

Rechnerische Auffälligkeitseinstufung und perzentilbasierte Referenzbereiche

Umstellung der Methodik im Verfahren *Perkutane Koronarintervention und Koronarangiographie (QS PCI)*

Stand: 7. Juli 2023; Ansprechpartner: Verfahrenssupport (verfahrenssupport@iqtig.org)

Inhalt

1	Warum wurde die Methodik angepasst?	3
2	Woher kommt die (zufallsbedingte) Variabilität bei Indikatorergebnissen?	4
3	Wie lässt sich die Variabilität bei Indikatorergebnissen messen?	5
4	Wie funktioniert die neue Methodik zur rechnerischen Auffälligkeitseinstufung?.....	7
5	Was ist der zugrunde liegende Indikatorwert?	8
6	Was bedeuten Unsicherheitsintervalle?.....	9
7	Was sind perzentilbasierte Referenzbereiche?	10
8	Was ändert sich bei perzentilbasierten Referenzbereichen durch die neue Methodik zur rechnerischen Auffälligkeitseinstufung?	10
9	Wie hängt die neue Methodik zur rechnerischen Auffälligkeitseinstufung mit der neuen Methodik für perzentilbasierte Referenzbereiche zusammen?	11
10	Warum werden nach der neuen Methodik zur Bestimmung der perzentilbasierten Referenzbereiche weniger Leistungserbringer auffällig?	12
11	Warum sind die perzentilbasierten Referenzbereiche nach der neuen Methodik strenger?.....	12
12	Warum führt die neue Methodik zu einer Verschiebung der Auffälligkeiten hin zu Leistungserbringern mit größeren Fallzahlen?	13
13	Werden durch die neue Methodik Leistungserbringer mit größeren Fallzahlen systematisch benachteiligt?.....	14

Einleitung

Der Begriff *rechnerische Auffälligkeit* beschreibt in der externen Qualitätssicherung solche Indikatorergebnisse, die darauf hindeuten, dass ein Leistungserbringer ein Qualitätsziel verfehlt. Bei einer rechnerischen Auffälligkeit sieht die Richtlinie zur datengestützten einrichtungsübergreifenden Qualitätssicherung (DeQS-RL)¹ vor, dass der betroffene Leistungserbringer einem Stellungnahmeverfahren unterzogen wird. Das Adjektiv *rechnerisch* bedeutet hier, dass bei der Einstufung als rechnerisch auffällig oder unauffällig ausschließlich die Zahlenwerte des Indikatorergebnisses (insbesondere die Anzahlen der Zähler- und Nennerfälle) als Informationen über den Leistungserbringer berücksichtigt werden. Demgegenüber stehen im Stellungnahmeverfahren weitere Informationen zur Verfügung, insbesondere die Stellungnahme des Leistungserbringers.

Zur Auswertung im Jahr 2022 hat das IQTIG die Methodik zur rechnerischen Auffälligkeitseinstufung im Verfahren *Perkutane Koronarintervention und Koronarangiographie (QS PCI)* geändert. Ebenso wurde die Methodik zur Bestimmung der perzentilbasierten Referenzbereiche angepasst. Dieses Dokument soll die wichtigsten Fragen in diesem Zusammenhang klären. Weiterführende Informationen stellt das IQTIG auf der Webseite <https://iqtig.org/das-iqtig/wie-wir-arbeiten/grundlagen/biometrische-grundlagen/biometrische-methodik-zur-auffaelligkeitseinstufung/> zur Verfügung.

Die dargestellte Methodik bezieht sich auf die klassischen Indikatoren des Verfahrens *QS PCI*, die auf der QS-Dokumentation der Leistungserbringer und den Sozialdaten bei den Krankenkassen beruhen. Seit der Auswertung im Jahr 2023 werden zusätzlich auch Indikatoren basierend auf einer Patientenbefragung ausgewertet. Diese beruhen auf Indexkonstruktionen, die Antworten aus Fragebögen aggregieren. Die Auswertungsmethodik des IQTIG für die Indikatoren der Patientenbefragung ist konsistent mit der hier vorgestellten Methodik. Die Indikatoren der Patientenbefragung sind jedoch komplexer und werden daher hier nicht thematisiert. Weitere Informationen zur Auswertung von Indikatoren der Patientenbefragung finden sich auf folgender Webseite des IQTIG: <https://iqtig.org/veroeffentlichungen/auswertungsmethodik-qi-patientenbefragungen/>.

¹ Richtlinie des Gemeinsamen Bundesausschusses zur datengestützten einrichtungsübergreifenden Qualitätssicherung. In der Fassung vom 19. Juli 2018, zuletzt geändert am 15. Dezember 2022, in Kraft getreten am 1. Januar 2023. URL: <https://www.g-ba.de/richtlinien/105/> (abgerufen am: 21.06.2023).

1 Warum wurde die Methodik angepasst?

Die methodischen Anpassungen verfolgen hauptsächlich drei Ziele:

1. Die Anzahl an rechnerischen Auffälligkeiten bei Leistungserbringern ohne Qualitätsdefizit soll verringert werden.
2. Qualitätsdefizite sowohl bei Leistungserbringern mit großer Fallzahl als auch bei Leistungserbringern mit kleiner Fallzahl sollen ausreichend im Fokus der Qualitätssicherung liegen. Die alte Methodik führt hingegen systematisch (d. h. aus rein statistischen Gründen) zu vielen rechnerischen Auffälligkeiten bei Leistungserbringern mit kleiner Fallzahl.
3. Der Anteil an rechnerischen Auffälligkeiten soll sich möglichst genau am gewählten Perzentil orientieren. Die alte Methodik der perzentilbasierten Referenzbereiche führt hingegen zu mehr rechnerischen Auffälligkeiten, als das Perzentil vorgibt.

Im Folgenden wird kurz erläutert, inwiefern die neue Methodik diese drei Ziele besser erreicht als die alte Methodik.

Bei der bislang verwendeten Methodik der rechnerischen Auffälligkeit wird nicht berücksichtigt, dass die beobachteten Indikatorergebnisse zufallsbedingter Variabilität unterliegen (siehe „Woher kommt die (zufallsbedingte) Variabilität bei Indikatorergebnissen?“; siehe auch „Was ist der zugrunde liegende Indikatorwert?“). Diese Variabilität ist bei kleinen Fallzahlen größer als bei großen Fallzahlen. Bei der neuen statistischen Methodik zur rechnerischen Auffälligkeit wird diese Fallzahlabhängigkeit explizit berücksichtigt. Dadurch werden Leistungserbringer, bei denen ein schlechtes beobachtetes Indikatorergebnis auch plausibel durch zufallsbedingte Variabilität erklärt werden kann, seltener rechnerisch auffällig (Ziel 1).

Da die zufallsbedingte Variabilität bei kleinen Fallzahlen größer ist, werden nach der alten Methodik Leistungserbringer mit kleiner Fallzahl mit größerer Wahrscheinlichkeit auffällig als Leistungserbringer mit großer Fallzahl. Mit der neuen statistischen Methodik zur Auffälligkeitseinstufung ist das Verhältnis von rechnerischen Auffälligkeiten bei Leistungserbringern mit kleiner Fallzahl und solchen mit großer Fallzahl in der Regel ausgeglichener (Ziel 2).

Durch die Anpassung der Methodik zur Bestimmung des perzentilbasierten Referenzbereichs wird sichergestellt, dass die Anzahl an rechnerischen Auffälligkeiten zum gewählten Perzentil passt (Ziel 3). Perzentilbasierte Referenzbereiche wie beispielsweise $\leq 95.$ Perzentil oder $\geq 5.$ Perzentil werden mit der neuen Methodik so berechnet, dass in etwa 95 % der Leistungserbringer rechnerisch unauffällig sind. Entsprechend werden in etwa 5 % der Leistungserbringer rechnerisch auffällig. Nach der alten Methodik war hingegen die Anzahl an rechnerischen Auffälligkeiten regelmäßig deutlich größer (siehe „Warum werden nach der neuen Methodik zur Bestimmung der perzentilbasierten Referenzbereiche weniger Leistungserbringer auffällig?“). Somit wird durch die neue Methodik bei gleichbleibendem Perzentil die Anzahl an ausgelösten Stellungsverfahren reduziert.

Als weiterer Vorteil stellt die neue Methodik die Konsistenz zur Auswertungsmethodik für die Indikatoren der Patientenbefragung sicher (siehe <https://iqtig.org/veroeffentlichungen/auswertungsmethodik-qi-patientenbefragungen/>).

2 Woher kommt die (zufallsbedingte) Variabilität bei Indikatorergebnissen?

Wie bei jedem technischen oder sozialen Prozess gibt es in der Gesundheitsversorgung (zufallsbedingte) Variabilität. Prozesse können nur innerhalb gewisser Toleranzen konstant gehalten werden. Dadurch kommt es auch zu Variabilität bei Ergebnissen. Hinzu kommen Unterschiede zwischen den behandelten Patientinnen und Patienten. Dies führt dazu, dass beobachtete Indikatorwerte variieren, auch wenn die Behandlungsqualität sich im Wesentlichen nicht ändert.

Diese Variabilität wird in der Statistik als Zufall modelliert. Dieser Zufall spielt sowohl bei Ergebnisindikatoren als auch bei Prozess- und Indikationsindikatoren eine Rolle.

Nicht immer kann der Grund für Variabilität objektiv einfach angegeben werden. Es gibt hauptsächlich drei Ursachen für Variabilität:

1. Verschiedene behandelte Patientinnen und Patienten unterscheiden sich in Eigenschaften, die für den Prozess oder das Ergebnis der Behandlung relevant sind. Insofern die Eigenschaften einfach erhebbar sind, können sie bei der Risikoadjustierung berücksichtigt werden. Nicht alle Risikofaktoren liegen jedoch in der Dokumentation vor, etwa weil der Aufwand der Erhebung unverhältnismäßig ist, der Risikofaktor sehr selten auftritt oder der Risikofaktor für die Therapie nicht relevant ist. Beispielsweise ist für die Behandlung einer Infektion die genaue Bestimmung des Erregers nicht immer nötig.
2. Die Prozesse beim Leistungserbringer unterliegen Schwankungen, auch wenn sich die Behandlungsqualität nicht grundsätzlich ändert. Dies kann beispielsweise zum einen an unterschiedlichen zeitlichen Abläufen (wie Wartezeiten) der Behandlung liegen. Zum anderen können verschiedene Patientinnen und Patienten von unterschiedlichem Personal behandelt werden, und die Zuordnung, welches Personal sich um welche Fälle kümmert, unterliegt dabei neben systematischen Einflüssen (wie Fallschwere sowie Spezialisierung und Erfahrung des Personals) auch zufälligen Einflüssen.
3. Schließlich gibt es Ursachen, die weder direkt bei der Patientin oder dem Patienten noch beim Leistungserbringer liegen. Dazu zählen beispielsweise Eigenschaften von Medizinprodukten (wie Messgeräte) oder medizinischer Präparate (wie beispielsweise Transplantationsorgane oder Herzschrittmacher).

3 Wie lässt sich die Variabilität bei Indikatorergebnissen messen?

Variabilität ist nur schwierig direkt zu beobachten, da dieselbe Behandlung bei derselben Patientin oder demselben Patienten nicht wiederholt werden kann. Indirekt lässt sich Variabilität zum einen durch den Vergleich von Indikatorergebnissen bei verschiedenen Leistungserbringern und zum anderen durch den Vergleich von Indikatorergebnissen im zeitlichen Verlauf darstellen. Wie im Folgenden dargestellt, ist bei beiden Vergleichen die Variabilität bei großen Fallzahlen kleiner als bei kleinen Fallzahlen.

Abbildung 1 zeigt in einem sogenannten *Funnelplot* die Indikatorergebnisse sämtlicher Leistungserbringer eines Jahres im Indikator „Erreichen des wesentlichen Interventionsziels bei PCI mit der Indikation ST-Hebungsinfarkt“ (ID 56014). Auf der x-Achse ist die Fallzahl aufgetragen, auf der y-Achse das Indikatorergebnis. Es ist deutlich zu erkennen, dass die Ergebnisse weiter links in der Abbildung stärker streuen als weiter rechts. Dies liegt nicht daran, dass die Behandlungsqualität bei Leistungserbringern mit kleiner Fallzahl heterogener ist. Vielmehr ist dies eine Konsequenz der fallzahlabhängigen Variabilität. Die gestrichelte Linie, die die rechnerisch auffälligen Ergebnisse von den unauffälligen Ergebnissen trennt, zeigt an, welche Abweichungen vom Referenzwert noch plausibel durch zufallsbedingte Variabilität erklärt werden können. Es ist in der Abbildung erkennbar, dass die beobachtete Variabilität nicht viel größer ist als die erwartbare zufallsbedingte Variabilität: Nur wenige Ergebnisse liegen unterhalb der gestrichelten Linie und sind damit rechnerisch auffällig. Auch liegen die meisten rechnerisch auffälligen Ergebnisse nahe an der gestrichelten Linie.

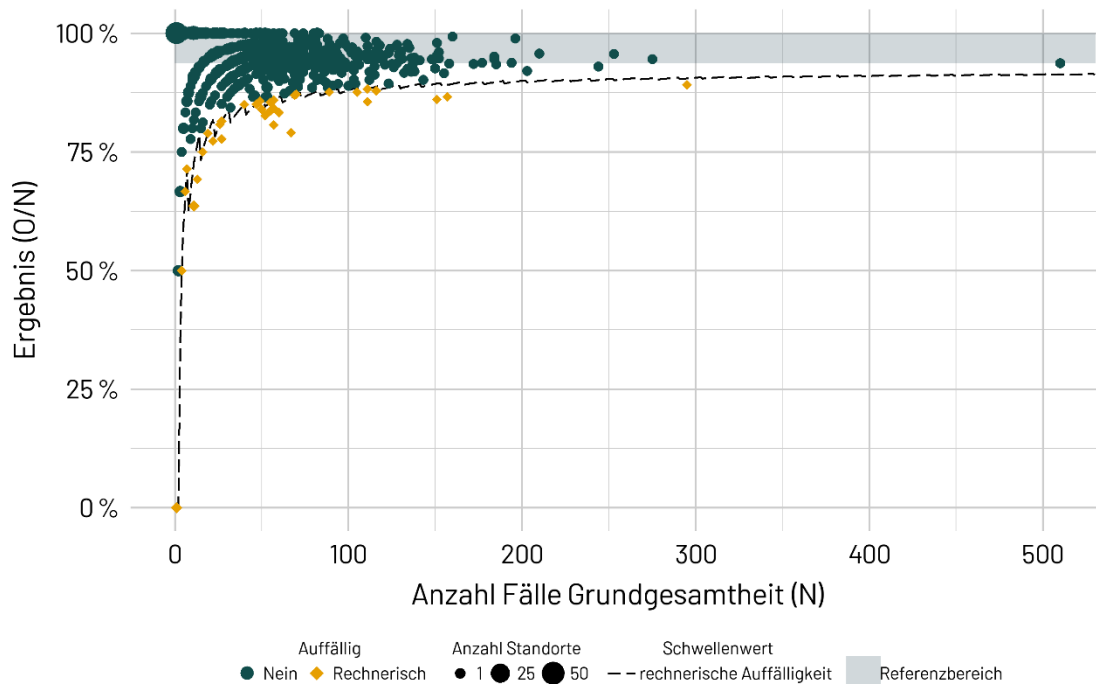


Abbildung 1: Funnelplot für den Indikator „Erreichen des wesentlichen Interventionsziels bei PCI mit der Indikation ST-Hebungsinfarkt“ (ID 56014)

Abbildung 2 zeigt einen Funnelplot für den risikoadjustierten Indikator „MACCE² innerhalb von 7 Tagen bei Patientinnen und Patienten mit isolierter Koronarangiographie“ (ID 56018). Im Unterschied zu Abbildung 1 ist hier auf der x-Achse die erwartete Anzahl (E) an interessierenden Ereignissen (in diesem Fall MACCE) aufgetragen. Dies liegt daran, dass die erwartete Anzahl an interessierenden Ereignissen direkter die Variabilität bestimmt als die Fallzahl. Da die erwartete Anzahl an interessierenden Ereignissen stark mit der Fallzahl korreliert, sind auch in dieser Abbildung Leistungserbringer mit großer Fallzahl tendenziell weiter rechts positioniert als solche mit kleiner Fallzahl. Auch in dieser Abbildung ist deutlich erkennbar, dass die Variabilität der Ergebnisse von links nach rechts abnimmt.

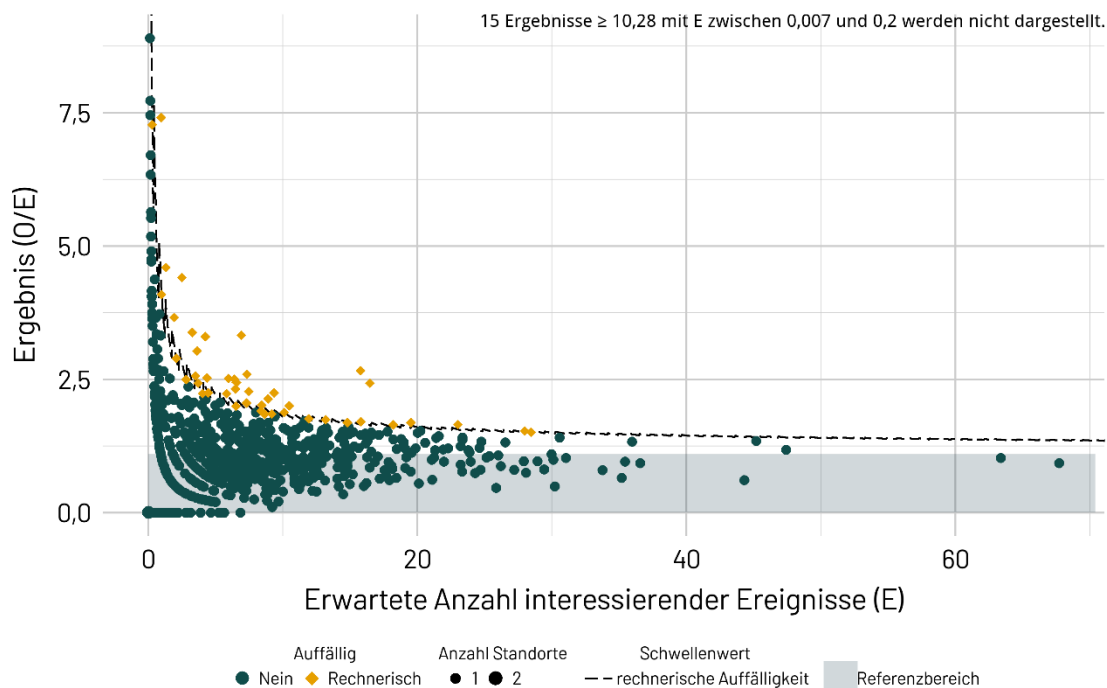


Abbildung 2: Funnelplot für den Indikator „MACCE innerhalb von 7 Tagen bei Patientinnen und Patienten mit isolierter Koronarangiographie“ (ID 56018)

Außerdem kann man betrachten, wie sich Indikatorergebnisse von Jahr zu Jahr ändern. Abbildung 3 zeigt beispielhaft Zeitreihen von Indikatorergebnissen verschiedener Leistungserbringer. Veränderungen am Indikatorergebnis von Jahr zu Jahr können entweder auf zufallsbedingte Variabilität zurückgeführt werden oder auf systematische Veränderungen im Behandlungsprozess (z. B. Veränderungen der Behandlungsqualität oder aber auch systematische Änderungen am Patientenkollektiv). Da die Variabilität umso größer ist, je geringer die Fallzahl ist, überwiegt bei Indikatorergebnissen mit kleiner Fallzahl in der Regel die zufallsbedingte Variabilität. Bei größeren Fallzahlen, insbesondere bei den Bundesergebnissen der meisten Indikatoren, ist hingegen die

² MACCE = Schwere kardiale und zerebrovaskuläre Komplikationen (major adverse cardiac and cerebrovascular events), intra- und postprozedural: Schlaganfall, Herzinfarkt, Tod.

Variabilität geringer, sodass systematische Trends einfacher nachgewiesen werden können. Die größere Variabilität der Indikatorergebnisse bei kleinen Fallzahlen ist in Abbildung 3 deutlich erkennbar.

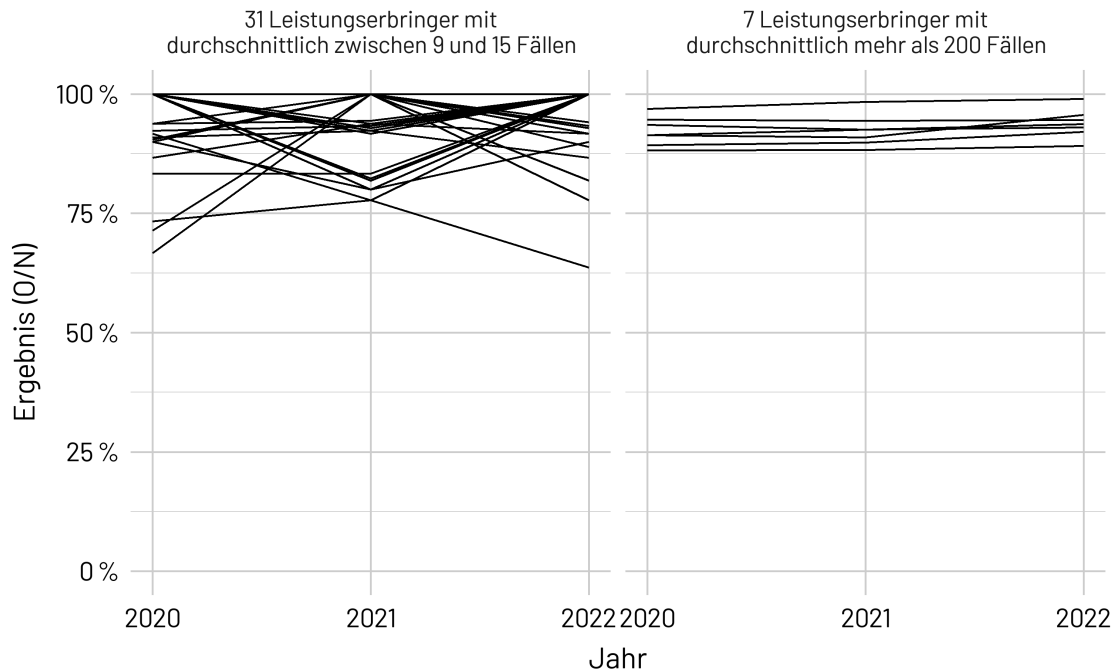


Abbildung 3: Zeitverläufe der Indikatorergebnisse von beispielhaften Leistungserbringern

4 Wie funktioniert die neue Methodik zur rechnerischen Auffälligkeitseinstufung?

Bei der neuen Methodik zur Auffälligkeitseinstufung wird für jedes beobachtete Indikatorergebnis, welches außerhalb des Referenzbereichs liegt, analysiert: Ist es plausibel, dass die Abweichung vom Referenzbereich durch zufallsbedingte Variabilität erklärt werden kann? Zur Quantifizierung dieser Plausibilität wird eine (Bayesianische) A-posteriori-Wahrscheinlichkeit berechnet, dass der zugrunde liegende Indikatorwert des Leistungserbringers im Referenzbereich liegt (siehe „Was ist der zugrunde liegende Indikatorwert?“). Die A-posteriori-Wahrscheinlichkeit wird mit einem Schwellenwert verglichen, der im Verfahren QS PCI auf 2,5 % gesetzt wurde: Ist die A-posteriori-Wahrscheinlichkeit kleiner gleich 2,5 %, so wird das Ergebnis als rechnerisch auffällig gewertet. Bei der alten Methodik zur Auffälligkeitseinstufung führt hingegen jedes beobachtete Indikatorergebnis außerhalb des Referenzbereichs zu einer rechnerischen Auffälligkeit.

Beispielsweise liegt ein Indikatorergebnis von $44/50 = 88\%$ außerhalb des Referenzbereichs von $\geq 93,65\%$ in Abbildung 1. Dieses Ergebnis besitzt eine A-posteriori-Wahrscheinlichkeit von 6,1%. Dieses Ergebnis wird also als unauffällig eingestuft und entspricht einem grünen Punkt in der Ab-

bildung. Bei einem Indikatorergebnis von $43/50 = 86\%$, das ebenfalls außerhalb des Referenzbereichs liegt, beträgt die A-posteriori-Wahrscheinlichkeit hingegen $2,2\%$. Dieses Ergebnis wird damit als auffällig eingestuft und entspricht einer gelben Raute.

Die A-posteriori-Wahrscheinlichkeit verhält sich in vielerlei Hinsicht ähnlich dem p -Wert aus der frequentistischen Statistik. Ein solcher p -Wert wird auch im Rahmen der Richtlinie zu planungsrelevanten Qualitätsindikatoren (plan. QI-RL)³ bei der *statistischen Auffälligkeitseinstufung* verwendet.⁴ Darüber hinaus hat sich das IQTIG unter anderem deswegen in der DeQS-RL für eine Bayesianische Methodik entschieden, weil dadurch Konsistenz mit der Methodik zur Auswertung der Patientenbefragungsindikatoren hergestellt wird.

5 Was ist der zugrunde liegende Indikatorwert?

Aufgrund von zufallsbedingter Variabilität ist das Indikatorergebnis eines Leistungserbringers kein direktes Abbild seiner Behandlungsqualität. Da die Variabilität mit zunehmender Fallzahl geringer wird, kann man sich im Gedankenexperiment fragen: Wie würde sich das Indikatorergebnis eines Leistungserbringers entwickeln, wenn er immer mehr Patientinnen und Patienten behandeln würde? Unter der Annahme, dass sich dabei die Behandlungsqualität des Leistungserbringers nicht verändert, würde die Variabilität immer kleiner werden und das Indikatorergebnis gegen einen festen Wert streben. Dieser Wert wird auch *zugrunde liegender Indikatorwert* genannt. Dieser zugrunde liegende Indikatorwert ist ein genaueres Abbild der Behandlungsqualität bzw. der Kompetenz des Leistungserbringers.⁵ Abbildung 4 illustriert diese Konvergenz des beobachteten Indikatorergebnis gegen den zugrunde liegenden Indikatorwert.

Wäre der zugrunde liegende Indikatorwert bekannt, wäre dieser eine gute Grundlage für die Auffälligkeitseinstufung. Da der zugrunde liegende Indikatorwert nicht bekannt ist, muss die Auffälligkeitseinstufung auf der Grundlage des beobachteten Indikatorwerts geschehen. Dabei sollte jedoch die Variabilität des beobachteten Indikatorwerts berücksichtigt werden.

³ Richtlinie zu planungsrelevanten Qualitätsindikatoren. In der Fassung vom 15. Dezember 2016, zuletzt geändert am 15. Dezember 2022, in Kraft getreten am 1. Januar 2023. URL: <https://www.g-ba.de/informationen/richtlinien/91/> (abgerufen am: 05.07.2023).

⁴ Genauer gesagt wird bei den planungsrelevanten Qualitätsindikatoren ein *exakter midp-Wert* verwendet und mit dem Schwellenwert 5% verglichen; siehe <https://iqtig.org/veroeffentlichungen/planqi-auswahl-umsetzung/>.

⁵ Der zugrunde liegende Indikatorwert ist jedoch nicht mit der Behandlungsqualität bzw. Kompetenz des Leistungserbringers gleichzusetzen. Wie gut der zugrunde liegende Indikatorwert tatsächlich die Behandlungsqualität misst, hängt von weiteren Eigenschaften des Indikators ab: Wie gut ist die Datenqualität? Ist der Indikator gut operationalisiert? Diese Eigenschaften werden genauer in Kapitel 13 der „Methodischen Grundlagen“ des IQTIG diskutiert (siehe <https://iqtig.org/das-iqtig/wie-wir-arbeiten/grundlagen/standard-titel/>).

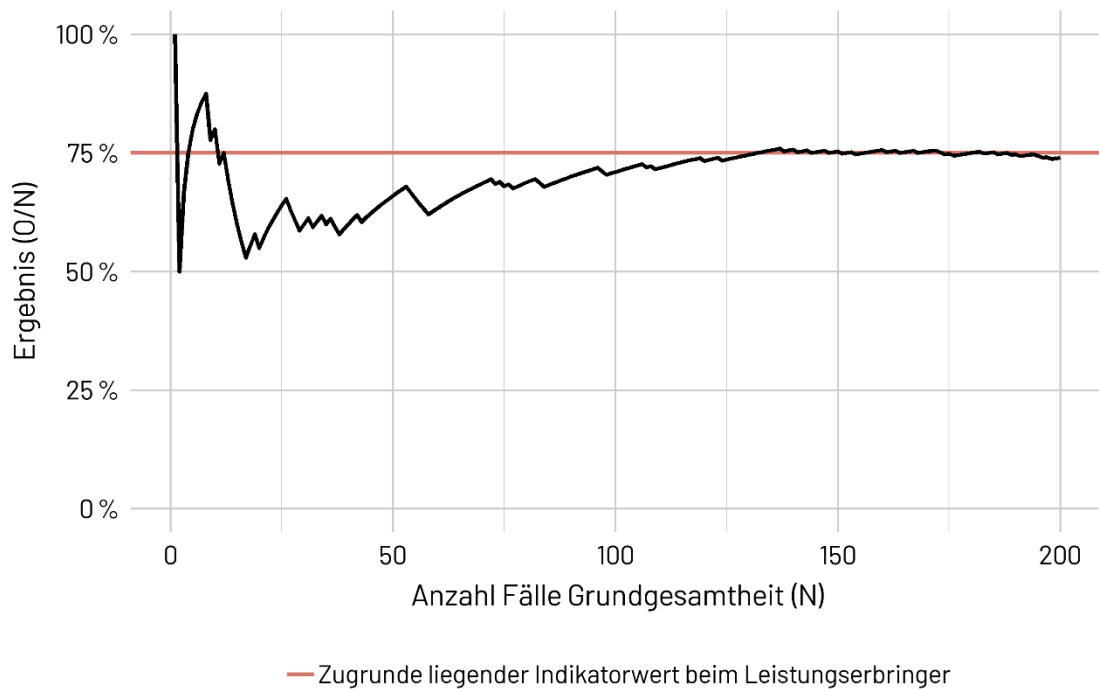


Abbildung 4: Das Indikatorergebnis eines beispielhaften Leistungserbringers im Verlauf, wenn das Ergebnis nach jedem Fall neu berechnet wird

6 Was bedeuten Unsicherheitsintervalle?

Der beobachtete Indikatorwert weicht in der Regel vom zugrunde liegenden Indikatorwert ab. Das Unsicherheitsintervall ist ein Intervall, von dem man aufgrund theoretischer Überlegungen erwartet, dass es den zugrunde liegenden Indikatorwert mit einer gewissen Sicherheit enthält (siehe auch „Was ist der zugrunde liegende Indikatorwert?“). Der Begriff „Unsicherheit“ bezieht sich dabei auf den Umstand, dass aufgrund von Variabilität der zugrunde liegende Indikatorwert nicht mit Sicherheit bekannt ist.

Die Sicherheit, mit der der zugrunde liegende Indikatorwert im Unsicherheitsintervall erwartet wird, kann durch einen Parameter namens *Glaubwürdigkeitsniveau* eingestellt werden. Im Verfahren QS PCI liegt dieser bei 95 %. Dies bedeutet: Bei in etwa 95 % der Leistungserbringer liegt der zugrunde liegende Indikatorwert im Unsicherheitsintervall. Größere Werte des Glaubwürdigkeitsniveaus führen zu breiteren Intervallen, denn breitere Intervalle enthalten den zugrunde liegenden Indikatorwert mit größerer Sicherheit.

Die Methodik der Unsicherheitsintervalle⁶ und das Glaubwürdigkeitsniveau sind so gewählt, dass sich der folgende Zusammenhang zur rechnerischen Auffälligkeitseinstufung ergibt (siehe auch

⁶ Es handelt sich um Bayesianische Glaubwürdigkeitsintervalle.

„Wie funktioniert die neue Methodik zur rechnerischen Auffälligkeitseinstufung?“): Ein Indikatorergebnis ist genau dann rechnerisch auffällig, wenn das Unsicherheitsintervall außerhalb des Referenzbereichs liegt.

7 Was sind perzentilbasierte Referenzbereiche?

Das Ziel von perzentilbasierten Referenzbereichen ist die Steuerung der Anzahl an rechnerischen Auffälligkeiten. Dazu dient das sogenannte Perzentil. Ist beispielsweise das 5. Perzentil (bei einem Referenzbereich $\geq 5.$ Perzentil) oder das 95. Perzentil (bei einem Referenzbereich $\leq 95.$ Perzentil) als Referenzwert bestimmt, so führt dies dazu, dass in etwa 5 % der Leistungserbringer rechnerisch auffällig werden. Bei perzentilbasierten Referenzbereichen ist die Lage des Referenzbereichs also nicht von vorneherein festgelegt: Der Referenzbereich kann erst berechnet werden, wenn für alle Leistungserbringer Indikatorergebnisse vorliegen.

Dazu werden die Leistungserbringer nach ihren Indikatorergebnissen sortiert. Die 5 % „auffälligsten“ Ergebnisse werden als rechnerisch auffällig deklariert. Alle anderen Ergebnisse sind rechnerisch unauffällig. Damit das gesetzte Perzentil eingehalten wird, muss sich die Sortierung der Leistungserbringer dabei nach der Methodik zur Auffälligkeitseinstufung richten (siehe „Was ändert sich bei perzentilbasierten Referenzbereichen durch die neue Methodik zur rechnerischen Auffälligkeitseinstufung?“).

8 Was ändert sich bei perzentilbasierten Referenzbereichen durch die neue Methodik zur rechnerischen Auffälligkeitseinstufung?

Bei Verwendung der neuen Methodik zur rechnerischen Auffälligkeitseinstufung wird bei der Sortierung der Leistungserbringer für die Perzentilbildung die Variabilität der Indikatorergebnisse berücksichtigt. Konkret bedeutet dies bei einem Referenzbereich von $\leq 95.$ Perzentil: Die Leistungserbringer werden aufsteigend nach der unteren Grenze ihres jeweiligen Unsicherheitsintervalls sortiert (siehe auch „Was bedeuten Unsicherheitsintervalle?“). Die 5 % Leistungserbringer mit den größten Werten werden rechnerisch auffällig.⁷ Der Referenzwert selbst wird nun so berechnet, dass er in der Mitte zwischen der unteren Grenze des Unsicherheitsintervalls des letzten unauffälligen Leistungserbringers und der unteren Grenze des Unsicherheitsintervalls des ersten auffälligen Leistungserbringers liegt.

⁷ Es ist nicht immer möglich, dass genau 5 % der Leistungserbringer rechnerisch auffällig werden. Dies liegt einmal daran, dass die Anzahl an Leistungserbringern nicht durch 20 teilbar sein muss. Außerdem kann es sein, dass verschiedene Leistungserbringer dasselbe Ergebnis haben. Daher wird der Referenzwert so gewählt, dass der Anteil an auffälligen Leistungserbringern möglichst genau 5 % beträgt.

Bei einem Referenzbereich von ≥ 5 . Perzentil wird analog die obere Grenze des Unsicherheitsintervalls betrachtet. Das heißt, die 5 % Leistungserbringer mit den niedrigsten Werten für die obere Grenze des Unsicherheitsintervalls werden rechnerisch auffällig.

9 Wie hängt die neue Methodik zur rechnerischen Auffälligkeitseinstufung mit der neuen Methodik für perzentilbasierte Referenzbereiche zusammen?

Nach der alten Methodik für perzentilbasierte Referenzbereiche wurden bei der Perzentilberechnung Indikatorergebnisse mit einer Fallzahl kleiner als 20 ausgeschlossen. Der Grund für den Ausschluss hat mit der fallzahlabhängigen Variabilität zu tun: Da die Variabilität bei kleinen Fallzahlen größer ist, treten extreme Indikatorergebnisse häufiger auf als bei großen Fallzahlen. Werden die Indikatorergebnisse ohne Berücksichtigung der Variabilität sortiert, befinden sich unter den 5 % besten Indikatorergebnissen überproportional viele Leistungserbringer mit kleiner Fallzahl. Dies ist beispielsweise in Abbildung 5 sichtbar: Sowohl unter den besten als auch unter den schlechtesten Indikatorergebnissen sind Leistungserbringer mit kleiner Fallzahl überrepräsentiert. Der Ausschluss von Leistungserbringern mit weniger als 20 Fällen bei der Berechnung von perzentilbasierten Referenzbereichen soll verhindern, dass nicht nur Leistungserbringer mit wenigen Fällen rechnerisch auffällig werden können. Da die zufallsbedingte Variabilität von Indikatorergebnissen bereits in der neuen Methodik zur Auffälligkeitseinstufung berücksichtigt wird, ist dieser Ausschluss nicht länger nötig. Der Verzicht auf einen Ausschluss führt dazu, dass das Perzentil genauer eingehalten wird (siehe „Was sind perzentilbasierte Referenzbereiche?“).

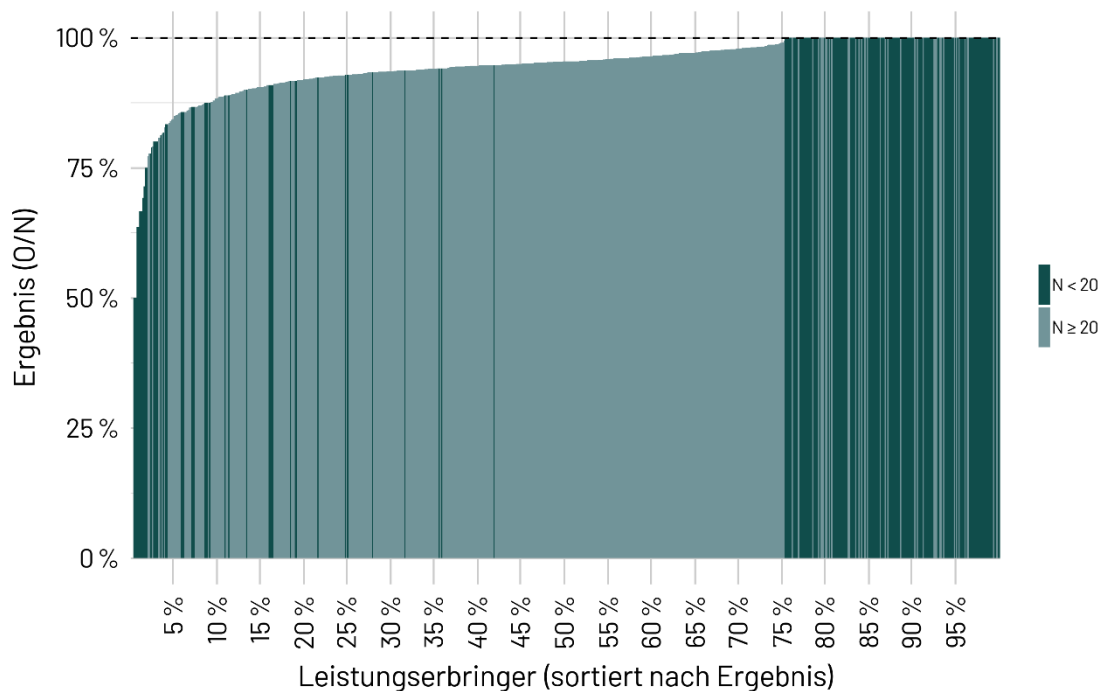


Abbildung 5: Die sortierten Leistungserbringerergebnisse im Indikator „Erreichen des wesentlichen Interventionsziels bei PCI mit der Indikation ST-Hebungsinfarkt“ (ID 56014)

10 Warum werden nach der neuen Methodik zur Bestimmung der perzentilbasierten Referenzbereiche weniger Leistungserbringer auffällig?

Nach der alten Methodik zur Bestimmung der perzentilbasierten Referenzbereiche wurden bei der Berechnung des Referenzwerts nicht sämtliche Indikatorergebnisse berücksichtigt: Indikatorergebnisse von Leistungserbringern mit weniger als 20 Fällen in der Grundgesamtheit wurden von der Berechnung ausgeschlossen (siehe „Wie hängt die neue Methodik zur rechnerischen Auffälligkeitseinstufung mit der neuen Methodik für perzentilbasierte Referenzbereiche zusammen?“). Der Referenzbereich, der bei den Leistungserbringern mit mindestens 20 Fällen berechnet wurde, wurde dann auch auf die anderen Leistungserbringer mit weniger Fällen angewendet.

In der alten Methodik ist also der Anteil an rechnerischen Auffälligkeiten unter den Leistungserbringern mit mindestens 20 Fällen kontrolliert: Unter diesen Leistungserbringern werden in etwa 5 % rechnerisch auffällig (wenn der Referenzwert auf dem 5. oder 95. Perzentil liegt). Unter den Leistungserbringern mit weniger als 20 Fällen ist der Anteil an rechnerischen Auffälligkeiten nicht kontrolliert und liegt in der Regel über 5 %. Dies liegt an der größeren zufallsbedingten Variabilität bei kleinen Fallzahlen. In der Praxis lag der Anteil an rechnerischen Auffälligkeiten bei den meisten Indikatoren im Verfahren QS PCI zwischen 5 und 9 %, je nachdem, wie die Fallzahlen verteilt waren. In der neuen Methodik ist hingegen der Anteil an rechnerischen Auffälligkeiten unter allen Leistungserbringern kontrolliert und liegt daher stets nahe bei 5 %.

11 Warum sind die perzentilbasierten Referenzbereiche nach der neuen Methodik strenger?

Perzentilbasierte Referenzbereiche werden jedes Jahr so neu bestimmt, dass der Anteil an rechnerischen Auffälligkeiten in etwa einem fest gegebenen Wert beträgt (in der Regel 5 %). Bei der neuen Methodik der rechnerischen Auffälligkeitseinstufung können Indikatorergebnisse außerhalb des Referenzbereichs als rechnerisch unauffällig bewertet werden (siehe „Wie funktioniert die neue Methodik zur rechnerischen Auffälligkeitseinstufung?“). Damit bei der Umstellung der Methodik zur rechnerischen Auffälligkeit der Anteil an rechnerischen Auffälligkeiten in etwa konstant bleibt, muss der Referenzbereich strenger werden.⁸ Abbildung 6 vergleicht den Wert des perzentilbasierten Referenzbereichs nach der alten und nach der neuen Methodik in einem Fun-

⁸ Genau genommen ist der Anteil an rechnerischen Auffälligkeiten nach der alten Methodik für perzentilbasierte Referenzbereiche in der Realität größer, als das Perzentil suggeriert (siehe „Warum werden nach der neuen Methodik zur Bestimmung der perzentilbasierten Referenzbereiche weniger Leistungserbringer auffällig?“). Dies kompensiert in gewissem Maße den in diesem Abschnitt beschriebenen Effekt. Diese Kompensation erfolgt jedoch in der Regel nur teilweise, sodass sich insgesamt eine Verschärfung des Referenzbereichs ergibt.

nelplot für den Indikator „Indikation zur isolierten Koronarangiographie – Anteil ohne pathologischen Befund“ (ID 56001). Form und Farbe der Punkte machen die Einstufung der Indikatorergebnisse nach der neuen und nach der alten Methodik sichtbar.

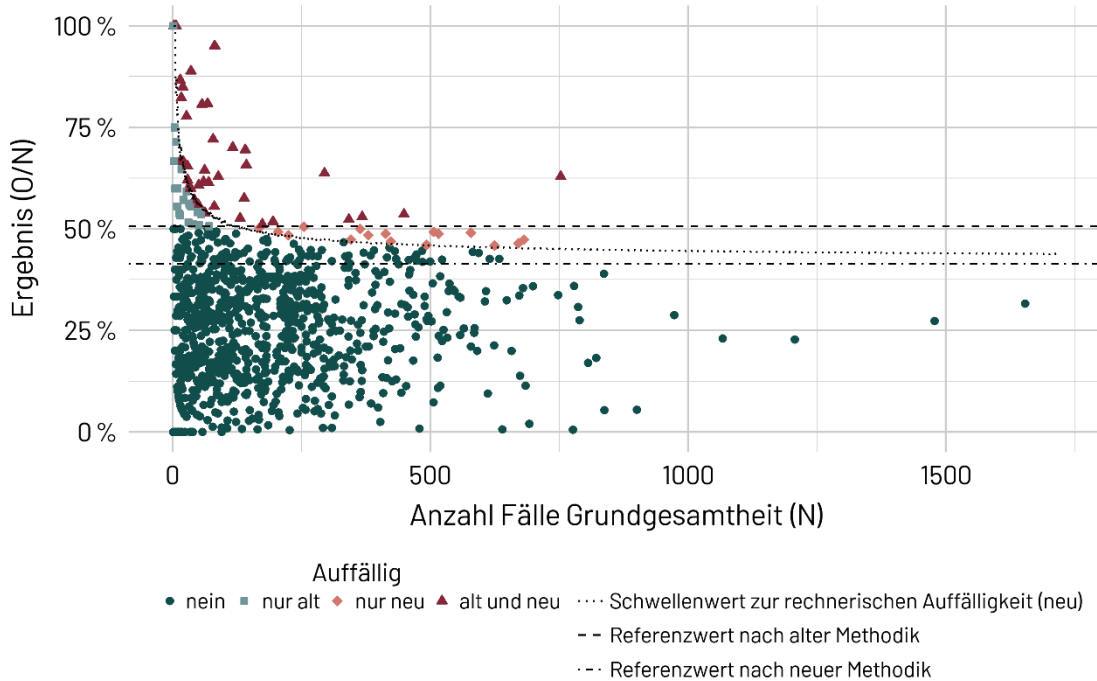


Abbildung 6: Ein Funnelpplot zum Indikator „Indikation zur isolierten Koronarangiographie – Anteil ohne pathologischen Befund“ (ID 56001). Die obere gestrichelte horizontale Linie ist der perzentilbasierte Referenzwert nach der alten Methodik. Die gepunktete Linie ist die Grenze zur rechnerischen Auffälligkeit nach der neuen Methodik bezüglich des perzentilbasierten Referenzbereichs, der ebenfalls nach der neuen Methodik berechnet wurde. Die untere gestrichelte horizontale Linie ist der perzentilbasierte Referenzwert nach der neuen Methodik.

12 Warum führt die neue Methodik zu einer Verschiebung der Auffälligkeiten hin zu Leistungserbringern mit größeren Fallzahlen?

Das Zusammenspiel der neuen Methodik zur Auffälligkeitseinstufung mit der neuen Methodik zur Bestimmung perzentilbasierter Referenzbereiche wirkt sich auf Leistungserbringer mit kleinerer und größerer Fallzahl unterschiedlich aus. Durch die neue Methodik zur Auffälligkeitseinstufung können auch Ergebnisse außerhalb des Referenzbereichs als unauffällig bewertet werden. Bei kleinen Fallzahlen werden dabei größere Abweichungen für plausibel erachtet als bei großen Fallzahlen. Andererseits führt die neue Methodik zur Bestimmung perzentilbasierter Referenzbereiche zu einer Verschärfung der Referenzbereiche (siehe „Warum sind die perzentilbasierten Referenzbereiche nach der neuen Methodik strenger?“). Bei kleinen Fallzahlen überwiegt der Effekt der Anpassung der Methodik zur Auffälligkeitseinstufung: Die Anzahl an Auffälligkeiten verringert

sich. Bei großen Fallzahlen überwiegt der Effekt der Verschärfung der Referenzbereiche: Es kommt zu mehr Auffälligkeiten.

Diese Verschiebung der Auffälligkeiten hin zu Leistungserbringern mit größeren Fallzahlen lässt sich in Abbildung 6 erkennen (siehe auch „Werden durch die neue Methodik Leistungserbringer mit größeren Fallzahlen systematisch benachteiligt?“).

13 Werden durch die neue Methodik Leistungserbringer mit größeren Fallzahlen systematisch benachteiligt?

Mit der neuen Methodik für rechnerische Auffälligkeit und der neuen Methodik für perzentilbasierte Referenzbereiche steigt zwar die Anzahl an rechnerischen Auffälligkeiten bei Leistungserbringern mit größerer Fallzahl. Dabei handelt es sich jedoch nicht um eine systematische Benachteiligung: Vielmehr wird dadurch korrigiert, dass nach der alten Methodik vorwiegend Leistungserbringer mit kleiner Fallzahl rechnerisch auffällig werden konnten. Nach der neuen Methodik sind die rechnerischen Auffälligkeiten gleichmäßiger über den Fallzahlbereich verteilt: Rechnerische Auffälligkeiten treten sowohl bei Leistungserbringern mit kleiner als auch bei solchen mit großer Fallzahl auf. Leistungserbringer mit kleiner und Leistungserbringer mit großer Fallzahl liegen damit im Fokus der Qualitätssicherung.

Beim Vergleich von Leistungserbringern mit unterschiedlichen Fallzahlen gilt generell: Die Bewertung von Leistungserbringern mit kleiner Fallzahl ist mit größerer Unsicherheit behaftet als die Bewertung von Leistungserbringern mit großer Fallzahl. Dies gilt im Wesentlichen unabhängig von der Einstufungsmethodik.⁹ In diesem Sinne ist es nicht möglich, durch eine Änderung an der Einstufungsmethodik völlig gleiche Bedingungen für alle Leistungserbringer herzustellen. Allerdings kann durch die Einstufungsmethodik der Fokus des Stellungsnahmeverfahrens gelenkt werden. Dabei ist es wichtig, dass sowohl Leistungserbringer mit kleiner Fallzahl als auch solche mit großer Fallzahl angemessen im Fokus der Qualitätssicherung sind.

⁹ Genauer gesagt gilt, dass das Risiko für eine Fehleinstufung umso größer ist, je kleiner die Fallzahl ist: Ein Leistungserbringer mit guter Qualität und kleiner Fallzahl hat ein größeres Risiko, als rechnerisch auffällig eingestuft zu werden, als ein Leistungserbringer mit gleichem zugrunde liegenden Indikatorwert und größerer Fallzahl. Ebenso wird ein Leistungserbringer mit schlechter Qualität (dessen zugrunde liegender Indikatorwert außerhalb des Referenzbereichs liegt) und kleiner Fallzahl mit einer größeren Wahrscheinlichkeit als rechnerisch unauffällig eingestuft, als ein Leistungserbringer mit gleichem zugrunde liegendem Indikatorwert und größerer Fallzahl. Dies wird beispielsweise in der folgenden Arbeit gezeigt: Vach, W; Wehberg, S; Güntert, B; Jakob, M; Luta, G (2022): Healthcare provider profiling: fixing observation period or fixing sample size? *BMJ: Open Quality* 11: e001588. DOI: 10.1136/bmjopen-2021-001588 (URL: <https://bmjopenquality.bmj.com/content/11/2/e001588>).