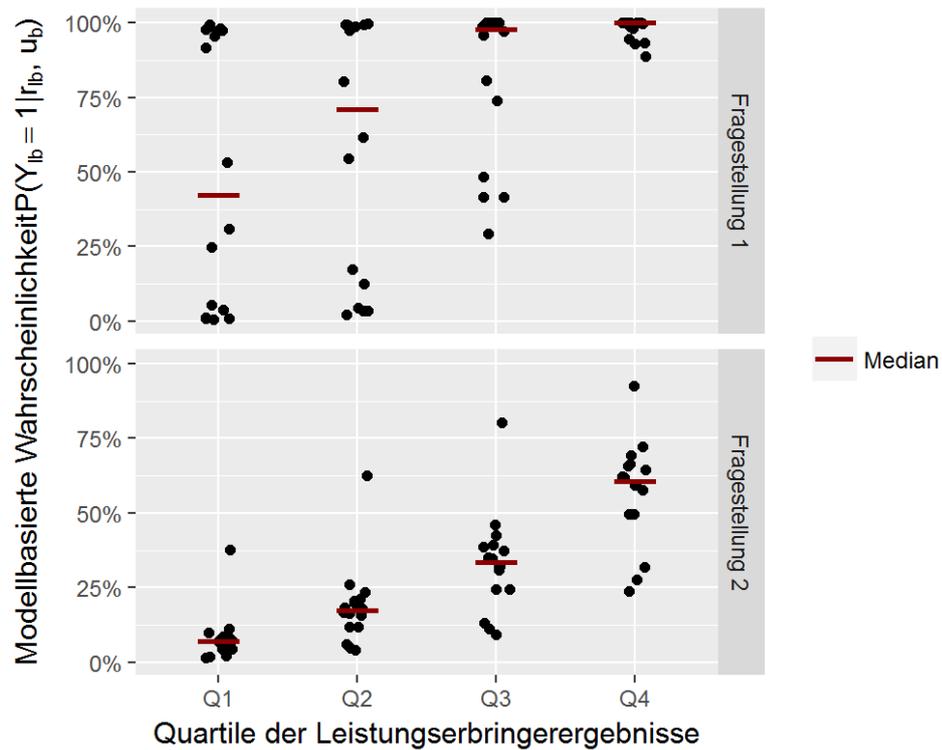


Abbildung 2: HEP QI 54003 (ratenbasierter Indikator). Streuung der geschätzten Bewertungswahrscheinlichkeit in Quartilen der Leistungserbringerergebnisse

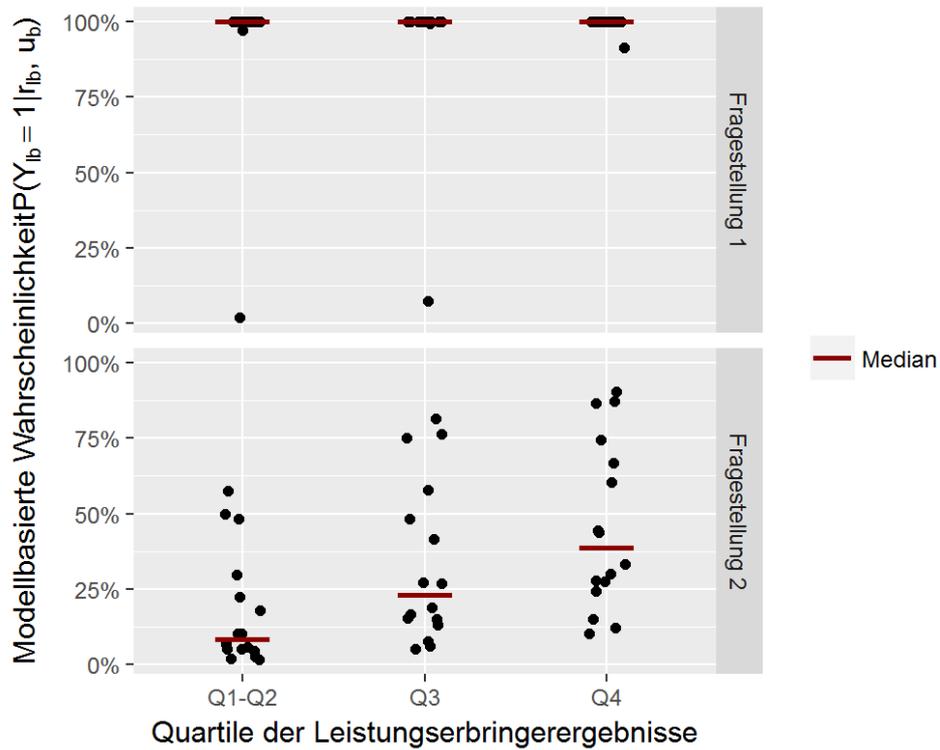


Abbildung 3: DEK QI 52010 (Sentinel-Event-Indikator). Streuung der geschätzten Bewertungswahrscheinlichkeit in Quartilen der Leistungserbringerergebnisse

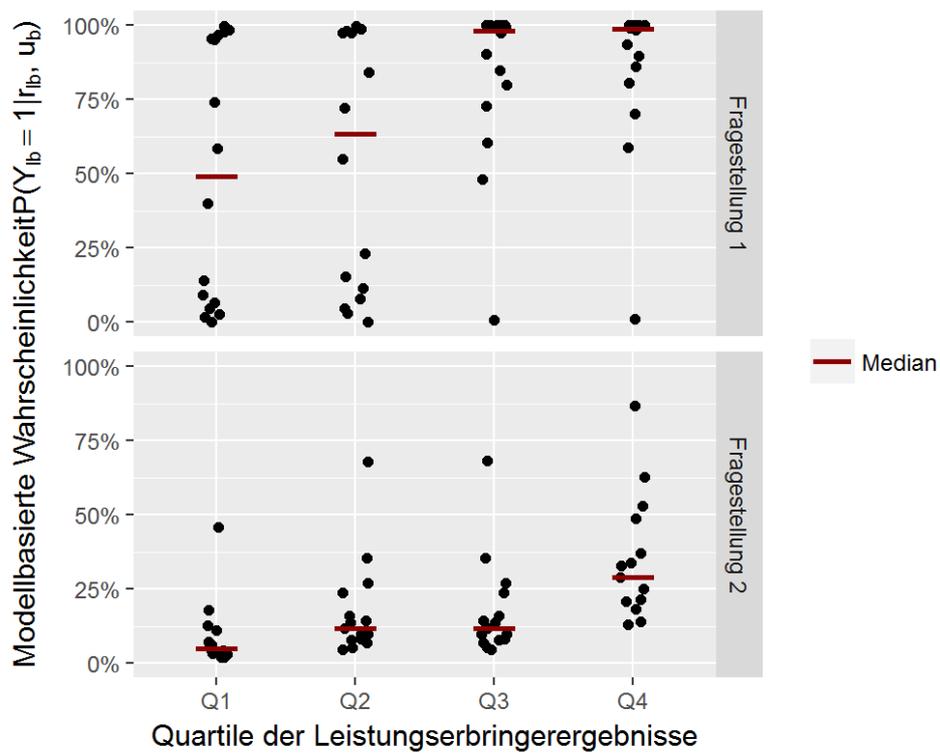


Abbildung 4: 09/1 QI 54139 (ratenbasierter Indikator). Streuung der geschätzten Bewertungswahrscheinlichkeit in Quartilen der Leistungserbringerergebnisse

4.3 Zusammenfassung und Diskussion

Es wurde eine Methodik vorgestellt, welche ermöglicht, die Heterogenität in der Aufnahme und Bewertungen des Strukturierten Dialoges quantitativ unter Berücksichtigung der rechnerischen Ergebnisse der Leistungserbringer zu untersuchen. Die Ergebnisse der Analyse zu beiden Fragestellungen belegen für die betrachteten Qualitätsindikatoren, dass auch unter Berücksichtigung der rechnerischen Leistungserbringerergebnisse eine starke Heterogenität zwischen den Bewertungsweisen der Landesstellen besteht. Dies gilt sowohl für die Frage nach der Aufnahme eines Stellungsnahmeverfahrens als auch für die finale Bewertung der Leistungserbringer als qualitativ auffällig. Dabei wird die Einstufung als qualitativ auffällig weniger heterogen gehandhabt als die Aufnahme eines Stellungsnahmeverfahrens.

Sensitivitätsanalysen mit Modellen, welche die rechnerischen Leistungserbringerergebnisse nicht berücksichtigen (vgl. Abschnitt 4.4.3), legen nahe, dass die Heterogenität in der Bewertung sogar *größer* wird, dadurch dass man die rechnerischen Ergebnisse der Leistungserbringer mit einbezieht. Die Hypothese, dass sich ein Teil der Heterogenität der Bewertungen des Strukturierten Dialogs aus tatsächlichen Unterschieden zwischen den Leistungserbringern erklärt, kann somit nicht bekräftigt werden.

Weiterhin zeigen Schätzungen der jeweiligen Parameter β_{Q_2} , β_{Q_3} und β_{Q_4} zu den Leistungserbringerergebnissen, dass die rechnerischen Ergebnisse der Leistungserbringer sowie ihre Fallzahl (implizit oder explizit) in die Bewertung durch die Landesstellen einbezogen werden. Insbesondere belegen weitere Sensitivitätsanalysen, dass bei der Bewertung der Leistungserbringerergebnisse durch die Bewertungsstellen neben dem rechnerischen Ergebnis auch die Fallzahl des Leistungserbringers eine Rolle spielt (vgl. Abschnitt 4.4.2). Dabei wurde allerdings nicht untersucht, wie genau welche Bewertungsstelle die rechnerischen Ergebnisse in ihre Bewertungen einfließen lässt. Beispielsweise berücksichtigen manche Bewertungsstellen Konfidenzintervalle, während andere Bewertungsstellen keine inferenzstatistischen Methoden anwenden.

Grundsätzlich ist die Rolle der rechnerischen Leistungserbringerergebnisse in der Definition der Modelle (1) je nach Fragestellungen eine andere. Für die Frage nach der Heterogenität bezüglich des Anforderns einer Stellungnahme, stellen die rechnerischen Ergebnisse der Leistungserbringer die wichtigste Grundlage dar, anhand derer sich die Bewertungsstellen orientieren können. Es ist daher sinnvoll zu betrachten, wie vergleichbare rechnerische Ergebnisse jeweils in den Bewertungsstellen bewertet werden. Eine andere Rolle haben die rechnerischen Ergebnisse bei der Analyse der finalen Qualitätsbewertungen als qualitativ auffällig. Hier dienen die rechnerischen Ergebnisse dazu tatsächliche Qualitätsunterschiede zwischen den Leistungserbringern zu approximieren, da kein anderes Maß um die Landesergebnisse quantitativ vergleichbar zu machen zur Verfügung steht.

Eine Limitation der vorliegenden Analyse liegt in der Einschränkung auf drei exemplarisch ausgewählte Qualitätsindikatoren: Nur wenige Qualitätsindikatoren weisen so viele rechnerisch auffällige Standorte und Stellungsnahmeverfahren auf, dass eine vergleichende statistische Analyse der Landesergebnisse auf Ebene einzelner Qualitätsindikatoren sinnvoll ist. Das Ergebnis

der Analyse ist daher nicht unbedingt repräsentativ für alle Qualitätsindikatoren. Es besteht allerdings die Möglichkeit zukünftig durch Verallgemeinerung der vorgestellten Modelle auch qualitätsindikatorenübergreifende Analysen durchzuführen. Dies kann etwa im Rahmen einer langfristigen Evaluation der getroffenen Maßnahmen zur Angleichung des strukturierten Dialoges sinnvoll sein. Die Varianz τ^2 der Bewertungsstelleneffekte aus dem Random-Effects-Modell wäre dabei auch für einen jahresübergreifenden Vergleich ein geeignetes Maß für die Heterogenität in der Bewertung des strukturierten Dialoges.

4.4 Details zur statistischen Modellierung

Wir gehen im Folgenden auf weitere Details zum in Gleichung (1) beschriebenen Modell ein und diskutieren die Modellwahl.

Sei Y_{lb} das betrachtete Bewertungsergebnis (je nach Fragestellung 1 oder 2) eines Leistungserbringers l durch Bewertungsstelle b . In Abschnitt 4.1 wurde folgendes logistisches Regressionsmodell mit bewertungsstellenspezifischen Effekten formuliert:

$$\text{logit}[P(Y_{lb} = 1|r_{lb}, u_b)] = \beta_0 + \beta_{Q_2} \cdot Q_2(r_{lb}) + \beta_{Q_3} \cdot Q_3(r_{lb}) + \beta_{Q_4} \cdot Q_4(r_{lb}) + u_b.$$

Hierbei sind $\beta_0, \beta_{Q_2}, \beta_{Q_3}, \beta_{Q_4}$ und u_b zu bestimmende Modellparameter, r_{lb} eine Transformation des rechnerischen Ergebnisses des Leistungserbringers. Mit $Q_2(r_{lb}), Q_3(r_{lb})$ und $Q_4(r_{lb})$ bezeichnen wir das zweite bis vierte Quartil von r_{lb} . Es handelt sich um ein hierarchisches Regressionsmodell bzw. Multilevelmodell [vgl. z. B. (Gelman und Hill 2007); (Stroup 2013)], da statistische Effekte auf zwei Ebenen (Bewertungsstelle, Leistungserbringer) berücksichtigt werden. Hierarchische Modelle werden insbesondere bei Analysen von Longitudinal- oder Cluster-Daten eingesetzt, bei denen Beobachtungen in bestimmten Gruppen auftreten. Die Beschreibung von Bewertungen des Strukturierten Dialogs, welche durch verschiedene Bewertungsstellen durchgeführt werden, ist somit ein Standardanwendungsfall hierarchischer Modelle.

4.4.1 Fixed- vs. Random-Effects-Modelle

Bei hierarchischen Regressionsmodellen unterscheidet man sog. Fixed- und Random-Effects-Parameter [vgl. z. B. Gelman und Hill (2007), Stroup (2013)]. Der Unterschied zwischen diesen beiden Arten von Parametern liegt dabei in der Vorgehensweise bei der statistischen Schätzung. Während bei der Schätzung von Fixed-Effects a priori keine Beschränkung der Parameter vorgesehen ist (und diese somit beliebig hohe Werte annehmen können), wird der mögliche Wertebereich von Random-Effects durch eine sog. Penalisierung beschränkt. Man spricht an dieser Stelle auch von Shrinkage, da die Modellkoeffizienten in einer penalisierten Schätzung tendenziell eine geringere Streuung aufweisen als ohne Penalisierung. Der Grad der Penalisierung muss dabei durch einen weiteren Parameter festgelegt werden. In Random-Effects-Modellen werden die sog. Penalisierungsparameter jeweils für eine Gruppe an Parametern (z. B. die Gruppe der u_b in Gleichung (1) anhand der Daten bestimmt. Dabei macht man die Annahme, dass die gruppenspezifischen Effekte (im Beispiel u_b) einer übergeordneten Normalverteilung $\mathcal{N}(0, \tau^2)$ folgen. Der Penalisierungsparameter wird dann als Varianz der gruppenspezifischen Effekte festgelegt [vgl. Wood (2013)]. Je größer die Varianz zwischen den Gruppen ausfällt, desto weniger wird die

Schätzung penalisiert. Im Gegensatz zu Fixed-Effects-Parametern, welche meist unter Verwendung der Maximum-Likelihood-Methode geschätzt werden, verwendet man für Random-Effects-Parameter sog. *BLUP*-Schätzer⁴ [vgl. z. B. Stroup (2013)]. In der Praxis führt die Verwendung von Random-Effects-Parametern mit Normalverteilungsannahme dazu, dass Ausreißereffekte im Vergleich zum Fixed-Effects-Modell reduziert und insbesondere die Parameter u_b von Gruppen mit geringer Fallzahl⁵ zum gruppenübergreifenden Mittelwert „geschrumpft“ werden. Eine detaillierte Diskussion zur Verwendung von Random-Effects-Modellen im Kontext der Qualitätsbewertung von Krankenhäusern findet sich bei Ash et al. (2012).

In der vorliegenden Analyse wurde ein Random-Effects-Modell verwendet, da einerseits die Anzahl an Leistungserbringern in einigen Bundesländern so gering ist, dass eine Stabilisierung der Schätzung zur Vermeidung von numerischen Konvergenzproblemen nötig ist. Andererseits ist genau die Quantifizierung der Heterogenität in der Bewertung durch die Landesstellen ein Ziel der Analyse. Der Varianz-Parameter τ^2 stellt hierfür ein geeignetes Maß zur Verfügung.

4.4.2 Operationalisierung des Begriffes vergleichbarer Leistungserbringer

Um zu analysieren wie unterschiedlich die Bewertung im Strukturierten Dialog für vergleichbare Leistungserbringer ausfällt, ist eine operative Definition des Begriffes „vergleichbare Leistungserbringer“ nötig. Dies ist unabhängig davon, ob man die Ergebnisse stratifiziert analysiert oder modellbasiert vorgeht. Auch wenn die Auslösung des strukturierten Dialoges nach QSKH-Richtlinie keine Differenzierung nach Fallzahl vorsieht, gehen wir im Folgenden davon aus, dass ein gleiches Ratenergebnis bei kleiner und großer Fallzahl (z. B. 2/4 und 50/100 bei einem Referenzbereich von $\leq 15\%$) als unterschiedliche Evidenz zu bewerten ist. Wir berücksichtigen daher in der Modellierung Fallzahl und Ratenergebnis der Leistungserbringer. Eine Ausnahme davon machen wir für Sentinel-Event-Indikatoren. Hier zählt allein die absolute Anzahl an beobachteten Sentinel-Events, da in diesem Fall die Auslösung des strukturierten Dialoges allein anhand der „Zähler“-Fälle (also nicht ratenbasiert) erfolgt.

Es gibt mehrere Möglichkeiten, die rechnerischen Ergebnisse der Leistungserbringer in das Modell (1) einzubeziehen. Sei $\hat{\pi}_{lb} = o_{lb}/n_{lb}$ das beobachtete Ratenergebnis (Zähler durch Nenner) von Leistungserbringer l in Bundesland bzw. Bewertungsstelle b . Wir verstehen $\hat{\pi}_{lb}$ als Schätzung einer zu Grunde liegenden Rate π_{lb} des Leistungserbringers anhand der beobachteten Daten. Sei t der Referenzwert des betrachteten Qualitätsindikators. Ohne Beschränkung der Allgemeinheit wird angenommen, dass der Qualitätsindikator so definiert ist, dass hohe Werte der Rate schlechter Versorgungsqualität entsprechen. Analog zu Spiegelhalter et al. (2012) bezeichnen wir mit $p(o_{lb}, n_{lb}, t)$ den p-Wert eines einseitigen exakten Binomialtests zur Null-Hypothese $H_0: \pi_{lb} \leq t$ („Der Leistungserbringer liegt im Referenzbereich“).⁶ $p(o_{lb}, n_{lb}, t)$ quantifiziert somit die Evidenz, dass ein Leistungserbringer den Referenzbereich einhält bzw. verletzt.⁷ Diesen

⁴ Engl. Best Linear Unbiased Predictors.

⁵ In diesem Anwendungsfall bedeutet „geringe Fallzahl“, dass nur wenige Leistungserbringer in einem Bundesland rechnerisch auffällig wurden.

⁶ Wir verwenden dabei auch die von Spiegelhalter et al. (2012) vorgeschlagene mid-P-Erweiterung.

⁷ So wie die Null-Hypothese definiert wurde, indiziert ein kleiner p-Wert eine Verletzung des Referenzbereiches.

p-Wert verwenden wir dabei in Modell (1) als Maß, wie ähnlich die Ergebnisse zweier Leistungserbringer sind. Genauer definieren wir die in Gleichung (1) einfließende Kovariable r_{lb} als

$$r_{lb} := p(o_{lb}, n_{lb}, t). \quad (2)$$

Eine anschauliche Interpretation dieser Definition von r_{lb} liefert ein nach Leistungserbringerfallzahl differenzierter Plot der Ergebnisse (vgl. Abbildung 5). Würde man nur die rechnerischen Ratenergebnisse ohne zusätzliche Berücksichtigung der Fallzahl verwenden um Leistungserbringer in vergleichbare Gruppen einzuteilen, würde man den Plot in (fallzahlunabhängige) horizontale Segmente einteilen (Abbildung 5 rechts). Dem entgegen entspricht die Verwendung der p-Werte einer Unterteilung des Plots in hyperbelförmige Segmente. Dies wird Abbildung 5 verdeutlicht. Die schwarzen Punkte zeigen die Ergebnisse hypothetischer Leistungserbringer. Gezeigt wird dabei nur der Ausschnitt rechnerisch auffälliger Leistungserbringer (oberhalb von 15 %). Leistungserbringerergebnisse in einem gleichen Farbsegment werden je nach Definition von r_{lb} als „ähnlich“ eingestuft.

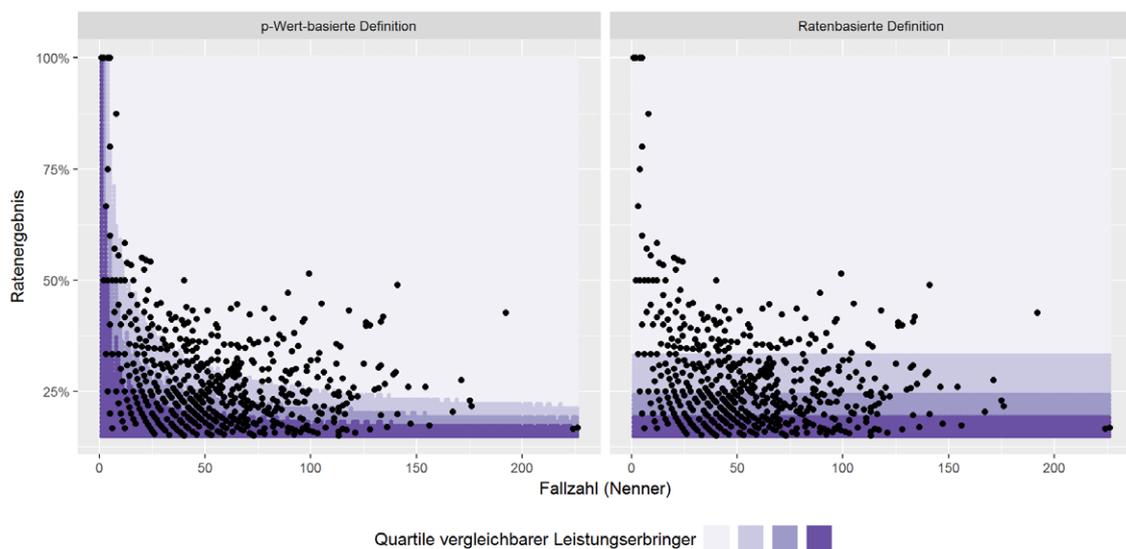


Abbildung 5: Illustration zu alternativen Definitionen von r_{lb} anhand einer nach Fallzahl differenzierten Darstellung der rechnerischen Ergebnisse. Segmente gleicher Farbe markieren Gruppen von Leistungserbringern deren Ergebnisse als vergleichbar eingestuft würden. Links die p-Wert-basierten Definition; rechts die ratenbasierte Definition.

Die obige Modellierung ist ohne Definition eines Referenzbereiches $t > 0$ für Sentinel-Event-Indikatoren nicht sinnvoll. Die Auslösung des Strukturierten Dialoges passiert hier ausschließlich durch das Zähler-Ergebnis $o_{lb} \geq 1$. Daher fließt in diesem Fall in die Modellierung auch nur das Zählerergebnis ein und wir definieren in (1) r_{lb} fallzahlunabhängig als

$$r_{lb} := o_{lb}.$$

Die Interpretation der Koeffizienten β_{Q_2} , β_{Q_3} und β_{Q_4} bleibt allerdings gleich: Sie quantifizieren, wie stark die Wahrscheinlichkeit $P(Y_{lb} = 1 | o_{lb}, n_{lb}, u_b)$ ansteigt, wenn ein Leistungserbringer im zweiten, dritten oder vierten Quartil der Ergebnisse r_{lb} liegt.

Alternativ zur Definition von r_{lb} in Gleichung (2) wurden zur Sensitivitätsanalyse Modelle mit unterschiedlichen Definitionen von r_{lb} berechnet. Konkret wurden dabei folgende Alternativ-Modelle mit Random-Effects berechnet:

1. Ausgangsmodell:

$$\text{logit}[P(Y_{lb} = 1 | o_{lb}, n_{lb}, u_b)] = \beta_0 + \beta_{Q_2} \cdot Q_2(r_{lb}) + \beta_{Q_3} \cdot Q_3(r_{lb}) + \beta_{Q_4} \cdot Q_4(r_{lb}) + u_b.$$

2. Variante 1: Modell mit stetiger Berücksichtigung des p-Wertes

$$\text{logit}[P(Y_{lb} = 1 | o_{lb}, n_{lb}, u_b)] = \beta_0 + \beta_1 \cdot \text{logit}[p(o_{lb}, n_{lb}, t)] + u_b.$$

3. Variante 2: Modell ohne Berücksichtigung der Fallzahl, mit stetiger Berücksichtigung des Ratenergebnisses

$$\text{logit}[P(Y_{lb} = 1 | o_{lb}, n_{lb}, u_b)] = \beta_0 + \beta_1 \cdot \text{logit}\left(\frac{o_{lb}}{n_{lb}}\right) + u_b.$$

4. Variante 3: Modell mit separatem dichotomen Fallzahl-Effekt ($n_{lb} \leq 20$ ja/ nein)⁸

$$\text{logit}[P(Y_{lb} = 1 | o_{lb}, n_{lb}, u_b)] = \beta_0 + \beta_1 \cdot \text{logit}\left(\frac{o_{lb}}{n_{lb}}\right) + \beta_2 \cdot I(n_{lb} \leq 20) + u_b.$$

5. Variante 4: Modell mit separatem stetigen Fallzahl-Effekt $f(n_{lb})$ in einem generalisierten gemischten additiven Modell (Wood 2006)

$$\text{logit}[P(Y_{lb} = 1 | o_{lb}, n_{lb}, u_b)] = \beta_0 + \beta_1 \cdot \text{logit}(\hat{\pi}_{lb}) + f(n_{lb}) + u_b.$$

Die logit-Transformation in Variante 1 und Variante 2 ist dabei ein gängiges Mittel die zwischen 0 und 1 liegenden p-Werte und Ratenergebnisse für die Regressionsmodellierung zu linearisieren [vgl. z. B. Hosmer et al. (2013)]. Die Modelle wurden für die beiden ratenbasierten Qualitätsindikatoren zu beiden Fragestellungen anhand des konditionalen AIC (*Aikake Information Criterion*) verglichen, einer Verallgemeinerung des AIC für Random-Effects-Modelle welches zur Modellselektion verwendet werden kann [vgl. Säfken et al. (2018)]. Somit wurde unter den betrachteten Modellen genau jenes Modell gewählt, welches am besten (nach AIC) zu den beobachteten Bewertungen im strukturierten Dialog passt. In fast allen Konstellationen hatte das Ausgangsmodell den niedrigsten AIC und wird daher favorisiert. Die AIC-Differenzen zwischen den Modellvarianten sind allerdings nicht in allen Konstellationen gleich ausgeprägt, was darauf schließen lässt, dass die Definition von r_{lb} je nach Qualitätsindikator und Fragestellung eine unterschiedliche Bedeutung haben kann. Die Tatsache, dass meist jenes Modell favorisiert wurde, welches den p-Wert-basierten Ansatz zur Definition von r_{lb} verwendet, kann man so interpretieren, dass einerseits viele Bewertungsstellen bei der Auslösung eines Stellungnahmeverfahrens explizit inferenzstatistische Methoden (etwa Konfidenzintervalle) berücksichtigen. Andererseits kann vermutet werden, dass Entscheidungen in den Bewertungsstellen implizit (also ohne konkrete Rechenvorschrift) die Fallzahl der Leistungserbringer in die Bewertung mit einbeziehen und dass dieses Verhalten wiederum durch den p-Wert-basierten Ansatz gut erklärt wird.

⁸ Wir bezeichnen mit $I(A)$ die Indikatorfunktion, die den Wert 1 annimmt, wenn die Bedingung A zutrifft und sonst den Wert 0 annimmt.

Von allen fünf Varianten wurden für die beide ratenbasierten Qualitätsindikatoren zu beiden Fragestellungen auch die Schätzungen der Varianzparameter τ^2 verglichen. Ähnlich wie beim AIC zeigt sich auch hier, dass je nach Fragestellung und Qualitätsindikator die Schätzungen von τ^2 unterschiedlich stark von der Definition von r_{lb} abhängen. Unter Berücksichtigung der Konfidenzintervalle ergeben sich jedoch keine relevanten Unterschiede zwischen den Modellvarianten. Die Schätzung von τ^2 hängt somit nicht maßgeblich von der genauen Definition von r_{lb} ab.

Für den Sentinel-Event-Indikator wurden ebenfalls alternative Modelld Definitionen mit und ohne Berücksichtigung der Fallzahl geprüft. Hier stellte sich heraus, dass eine Berücksichtigung der Fallzahl in der Definition von r_{lb} keine Verringerung des AIC in den Modellen zu Fragestellung 1 (Anforderung einer Stellungnahme) ergab. Die Modellierung von Fragestellung 2 (Bewertung als qualitativ auffällig) mit einem generalisierten additiven Modell mit stetigem Fallzahl-Effekt $f(n_{lb})$ analog zu Variante 3 ergab eine geringfügige Verringerung des AIC. In Fragestellung 2 ist also auch für den Sentinel-Event-Indikator nicht auszuschließen, dass die Fallzahl des Leistungserbringers in der abschließenden Qualitätsbewertung eine Rolle spielt. Um die Interpretation des Modelles zu erleichtern, wurde allerdings davon abgesehen diesen nichtlinearen Effekt durch die Fallzahl des Leistungserbringers zu modellieren.

4.4.3 Modelle ohne Berücksichtigung der Leistungserbringer-Ergebnisse

Eine weitere wichtige Sensitivitätsanalyse besteht im Vergleich von Modellen mit und ohne Berücksichtigung der Leistungserbringer-Ergebnisse. Im Folgenden stellen wir pro Indikator und Fragestellung jeweils vier Modelle gegenüber:

- M_1 : Random-Effect Modell ohne Berücksichtigung der Leistungserbringer (LE)-Ergebnisse

$$\text{logit}[P(Y_{lb} = 1|u_b)] = \beta_0 + u_b$$

- M_2 : Random-Effect Modell mit Berücksichtigung der LE-Ergebnisse

$$\text{logit}[P(Y_{lb} = 1|r_{lb}, u_b)] = \beta_0 + \beta_1 \cdot r_{lb} + u_b$$

- M_3 Berücksichtigung der LE-Ergebnisse ohne Random-Effects

$$\text{logit}[P(Y_{lb} = 1|r_{lb})] = \beta_0 + \beta_1 \cdot r_{lb}$$

- M_4 Triviales Modell ohne Random-Effects und ohne LE-Ergebnisse

$$\text{logit}[P(Y_{lb} = 1)] = \beta_0$$

mit $u_b \sim \mathcal{N}(0, \tau^2)$. Zum Vergleich der Modelle betrachten wir erneut die AIC-Werte. In der Situation von Random-Effects-Modellen ist zu beachten, dass man den konditionalen AIC betrachtet [vgl. Säfken et al. (2018)]. Die Ergebnisse finden sich in Tabelle 7 und Tabelle 8: (Konditionaler) AIC der Modellvarianten M_1 - M_4 zu Frage 2. Es überrascht nicht, dass das flexibelste Modell M_2 für alle drei Qualitätsindikatoren und beide Fragestellungen die Daten am besten erklärt.

Sehr überraschend sind allerdings die Ergebnisse der Parameter-Schätzungen in der Modellvariante M_1 ohne Berücksichtigung der rechnerischen Leistungserbringerergebnisse. Die Parameter-Schätzungen zu Fragestellung 1 und 2 finden sich in Tabelle 9 und Tabelle 10. Vergleicht man

dies mit den Ergebnissen aus Tabelle 5 und Tabelle 6, so fällt auf, dass die Heterogenität zwischen den Bewertungsstellen *geringer* ausfällt, wenn man die Ergebnisse der Leistungserbringer nicht berücksichtigt.

Tabelle 7: (Konditionaler) AIC der Modellvarianten M_1 - M_4 zu Frage 1

	AIC mit Random-Effects	AIC ohne Random-Effects
HEP: QI 54003		
ohne LE-Ergebnisse (M_1 und M_4)	469	757
mit LE-Ergebnissen (M_2 und M_3)	238	620
DEK: QI 52010		
ohne LE-Ergebnisse (M_1 und M_4)	120	406
mit LE-Ergebnissen (M_2 und M_3)	80	382
09/1: QI 54139		
ohne LE-Ergebnisse (M_1 und M_4)	267	600
mit LE-Ergebnissen (M_2 und M_3)	179	581

Tabelle 8: (Konditionaler) AIC der Modellvarianten M_1 - M_4 zu Frage 2

	AIC mit Random-Effects	AIC ohne Random-Effects
HEP: QI 54003		
ohne LE-Ergebnisse (M_1 und M_4)	439	473
mit LE-Ergebnissen (M_2 und M_3)	389	421
DEK: QI 52010		
ohne LE-Ergebnisse (M_1 und M_4)	354	443
mit LE-Ergebnissen (M_2 und M_3)	324	414
09/1: QI 54139		
ohne LE-Ergebnisse (M_1 und M_4)	188	203
mit LE-Ergebnissen (M_2 und M_3)	185	202

Tabelle 9: Parameterschätzungen Modellvariante M_1 der drei Modelle zu Fragestellung 1: Heterogenität bezüglich des Anforderns einer Stellungnahme

Parameter	Schätzung	(Median-)Odds-Ratio	Konfidenzintervall (Median-)Odds-Ratio	
			2,5 %	97,5 %
HEP: QI 54003				
β_0	2,2	-	-	-
τ	2,7	13	4,7	140
DEK: QI 52010				
β_0	10,3	-	-	-
τ	54	170	8,7	9,5 · 10 ⁹
09/1: QI 54139				
β_0	1,9	-	-	-
τ	3,4	26	7,1	510

Tabelle 10: Parameterschätzungen Modellvariante M_1 der drei Modelle zu Fragestellung 2: Heterogenität bezüglich der Bewertung als qualitativ auffällig

Parameter	Schätzung	(Median-)Odds-Ratio	Konfidenzintervall (Median-)Odds-Ratio	
			2,5 %	97,5 %
HEP: QI 54003				
β_0	-0,58	-	-	-
τ	0,92	2,4	1,6	4,7
DEK: QI 52010				
β_0	-1,23	-	-	-
τ	1,48	4,1	2,5	10
09/1: QI 54139				
β_0	-1,51	-	-	-
τ	0,95	2,5	1	8

4.4.4 Implementation

Alle Berechnungen wurden in der Statistiksoftware R Version 3.5.0 durchgeführt. Zur Berechnung der logistischen Regressionsmodelle mit Random-Effects wurde das Paket lme4 (Bates et al. 2015) in der Version 1.1-17 verwendet. Die Integralapproximation der log-Likelihood erfolgte durch eine adaptive Gauss-Hermite-Approximation mit jeweils 25 Stützpunkten pro Achse. Zur Optimierung der log-Likelihood wurde der BOBYQA-Algorithmus aus dem minqa-Paket, Version

1.2.4, genutzt. Die Konfidenzintervalle der Parameter-Schätzungen wurden durch die Likelihood-Profile-Methode berechnet, implementiert durch das lme4-Paket. Zur Berechnung der p-Werte für die einseitigen Binomialtests in der Definition von r_{lb} in Gleichung (2) wurde die Implementation des IQTIG aus dem IQTIGpvcI-Paket in der Version 1.0.0 verwendet (IQTIG 2018). Die Modellvergleiche anhand des konditionalen AIC wurden mithilfe des cAIC4-Paketes Version 0.4 berechnet Säfken et al. (2018).

4.5 Deskriptive Auswertungen zu den Ergebnissen des Strukturierten Dialogs zum Erfassungsjahr 2016 für ausgewählte Indikatoren

Im Folgenden werden als Hintergrund deskriptive Auswertungen zu den Ergebnissen des strukturierten Dialoges der drei ausgewählten Qualitätsindikatoren dargestellt, welche die modellbasierten Auswertungen aus Abschnitt 4.2 ergänzen. Ziel dieser Auswertungen ist es Transparenz über die Datengrundlage der vorgestellten modellbasierten Analysen zu schaffen. Dies soll eine Einordnung der Heterogenität in den Bewertungen des strukturierten Dialoges ermöglichen, die anders als die Analysen aus Abschnitt 4.2 keinerlei zusätzliche Modellannahmen erfordert. Insbesondere die nach Bewertungsstelle, Fallzahl und Indikatorergebnis stratifizierte Darstellungen zeigen, dass Bewertungen im strukturierten Dialog sowohl abhängig vom Indikatorergebnis als auch abhängig von der Fallzahl getroffen werden.

4.5.1 HEP: QI 54003 Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur

Anteil an Stellungnahmen an allen rechnerisch auffälligen Standorte

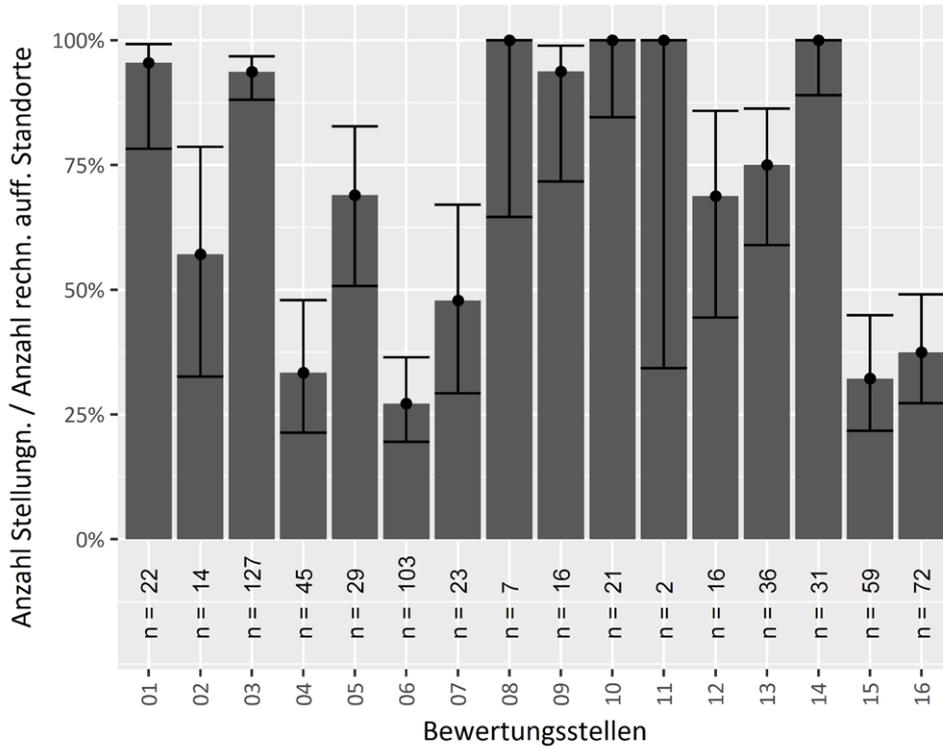


Abbildung 6: Qualitätsindikator HEP: 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“. Anfordern einer Stellungnahme nach rechnerischer Auffälligkeit

Anteil an qualitativ auffälligen Bewertungen an allen Stellungnahmen

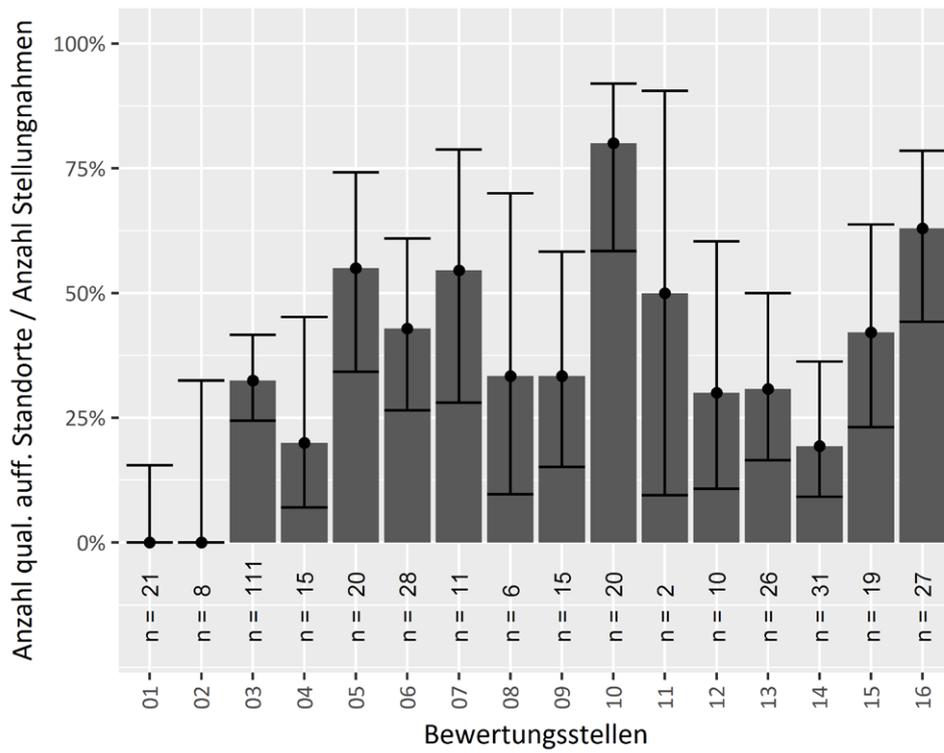


Abbildung 7: Qualitätsindikator HEP: 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“. Bewertungen nach Einleitung des Stellungnahmeverfahrens

Darstellung der Standortergebnisse nach Bewertungsstelle und Fallzahl

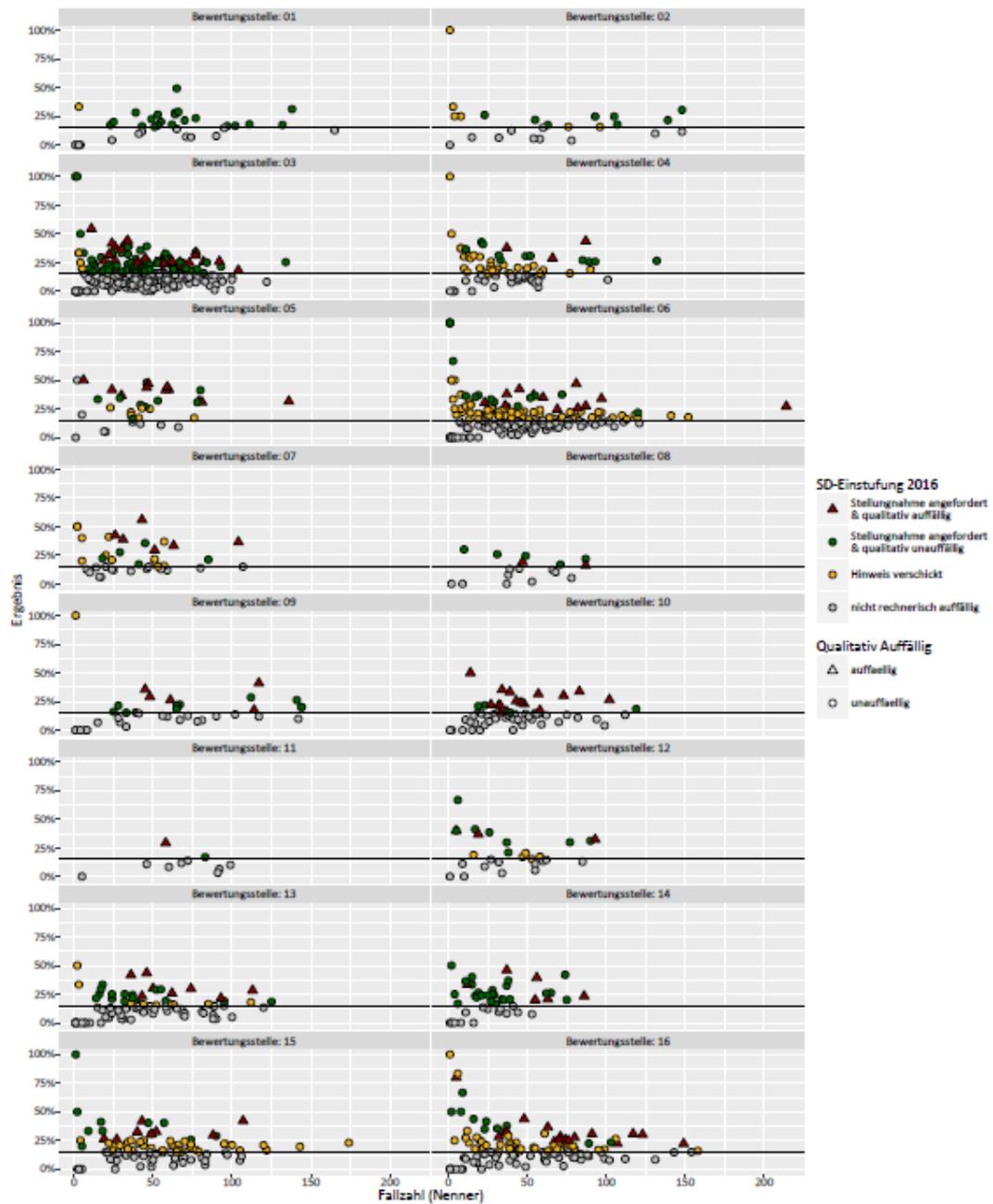


Abbildung 8: Ergebnisse des strukturierten Dialoges zum Qualitätsindikator HEP: 54003 „Präoperative Verweildauer bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur“ für das Erfassungsjahr 2016 differenziert nach Fallzahl und Bewertungsstelle. Diese Daten sind die Grundlage für die modellbasierten Auswertungen in Abschnitt 2.2.

4.5.2 09/1: QI 54139 Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen

Anteil an Stellungnahmen an allen rechnerisch auffälligen Standorten

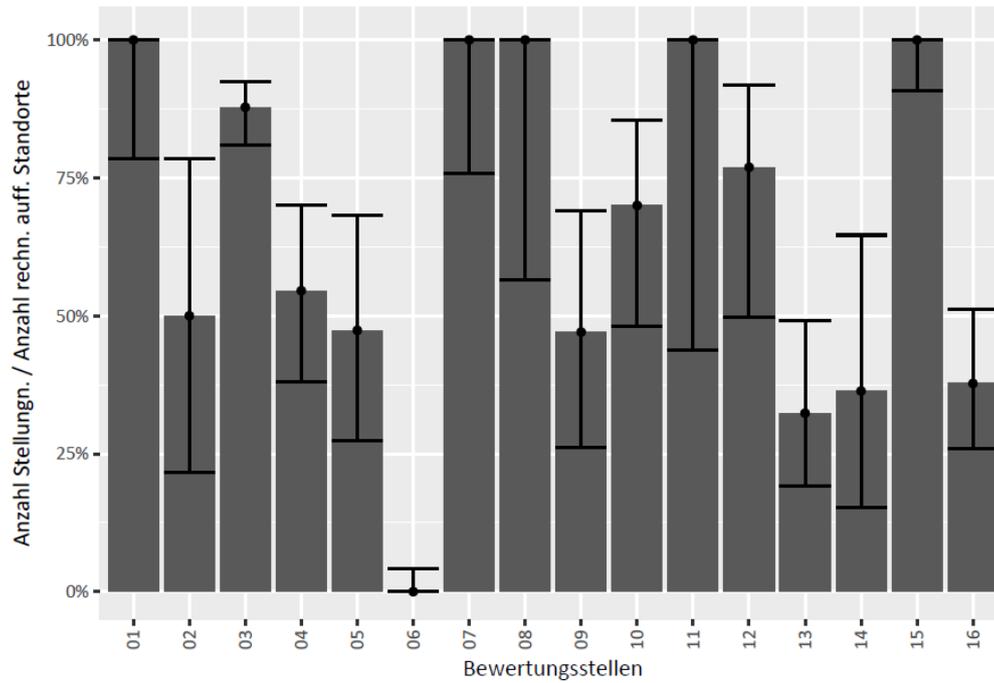


Abbildung 9: Qualitätsindikator 09/1: 54134 „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“. Anfordern einer Stellungnahme nach rechnerischer Auffälligkeit

Anteil an qualitativ auffälligen Bewertungen an allen Stellungnahmen

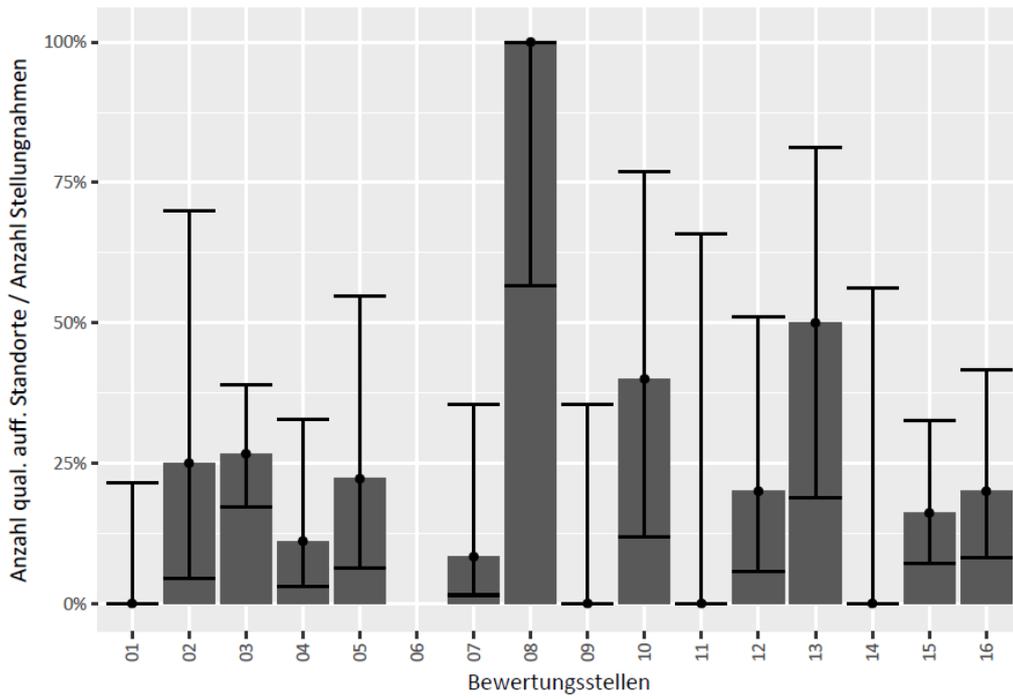


Abbildung 10: Qualitätsindikator 09/1: 54134 „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“. Bewertungen nach Einleitung des Stellungnahmeverfahrens

Darstellung der Standortergebnisse nach Bewertungsstelle und Fallzahl

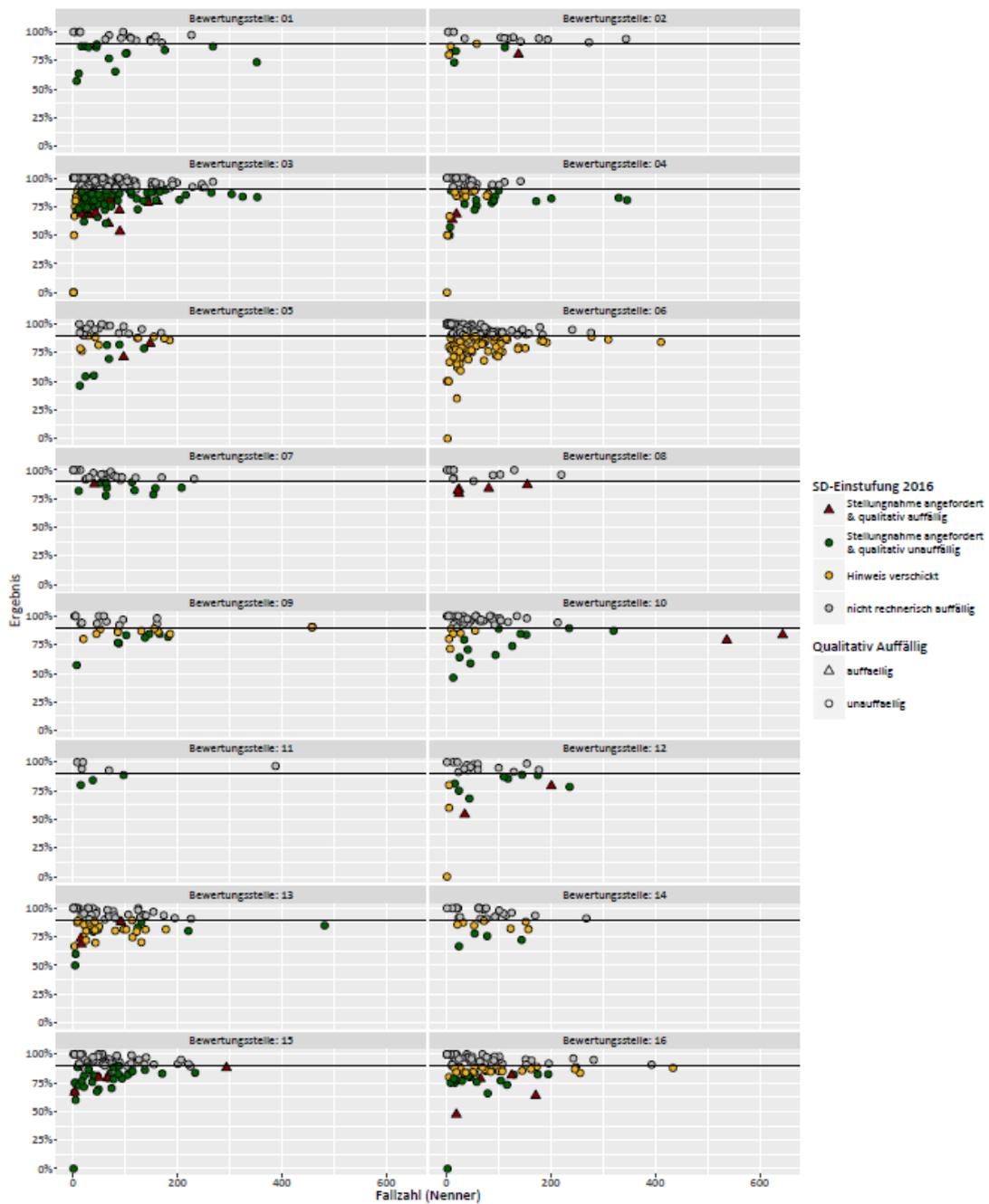


Abbildung 11: Ergebnisse des Strukturierten Dialogs zum Qualitätsindikator 09/1: 54134 „Leitlinienkonforme Indikation bei bradykarden Herzrhythmusstörungen“ für das Erfassungsjahr 2016 differenziert nach Fallzahl und Bewertungsstelle. Diese Daten sind die Grundlage für die modellbasierten Auswertungen in Abschnitt 2.2.

4.5.3 DEK: QI 52010 Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4

Anteil an Stellungnahmen an allen rechnerisch auffälligen Standorten

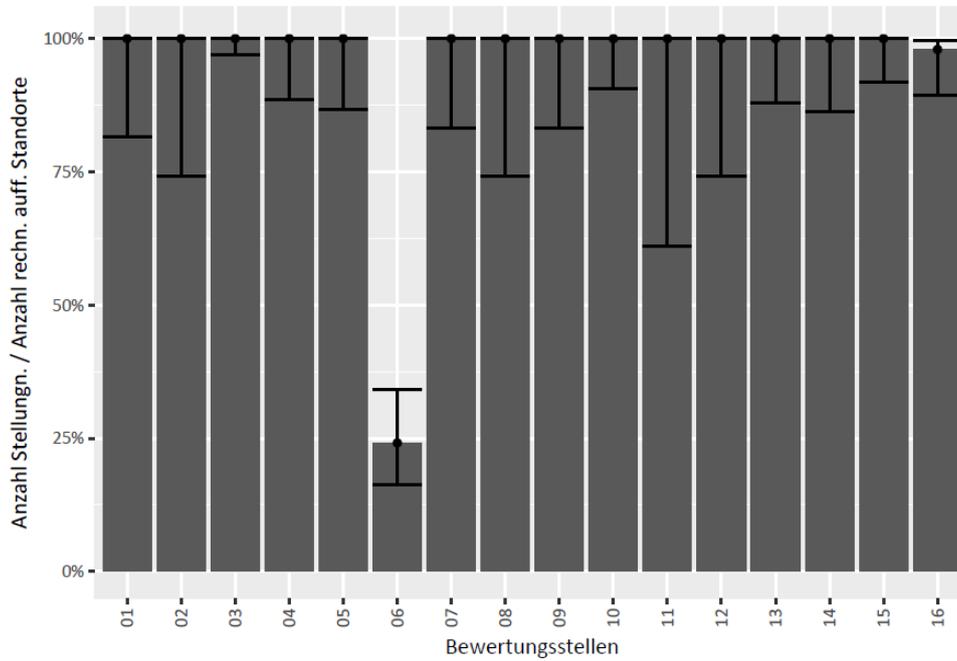


Abbildung 12: Qualitätsindikator DEK: 52010 „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“. Anfordern einer Stellungnahme nach rechnerischer Auffälligkeit

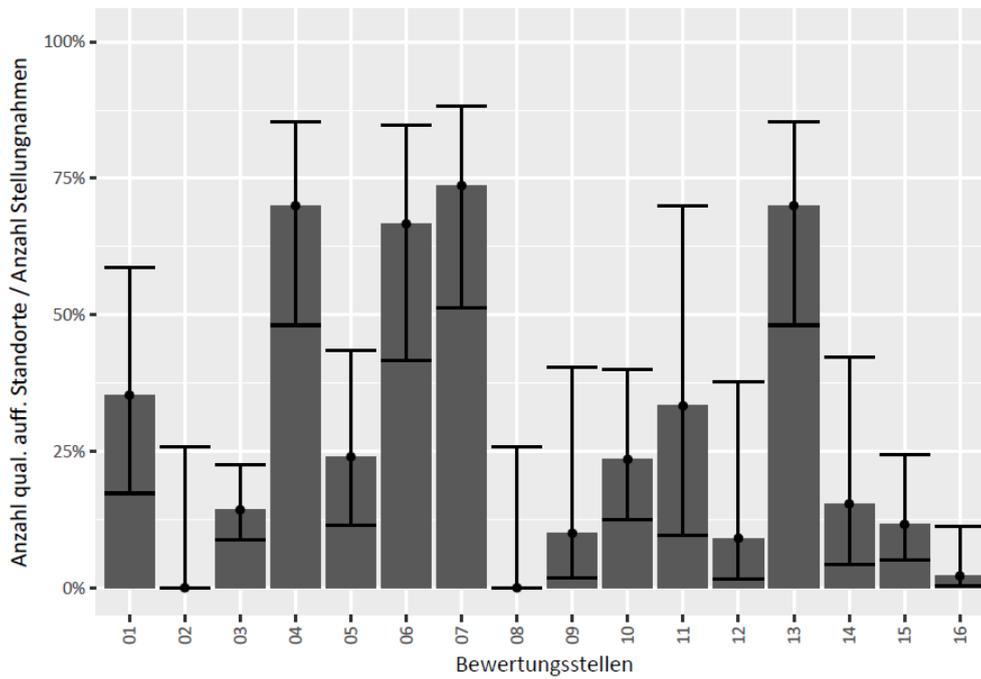
Anteil an qualitativ auffälligen Bewertungen an allen Stellungnahmen

Abbildung 13: Qualitätsindikator DEK: 52010 „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“. Bewertungen nach Einleitung des Stellungnahmeverfahrens

Darstellung der Standortergebnisse nach Bewertungsstelle und Fallzahl



Abbildung 14: Ergebnisse des Strukturierten Dialogs zum Qualitätsindikator DEK: 52010 „Alle Patienten mit mindestens einem stationär erworbenen Dekubitalulcus Grad/Kategorie 4“ für das Erfassungsjahr 2016 differenziert nach Fallzahl und Bewertungsstelle. Diese Daten sind die Grundlage für die modellbasierten Auswertungen in Abschnitt 2.2.

5 Beispielfälle für die Qualitätsbewertung im Stellungnahmeverfahren

In seiner Stellungnahme kann ein Leistungserbringer Gründe anführen, die aus seiner Sicht das Verfehlen des Referenzbereichs erklären. Diese können sich auf einen oder mehrere Einflussfaktoren auf das Indikatorergebnis beziehen, die in der Berechnungsvorschrift des Indikators bisher nicht berücksichtigt sind. Im Folgenden werden fiktive Beispielfälle zum Stellungnahmeverfahren dargestellt, die veranschaulichen, wie diese Einflussfaktoren in der fachlichen Bewertung klassifiziert werden.

5.1 Fachlich-inhaltliche Einflussfaktoren

Beispiel 1: Kontraindikationen

Es wird ein Qualitätsindikator zur Thromboseprophylaxe genutzt, der abbilden soll, ob diese Behandlung in allen Fällen durchgeführt wurde, in denen sie angezeigt ist. In die Rechenregel des Indikators gehen Begleiterkrankungen der Patientinnen und Patienten nicht ein.

In einer Stellungnahme wird der Verzicht auf eine Thromboseprophylaxe bei einem Teil der Behandlungsfälle damit begründet, dass einige Patientinnen und Patienten ein erhöhtes Blutungsrisiko aufwiesen (z. B. wegen einer Thrombozytopenie).

Analyse:

Das Qualitätsziel des Indikators bezieht sich nicht auf Patientinnen und Patienten, die eine Kontraindikation zur Gabe von gerinnungshemmenden Medikamenten aufweisen. Die Verantwortung dafür, dass diese Kontraindikation gegeben ist, liegt auch nicht beim Leistungserbringer. Idealerweise wären Fälle mit erhöhtem Blutungsrisiko bei der Indikatorberechnung aus der Patientengrundgesamtheit des Indikators ausgeschlossen worden, da sie nicht zur Zielpopulation gehören. Der Einflussfaktor „Kontraindikation zur Thromboseprophylaxe“ kann bei der partiellen Nachberechnung in der fachlichen Bewertung berücksichtigt werden, indem geklärt wird, bei welchen der Behandlungsfälle dieser Einflussfaktor vorlag.

Vorgehen:

Diejenigen Fälle, bei denen die Fachkommission das Vorliegen einer „Kontraindikation zur Thromboseprophylaxe“ feststellt, werden bei der partiellen Nachberechnung aus der Grundgesamtheit ausgeschlossen.

Beispiel 2: Notfallbehandlung

Ein Qualitätsindikator beschreibt, ob bei einer diagnostischen Darmspiegelung (Koloskopie) der komplette Dickdarm beurteilt wurde (Erreichen des terminalen Ileums). In der Rechenregel des Indikators ist der Grund für die Untersuchung nicht berücksichtigt.

In einer Stellungnahme wird der Verzicht auf eine komplette Untersuchung des Dickdarms vom Leistungserbringer damit begründet, dass es sich bei einigen Patientinnen und Patienten um eine Notfalluntersuchung wegen einer akuten Blutung gehandelt habe.

Analyse:

Der Indikator ist nur auf elektive Untersuchungen anzuwenden. In akuten Notfallsituationen ist die Untersuchung des kompletten Dickdarms (in der Regel) kein Qualitätsmerkmal. Die Berücksichtigung bei der fachlichen Bewertung erfolgt dadurch, dass geklärt wird, bei welchen Fällen des Leistungserbringers der Einflussfaktor „akute Blutung als Indikation zur umgehenden Darmspiegelung“ Nachberechnung vorlag.

Vorgehen:

Diejenigen Fälle, bei denen LAG bzw. Bundesstelle und Fachkommission das Vorliegen einer „Notfalluntersuchung wegen akuter Blutung“ feststellen, werden bei der partiellen Nachberechnung aus der Grundgesamtheit ausgeschlossen.

Beispiel 3: externe Verantwortlichkeit

Ein Qualitätsindikator misst die In-Hospital-Sterblichkeit von Patientinnen und Patienten mit ambulant aufgetretenem akutem Herzinfarkt. Das Risikoadjustierungsmodell des Indikators berücksichtigt u. a. die wesentlichen Vorerkrankungen der Patientinnen und Patienten.

In einer Stellungnahme wird angeführt, dass aufgrund der regionalen Infrastruktur die Transportzeiten des Rettungsdienstes im Durchschnitt deutlich höher sind als bei anderen Einrichtungen. Als Beleg werden Studien zur mittleren Transportzeit angeführt und den eigenen, vom Leistungserbringer erfassten Daten gegenübergestellt.

Analyse:

Der Einflussfaktor „erhöhte Transportzeit des Rettungsdienstes“ ist vom Leistungserbringer nicht zu verantworten. Das Ausmaß, in dem die erhöhte Transportzeit zu schlechteren Ergebnissen (auch im Vergleich zu anderen Leistungserbringern) geführt hat, lässt sich nicht eindeutig angeben, wenn die Transportzeit nicht bei allen Leistungserbringern dokumentiert und für die Qualitätssicherung erfasst wird. Bei der fachlichen Bewertung wird dieser Einflussfaktor daher als „nicht für die partielle Nachberechnung geeignet“ klassifiziert.

Vorgehen:

Die „erhöhte Transportzeit des Rettungsdienstes“ wird nur im Rahmen der heuristischen Beurteilung durch die Fachkommission berücksichtigt. Kommt diese Beurteilung zum Schluss, dass bei den Fällen des betreffenden Leistungserbringers ein deutlich erhöhtes Risikoprofil in Bezug auf diesen Faktor im Vergleich zu den anderen Leistungserbringern vorliegt, kann dies als besondere Konstellation anerkannt werden.

Beispiel 4: bekannter Risikofaktor

Ein Qualitätsindikator beschreibt das Auftreten von Behandlungskomplikationen nach Implantation eines künstlichen Gelenks. Das Risikoadjustierungsmodell berücksichtigt die wichtigen Begleiterkrankungen, die zu einer erhöhten Häufigkeit von Komplikationen führen.

In seiner Stellungnahme weist ein Leistungserbringer auf ein erhöhtes Risikoprofil der von ihm behandelten Patientinnen und Patienten hin. In der Stellungnahme und den begleitenden Einzelfallbeschreibungen werden prolongierte, komplizierte Behandlungsverläufe geltend gemacht. Die darin genannten Risikofaktoren sind im Risikoadjustierungsmodell des Indikators berücksichtigt.

Analyse:

Es werden nur Einflussfaktoren genannt, die entweder vom Leistungserbringer zu verantworten sind oder bereits in der Berechnungsvorschrift des Indikators berücksichtigt sind. Der Einflussfaktor „erhöhtes Risikoprofil“ ist daher für eine partielle Nachberechnung nicht relevant. Im Rahmen der heuristischen Beurteilung durch die Fachkommission soll das erhöhte Risikoprofil ebenfalls nicht anerkannt werden, da im Vergleich zu den anderen Leistungserbringern nach Risikoadjustierung keine besondere Konstellation vorliegt.

Vorgehen:

Der Einflussfaktor „erhöhtes Risikoprofil“ wird nicht als Erklärung für das Verfehlen des Referenzbereichs anerkannt.

Beispiel 5: strukturelle Sondersituation, planbar

Ein Qualitätsindikator beschreibt Ergebnisse (z. B. perioperative Infektionen) oder durchzuführende Prozesse (z. B. postoperative Überwachung der Vitalfunktionen) in einem operativen Fachgebiet.

Die Stellungnahme eines Leistungserbringers verweist auf die Renovierung seines OP-Bereiches. Die operative Behandlung der Patientinnen und Patienten, die von dem Qualitätsindikator adressiert wird, fand vorübergehend in einem während der Renovierungsphase genutzten OP-Container statt.

Analyse:

Der Sachverhalt war planbar: Die Entscheidung, während der Renovierungsphase Patienten zu behandeln, impliziert die Verantwortlichkeit des Krankenhauses dafür, weiterhin eine ausreichende Qualität der Behandlung zu gewährleisten.

Vorgehen:

Der Einflussfaktor „Renovierung“ kann nicht als besondere Konstellation, die das Verfehlen des Referenzbereiches erklärt, geltend gemacht werden.

Beispiel 6: unvorhergesehene strukturelle Situation

Ein Qualitätsindikator beschreibt, ob eine gemäß Leitlinie geforderte radiologische CT-Untersuchung (z. B. zur Diagnostik einer Lungenembolie) durchgeführt wird.

In einer Stellungnahme wird geltend gemacht, dass die CT-Untersuchung bei einer relevanten Anzahl an Patientinnen und Patienten nicht durchgeführt werden konnte, da das Gerät unvorhersehbar ausgefallen war (z. B. durch einen technischen Defekt). Kurzfristige Versuche zur Behebung des Problems scheiterten.

Analyse:

Das Vorhalten eines zweiten Geräts kann nicht verlangt werden, ein Aufschub der Behandlung der Patientinnen und Patienten war wegen der Dringlichkeit der Versorgung nicht möglich. Für die Behandlungsfälle, die während des Ausfalls behandelt wurden, kann keine Qualitätsaussage getroffen werden, wenn die Situation nicht vom Leistungserbringer beeinflussbar war.

Vorgehen:

Der Einflussfaktor „unerwarteter Geräteausfall“ wird bei der partiellen Nachberechnung durch Ausschluss der betreffenden Behandlungsfälle berücksichtigt.

Beispiel 7: Patientenpräferenz

Ein Qualitätsindikator bildet ab, ob bei Patientinnen und Patienten mit Blutung aus dem Magen-Darm-Trakt die empfohlenen Untersuchungen zur Ermittlung der Blutungsquelle (endoskopische Diagnostik) durchgeführt werden.

In einer Stellungnahme wird der Verzicht auf die endoskopische Untersuchung bei einigen der Behandlungsfälle damit begründet, dass die Patientinnen und Patienten die Untersuchung abgelehnt haben. Es wird vom Leistungserbringer dargestellt, dass eine adäquate Aufklärung über Folgen der fehlenden Diagnostik erfolgt ist.

Analyse:

Die Verantwortung für den Einflussfaktor „Patientenpräferenz“ liegt nicht beim Leistungserbringer, sofern eine angemessene Aufklärung der Patientinnen und Patienten nachgewiesen ist. Im Rahmen der fachlichen Bewertung muss geklärt werden, für welche Behandlungsfälle eine entsprechende Aufklärung nachvollziehbar belegt ist.

Vorgehen:

Die Behandlungsfälle, bei denen LAG bzw. Bundesstelle und Fachkommission eine „Ablehnung der Behandlung durch die Patientinnen/Patienten trotz angemessener Aufklärung“ feststellen, werden bei der partiellen Nachberechnung berücksichtigt, indem sie aus der Grundgesamtheit des Indikators ausgeschlossen werden.

5.2 Datenfehler

Beispiel 8: unentdeckte Softwarefehler

Ein Qualitätsindikator misst die In-Hospital-Sterblichkeit bei einer Herzklappenoperation. Das Risikoadjustierungsmodell berücksichtigt u. a. die Nierenfunktion, gemessen an der Kreatinin-Konzentration im Blutserum.

In einer Stellungnahme wird vorgebracht, dass durch einen Softwarefehler beim Export der QS-Daten die Kreatinin-Konzentrationen in einer falschen Einheit übermittelt wurden, ohne dass dies in der Benutzeroberfläche der QS-Software sichtbar war.

Analyse:

Wenn der Softwarefehler vom Leistungserbringer nicht bemerkt werden konnte, handelt es sich bei dem Datenfehler um einen nicht vom Leistungserbringer zu verantwortenden Faktor. Die Auswirkung auf das Indikatorergebnis lässt sich jedoch nicht durch einfachen Ausschluss der Behandlungsfälle im Rahmen der partiellen Nachberechnung angemessen beschreiben.

Vorgehen:

Ob die „Fehlmessung des Risikoprofils“ der betroffenen Behandlungsfälle so bedeutend war, dass sie das Verfehlen des Referenzbereichs erklärt, wird im heuristischen Beurteilungsschritt von LAG bzw. Bundesstelle und Fachkommission untersucht.

Beispiel 9: Fehldokumentation

Ein Qualitätsindikator beschreibt das Auftreten von Komplikationen nach einem operativen Eingriff.

In seiner Stellungnahme macht der Leistungserbringer geltend, dass bei einem Teil der Behandlungsfälle tatsächlich keine Komplikationen vorgelegen hatten, sondern dass es sich um fehlerhafte Eingaben in der QS-Dokumentationssoftware handelt. Als Belege werden Auszüge der Behandlungsdokumentation (Patientenakte) beigefügt.

Analyse:

Soll das Ausmaß des vom Leistungserbringer zu verantwortenden Dokumentationsfehlers abgeschätzt werden, können bei einer Nachberechnung die betreffenden Behandlungsfälle als Fälle ohne Komplikation gezählt werden, z. B. durch Ausschluss aus dem Zähler des Indikators. Da die fehlerhaften Angaben in den QS-Daten jedoch durch Dokumentationsfehler des Leistungserbringers bedingt sind, handelt es sich um vom Leistungserbringer zu verantwortende Datenfehler. Eine Berücksichtigung im informativen Indikatorwert, der nur um Faktoren bereinigt ist, die nicht vom Leistungserbringer zu verantworten sind, ist nicht sachgemäß.

Vorgehen:

Bei der Ermittlung des informativen Indikatorwerts im Rahmen der partiellen Nachberechnung werden die Angaben zu den Komplikationen so berücksichtigt, wie sie ursprünglich dokumentiert waren. Die Dokumentationsfehler können nicht als besondere Konstellation, die den Hinweis auf ein Qualitätsdefizit entkräftet, geltend gemacht werden.

6 Textvorschläge Einholung Stellungnahmen

Für eine weitere Standardisierung im Vorgehen werden hier Textvorschläge für die Einholung einer Stellungnahme gemacht. Das IQTIG empfiehlt auch, aus Transparenzgründen die Methodik des Stellungnahmeverfahrens im Internet zu veröffentlichen und einen Link zu der Webseite in das Schreiben zur Einholung einer Stellungnahme einzufügen. Eine solche Transparenz ist einerseits aus wissenschaftlicher Sicht geboten und kann außerdem die Akzeptanz des Verfahrens steigern.

Wie aus dem Protokoll des Workshops mit den LAG- und LQS-Vertreterinnen und -Vertretern hervorgeht, nutzen einige Fachgruppen selbst entwickelte Fragebögen für die Durchführung des Stellungnahmeverfahrens oder des Strukturierten Dialogs (siehe Kapitel 1 in diesem Anhang). Da zum Zeitpunkt der Berichtslegung kein solcher Fragebogen vorliegt, kann nicht eingeschätzt werden, ob diese für die Prüfung auf das Vorliegen von nicht vom Leistungserbringer zu vertretene Einflussfaktoren geeignet sind.

Sehr geehrte Damen und Herren,

auf Basis der von Ihnen übermittelten Daten wurde für Ihre Einrichtung das Ergebnis für den Qualitätsindikator [NN] berechnet. Das Ergebnis Ihrer Einrichtung liegt statistisch signifikant außerhalb des Referenzbereichs. Dies bedeutet, dass der Indikator auf ein Qualitätsdefizit in der Versorgung hinweist. Sie erhalten hiermit die Möglichkeit, zu diesem Indikatorergebnis Stellung zu nehmen. Wenn Sie Gründe anführen möchten, die das Ergebnis erklären und nicht von Ihrer Einrichtung zu verantworten waren, können Sie diese innerhalb von fünf Wochen nach Zugang dieses Schreibens durch eine qualifizierte Stellungnahme nach untenstehenden Kriterien geltend machen. Ihre Stellungnahme wird dann einer Fachkommission vorgelegt, die bewerten wird, ob die von Ihnen angeführten Gründe anerkannt und das Indikatorergebnis mit „kein Hinweis auf Qualitätsdefizit“ eingestuft werden kann. Andernfalls wird das Indikatorergebnis als „Qualitätsdefizit“ bewertet. Erreicht uns innerhalb der Frist keine den untenstehenden Kriterien entsprechende Stellungnahme, wird das Ergebnis Ihrer Einrichtung als „Qualitätsdefizit“ bewertet.

Neues Stellungnahmeverfahren

Bitte beachten Sie, dass sich das Vorgehen bei der Bewertung und im Stellungnahmeverfahren für diesen Qualitätsindikator aus der DeQS-RL von dem Ihnen bisher bekannten Vorgehen gemäß der QSKH-RL unterscheidet. Die Methodik für die Bewertung von Stellungnahmen und die abschließende Qualitätsbewertung können Sie hier nachlesen: [[Link zur Veröffentlichung der Methodik einfügen](#)]

Formale Kriterien

Bitte beachten Sie beim Verfassen Ihrer Stellungnahme folgende Hinweise:

- Für jedes auffällige Indikatorergebnis ist eine separate Stellungnahme abzugeben.

- Die Stellungnahme ist einschließlich eventuell notwendiger Anhänge in Schriftform zu erstellen und einzureichen.
- Wenn Sie Belege haben, die Ihre Argumentation unterstützen, können Sie diese in angemessenem Umfang einreichen. Dazu zählen beispielsweise Einzelfalldokumente wie Arztbriefe, Epikrisen, Verlegungsberichte, Operationsberichte und Befunde. Belege können jedoch nur als Ergänzung des Erläuterungstextes berücksichtigt werden und ersetzen diesen nicht. Belege ohne erläuternden, übergreifenden Text (s. u.) werden nicht anerkannt.
- Die fachlich verantwortliche Instanz Ihrer Einrichtung ist über den indikatorbasierten Hinweis auf ein Qualitätsdefizit in Ihrer Einrichtung zu informieren und hat die angefertigte Stellungnahme zu autorisieren. Bitte bestätigen Sie in Ihrer Stellungnahme, dass dies geschehen ist.
- Bitte wahren Sie in Ihrer Stellungnahme die Anonymität sowohl Ihrer Patientinnen und Patienten (personenidentifizierende Daten) und Ihrer Einrichtung (einrichtungsidifizierende Daten) als auch ggf. weiterer Beteiligter. Liegt eine Verletzung des Datenschutzes vor, wird die Stellungnahme umgehend gelöscht und gilt als nicht abgegeben.
- Bitte beantworten Sie am Anfang Ihrer Stellungnahme folgende Frage: Sind Ihnen Fehler in den von Ihnen übermittelten Daten bekannt? Wenn ja, nennen Sie bitte den/die Fehler sowie – sofern zutreffend – die Gründe, warum der/die Fehler in der Datengrundlage nicht von Ihrer Einrichtung zu verantworten war/en.

Inhaltliche Kriterien (für jeden Qualitätsindikator)

Neben den obigen formalen Kriterien möchten wir Sie bitten, folgende inhaltliche Anforderung an Ihre Stellungnahme zu berücksichtigen. Wenn Sie eine Stellungnahme abgeben, soll diese eine Analyse enthalten, die erklärt, weshalb das Qualitätsindikatorergebnis vom Referenzbereich abweicht, obwohl Ihrer Meinung nach die Versorgungsqualität in Ihrer Einrichtung gut ist. In Ihrer Stellungnahme können zum einen Gründe angegeben werden, die sich nur auf einzelne oder weniger Behandlungsfälle beziehen (z. B. dass ein Notfall ein Abweichen vom üblichen Vorgehen erforderlich machte), und zum anderen Gründe, die alle Behandlungsfälle ihrer Einrichtung für diesen Indikator betreffen (z. B. dass wegen einer Spezialisierung Ihrer Einrichtung auf bestimmte Behandlungen die Patientinnen und Patienten ein besonderes Risikoprofil haben). Wenn möglich, sollte Ihre Stellungnahme jedoch auf eine behandlungsfallübergreifende Analyse fokussieren, um systemische Ursachen und ggf. systemische Verbesserungsansätze darzustellen.

- **Bei einem quantitativ auffälligen Ergebnis in einem Prozessindikator:**⁹
Bitte schildern Sie in Ihrer Stellungnahme, warum die erforderlichen Prozessstandards (z. B. Durchführung, Anwendung oder Nicht-Durchführung einer Maßnahme, zeitliche Anforderungen, Dosierungen) nicht in dem Maße eingehalten wurden, wie durch den Referenzbereich des Qualitätsindikators gefordert wird, und warum dies dennoch nicht als Qualitätsproblem einzustufen ist.
- **Bei einem quantitativ auffälligen Ergebnis in einem risikoadjustierten Prozessindikator:**

⁹ Bei der Erstellung der Anschreiben soll je nach betroffenem Qualitätsindikator nur der jeweils passende der hier aufgelisteten Textblöcke verwendet werden.

Da bei diesem Indikator patientenseitige Faktoren den Behandlungsprozess beeinflussen können, wird dieser Indikator risikoadjustiert. Bitte schildern Sie in Ihrer Stellungnahme, warum die erforderlichen Prozessstandards nicht eingehalten wurden und warum dies trotz Berücksichtigung patientenseitiger Faktoren in dem Indikator nicht als Qualitätsproblem einzustufen ist.

▪ **Bei einem quantitativ auffälligen Ergebnis in einem Indikationsindikator:**

Bitte schildern Sie in Ihrer Stellungnahme, warum die Prozedur häufiger durchgeführt wurde, als dies nach der Rechenvorschrift des Qualitätsindikators indiziert war, respektive warum die Prozedur trotz bestehender Indikation laut Qualitätsindikator nicht häufig genug durchgeführt wurde, und warum dies dennoch nicht als Qualitätsdefizit einzustufen ist.

▪ **Bei einem quantitativ auffälligen Ergebnis in einem risikoadjustierten Ergebnisindikator:**

Bei der Durchführung medizinischer Behandlungen sollen Komplikationen jeglicher Art und Schwere möglichst selten auftreten und Therapieziele möglichst häufig erreicht werden. Allerdings gibt es Umstände, in denen Komplikationen nicht gänzlich vermeidbar sind und Therapieziele nicht vollständig erreicht werden können. Um dies angemessen zu berücksichtigen, wird dieser Qualitätsindikator risikoadjustiert. Bitte schildern Sie in Ihrer Analyse, warum das Ergebnis Ihrer Einrichtung trotz Risikoadjustierung außerhalb des Referenzbereichs liegt und welche/r Einflussfaktor/en ggf. bei Ihren Patientinnen und Patienten so häufig auftraten, dass diese nicht mehr adäquat durch die Risikoadjustierung berücksichtigt wurde/n.

▪ **Bei einem quantitativ auffälligen Ergebnis in einem nicht risikoadjustierten Ergebnisindikator:**

Bei der Durchführung medizinischer Behandlungen sollen Komplikationen jeglicher Art und Schwere möglichst selten auftreten und Therapieziele möglichst häufig erreicht werden. Allerdings gibt es Umstände, in denen sie nicht gänzlich vermeidbar sind und Therapieziele nicht vollständig erreicht werden können. Bitte schildern Sie in Ihrer Analyse, welche Einflussfaktoren bei Ihren Patientinnen und Patienten häufiger auftraten als im Bundesdurchschnitt und warum das Verfehlen des Referenzbereichs daher nicht als Qualitätsproblem einzustufen ist.

Werden organisatorische Gründe für das quantitativ auffällige Indikatorergebnis benannt, ist nachzuweisen, dass Ihre Einrichtung mit der gebotenen Sorgfaltspflicht alles Zumutbare getan hat, um diese organisatorischen Mängel aufzufangen (z. B. die Einplanung von genügend Ressourcen und genügend Zeit).

Angabe von Vorgangsnummern zu geltend gemachten Einflussfaktoren

Wenn Sie in Ihrer Stellungnahme einen nicht von Ihrer Einrichtung zu verantwortenden Einflussfaktor geltend machen möchten, der in den Daten der Qualitätssicherung bisher nicht dokumentiert ist, geben Sie bitte in Ihrer Stellungnahme für alle Behandlungsfälle Ihrer Einrichtung zum jeweiligen Indikator an, ob dieser Einflussfaktor vorliegt oder nicht. Dazu sind die entsprechenden Vorgangsnummern mit der Stellungnahme zu übermitteln. Diese Angabe ist auch für diejenigen Behandlungsfälle zu machen, bei denen kein interessierendes Ereignis aufgetreten ist, d. h. die nicht im Zähler des Indikators erschienen sind. Eine detaillierte Einzelfalldarstellung aller Behandlungsfälle ist nicht erforderlich, sofern aus Ihrer Stellungnahme nachvollziehbar ist, warum der angeführte Einflussfaktor bei den entsprechenden Behandlungsfällen vorliegt.

Auf die Angabe der Vorgangsnummern für einen Einflussfaktor kann verzichtet werden, wenn anhand Ihrer Stellungnahme nachvollziehbar ist, dass dieser Einflussfaktor bei allen Fällen des Qualitätsindikators vorlag.

Information über das Vorgehen bei Nichteinhaltung der formalen Kriterien

Bei Nichteinhaltung der formalen Kriterien gilt die Stellungnahme als nicht abgegeben. Bei Nichteinhaltung der Vorgaben zur Wahrung der Anonymität Ihrer Patientinnen und Patienten, Ihrer Einrichtung oder weiterer Beteiligter wird die Stellungnahme darüber hinaus aus Datenschutzgründen vernichtet. Sie werden über diesen Vorgang benachrichtigt.

7 Beispiel zur Übermittlung von Vorgangsnummern in Stellungnahmen

Möchte ein Leistungserbringer in seiner Stellungnahme zu einem Indikatorergebnis geltend machen, dass ein oder mehrere nicht von ihm zu verantwortende Faktoren das Indikatorergebnis beeinflusst haben, so müssen die Vorgangsnummern der betreffenden Behandlungsfälle übermittelt werden (siehe Abschnitt 6.5.2 im Abschlussbericht). Dies wird im Folgenden an einem fiktiven Beispiel veranschaulicht¹⁰.

Als Beispiel wird ein Qualitätsindikator angenommen, der den Anteil der Patientinnen und Patienten misst, die bei einem ambulant aufgetretenen akuten Herzinfarkt innerhalb von 2 Stunden eine Koronarangiographie erhalten haben. Ein Leistungserbringer macht in seiner Stellungnahme folgende 4 Einflussfaktoren auf sein Indikatorergebnis geltend:

- Faktor 1: ein nicht vorhersehbarer Geräteausfall im einzigen Katheterlabor der Einrichtung
- Faktor 2: Ablehnung der Koronarangiographie durch die Patientin/den Patienten
- Faktor 3: ein fälschlich übermittelter Datensatz, der nicht zur Indikatorgrundgesamtheit gehört, aber wegen einer fehlerhaften Implementation des QS-Filters in der Software ausgelöst und dokumentiert wurde
- Faktor 4: lange Anfahrtszeit des Rettungsdienstes wegen geographischer Lage

Die folgende Tabelle 11 zeigt beispielhaft, in welcher Form die Übermittlung der Vorgangsnummern durch den Leistungserbringer erfolgen sollte. Dabei wird empfohlen, eine Übermittlung in digitaler Form zu etablieren, um Fehler bei der Weiterverarbeitung der Informationen zu vermeiden.

Tabelle 11: Fiktives Beispiel für die Angabe von Vorgangsnummern zur Stellungnahme zu einem Indikator

Vorgangsnummer des Behandlungsfalls	Faktor 1: Geräteausfall	Faktor 2: kein Therapiewunsch	Faktor 3: fälschlicher Datenexport wegen fehlerhafter QS-Filter-Implementation
12	1	1	
13	1		
14	1	1	
208			1
456		1	
77		1	

¹⁰ Das Beispiel erhebt keinen Anspruch auf Korrektheit der medizinischen Inhalte oder der Beurteilung der Stellungnahme, sondern dient der Verdeutlichung des grundsätzlichen Vorgehens.

Vorgangsnummer des Behandlungsfalls	Faktor 1: Geräteausfall	Faktor 2: kein Therapiewunsch	Faktor 3: fälschlicher Datenexport wegen fehlerhafter QS-Filter-Implementation
79		1	
558		1	

In dem gezeigten Beispiel liegen aus Sicht des Leistungserbringers drei nicht von ihm zu verantwortende Einflussfaktoren vor, die sich nur auf einen Teil der Behandlungsfälle beziehen. Das Vorliegen des Faktors wird durch jeweils durch die 1 gekennzeichnet. Zu beachten ist, dass in einem Behandlungsfall ggf. mehrere Einflussfaktoren vorliegen können. Bei Vorgangsnummern, die in der Tabelle nicht aufgeführt sind, wird davon ausgegangen, dass keiner der in der Tabelle genannten Einflussfaktoren bei dem betreffenden Behandlungsfall vorlag. Zusätzlich sei für das Beispiel angenommen, dass der Leistungserbringer für den vierten Faktor „lange Anfahrtszeit des Rettungsdienstes wegen geographischer Lage“ in seiner Stellungnahme erläutert hat, dass dieser auf alle Behandlungsfälle zutreffe, und dass für diesen Faktor daher keine Vorgangsnummern übermittelt wurden.

8 Beispiel zur Berücksichtigung mehrerer Einflussfaktoren bei der partiellen Nachberechnung

Die partielle Nachberechnung im Rahmen der fachlichen Bewertung basiert auf der Kategorisierung der in einer Stellungnahme angeführten Einflussfaktoren hinsichtlich Verantwortlichkeit, Art des Einflussfaktors (fachlich-inhaltlich vs. Datenfehler) und Auswirkung auf das Indikatorergebnis (siehe Abschnitt 6.4.1 im Abschlussbericht). Am Beispiel der übermittelten Einflussfaktoren aus Kapitel 7 wird im Folgenden dargestellt, wie der informative Indikatorwert berechnet werden kann.

Für das Beispiel sei von folgenden Einschätzungen ausgegangen:

- Die Fachkommission erkennt alle vier in der Stellungnahme genannten Faktoren als nicht vom Leistungserbringer zu verantwortende Einflussfaktoren an.
- Die Faktoren 1, 2 und 4 werden als fachlich-inhaltliche Faktoren kategorisiert, der Faktoren 3 als Datenfehler.
- Die Auswirkung der Faktoren 1, 2 und 3 wird so eingeschätzt, dass die betreffenden Behandlungsfälle nicht in die Indikatorberechnung hätten eingehen sollen. Die Auswirkung von Faktor 4 auf das Indikatorergebnis wird als graduell bewertet, d. h. die betreffenden Fälle können nicht aus der Grundgesamtheit des Indikators ausgeschlossen werden.

Außerdem sei davon ausgegangen, dass bei der Berechnung des originalen Indikatorwerts die Behandlungsfälle mit den Vorgangsnummern 12, 13, 208, 77 und 79 ein interessierendes Ereignis aufgewiesen haben, sowie 5 weitere Fälle, die in der Tabelle der Stellungnahme (Tabelle 11) nicht aufgeführt sind (Vorgangsnummern 1 bis 5). Das bedeutet, dass der Leistungserbringer für letztere 5 Fälle mit interessierendem Ereignis keinen besonderen Einflussfaktor geltend gemacht, und des Weiteren, dass bei den Fällen mit den Vorgangsnummern 14, 456 und 558 zwar mindestens einer der geltend gemachten Einflussfaktoren vorgelegen hat, aber trotzdem kein interessierendes Ereignis aufgetreten ist. Tabelle 12 visualisiert die Berücksichtigung der Einflussfaktoren bei der partiellen Neuberechnung:

Tabelle 12: Fiktives Beispiel für die fallbezogene Berücksichtigung von Einflussfaktoren bei der partiellen Neuberechnung

Vorgangsnummer des Behandlungsfalls	Faktor 1: Geräteausfall	Faktor 2: kein Therapiewunsch	Faktor 3: fälschlicher Datenexport wegen fehlerhafter QS-Filter-Implementation	interessierendes Ereignis	partielle Nachberechnung	
					aus Nenner auszu-schließen	aus Zähler auszu-schließen
12	1	1		1	ja	ja

Vorgangsnummer des Behandlungsfalls	Faktor 1: Geräteausfall	Faktor 2: kein Therapiewunsch	Faktor 3: fälschlicher Datenexport wegen fehlerhafter QS-Filter-Implementation	interessierendes Ereignis	partielle Nachberechnung	
					aus Nenner auszuschließen	aus Zähler auszuschließen
13	1			1	ja	ja
14	1	1			ja	
208			1	1	ja	ja
456		1			ja	
77		1		1	ja	ja
79		1		1	ja	ja
558		1			ja	
1				1		
2				1		
3				1		
4				1		
5				1		

Aus diesen Beispielannahmen ergibt sich für die partielle Nachberechnung, dass 8 Fälle aus der Grundgesamtheit des Indikators (Nenner) ausgeschlossen werden, da $a = 7$ Fälle aus fachlich-inhaltlichen Gründen (Faktoren 1 oder 2) und 1 Fall wegen eines Datenfehlers (Faktor 3) nicht für die Erfüllung des Qualitätsmerkmals des Indikators von Interesse sind. Der Zähler des Indikators wird um die Anzahl derjenigen aus dem Nenner auszuschließenden Behandlungsfälle vermindert, die auch ein Zählerereignis aufweisen ($o_a = 5$ Fälle). Faktor 4 kann nicht in die partielle Nachberechnung eingehen, da sein Beitrag zum Indikatorergebnis nicht eindeutig anzugeben ist, und muss deshalb im heuristischen Beurteilungsschritt bewertet werden.

9 Berechnungsvorschrift für die partielle Nachberechnung

9.1 Berechnung des informativen Indikatorwerts

Im Folgenden wird die Berechnung des informativen Indikatorwerts für anteilsbasierte und risikoadjustierte Indikatoren beschrieben.

Anteilsbasierte Qualitätsindikatoren

Betrachtet man den einfachen Fall eines anteilsbasierten Qualitätsindikators, so liegt im betrachteten Zeitraum ein Indikatorergebnis vor mit J Fällen des Leistungserbringers in der Grundgesamtheit und o interessierenden Vorgängen im Zähler des Indikators. Ohne Beschränkung der Allgemeinheit sei hier angenommen, dass der optimale Indikatorwert bei $o = 0$ also 0 % liegt.

Es bezeichne R den Referenzbereich des Indikators. Nach der quantitativen Bewertungsvorschrift $s(o, J, R)$ (vgl. Abschlussbericht, Kapitel 5) wurde der Leistungserbringer als quantitativ auffällig bewertet.

Die fachliche Bewertung ergibt, dass eine gewisse Anzahl a von Fällen nicht in der Grundgesamtheit des Indikators hätte berücksichtigt werden sollen. Von diesen a Fällen hatten o_a Fälle ein interessierendes Ereignis. Unter den restlichen $J - a$ Fällen wurde bei b Fällen festgestellt, dass für diese kein interessierendes Ereignis zu zählen ist (d. h. sie nicht in den Zähler des Indikators einfließen sollen).

Für die Beurteilung des partiell nachberechneten Indikatorwertes (informativer Indikatorwert) wird ein bereinigter Referenzbereich benötigt (siehe Abschlussbericht, Abschnitt 6.4). Dieser wird im Folgenden mit \tilde{R} bezeichnet.

Der Leistungserbringer wird nun nach der quantitativen Bewertungsvorschrift $s(o - o_a - b, J - a, \tilde{R})$ erneut bewertet. In einem rein enumerativen Bewertungsschema (aktuelle „rechnerische Auffälligkeit“ der QSKH-Richtlinie) wäre die Entscheidungsregel

$$\frac{o - o_a - b}{J - a} > \tilde{R}.$$

Risikoadjustierte Qualitätsindikatoren

Im Falle risikoadjustierter Qualitätsindikatoren beinhaltet die quantitative Bewertung auch die Berücksichtigung der geschätzten Wahrscheinlichkeiten e_j für alle Patienten $j = 1, \dots, J$. Die quantitative Bewertung ist dann eine Funktion $s(o, e, R)$, wobei $e = \sum_{j=1}^J e_j$. Die fachliche Bewertung ergibt, dass eine gewisse Anzahl a von Fällen nicht in der Grundgesamtheit des Indikators hätte berücksichtigt werden sollen. Von diesen a Fällen hatten o_a Fälle ein interessierendes Ereignis. Unter den restlichen $J - a$ Fällen wurde bei b Fällen festgestellt, dass für diese kein interessierendes Ereignis zu zählen ist. Die erneute quantitative Bewertung erfolgt dann nach

der Regel $s(o - o_a - b, \sum_{j=1}^{J-a} e_j, \tilde{R})$. An Stelle von e fließt somit nur noch die Summe der geschätzten Wahrscheinlichkeiten aller verbleibenden Fälle ein¹¹. In einem rein enumerativen Bewertungsschema wäre die Entscheidungsregel dann:

$$\frac{o - o_a - b}{\sum_{j=1}^{J-a} e_j} > \tilde{R}.$$

Zu beachten ist, dass für diese Form der Nachberechnung die geschätzten Wahrscheinlichkeiten e_j für alle Behandlungsfälle des Leistungserbringers zum Zeitpunkt der Nachberechnung zur Verfügung gestellt werden müssen.

Abwägung von analytischer und enumerativer Herangehensweise der Nachberechnung

Gemäß Abschlussbericht, Abschnitt 5.1.1 („Analytische Herangehensweise“) ist in der quantitativen Bewertung von Qualitätsindikatoren abzuwägen, in welchen Fällen eine enumerative bzw. eine analytische Herangehensweise an die Auswertung verfolgt wird. Für die quantitative Bewertung von anteilsbasierten und risikoadjustierten Indikatoren wird die analytische Herangehensweise empfohlen. Erstens wird diese durch in der Datenerhebung nicht quantifizierte Einflussfaktoren begründet, die eine Berücksichtigung statistischer Unsicherheit in der Einstufung erforderlich machen. Zweitens wird die analytische Herangehensweise gewählt, da im methodischen Verständnis des IQTIG ein Qualitätsindikator eine Messung der Kompetenz eines Leistungserbringers in einem speziellen Qualitätsmerkmal darstellt. Zur Messung dieser Kompetenz werden i. d. R. alle durch den QS-Filter spezifizierten Fälle eines Erfassungsjahres herangezogen. Zwar handelt es sich um eine Vollerhebung im Sinne des QS-Filters, jedoch im Sinne der Kompetenzmessung handelt es sich weiterhin um eine Stichprobe von Behandlungsfällen.

Unter der Annahme, dass durch die Stellungnahme und die fachliche Bewertung die nicht quantifizierten Einflussfaktoren sichtbar gemacht werden, stellt sich die Frage, inwiefern nach der Nachberechnung weiterhin eine analytische Herangehensweise gerechtfertigt ist. Stochastizität wird allerdings auch durch das Stellungnahmeverfahren nicht entfernt, da auch nach Entfernung der nun bekannten, gerichteten Einflussfaktoren noch unbekannte oder zufällige Effekte bleiben können. Des Weiteren beruht das Indikatorergebnis weiterhin nur auf einer Stichprobe von Behandlungsfällen, die zur Messung der Kompetenz des Leistungserbringers für das jeweilige Qualitätsmerkmal herangezogen werden. Für die Bewertung im Rahmen der partiellen Nachberechnung wird daher ebenfalls die analytische Herangehensweise gewählt.

9.2 Berechnung des bereinigten Referenzbereiches

Es bezeichne \tilde{R} den bereinigten Referenzbereich des Indikators. Dieser wird berechnet, indem der originale Referenzbereich R in zwei Anteile aufgeteilt wird (siehe Abschlussbericht, Abschnitt 6.4.2): Dabei bezeichne R_1 den Anteil gerichteter, nicht vom Leistungserbringer zu beeinflussender Faktoren auf die Indikatorwerte und R_2 das tolerierte Qualitätsdefizit, wobei gilt:

¹¹ Es wird davon ausgegangen, dass die Wahrscheinlichkeiten e_j so nummeriert sind, dass die letzten $J - a$ Fälle genau die aus der Grundgesamtheit auszuschließenden Fälle darstellen.

$$R = R_1 + R_2$$

Der bereinigte Referenzbereich berechnet sich dann als

$$\tilde{R} = \frac{R_2}{1 - R_1}$$

Die Korrektur im Nenner (gegenüber einer einfachen Setzung von $\tilde{R} = R_2$) begründet sich dadurch, dass sich der Toleranzanteil R_2 im originalen Referenzbereich nicht auf alle Behandlungsfälle der Grundgesamtheit des Indikators bezieht: Die Toleranz ist auf die Behandlungsfälle bezogen, deren Versorgungsqualität beurteilt werden soll, und damit nicht auf den Anteil R_1 der Behandlungsfälle, der in der Grundgesamtheit des Indikators enthalten ist, weil ein Teil der vom Leistungserbringer nicht zu verantwortenden Faktoren nicht in der Berechnung berücksichtigt wird. Der Anteil der eigentlich für die Qualitätsbeurteilung interessierenden Behandlungsfälle beträgt somit $1 - R_1$. Durch den Korrekturfaktor wird sichergestellt, dass die relative Höhe des tolerierten Qualitätsdefizits auch nach der partiellen Nachberechnung gleichbleibt.

9.3 Retrospektive Schätzung der in Referenzbereichen enthaltenen Toleranz für Qualitätsdefizite

Für die Berechnung des bereinigten Referenzbereiches muss für den originalen Referenzbereich die Größe des Anteils gerichteter, nicht vom Leistungserbringer zu beeinflussender Faktoren einerseits und des Anteils tolerierter Qualitätsdefizite andererseits bekannt sein. Diese Größen wurden für die bisher in der externen Qualitätssicherung eingesetzten Referenzbereiche nicht gesondert ausgewiesen. Im Folgenden ist ein Vorschlag beschrieben, wie sie retrospektiv geschätzt werden können, damit sie für die Berechnung des bereinigten Referenzbereichs verwendet werden können.

Es wird empfohlen, die Differenz zwischen dem Idealwert des Indikators (0 % bzw. 100 %) und dem empirischen bundesweiten Wert (der Wert, der bei Berechnung des Indikators auf Grundlage aller Behandlungsfälle bundesweit resultiert) als Näherung für den Anteil gerichteter Einflussfaktoren zu verwenden (höchstens aber die Breite des Referenzbereichs). Dieses Vorgehen folgt der Überlegung, dass der Bundeswert einerseits per Definition auf dem durchschnittlichen Risikoprofil der Patientinnen und Patienten basiert, dass also alle gerichteten Einflussfaktoren mit ihrem mittleren Ausmaß in diesen Wert eingehen. Andererseits kann aus der Tatsache, dass der Indikator Qualitätsverbesserungen bewirken soll, abgeleitet werden, dass die bundesweite Versorgungsqualität nicht optimal ist. Für den Abstand des empirischen Bundeswertes vom Idealwert kann also angenommen werden, dass er mindestens so groß ist wie der mittlere Einfluss der gerichteten Faktoren auf den Indikatorwert¹², und dass er zusätzlich einen Anteil enthält, der einer nicht optimalen bundesweiten Versorgungsqualität geschuldet ist

¹² Da es sich beim bundesweiten Wert um eine empirisch bestimmte Größe handelt, könnte es zufallsbedingt dazu kommen, dass diese Annahme nicht zutrifft. Die bundesweiten Fallzahlen werden jedoch als ausreichend groß angesehen, um diese Möglichkeit vernachlässigen zu können.

Es bezeichne R den originalen Referenzbereich des Indikators und $\hat{\Pi}$ den empirischen Bundeswert des Indikators. Ohne Beschränkung der Allgemeinheit sei hier angenommen, dass der optimale Indikatorwert bei $o = 0$ also 0 % liegt. Der Anteil gerichteter, nicht vom Leistungserbringer zu beeinflussender Faktoren R_1 am originalen Referenzbereich wird dann geschätzt als

$$R_1 = \hat{\Pi}$$

falls $R > \hat{\Pi}$. Andernfalls wird $R_1 = R$ gesetzt.

Der Anteil der tolerierten Qualitätsabweichung R_2 am originalen Referenzbereich ergibt sich dann als Differenz aus dem originalen Referenzbereich und dem anhand des empirischen Bundeswerts geschätzten Anteils gerichteter Einflussfaktoren:

$$\begin{aligned} R_2 &= R - R_1 \\ &= R - \hat{\Pi} \end{aligned}$$

falls $R > \hat{\Pi}$. Andernfalls $R_2 = 0$.

Für risikoadjustierte Qualitätsindikatoren z. B. in O/E-Darstellung kann das beschriebene Vorgehen sinngemäß angewendet werden.

Auswirkungen des Schätzverfahrens auf die Qualitätsbewertung

Bei der vorgeschlagenen Schätzung anhand des empirischen Bundeswertes wird der Anteil R_1 gerichteter, nicht vom Leistungserbringer zu beeinflussender Faktoren an den Indikatorwerten tendenziell zu groß eingeschätzt. Dies führt umgekehrt dazu, dass der geschätzte Toleranzanteil R_2 und damit der bereinigte Referenzbereich für die partielle Nachberechnung umso schmaler (strenger) sind, je weiter der empirische Bundeswert vom eigentlichen (unbekannten) erreichbaren Optimum entfernt ist und je näher er an der originalen Referenzbereichsgrenze liegt. Die in der Tendenz höhere Schätzung von R_1 und niedrigere Schätzung von R_2 führt dazu, dass Leistungserbringer in ihrer Stellungnahme einen größeren Anteil ihrer Abweichung vom Referenzbereich durch nicht von ihnen zu verantwortende Faktoren erklären müssen und ein geringerer Anteil als Qualitätsdefizit toleriert wird. Es kommt also in der fachlichen Bewertung umso leichter zu Bewertungen als „Qualitätsdefizit“, je näher empirischer Bundeswert und originale Referenzbereichsgrenze zusammenliegen.

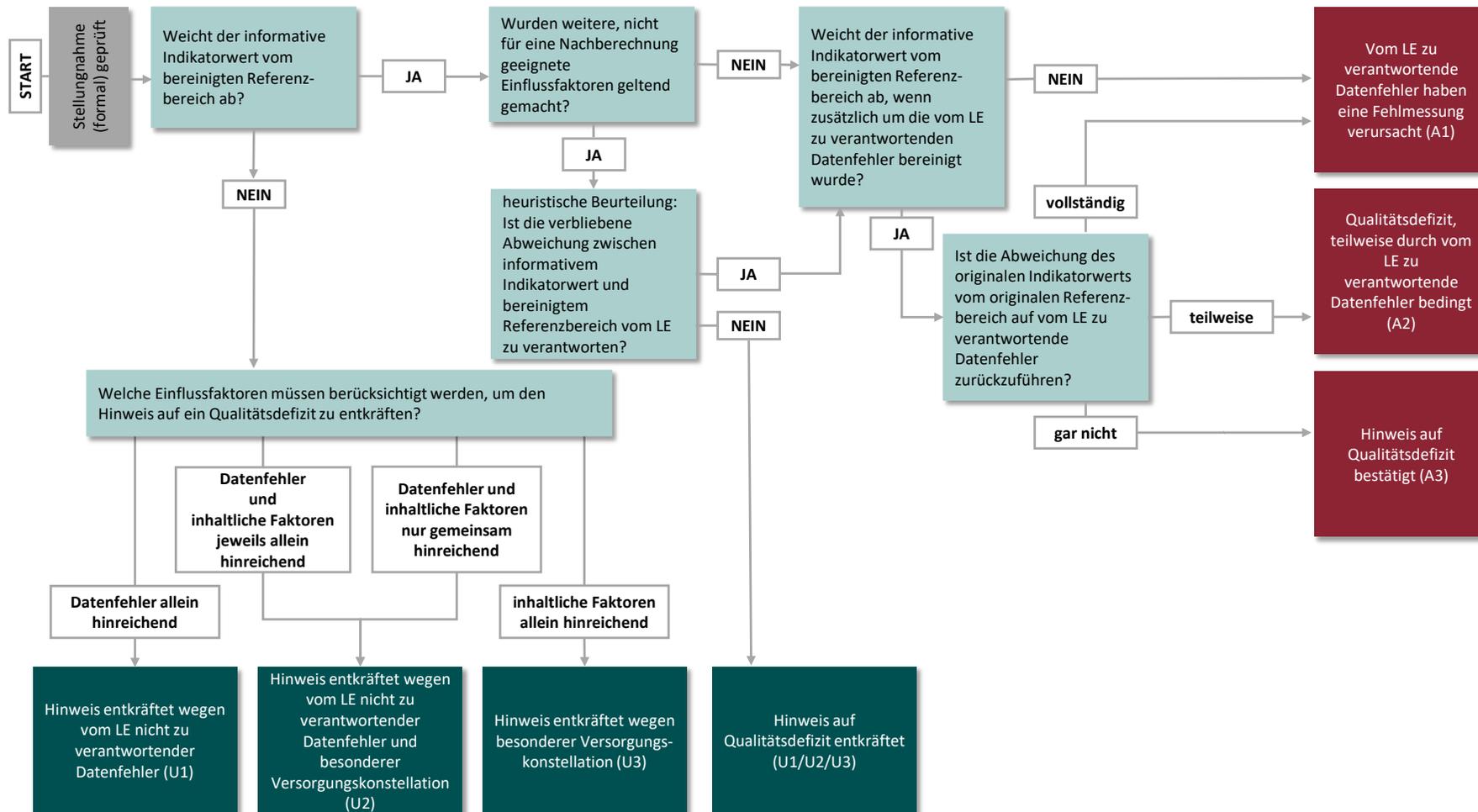
Geht man davon aus, dass der Referenzbereich gezielt und unter Berücksichtigung der verfügbaren Informationen zum Ist-Zustand der Versorgung festgelegt wurde, so spiegelt sich in einem geringen Unterschied zwischen empirischem Bundeswert und Referenzbereichsgrenze die Absicht wieder, nur geringe Abweichungen vom Ist-Zustand zum Schlechteren hin zu tolerieren. Dabei bleibt zwar offen, ob bei der Festlegung des Referenzbereichs eine geringe Toleranz von Qualitätsdefiziten intendiert war oder ob davon ausgegangen wurde, dass bereits der Ist-Zustand, gemessen am empirischen Bundeswert, deutlich vom Optimum abweicht. Da jedoch bei den betreffenden Indikatoren offensichtlich von einem größeren Verbesserungspotenzial und damit Verbesserungsbedarf auszugehen ist, kann der oben beschriebene Effekt, nämlich eine eher niedrig geschätzte Toleranz, unter der Perspektive der Qualitätsverbesserung als gewünschter Effekt interpretiert werden.

Ein Nachteil des vorgeschlagenen Vorgehens ist umgekehrt, dass – bei unverändertem Referenzbereich – umso weniger Abweichung vom Referenzbereich nach der partiellen Nachberechnung durch besondere Konstellationen erklärt werden muss, je besser der empirische Bundeswert ist. Dies könnte zu einer unangemessenen Toleranz von Qualitätsdefiziten bei einzelnen Leistungserbringern führen, wenn alle anderen Leistungserbringer deutlich bessere Versorgungsqualität aufweisen oder wenn es im Zeitverlauf zu deutlichen Verbesserungen im empirischen Bundeswert kommt. Wenn der empirische Bundeswert eines Qualitätsindikators weit innerhalb des Referenzbereichs liegt, sollte daher die Angemessenheit des Referenzbereichs überprüft werden.¹³

Grundsätzlich könnte der Toleranzanteil in den Referenzbereichen, die bisher in der externen Qualitätssicherung eingesetzt werden, auch auf andere Weise als hier vorgeschlagen geschätzt werden. Zum Zeitpunkt der Berichtslegung entwickelt das IQTIG eine Methodik, wie bei Festlegung von Referenzbereichen die Größe der beiden Anteile des Referenzbereichs angemessen und nachvollziehbar festgelegt werden kann.

¹³ Die Prüfung kann auch ergeben, dass die Ergebnisse bei diesem Indikator bundesweit so gut sind, dass kein nennenswertes Verbesserungspotenzial mehr besteht. In diesem Fall muss die Option diskutiert werden, statt einer Verschärfung des Referenzbereichs die Verwendung des Indikators auszusetzen.

10 Gesamtalgorithmus für die fachliche Bewertung



11 Zielvereinbarungen

11.1 Formular zum Sachstandsbericht einzelner Zielvereinbarungen

Der Sachstandsbericht ist ein Controlling-Instrument zum Management von Zielvereinbarungen. Er dient der Kommunikation zwischen Geschäftsstelle¹⁴ und Lenkungsgremium und stellt den Abschluss sowie den Erfüllungsgrad von Zielvereinbarungen dar.

Mitteilung zu LE <Pseudonym>			
Stand: <i>[Datum]</i>			
Ansprechpartner Geschäftsstelle: <i>[Name]</i>			
Anlass-Kategorie der Mitteilung: Wählen Sie ein Element aus. <i>[Dropdown-Menü:</i>			
<ul style="list-style-type: none"> ▪ <i>Persistierendes Qualitätsproblem,</i> ▪ <i>Kritische Compliance bezüglich der Zusammenarbeit im Rahmen der Qualitätssicherung,</i> ▪ <i>mangelnde Einhaltung der formalen Kriterien der Zielvereinbarung,</i> ▪ <i>mangelnde Einhaltung der qualitativen Kriterien der Zielvereinbarung,</i> ▪ <i>Sonstiges]</i> 			
Betrifft das Erfassungsjahr			
QS-Verfahren			
Betroffene(r) QI / QIs			
Begründete Einschätzung des Handlungsbedarfs			
<i>[Begründung für die Involvierung des Lenkungsgremiums]</i>			
....			
Quantitative Ergebnisse und fachliche Bewertung der Erfassungsjahre <i>[XY]</i> bis <i>[YZ]</i>			
Erfassungsjahr	QI-Ergebnis (%)	Ergebnis außerhalb des Referenzbereichs	Bewertung nach Abschluss des Stellungnahmeverfahrens
<i>[QI-ID und QI-Bezeichnung]</i>			
...			
<i>[QI-ID und QI-Bezeichnung]</i>			

¹⁴ Gemeint ist die Geschäftsstelle der LAG bzw. das IQTIG.

Quantitative Ergebnisse und fachliche Bewertung der Erfassungsjahre [XY] bis [YZ]			
...			
<i>[Erläuterung der dargestellten Ergebnisse und Entwicklungen]</i>			
....			

Zielvereinbarung(en)	
Frist für Bericht des LE zur Zielerreichung	<i>[Datum]</i>
Unterschiedene Zusendung erfolgt	<i>[ja / nein]</i>
<u>Vereinbartes Qualitätsziel 1: [Ziel]</u>	
Betrifft den QI	<i>[QI-ID und QI-Bezeichnung]</i>
Begründung des Ziels	<i>[Text]</i>
Frist zur Einhaltung des Ziels	<i>[Datum oder Erfassungsjahr]</i>
Fristgerecht erfolgt	<i>[ja / nein]</i>
<u>Vereinbartes Qualitätsziel 2 - n: [Ziel]</u>	
Betrifft den QI	<i>[QI-ID und QI-Bezeichnung]</i>
Begründung des Ziels	<i>[Text]</i>
Frist zur Einhaltung des Ziels	<i>[Datum oder Erfassungsjahr]</i>
Fristgerecht erfolgt	<i>[ja / nein]</i>
<u>Vereinbarte Maßnahme 1: [Maßnahme]</u>	
Betrifft den QI	<i>[QI-ID und QI-Bezeichnung]</i>
Begründung der Maßnahme	<i>[Text]</i>
Frist für den schriftlichen Nachweis	<i>[Datum]</i>
Fristgerecht erfolgt	<i>[ja / nein]</i>
<u>Vereinbarte Maßnahme 2 - n: [Maßnahme]</u>	
Betrifft den QI	<i>[QI-ID und QI-Bezeichnung]</i>
Begründung der Maßnahme	<i>[Text]</i>
Frist für den schriftlichen Nachweis	<i>[Datum]</i>
Fristgerecht erfolgt	<i>[ja / nein]</i>

Zielvereinbarung(en)	
<u>Empfohlene, nicht konsentierete Maßnahme 1: [Maßnahme]</u>	
Betrifft den QI	[QI-ID und QI-Bezeichnung]
Begründung der Maßnahme	[Text]
Begründung des Leistungserbringers für die Nicht-Konsentierung	[Text]
<u>Empfohlene, nicht konsentierete Maßnahme 2 - n: [Maßnahme]</u>	
Betrifft den QI	[QI-ID und QI-Bezeichnung]
Begründung der Maßnahme	[Text]
Begründung des Leistungserbringers für die Nicht-Konsentierung	[Text]
Abschließende Einschätzung der Geschäftsstelle	
[Text]	
...	
Vorschlag der Geschäftsstelle zum weiteren Vorgehen	
[Text]	
...	

<Ende des Formularinhalts>

11.2 Erläuterung zum Formular

Anlass-Kategorie der Mitteilung:

Persistierendes Qualitätsproblem

Wurde bei einem Leistungserbringer in einem Indikator wiederholt ein Qualitätsdefizit festgestellt und zeigt sich trotz durchgeführter qualitätsfördernder Maßnahmen keine Verbesserung der Qualität, so stellt dies einen Anlass für eine Mitteilung an das Lenkungsgremium dar.

Kritische Compliance bezüglich der Zusammenarbeit im Rahmen der Qualitätssicherung

Eine mangelnde Compliance bezüglich der Zusammenarbeit im Rahmen der Qualitätssicherung ist zum Beispiel gegeben, wenn ein Leistungserbringer ohne Nennung eines nachvollziehbaren Grundes die Teilnahme an qualitätsfördernden Maßnahmen verweigert (z. B. Kollegiales Gespräch, Vor-Ort-Begehung oder Unterschrift einer Zielvereinbarung).

mangelnde Einhaltung der formalen Kriterien der Zielvereinbarung

Werden formale Kriterien (z. B. Einhaltung von Fristen) systematisch oder mehrfach nicht eingehalten, so stellt auch dies einen Anlass für eine Mitteilung an das Lenkungsgremium dar.

mangelnde Einhaltung der qualitativen Kriterien der Zielvereinbarung

Werden die vereinbarten Ziele und Maßnahmen nicht in ausreichender Qualität erfüllt und umgesetzt, so stellt dies einen Anlass für eine Mitteilung an das Lenkungsgremium dar.

Die Bewertung, ob die Kriterien qualitativ erfüllt werden, erfolgt durch die Geschäftsstelle nach fachlicher Beratung der Fachkommission.

Sonstiges

Da eine abschließende Auflistung der möglichen Gründe, die einen Anlass für eine Mitteilung an das Lenkungsgremium darstellen, nicht möglich ist, gibt es die Kategorie „Sonstiges“. Hierunter werden diejenigen Gründe gefasst, die über die bereits beschriebenen Anlass-Kategorien nicht abbildbar sind.

Begründete Einschätzung des Handlungsbedarfs

Zur weiteren Verdeutlichung der bereits genannten Anlass-Kategorie, wird an dieser Stelle näher beschrieben, welche Überlegungen oder Ereignisse zur Notwendigkeit einer Involvierung des Lenkungsgremiums geführt haben. Hier kann ebenso eine Einschätzung zur Dringlichkeit erfolgen.

Ergebnisse der quantitativen Auffälligkeitseinstufung und fachliche Bewertung der Erfassungsjahre [XY] bis [YZ]

Die tabellarische Zusammenstellung soll die quantitativen Ergebnisse zu den relevanten Qualitätsindikatoren, die Hinweise ob das Ergebnis entsprechend der Rechnungsvorschrift stat. signifikant vom Referenzbereich abweicht und ihre Bewertung nach Abschluss des Stellungnahmeverfahrens darstellen.

Zielvereinbarung (-en)

Es erfolgt eine detaillierte und zeitlich geordnete Auflistung aller Zielvereinbarungen, die den entsprechenden Indikator betreffen. Wurden in der Vergangenheit bereits Zielvereinbarungen getroffen, so werden auch diese hier dargestellt. Neben der Information, zu wann eine Unterschrift der gesamten Zielvereinbarung durch den Leistungserbringer erfolgen sollte und ob diese erfolgt ist, werden für jedes vereinbarte Ziel bzw. jede vereinbarte Maßnahme folgende Inhalte beschreiben:

1. Betrifft den QI:

Hier wird der QI (ID und Bezeichnung) genannt, für den das jeweilige Ziel / die jeweilige Maßnahme gilt.

2. Begründung des Ziels / der Maßnahme:

Es wird begründet, weshalb das Ziel / die Maßnahme aufgenommen wurde.

3. Frist für den schriftlichen Nachweis:

Für jedes Ziel / jede Maßnahme wird in der Zielvereinbarung eine Frist genannt, zu der die Erfüllung nachgewiesen werden muss. Diese wird hier aufgeführt.

4. Fristgerecht erfolgt:

Es erfolgt die Angabe (ja / nein), ob die geforderte Frist eingehalten wurde.

Wurden dem Leistungserbringer von der Geschäftsstelle weitere Maßnahmen vorgeschlagen, die Aufnahme dieser Maßnahmen in die Zielvereinbarung jedoch seitens des Leistungserbringers abgelehnt, werden diese ebenfalls aufgelistet. Dargestellt werden für diese Maßnahmen folgende Informationen:

5. Betrifft den QI:

Hier wird der QI (ID und Bezeichnung) genannt, für den das jeweilige Ziel / die jeweilige Maßnahme gelten soll.

6. Begründung des Ziels / der Maßnahme:

Es wird begründet, weshalb das Ziel / die Maßnahme empfohlen wurde.

7. Begründung des Leistungserbringers für die Nicht-Konsentierung:

Es wird die vom Leistungserbringer ggf. angeführte Begründung, weshalb einer Aufnahme in die Zielvereinbarung nicht zugestimmt wurde, genannt.

Abschließende Einschätzung der Geschäftsstelle

Es erfolgt eine kurze Gesamteinschätzung des Sachverhalts aus Sicht der Geschäftsstelle.

Vorschlag der Geschäftsstelle zum weiteren Vorgehen (optional)

Können dem Lenkungsgremium konkrete weitere Maßnahmen außerhalb der Handlungskompetenz der Geschäftsstelle empfohlen werden, werden diese hier benannt.

Literatur

- Agresti, A (2013): *Categorical Data Analysis*. Third Edition. (Wiley Series in Probability and Statistics). Hoboken, US-NJ: Wiley. ISBN: 978-0-470-46363-5.
- Ash, AS; Fienberg, SE; Louis, TA; Normand, SLT; Stukel, TA; Utts, J (2012): *Statistical Issues in Assessing Hospital Performance*. Commissioned by the Committee of Presidents of Statistical Societies [*White paper*]. Revised: 27.01.2012. Baltimore, US-MD: Centers for Medicare and Medicaid Services; Committee of Presidents of Statistical Societies. URL: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf> (abgerufen am: 09.10.2019).
- Bates, D; Mächler, M; Bolker, B; Walker, S (2015): Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). DOI: 10.18637/jss.v067.i01.
- Gelman, A; Hill, J (2007): *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, GB: Cambridge University Press. ISBN: 978-0-521-86706-1.
- Hosmer, DW Jr.; Lemeshow, S; Sturdivant, RX (2013): *Applied Logistic Regression*. Third Edition. Hoboken, US-NJ: John Wiley & Sons. ISBN: 978-0-470-58247-3.
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2017): Bericht zum Strukturierten Dialog 2016. Erfassungsjahr 2015. Anhang. Stand: 15.05.2017. Berlin: IQTIG. URL: https://iqtig.org/downloads/berichte/2015/IQTIG_Bericht-zum-Strukturierten-Dialog-2016_Anhang.pdf (abgerufen am: 27.02.2019).
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2018): IQTIG – R functions for hospital profiling. Package 'IQTIGpvc'. [Stand:] 06.03.2018. Berlin: IQTIG. URL: https://iqtig.org/dateien/dasiqtig/grundlagen/biometrische_methoden/IQTIGpvc.pdf (abgerufen am: 09.10.2019).
- Larsen, K; Petersen, JH; Budtz-Jørgensen, E; Endahl, L (2000): Interpreting Parameters in the Logistic Regression Model with Random Effects. *Biometrics* 56(3): 909-914. DOI: 10.1111/j.0006-341X.2000.00909.x.
- Säfken, B; Rügamer, D; Kneib, T; Greven, S (2018): Conditional Model Selection in Mixed-Effects Models with cAIC4. [Stand:] 17.03.2018. [Dokument Nr.:] arXiv:1803.05664v2 [stat.CO]. URL: <https://arxiv.org/pdf/1803.05664.pdf> (abgerufen am: 25.06.2018).
- Spiegelhalter, D; Sherlaw-Johnson, C; Bardsley, M; Blunt, I; Wood, C; Grigg, O (2012): Statistical methods for healthcare regulation: rating, screening and surveillance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 175(1): 1-47. DOI: 10.1111/j.1467-985X.2011.01010.x.
- Stroup, WW (2013): *Generalized Linear Mixed Models. Modern Concepts, Methods and Applications*. (Texts in Statistical Science). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC. ISBN: 978-1-4398-1512-0.

- Winkler-Komp, G; Misselwitz, B; Kupfernagel, F; van Emmerich, C; Döbler, K (2014):
Maßnahmen zur Qualitätssicherung in Krankenhäusern: Strukturierter Dialog – Strukturen
und Prozesse. Ergebnis einer Umfrage des Gemeinsamen Bundesausschusses bei den auf
Landesebene beauftragten Stellen und der Institution nach § 137 a SGB V. *Das Krankenhaus*
2014(3): 198-205.
- Wood, SN (2006): Generalized Additive Models. An Introduction with R. (Texts in Statistical
Science). Boca Raton, US-FL [u. a.]: Chapman & Hall/CRC. ISBN: 978-1-58488-474-3.
- Wood, SN (2013): A simple test for random effects in regression models. *Biometrika* 100(4):
1005-1010. DOI: 10.1093/biomet/ast038.