

Wahl des Schwellenwerts α zur rechnerischen Auffälligkeitseinstufung

Stand: 4. Juni 2026; Zuständigkeit: Fachbereich Biometrie

Inhalt

| | | |
|-------|--|----|
| 1 | Hintergrund | 2 |
| 2 | Erläuterung des Schwellenwerts α | 2 |
| 3 | Wahl des Schwellenwerts α | 3 |
| 3.1 | Auswirkung und Interpretation | 3 |
| 3.2 | Vorgehen | 5 |
| 3.2.1 | Feste Referenzbereiche | 5 |
| 3.2.2 | Perzentilbasierte Referenzbereiche | 6 |
| 3.2.3 | Gesamtschau | 6 |
| 3.3 | Ergebnisse | 7 |
| 3.4 | Beispiele | 8 |
| 3.4.1 | Wahl bei festem Referenzbereich: QI 51846 | 8 |
| 3.4.2 | Wahl bei perzentilbasiertem Referenzbereich: QI 212001 | 9 |
| 4 | Ausblick | 10 |
| | Literatur | 11 |

1 Hintergrund

Zu den Auswertungsjahren 2025 und 2026 wurde in fünf QS-Verfahren (QS GYN-OP, QS CHE, QS MC, QS KAROTIS, QS DEK) die Methodik zur rechnerischen Auffälligkeitseinstufung umgestellt. Dadurch wurde die bisherige Methodik abgelöst, die sich noch an der außer Kraft getretenen Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern (QSKH-RL)¹ orientierte. Zusammen mit QS PCI (seit Auswertungsjahr 2022) werden damit derzeit die rechnerischen Auffälligkeiten der Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung (DeQS-RL)² aus insgesamt sechs QS-Verfahren nach der neuen Methodik ermittelt.³ Das IQTIG plant, in den kommenden Auswertungsjahren die Methodik auch in allen weiteren QS-Verfahren der DeQS-RL schrittweise umzustellen.

Eine detaillierte Beschreibung der Methodik findet sich in dem Begleitdokument zu den endgültigen Rechenregeln zum Auswertungsjahr 2026 „Umstellung der statistischen Methodik für die Ermittlung rechnerischer Auffälligkeiten und perzentilbasierter Referenzbereiche“. Das Dokument ist zusammen mit weiteren Materialien zum Thema auf der IQTIG-Webseite abrufbar.⁴ Kurz gesagt berücksichtigt die neue Methodik zur rechnerischen Auffälligkeitseinstufung die von der Fallzahl abhängige statistische Unsicherheit.⁵ Sie stellt sicher, dass nicht allein aus stochastischen Gründen bei Leistungserbringern mit geringer Fallzahl viele Auffälligkeiten auftreten. Es wird außerdem die Berechnung perzentilbasierter Referenzbereiche angepasst, sodass möglichst genau der vom Perzentil vorgegebene Anteil an Leistungserbringern rechnerisch auffällig wird.

2 Erläuterung des Schwellenwerts α

Ein Steuerungselement für die rechnerische Auffälligkeitseinstufung ist der Schwellenwert α . Vereinfacht lässt sich sagen: Er legt für einen Qualitätsindikator fest, wie gravierend die quantitative Evidenz, dass ein Leistungserbringer den Referenzbereich verletzt, sein muss. Die Evidenz ist dabei operationalisiert als die sogenannte A-posteriori-Wahrscheinlichkeit: die Wahrscheinlichkeit, dass der Leistungserbringer den Referenzwert einhält, berechnet unter Berücksichti-

¹ Richtlinie des Gemeinsamen Bundesausschusses gemäß § 136 Abs. 1 SGB V i. V. m. § 135a SGB V über Maßnahmen der Qualitätssicherung für nach § 108 SGB V zugelassene Krankenhäuser. In der Fassung vom 15. August 2006, zuletzt geändert am 14. Mai 2020. Außer Kraft getreten. URL: <https://www.g-ba.de/informationen/richtlinien/38/> (abgerufen am: 04.06.2026)

² Richtlinie des Gemeinsamen Bundesausschusses zur datengestützten einrichtungsübergreifenden Qualitätssicherung. In der Fassung vom 19. Juli 2018, zuletzt geändert am 18. Dezember 2025, in Kraft getreten am 1. Januar 2026. URL: <https://www.g-ba.de/informationen/richtlinien/105/> (abgerufen am: 04.06.2026).

³ Betroffen sind alle Qualitätsindikatoren in den QS-Verfahren, außer solche mit Sentinel-Referenzbereich.

⁴ <https://iqtig.org/das-iqtig/wie-wir-arbeiten/grundlagen/biometrische-grundlagen/biometrische-methodik-zur-auffaelligkeitseinstufung/>

⁵ Für risikoadjustierte Ratenindikatoren („O/E-Indikatoren“) ist die statistische Unsicherheit von der Anzahl erwarteter Ereignisse (E) abhängig, und damit nur indirekt von der Fallzahl (N). Allerdings besteht empirisch eine besonders starke Korrelation zwischen N und E . Im Spezialfall von vollständig ausgeglichenen Risikoprofilen zwischen Leistungserbringern wäre die Korrelation perfekt. Zur Einfachheit der Darstellung wird auf diese Unterscheidung in diesem Dokument verzichtet.

gung seines Indikatorergebnisses. Der Leistungserbringer wird rechnerisch auffällig, wenn diese A-posteriori-Wahrscheinlichkeit mit dem Schwellenwert α übereinstimmt oder ihn unterschreitet. Formal ausgedrückt ist ein Leistungserbringer dann auffällig, wenn gilt:

$$P(\theta \in \text{Referenzbereich} \mid \text{Indikatorergebnis}) \leq \alpha,$$

wobei hier θ der zugrunde liegenden Parameter für die Kompetenz des Leistungserbringers im Qualitätsindikator ist und P für die Wahrscheinlichkeit steht. Mit Indikatorergebnis ist nicht nur der Indikatorwert (z. B. O/N) gemeint, sondern alle beobachteten Daten eines Leistungserbringers, die ins Wahrscheinlichkeitsmodell eingehen, also insbesondere z. B. auch die Fallzahl N . Der beobachtete Indikatorwert ist nur als eine statistische Schätzung für den nicht beobachtbaren zugrunde liegenden Parameter θ des Leistungserbringers zu verstehen (Cederbaum et al. 2025).

Bei einem kleineren Schwellenwert α fordert die Einstufungsmethodik somit mehr Evidenz für ein Verletzen des Referenzbereichs; beispielsweise in Form stärkerer Abweichungen des beobachteten Indikatorergebnisses vom Referenzbereich oder einer größeren Zahl von beobachteten Fällen. In dieser Funktion hat der Schwellenwert α Ähnlichkeit mit einem statistischen Signifikanzniveau. Dementsprechend liegt es nahe, sich bei α an gängigen Konventionen für das Signifikanzniveau aus wissenschaftlichen Studien zu orientieren (z. B. $\alpha = 0,05$). Es gibt allerdings gute Gründe, bei der Wahl des Schwellenwerts α für die Qualitätssicherung von solch einer strikten Konventionen abzuweichen. Stattdessen wird beispielsweise empfohlen, Ressourcen für Handlungsanschlüsse und Abwägungen zur Patientensicherheit in die Wahl einfließen zu lassen (Austin und Brunner 2008, Hengelbrock et al. 2023, IQTIG 2020, Kapitel 5, WIdO 2007, Kapitel 3). Die Flexibilität bei der Wahl stellt einen Vorteil gegenüber der bisher verwendeten Methodik für die rechnerische Auffälligkeitseinstufung dar. Die bisher verwendete Methodik lässt sich als Spezialfall der neuen Methodik sehen, bei der ein sehr hoher Schwellenwert $\alpha \approx 0,5$ gewählt wird. Als Spezialfall bleibt die Möglichkeit für ein Vorgehen nach der bisher verwendeten Methodik somit mit Umstellung erhalten.

3 Wahl des Schwellenwerts α

3.1 Auswirkung und Interpretation

Da das beobachtete Indikatorergebnis unter anderem auch von stochastischen Einflüssen abhängt, ist die Einstufung in rechnerisch auffällige und unauffällige Leistungserbringer anhand der Indikatorergebnisse eine Entscheidung unter Unsicherheit (Cederbaum et al. 2025). Jede mögliche Methodik ist daher imperfekt. Die Einstufung kann aber mehr oder weniger nützlich für die Qualitätssicherung sein. Sie ist dann nützlich, wenn sie vorwiegend Leistungserbringer mit geringer zugrunde liegender Kompetenz θ identifiziert. Für solche Leistungserbringer ist es naheliegend, dass bei ihnen tatsächliches Potenzial zur Qualitätsverbesserung besteht. Sie eignen sich daher als Ziel von Aufwänden für die Qualitätsförderung, beispielsweise in Form von Stellungnahmeverfahren nach § 17 Abs. 6 DeQS-RL und qualitätsverbessernden Maßnahmen nach § 17 Abs. 8 DeQS-RL. Die neue Methodik der rechnerischen Auffälligkeitseinstufung und die Wahl des

zugehörigen Schwellenwerts α dienen somit einem zielgerichteten Einsatz der Aufwände für die Qualitätsförderung. Wegen der Komplexität der betrachteten Prozesse und der Schwierigkeit der Messung der zugrunde liegenden Kompetenz θ lässt sich der zielgerichtete Einsatz allerdings nicht objektiv quantifizieren. Aus Sicht des IQTIG stellt der Schwellenwert α damit keinen Parameter dar, für den es *per se* eine „korrekte“ Wahl gibt, sondern lediglich mehr oder weniger angemessene Werte. Dies verhält sich analog zur Wahl des Signifikanzniveaus in wissenschaftlichen Studien (Maier und Lakens 2022).

Das IQTIG verwendet für die rechnerische Auffälligkeitseinstufung sowohl feste, als auch percentilbasierte Referenzbereiche (IQTIG 2024). Die Auswirkung und Interpretation des Schwellenwerts α kann man je nach Art des Referenzbereichs eines Qualitätsindikators unterschiedlich zusammenfassen.

Bei **festen Referenzbereichen** gilt: Je kleiner der Schwellenwert α gewählt wird, desto weniger rechnerische Auffälligkeiten treten auf. Das liegt daran, dass bei einem kleineren Schwellenwert α mehr Evidenz für ein Verletzen des festen Referenzbereichs gefordert wird und es tendenziell weniger Leistungserbringer gibt, die diesen Evidenzgrad erfüllen. Einerseits gilt daher: Je kleiner der Schwellenwert, desto weniger Leistungserbringer mit hoher zugrunde liegender Kompetenz θ werden auffällig. Andererseits werden auch weniger Leistungserbringer mit niedriger Kompetenz θ auffällig. Der Schwellenwert α lässt sich also für feste Referenzbereiche als Abwägung von Aufwänden interpretieren (Hengelbrock et al. 2023, IQTIG 2020): Dem gesamtgesellschaftlichen Aufwand für eine rechnerische Auffälligkeit (z. B. Auswirkungen auf Ressourcen der zuständigen Stellen für Stellungnahmeverfahren oder die Reputation eines Leistungserbringers) und dem Aufwand für eine verpasste rechnerische Auffälligkeit trotz geringer zugrunde liegender Kompetenz θ beim Leistungserbringer (z. B. Auswirkungen auf die Patientensicherheit). Da die gesamtgesellschaftlichen Aufwände nicht quantifizierbar sind, ergibt sich aus dieser Interpretation allerdings keine praktisch umsetzbare Regel zur indicatorspezifischen Wahl des Schwellenwerts α . Stattdessen können bei der Wahl die Ergebnisse von Proberechnungen auf historischen Daten evaluiert werden, um den Schwellenwert α aus einer Vorauswahl von möglichen Werten zu setzen (siehe Abschnitt 3.2).

Bei **percentilbasierten Referenzbereichen** wirkt sich der Schwellenwert α *nicht* darauf aus, wie viele rechnerische Auffälligkeiten resultieren. Dies ist bereits durch das festgelegte Perzentil gesteuert (z. B. werden bei einem Referenzbereich am 95. Perzentil etwa 5 % der Leistungserbringer auffällig). Der Schwellenwert α hat demnach nicht dieselbe Bedeutung im Sinne einer Aufwandsabwägung wie bei festen Referenzbereichen. Anstelle dessen wirkt sich α darauf aus, bei *welchen* Leistungserbringern die rechnerischen Auffälligkeiten entstehen. Generell ist es so, dass beim percentilbasierten Referenzbereich nach der neuen Methodik der Auffälligkeitseinstufung nur solche Indikatorergebnisse zur rechnerischen Auffälligkeit führen, die bundesweit unterdurchschnittlich sind und zudem für ihren Fallzahlbereich zu den schlechtesten Ergebnissen gehören. Für einen *hohen* Schwellenwert α wird sich mehr auf diejenigen schlechten Indikatorergebnisse fokussiert, die *stärker* vom Zufall beeinflusst sind, also *weniger verlässliche* Aussagen

über die zugrunde liegende Kompetenz θ machen: Es treten relativ mehr rechnerische Auffälligkeiten im *niedrigen* Fallzahlbereich auf, also bei den für eine *niedrige* Fallzahl schlechtesten, unterdurchschnittlichen Indikatorergebnissen. Für einen niedrigen Schwellenwert α ist es umgekehrt. Es wird sich mehr auf diejenigen schlechten Indikatorergebnisse fokussiert, die *verlässlichere* Aussagen über die zugrunde liegende Kompetenz θ machen: Es treten relativ mehr rechnerische Auffälligkeiten im *hohen* Fallzahlbereich auf, also bei den für eine *hohe* Fallzahl schlechtesten, unterdurchschnittlichen Indikatorergebnissen. Auch hier eignen sich Proberechnungen auf historischen Daten um den Schwellenwert α auf typischen Fallzahl- und Ergebnisverteilungen eines Qualitätsindikators zu evaluieren (siehe Abschnitt 3.2).

3.2 Vorgehen

Im Auswertungsjahr 2022 wurde für QS PCI indikatorübergreifend der Schwellenwert $\alpha = 0,025$ als Startwert gewählt.⁶ Bei der Umstellung der Methodik in weiteren QS-Verfahren für die Auswertungsjahre 2025 und 2026 werden hingehen bei der Setzung des Schwellenwerts α verfahrens- oder indikatorspezifische Besonderheiten berücksichtigt. Ausgewählt wird der Schwellenwert für einen Indikator dabei aus einer Vorauswahl von kanonischen Werten

$$\alpha \in \{0,025; 0,05; 0,1; 0,15; 0,2; 0,3; 0,4; 0,5\}.$$

Es werden Proberechnungen für jeden Indikator herangezogen, um einen oder mehrere angemessene Schwellenwerte α auszuwählen. Die Proberechnungen werden pro Indikator mit den Schwellenwerten aus der oben genannten Vorauswahl als Kandidaten und auf historischen Daten durchgeführt (zumeist Vorjahr und Vor-Vorjahr). Die Schwellenwert-Kandidaten werden anhand der folgenden Kriterien beurteilt und danach in der Gesamtschau abgewogen.

3.2.1 Feste Referenzbereiche

Bei festen Referenzbereichen wird in den Proberechnungen die **Gesamtzahl an rechnerischen Auffälligkeiten** und die **Sensitivität** betrachtet. Dass nach bisheriger Methodik allein aus stochastischen Gründen bei Leistungserbringern mit geringer Fallzahl viele Auffälligkeiten auftreten führt zu unnötigen Aufwänden der Qualitätssicherung und spricht für einen tendenziell kleineren Schwellenwert α .

Der Schwellenwert α sollte allerdings nicht so klein sein, dass die bundesweite Gesamtzahl an rechnerischen Auffälligkeiten die Anzahl an qualitativen Auffälligkeiten aus dem Stellungnahmeverfahren unterschreitet. Die Anzahl an qualitativen Auffälligkeiten dient als grober Anhaltspunkt für den bundesweit durch die zuständigen Stellen leistbaren Aufwand für die Qualitätsförderung (Stellungnahmeverfahren zuzüglich weitere Folgemaßnahmen gemäß § 17 Abs. 8 DeQS-RL). Die Anzahl an qualitativen Auffälligkeiten wird also primär als Indiz dafür verstanden, wie viel

⁶ Die Setzung von $\alpha = 0,025$ stellt Konsistenz her zwischen einer rechnerischen Auffälligkeitseinstufung über die A-posteriori-Wahrscheinlichkeit und einer Auffälligkeitseinstufung über die Grenzen der 95 %-Vertrauensbereiche, die vom IQTIG standardmäßig zusammen mit Indikatorergebnissen berichtet werden. Für andere Werte von α für die rechnerische Auffälligkeitseinstufung besteht diese Konsistenz nicht.

Aufwand für die Prüfung, Bewertung und Verbesserung von Strukturen und Prozessen im Stellungnahmeverfahren für einen Indikator anfällt und leistbar ist. Eine Reduktion der bundesweiten rechnerischen Auffälligkeiten deutlich unter diesen Wert ist für die Qualitätsförderung nicht angebracht.

Des Weiteren sollte der Schwellenwert α nicht so klein sein, dass die Sensitivität im Qualitätsindikator zu gering wird. Mit Sensitivität ist die Wahrscheinlichkeit gemeint, dass ein Leistungserbringer, dessen zugrunde liegende Kompetenz außerhalb des Referenzbereichs liegt, rechnerisch auffällig wird. Die Sensitivität wird neben dem Schwellenwert auch von Fallzahl und Prävalenz des interessierenden Ereignisses beeinflusst. Außerdem bezieht sie sich immer auf eine bestimmte angenommene Kompetenz θ . Sie wird daher indicatorspezifisch berechnet. Konkret betrachtet das IQTIG die Sensitivität bei der medianen Fallzahl und im Durchschnitt über die Fallzahlverteilung. Als angenommene Kompetenz θ verwendet das IQTIG eine besonders geringe Kompetenz, nämlich halb so hoch wie vom Referenzbereich gefordert. Als Mindestziel für die Sensitivität wird ein Wert von ca. $\frac{1}{2}$ verwendet. Diese theoretische Wahrscheinlichkeit von ca. $\frac{1}{2}$ rechnerisch auffällig zu werden kann man folgendermaßen interpretieren: Ein Leistungserbringer mit besonders geringer zugrunde liegender Kompetenz würde ca. in jedem zweiten Auswertungsjahr auffällig werden.

3.2.2 Perzentilbasierte Referenzbereiche

Bei perzentilbasierten Referenzbereichen wird in den Proberechnungen der **Anteil an Auffälligkeiten in Abhängigkeit der Fallzahl** betrachtet.

Der Schwellenwert α sollte weder eine starke Konzentration von rechnerischen Auffälligkeiten bei Leistungserbringern mit höherer zufallsbedingter Variabilität der Ergebnisse (also kleineren Fallzahlen), noch bei Leistungserbringern weniger zufallsbedingter Variabilität der Ergebnisse (also höhere Fallzahlen) verursachen. Stattdessen sollte er annäherungsweise eine ausgeglichene Verteilung ermöglichen. Das bedeutet: Die für ihre Fallzahl schlechtesten, unterdurchschnittlichen Indikatorergebnisse sollten auffällig werden können – ohne dass es dabei eine große Rolle spielt, ob sie eher im niedrigen oder eher im hohen Fallzahlbereich zu verorten sind.

Von diesem Ziel ist abzuweichen, wenn für einen Indikator mit perzentilbasiertem Referenzbereich Evidenz für einen negativen Zusammenhang zwischen Fallzahl und Behandlungsqualität (sog. Volume-Outcome-Zusammenhang) vorliegt. Der Schwellenwert α ließe sich durch die Wahl eher hoher Werte dann so ausrichten, dass der Fokus überproportional auf den niedrigeren Fallzahlbereich gerichtet würde.

3.2.3 Gesamtschau

Die letztendliche Wahl des Schwellenwerts wird durch eine qualitative Gesamtschau der Ergebnisse aus den Proberechnungen über alle Indikatoren eines QS-Verfahrens vorgenommen. Dabei wird als weiteres Kriterium berücksichtigt, dass möglichst einheitliche Schwellenwerte innerhalb eines QS-Verfahrens aus praktischen Gründen zu bevorzugen sind. Wenn beispielsweise zwei benachbarte Optionen für den Schwellenwert zu ähnlichen Proberechnungs-Ergebnissen führen,

wird im Sinne der Einheitlichkeit entschieden. Dies ergibt auch Sinn wegen des näherungsweise Charakters der Proberechnungen auf historischen Daten, bei denen kleine Unterschiede nicht überinterpretiert werden sollten. Einheitlichkeit im QS-Verfahren ist auch dann ein führendes Kriterium, wenn bereits bekannte relevante Änderungen in den Indikator-Rechenregeln zwischen den historischen Daten der Proberechnungen und künftigen Auswertungsjahren dazu führen, dass Proberechnungen wenig aussagekräftig sind.

3.3 Ergebnisse

Die bis einschließlich Auswertungsjahr 2026 gewählten Schwellenwerte sind in Tabelle 1 dargestellt. Das oben beschriebene Vorgehen (Abschnitt 3.2) wurde zur Wahl der Schwellenwerte für diejenigen QS-Verfahren gesetzt, die ab Auswertungsjahr 2025 auf die neue rechnerische Auffälligkeit umgestellt wurden.

Tabelle 1: Schwellenwerte für alle QS-Verfahren mit neuer Methodik der rechnerischen Auffälligkeitseinstufung bis Auswertungsjahr 2026

| QS-Verfahren | Indikatoren (außer Sentinel) | Schwellenwert | Neue Methodik zur rechnerischen Auffälligkeitseinstufung ab Auswertungsjahr |
|---|--|------------------|---|
| Perkutane Koronarintervention (PCI) und Koronarangiographie (QS PCI) | Alle | $\alpha = 0,025$ | 2022 |
| Gynäkologische Operationen (QS GYN-OP) | Alle | $\alpha = 0,1$ | 2025 |
| Cholezystektomie (QS CHE) | Alle | $\alpha = 0,05$ | 2025 |
| Mammachirurgie (QS MC) | QI 50719: Lymphknotenentnahme bei DCIS und brusterhaltender Therapie | $\alpha = 0,15$ | 2025 |
| | Alle verbleibenden | $\alpha = 0,05$ | |
| Dekubitusprophylaxe (QS DEK) | QI 52009: Stationär erworbener Dekubitalulcus (ohne Dekubitalulcera Stadium/Kategorie 1) | $\alpha = 0,3$ | 2026 |
| Karotis-Revaskularisation (QS KAROTIS) | QI 161800: Postprozedurale fachneurologische Untersuchung | $\alpha = 0,15$ | 2026 |
| | Alle verbleibenden | $\alpha = 0,05$ | |

3.4 Beispiele

Zur weiteren Veranschaulichung wird die Wahl zweier Schwellenwerte aus *QS MC* genauer dargestellt; einmal für einen Indikator mit festem Referenzbereich (Abschnitt 3.4.1) und einmal für einen Indikator mit perzentilbasiertem Referenzbereich (Abschnitt 3.4.2).

3.4.1 Wahl bei festem Referenzbereich: QI 51846

QI 51846 betrachtet, wie häufig bei Ersteingriffen wegen invasivem Mammakarzinom oder DCIS eine prätherapeutische histologische Diagnosesicherung durch Stanz- oder Vakuumbiopsie erfolgt. Dieses Ereignis sollte möglichst häufig vorkommen. Der Indikator hat einen festen Referenzbereich, der bei einem Anteil von $\geq 97,00\%$ liegt. Abbildung 1 enthält einen Funnelplot mit den Ergebnissen dieses Indikators zum Auswertungsjahr 2024.

In den Proberechnungen auf Daten des Auswertungsjahres 2024 zeigte sich: Für $\alpha = 0,05$ reduzierte sich die Anzahl an rechnerischen Auffälligkeiten von 85 auf 37 relativ zur alten Methodik. Es fielen, wie erwartet (siehe Abschnitt 3.1), hauptsächlich Leistungserbringer mit kleineren Fallzahlen, also größerer zufallsbedingter Variabilität der Ergebnisse, aus der rechnerischen Auffälligkeit heraus; nämlich die grünen Punkte in Abbildung 1, die *unterhalb* des Referenzbereichs, also des hellgrünen Rechtecks, liegen, aber *oberhalb* der Funnel-Linie, also der schwarzen gestrichelten Linie. Zum Vergleich ist in Abbildung 1 als rote Linie die hypothetische Funnel-Linie bei einem Schwellenwert $\alpha = 0,025$ eingezeichnet. Man sieht: Es wären für diesen kleineren Schwellenwert 11 weitere, eher kleinere Leistungserbringer mit Ergebnis außerhalb des Referenzbereichs nicht mehr rechnerisch auffällig geworden.

Gleichzeitig fiel die Zahl der rechnerischen Auffälligkeiten nicht unter die Anzahl an qualitativen Auffälligkeiten im selben Auswertungsjahr (12), also nicht unter einen offensichtlich leistbaren Aufwand für die Qualitätsförderung. Der Schwellenwert $\alpha = 0,05$ war für diesen Indikator auch hoch genug, um eine angemessene Sensitivität (ca. $\frac{1}{2}$) zu gewährleisten: Die Wahrscheinlichkeit, einen Leistungserbringer mit besonders geringer Kompetenz θ (zugrunde liegende Häufigkeit des unerwünschten Ereignisses nicht bei $100\% - 97\% = 3\%$, wie vom Referenzbereich gefordert, sondern bei $2 \cdot 3\% = 6\%$) rechnerisch auffällig einzustufen, war ausreichend: im Mittel über alle Fallzahlen lag sie bei 47% (zum Vergleich: für $\alpha = 0,025$ läge sie bei lediglich 38%) und für die mediane Fallzahl im Indikator bei 43% , also jeweils nahe bei $\frac{1}{2}$.

Ein Schwellenwert $\alpha = 0,05$ war also ein angemessener Kandidat für den QI 51846 und fügte sich auch gut in ein weitestgehend einheitliches Vorgehen im *QS*-Verfahren *QS MC* ein (in fast allen Indikatoren im *QS*-Verfahren erschien $\alpha = 0,05$ als angemessener Kandidat für den Schwellenwert). Der Schwellenwert α wurde daher für das Auswertungsjahr 2025 im QI 51846 so festgelegt.

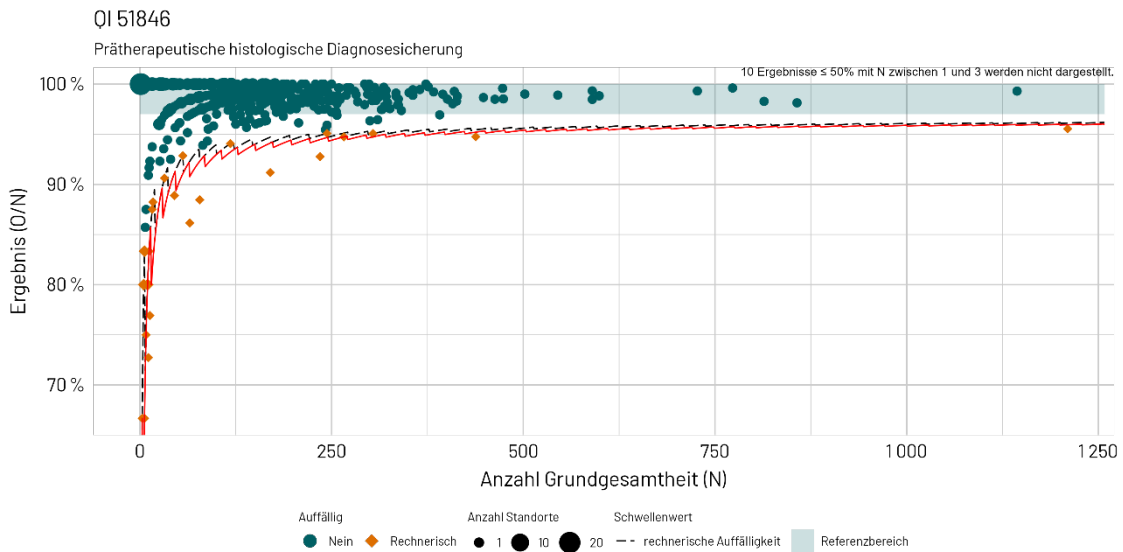


Abbildung 1: Funnelplot mit Daten aus dem Auswertungsjahr 2024 für QI 51846 mit festem Referenzbereich ($\geq 97,00\%$). Die rechnerische Auffälligkeit (Punkte in Orange) wird anhand der Funnel-Linie für Schwellenwert $\alpha = 0,05$ (schwarze gestrichelte Linie) bestimmt. Zum Vergleich ist ebenfalls eine Funnel-Linie für einen möglichen Schwellenwert $\alpha = 0,025$ eingezeichnet (rote Linie).

3.4.2 Wahl bei perzentilbasiertem Referenzbereich: QI 212001

QI 212001 betrachtet wie häufig Operationen mit präoperativer Drahtmarkierung durch Mammografie, Sonografie oder MRT in der Grundgesamtheit von brusterhaltenden Ersteingriffen vorkommen. Dieses Ereignis sollte möglichst häufig vorkommen. Der Indikator hat einen perzentilbasierten Referenzbereich am 5. Perzentil. Es werden also näherungsweise 5 % der Leistungserbringer rechnerisch auffällig, unabhängig vom Schwellenwert α . Der Schwellenwert α wirkt sich aber darauf aus, welche Leistungserbringer auffällig werden.

In den Proberechnungen auf Daten des Auswertungsjahres 2024 zeigte sich, dass für $\alpha = 0,05$ eine deutlich ausgeglichenerere Fallzahlverteilung bei den rechnerischen Auffälligkeiten im Indikator erreicht wird (grüne Balken in Abbildung 2), als mit der bisherigen Methodik der rechnerischen Auffälligkeit (rote Balken in Abbildung 2). Zudem ergibt sich eine Aufwandsreduktion durch genaueres Einhalten des vorgegebenen Perzentils, mit nun 5,1 % Anteil an Auffälligkeiten statt 12,3 % nach alter Methodik (gestrichelte Linien in Abbildung 2).⁷ Mit $\alpha = 0,05$ ist die Verteilung der Auffälligkeiten über den Fallzahlbereich näherungsweise ausgeglichen: In jedem Fallzahl-Quintil lag mit Umstellung auf die neue Methodik und $\alpha = 0,05$ der Anteil an rechnerischen

⁷ Die extreme Abweichung vom durch das 5. Perzentil geforderten Anteil an Auffälligkeiten (5 %) nach der bisherigen Methodik (12,3 %) zeigt deutlich die Problematik der nicht eingehaltenen Perzentile: der gewünschte bundesweite Aufwand für Stellungnahmeverfahren wird deutlich verfehlt. Der Grund ist, dass besonders kleine Leistungserbringer zwar für die Bestimmung des perzentilbasierten Referenzbereichs ausgeschlossen, allerdings für die Ermittlung der rechnerischen Auffälligkeiten daraufhin wieder eingeschlossen werden (siehe IQTIG 2020, Abschnitt 2.6.2). Es entsteht dadurch systematisch eine übermäßige Konzentration von Auffälligkeiten bei kleinen Leistungserbringern (siehe Abbildung 2). Mit der Umstellung auf die neue rechnerische Auffälligkeit kann dem entgegengewirkt werden.

Auffälligkeiten zwischen ca. 3 % und 9 % (Abbildung 2). Zum Vergleich: Eine weitere Reduktion des Schwellenwerts auf $\alpha = 0,025$ hätte nur unwesentlich eine noch ausgeglichene Verteilung auf den historischen Daten produziert (Anteil Auffälligkeiten in den Quintilen zwischen ca. 3 % und 7 %). Im Sinne der Einheitlichkeit innerhalb von QS MC (siehe Analysen zu QI 51846 in Abschnitt 3.4.1) wurde sich im QI 212001 daher für $\alpha = 0,05$ entschieden.

Lediglich bei einem Indikator, QI 50719, wurde im Einklang mit dem in diesem Dokument erläuterten Vorgehen von der Einheitlichkeit des Schwellenwert α innerhalb von QS MC abgewichen. Der Indikator mit festem Referenzbereich betrachtet nur eine sehr eingeschränkte Untergruppe an Patientinnen und Patienten des QS-Verfahrens. Dies führt für viele Leistungserbringer zu niedrigen Fallzahlen in der Grundgesamtheit. Für den Indikator war deshalb der höhere Schwellenwert $\alpha = 0,15$ notwendig, um die angemessene Sensitivität zu gewährleisten.

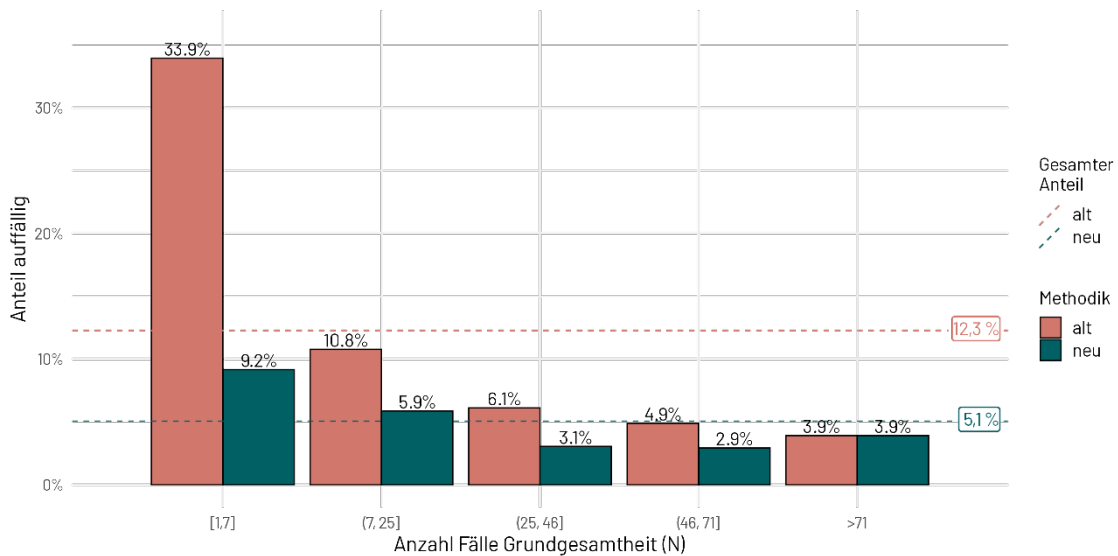


Abbildung 2: Relative Häufigkeit von rechnerischen Auffälligkeiten pro Fallzahlquintil mit Daten aus dem Auswertungsjahr 2024 für QI 212001 mit perzentilbasiertem Referenzbereich (5. Perzentil). Rote Balken stellen die Häufigkeit nach alter Methodik der rechnerischen Auffälligkeitseinstufung, grüne Balken stellen die Häufigkeit nach neuer Methodik der rechnerischen Auffälligkeitseinstufung mit Schwellenwert $\alpha = 0,05$ dar. Die gestrichelten Linien zeigen jeweils die Gesamthäufigkeit über alle Fallzahlquintile.

4 Ausblick

Das IQTIG plant, die festgelegten Schwellenwerte und auch das Vorgehen zur Wahl des Schwellenwerts α über die Zeit zu evaluieren und ggf. anzupassen. Beispielsweise ist eine Überprüfung des Schwellenwerts sinnvoll, wenn sich die rechnerische Definition oder der Referenzbereich eines Indikators substantiell verändert. Zudem sei daran erinnert, dass es in der Praxis keinen *per se* „korrekten“ Schwellenwert α eines Indikators gibt, sondern nur Setzungen, die mehr oder weniger angemessen für die Qualitätssicherung sind. Außerdem erarbeitet das IQTIG ein standardmäßiges Vorgehen für die Wahl des Schwellenwerts α bei neu eingeführten QS-Verfahren, die nicht auf historische Daten für Proberechnungen zurückgreifen können.

Literatur

- Austin, PC; Brunner, LJ (2008): Optimal Bayesian probability levels for hospital report cards. *Health Services and Outcomes Research Methodology* 8(2): 80–97. DOI: 10.1007/s10742-007-0025-4.
- Cederbaum, J; Rauh, J; Boywitt, D; Holleck-Weithmann, S; Zander-Jentsch, B (2025): Berücksichtigung und Reduktion statistischer Unsicherheit bei Auswertungen von Qualitätssicherungsdaten. Kapitel 13. In: Heidecke, C-D; Dingelstedt, A; Klein, S; Hrsg.: *Weißbuch datengestützte Qualitätssicherung im Gesundheitswesen*. Berlin: MWV [Medizinisch Wissenschaftliche Verlagsgesellschaft], 87–94. DOI: 10.32745/9783954669431-13.
- Hengelbrock, J; Rauh, J; Cederbaum, J; Kähler, M; Höhle, M (2023): Hospital profiling using Bayesian decision theory. *Biometrics* 79(3): 2757–2769. DOI: 10.1111/biom.13798.
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2020): Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern. Abschlussbericht zu Stufe 1 und Stufe 2. Stand: 11.02.2020. Berlin: IQTIG. URL: https://iqtig.org/downloads/berichte/2020/IQTIG_Weiterentwicklung-Strukturierter-Dialog-Stufe-1-u-2-Abschlussbericht_2020-02-11.pdf. [20.10.2022].
- IQTIG [Institut für Qualitätssicherung und Transparenz im Gesundheitswesen] (2024): Methodische Grundlagen. Version 2.1. [Stand:] 27.11.2024. Berlin: IQTIG. URL: https://iqtig.org/downloads/berichte/2024/IQTIG_Methodische-Grundlagen_Version-2.1_2024-11-27.pdf (abgerufen am: 07.01.2026).
- Maier, M; Lakens, D (2022): Justify Your Alpha: A Primer on Two Practical Approaches. *Advances in Methods and Practices in Psychological Science* 5(2): 25152459221080396. DOI: 10.1177/25152459221080396.
- WIdO [Wissenschaftliches Institut der AOK] (2007): Qualitätssicherung der stationären Versorgung mit Routinedaten (QSR). Abschlussbericht. Bonn: AOK-Bundesverband. ISBN: 978-3-922093-42-8. URL: https://www.wido.de/fileadmin/Dateien/Dokumente/Publikationen_Produkte/QSR/wido_qsr_abschlussbericht_0407.pdf (abgerufen am: 07.01.2026).