

Prof. Dr. Andreas Becker

# Grenzen der Objektivität – Qualität von Qualitätsvergleichen

*Ob Krankenhauswahl, Bedarfsplanung, Finanzierung oder Modellverträge – im Augenblick scheint nichts ohne „Qualität“ zu gehen. Eine besondere Rolle soll nach den Vorgaben des Gesetzgebers dabei die Messung und Darstellung der Versorgungsqualität spielen, die maßgebliche, international anerkannte Standards der Wissenschaft erfüllen soll. Dem Gesetzgeber ist also bewusst, dass die „Qualität der Qualitätsvergleiche“ eine entscheidende Voraussetzung für die Praxistauglichkeit und Akzeptanz der Qualitätsvergleiche darstellt. Die Objektivität als methodisches Gütekriterium eines Qualitätsvergleichs spielt hierbei eine besondere Rolle. Der Autor erläutert in seinem Aufsatz zum gleichnamigen Referat anlässlich des Qualitätstages der Krankenhäuser der Deutschen Krankenhausgesellschaft am 9. Mai 2014 in Berlin, was Objektivität für einen fairen Qualitätsvergleich bedeutet und zeigt die Grenzen der Vergleiche auf. Hieraus formuliert er Anforderungen an die Qualität der Qualitätsvergleiche, die künftig festgelegt und eingehalten werden sollen.*

**T**ransparenz über die Behandlungsqualität ist ein legitimes Patienteninteresse und notwendige Voraussetzung für einen Qualitätswettbewerb. Mit einer Erhöhung der Transparenz wird einer stärker werdenden gesellschaftlichen Forderung Rechnung getragen. Ein Benchmark ist darüber hinaus ein intrinsischer Anreiz für jeden Teilnehmer, die Qualität seiner Leistungen im Vergleich zu bewerten und ggf. seine Prozesse zu verbessern (DKG 2014).

Auf dem Weg hin zu mehr Transparenz und einer öffentlichen Berichterstattung über die Versorgungsqualität wurde ein wichtiger Schritt getan: Mit dem GKV-Finanzstruktur- und Qualitäts-Weiterentwicklungsgesetz (GKV-FQWG 2014) beauftragte der Gesetzgeber den Gemeinsamen Bundesausschuss (G-BA) mit der Gründung eines fachlich unabhängigen, wissenschaftlichen Instituts für Qualität und Transparenz im Gesundheitswesen. Das Institut arbeitet im Auftrag des G-BA an Maßnahmen zur Qualitätssicherung und zur Darstellung der Versorgungsqualität im Gesundheitswesen. Es soll insbesondere beauftragt werden, „für die Messung und Darstellung der Versorgungsqualität möglichst sektorenübergreifend abgestimmte risikoadjustierte Indikatoren und Instrumente ... zu entwickeln“ (§ 137 a Absatz 3 Nr. 1).

Mit der Vorgabe der Anwendung von Verfahren zur Risikoadjustierung folgt der Gesetzgeber der in Klinik und Wissenschaft gesicherten Erkenntnis, dass der Vergleich von Daten zur Versorgungsqualität (und somit von Einrichtungen im Gesundheitswesen) nur dann Sinn macht und fair ist, wenn die individuellen Risiken der behandelten Patienten bzw. Patientengruppen angemessen berücksichtigt werden.

Wären alle Patienten und damit zum Beispiel auch ihr Alter, unterschiedliche Krankheitsschweregrade, Komorbiditäten und Präferenzen für bestimmte Behandlungsmethoden zufällig (und somit gleichmäßig) über alle Krankenhäuser verteilt,

wäre eine Berücksichtigung des sogenannten Casemix nicht erforderlich. Eine solche Situation liegt in klinischen Studien vor, jedoch nicht in der realen klinischen Welt, denn hier entzieht sich der Casemix der Kontrolle eines Krankenhauses.

Unter der zulässigen Annahme, dass ein Krankenhaus das Outcome durch geeignete Strukturen und Prozesse beeinflussen – und in einem gewissen Umfang kontrollieren – kann, ergibt sich für einen fairen Vergleich folglich die Notwendigkeit, die nicht beeinflussbaren unterschiedlichen Risiken behandelter Patienten bei der Messung und Darstellung der Versorgungsqualität zu berücksichtigen (Katzan et al. 2014, Iezzoni 2013).

Die Anwendung von mathematischen Verfahren zur Risikoadjustierung ist also für eine öffentliche Berichterstattung (public reporting) von elementarer Bedeutung. Darüber hinaus werden risikoadjustierte Daten bzw. Indikatoren künftig auch eine notwendige, wenn auch nicht hinreichende Voraussetzung bei der rechtssicheren Gestaltung von Mindestmengen und natürlich auch bei der Entwicklung von qualitätsorientierten Vergütungssystemen (pay for performance) spielen.

In Anbetracht der Bedeutung einer öffentlichen Berichterstattung für alle Interessengruppen ist die Forderung nach einer größtmöglichen Objektivität nachvollziehbar und berechtigt, da die Ergebnisse möglicherweise nachhaltige Auswirkungen auf die Empfehlung der Einweiser, die Patientenentscheidung für oder gegen ein Krankenhaus und künftig sogar auf die Vergütung oder die Krankenhausplanung haben können. Einrichtungintern geht es um die Entscheidung, ob Analysen durchgeführt werden und ggf. ein Qualitätsverbesserungsprojekt initiiert wird. Da die dadurch gebundenen Ressourcen somit nicht an anderer Stelle eingesetzt werden können, muss sich die Entscheidung hierzu auf möglichst sichere Daten der öffentlichen Berichterstattung stützen können.

**Abbildung 1: Quellen der Variabilität (V) der Ergebnisqualität**

$$V(\text{Ergebnisqualität}) = V(\text{Datenqualität}) + V(\text{Definitionsqualität}) + V(\text{Casemix}) + V(\text{Strukturqualität}) + V(\text{Prozessqualität}) + V(\text{Fallzahlen}) + V(\text{Zufall})$$

In Anlehnung an: Lilford et al. 2004, Iezzoni 2013

**Ergebnisqualität ist variabel**

Der Gesetzgeber hat richtigerweise erkannt, dass eine wissenschaftlich fundierte Vorgehensweise erforderlich ist, die der Komplexität der Aufgabe und den verschiedenen Einflussgrößen auf

die zu messende und zu kommunizierende Ergebnisqualität gerecht wird. Da diese Einflussgrößen nicht nur die Ergebnisqualität und deren Variabilität, sondern auch den Qualitätsvergleich an sich beeinflussen, lohnt sich an dieser Stelle ein Blick auf sie in ► **Abbildung 1**.

**Objektivität**

Objektivität als ein Aspekt der methodischen Güte von Qualitätsvergleichen kann als Unabhängigkeit einer Aussage von subjektiven Einschätzungen und Bewertungen verstanden werden. In der sogenannten Teststatistik unterscheidet man die „Durchführungs-, Ausführungs- und Interpretationsobjektivität“ (Lienert & Raatz 1989).

Übertragen auf eine öffentliche Berichterstattung und die ihr zugrunde liegenden Indikatoren würde man erwarten, dass die bereits oben erwähnten patientenseitigen Faktoren (Case-mix) berücksichtigt werden, die Verfahren korrekt angewendet und die Ergebnisse nach definierten Standards interpretiert werden. Hinzu kommt noch die Forderung an einen entsprechenden Standard bei der Präsentation der Ergebnisse, zum Beispiel durch Grafiken und Tabellen ohne Verfälschung der Ergebnisse. Dies wäre dann aus der Sicht des Autors als „Präsentationsobjektivität“ zu bezeichnen.

Wenn ein einzelnes Krankenhaus oder ein Verbund von Krankenhäusern Daten zur Versorgungsqualität innerhalb der organisationalen bzw. Verbundgrenzen veröffentlicht, so sollten die Mitarbeiter bzw. angeschlossenen Krankenhäuser auf die Qualität des Qualitätsvergleichs vertrauen können und wir könnten hier die zugesicherte Qualität als ein „internes Qualitätsversprechen“ bezeichnen. Werden die Daten auch einer externen Öffentlichkeit zugänglich und bekannt gemacht, so wäre die zugesicherte Qualität des Qualitätsvergleichs nun auch ein „externes Qualitätsversprechen“. In beiden Fällen können die durchführenden Organisationen frei über die Qualitätsanforderungen und deren Erfüllungsgrad entscheiden.

Ganz anders ist die Situation jedoch, wenn die öffentliche Berichterstattung auf einem gesetzlichen Auftrag basiert und die Berichterstattung infolge dieses Auftrages durch einen Dritten (in diesem Fall das Institut für Qualität und Transparenz im Gesundheitswesen) umgesetzt wird. In einem solchen gesetzlich geregelten Raum reicht ein Qualitätsversprechen nicht mehr aus, vielmehr ist eine verbindliche Qualitätszusage (auch aus Gründen der Rechtssicherheit) einzufordern bzw. kann erwartet und vorausgesetzt werden.

Dieser Forderung wird im GKV-FQWG Rechnung getragen, da das Institut als wissenschaftliches Institut fachlich unabhängig sein soll (§ 137 a Absatz 1), seine Aufgaben auf der Basis der maßgeblichen, international anerkannten Standards der Wissenschaft erfüllen soll (§ 137 a Absatz 5), ein wissenschaftlicher Beirat aus unabhängigen Sachverständigen eingerichtet werden soll (§ 137 a Absatz 5) und Forschungs- und Entwicklungsaufträge an externe Sachverständige vergeben werden können (§ 137 a Absatz 6).

**Datenqualität**

Daten sind die Grundlage für die Messung und Darstellung der Versorgungsqualität. Das gilt natürlich auch für risikoadjustierte Indikatoren und andere Instrumente. Hierbei spielt es keine Rolle, ob es sich um administrative Routinedaten oder Daten handelt, die im Rahmen der Qualitätssicherung erhoben werden (zum Beispiel externe Qualitätssicherung, klinische Studien, Register). Die Datenqualität kann anhand verschiedener Gütekriterien beschrieben werden. Hierzu gehören ihre Vollständigkeit, Korrektheit und natürlich die Frage, ob die Daten für ein bestimmtes Vorhaben (zum Beispiel die Entwicklung eines Indikators) geeignet sind. Dies gilt insbesondere für administrative Routinedaten, sie sollten nur unter Kenntnis der möglichen Limitierungen und mit dem erforderlichen Fachwissen interpretiert und angewendet werden. Für die meisten Verfahren gilt, dass die patientenbezogene Dokumentation als Grundlage für die Erfassung von Daten herangezogen wird. Eine insuffiziente Dokumentation, die beispielsweise lückenhaft ist oder einen hohen Grad an Interpretationsvermögen erfordert, führt zwangsläufig zu einer eingeschränkten Datenqualität und wirkt sich so auch negativ auf die Qualität der Risikoadjustierung aus. Auf die Bedeutung der Datenqualität für die Qualitätsbewertung und öffentliche Berichterstattung gehen viele Autoren ein, an dieser Stelle sollen exemplarisch die Arbeiten von Goto et al. (2014), Jakob et al. (2014) und Meddings et al. (2013) genannt werden.

**Definitionsqualität**

Bei der Erhebung von Daten spielen Definitionen eine große Rolle, beispielsweise Definitionen zu Begleiterkrankungen oder Komplikationen. So weisen Talbot et al. (2013) in einer Arbeit zur öffentlichen Berichterstattung von krankenhausbezogenen Raten nosokomialer Infektionen nachdrücklich auf die Bedeutung der zugrunde liegenden Definitionen hin. Die Autoren warnen in diesem Zusammenhang auch vor einer Fehlallokation finanzieller Mittel im Rahmen einer qualitätsorientierten Vergütung als unbeabsichtigte Folge methodischer Mängel.

Zu diesen medizinischen Definitionen kommen bei den administrativen Routinedaten noch weitere Definitionen aus

der ICD- und OPS-Klassifikation und den Deutschen Kodierrichtlinien (DKR) hinzu. Die hier enthaltenen Regeln bestimmen, welche Daten aus der Patientendokumentation im Rahmen der Kodierung in den administrativen Routinedatensatz transferiert werden dürfen. Hieraus folgt: Bei der Berechnung risikoadjustierter Indikatoren unter Verwendung administrativer Routinedaten können also nur die Daten verwendet werden, die die DKR und die ICD- bzw. OPS-Regelwerke passieren lassen.

Es leuchtet ein, dass der Kodierungsprozess somit auch von der Komplexität und Verständlichkeit der Regeln abhängt; dies zeigt schon die umfangreiche Berichterstattung zu unterschiedlichen Ansichten von Krankenhäusern und Krankenkassen zu Abrechnungsfragen. Ein weiterer Punkt spielt bei der datengestützten Qualitätsbewertung mittels administrativer Routinedaten eine mitunter entscheidende Rolle. Es geht hierbei um die Frage, ob eine bestimmte Nebendiagnose zum Zeitpunkt der Krankenhausaufnahme bereits vorlag oder nicht. Diese Entscheidung kann auf der Basis der patientenbezogenen Dokumentation meist zwar entschieden werden. Problematisch ist aber, dass eine entsprechende Abbildung bei vielen relevanten ICD-Kodes nicht möglich ist, da ein Kennzeichen zur Markierung *bei Aufnahme vorhanden ja/nein* in den Datensätzen und Regelwerken nicht vorgesehen ist. Im englischen Sprachraum spricht man vom *Present On Admission (POA-)* Kennzeichen, das die Validität von Kennzahlen- bzw. Indikatorergebnissen und auch die Validität risikoadjustierter Indikatoren erhöhen soll (Becker 2012, Cram et al. 2014).

Im Zusammenhang mit der Nutzung administrativer Routinedaten ist unter der Definitionsqualität auch die Qualität von Algorithmen zu berücksichtigen, die beispielsweise bestimmte Patientengruppen definieren sollen.

- So beeinflusst die Güte solcher Algorithmen die Anzahl der erkannten Fälle mit schwerer Sepsis, wie Gaieski et al. (2013) berichtet haben.
- Mull et al. (2013) zeigten, dass unterschiedliche Methoden zur Messung der 30-Tages-Wiederaufnahmeraten zu höchst unterschiedlichen Ergebnissen im Krankenhausranking führen.
- Ebenso ist die Frage zu stellen, wie mit Aufnahmen aus anderen Einrichtungen umgegangen wird (Katzan et al. 2014). Welchem Krankenhaus wird der Todesfall zugeordnet, wenn ein Patient nach der Verlegung verstirbt: Dem verlegenden oder dem aufnehmenden Krankenhaus?

Es wird ersichtlich, welche Rolle die Definitionsqualität für die Datenqualität und folglich die Messung und Darstellung der Versorgungsqualität hat. Weitere intrinsische Limitierungen der Klassifikationen, zum Beispiel fehlende eindeutige ICD-Kodes für relevante Komplikationen, sollen im Rahmen dieses Artikels nicht weiter erläutert werden.

## Casemix und Risikoadjustierung

Da die Patientenstruktur in verschiedenen Krankenhäusern mitunter sehr unterschiedlich sein kann, ist für einen fairen

Klinikvergleich eine risikoadjustierte Betrachtungsweise unabdingbar. Verschiedene Autoren haben über deutliche Verschiebungen im Krankenhausranking berichtet, wenn die Rankings ohne bzw. mit Risikoadjustierung durchgeführt wurden. Zur Risikoadjustierung werden verschiedene mathematische Verfahren angewendet, die hier nicht einzeln aufgeführt und erläutert werden sollen. Erwähnt werden soll aber, dass Glance et al. (2006) auf das *Risiko der Risikoadjustierung* hingewiesen haben, das sich bereits aus der Wahl der Methode ergibt und Auswirkungen auf die Position eines Krankenhauses in einem Ranking hat.

Von großer Bedeutung bei der Modellentwicklung ist, dass wesentliche in der Literatur formulierte Qualitätsanforderungen eingehalten und die statistischen Verfahren sowie die hierzu verwendeten Prozeduren transparent dargestellt werden. Ebenso sollen sogenannte Gütekriterien, welche die Qualität entsprechender Modelle zur Risikoadjustierung bzw. Indikatoren beschreiben, transparent und somit für alle Interessengruppen nachvollziehbar sein (Becker & Eissler 2014, Benchimol et al. 2011, Harrison et al. 2006, Katzan et al. 2014, Krumholz et al. 2006, Reiter et al. 2007, Steyerberg & Vergouwe 2014, Ziegler et al. 2001).

Es gilt auch zu bedenken, dass selbst hochwertige Risikoadjustierungsmodelle nicht alle Einflussgrößen berücksichtigen können (Lilford et al. 2004); dies gilt insbesondere für solche, die auf administrativen Routinedaten basieren. Daher ist eine Kombination mit klinischen Daten bei bestimmten Fragestellungen sinnvoller, wie Siregar et al. (2014) anhand von niederländischen Daten zur standardisierten Krankenhaussterblichkeit bei kardiochirurgischen Operationen zeigen konnten. So ist es beispielsweise nicht möglich, über die ICD-Klassifikation den Schweregrad eines ischämischen Hirninfarktes direkt und in Anlehnung an etablierte Skalen zu beschreiben. Somit dürfte ein solches Modell zur Risikoadjustierung bestenfalls ein gutes Surrogat für ein Modell sein, welches auch Daten zum Schweregrad berücksichtigt (Katzan et al. 2014, Fonarow et al. 2012).

Ein weiteres Beispiel illustriert die Auswirkung der Entlassungs- und Verlegungspraxis. Werden Patienten vor dem Versterben in die häusliche Umgebung entlassen oder in eine andere Einrichtung verlegt (zum Beispiel Pflegeeinrichtung oder Hospiz), so hat dies bei bestimmten Patientengruppen mitunter deutliche Auswirkungen auf die Anzahl der im Krankenhaus verstorbenen Patienten. Es stellt sich also die Frage nach der *Zählung der Todesfälle*, zu der Kozar et al. (2014) am Beispiel traumatologischer Fälle näher ausführen.

Darüber hinaus muss ein Modell zur Risikoadjustierung nach seiner Berechnung natürlich auch richtig angewendet werden. Auf die hierbei möglichen systematischen und zufälligen Fehler soll im Rahmen dieser Arbeit nicht weiter eingegangen werden.

Es gilt also: Die im Rahmen von Klinikvergleichen beschriebene Ergebnisqualität hängt in relevantem Umfang auch von der Qualität der Modelle zur Risikoadjustierung, also deren Gütekriterien, ab. Es ist daher im Interesse aller Beteiligten, wenn die Entwicklung und Gütekriterien solcher Modelle

transparent und nachvollziehbar offengelegt werden. Sogenannte *geheime Blackboxen* erhöhen nicht die Akzeptanz, sie führen eher zu berechtigtem Misstrauen. Eine ausführliche Beschreibung und Anwendung entsprechender Methoden und Gütekriterien ist zu finden bei Becker & Eissler (2014).

## Struktur- und Prozessqualität

Ein Krankenhaus kann das Outcome durch geeignete Strukturen und Prozesse beeinflussen. Personelle Ressourcen (quantitativ und qualitativ), die Infrastruktur und die Arbeitsumgebung können sich nachhaltig auf die Prozessqualität und somit die Ergebnisqualität auswirken. Bei der Prozessqualität geht es insbesondere um die Frage, ob die beste Evidenz (zum Beispiel in Form von Leitlinien) am *scharfen Ende der Prozesse* an die Patienten kommt. Gelingt es zum Beispiel regelmäßig, bei Pneumonienpatienten eine passende antibiotische Therapie einzuleiten, den Therapieverlauf zu überwachen und die Therapie ggf. anzupassen, wird sich dies positiv auf die Ergebnisqualität auswirken.

## Fallzahlen

Beim Münzenwerfen kommt es gar nicht so selten vor, dass nacheinander mehrfach Kopf oder Zahl geworfen wird. Aber wie auch bekannt ist: Je öfter eine Münze geworfen wird, umso wahrscheinlicher ist es, dass am Ende gleich oft Kopf und Zahl geworfen wird. Würde man also die Ergebnisse vieler Wurfserien miteinander vergleichen, so würde man feststellen, dass sich die Anzahl von Kopf und Zahl von den Wurfserien mit wenigen Würfeln zu denen mit sehr vielen Würfeln immer mehr angleichen. Mit anderen Worten: Die Ergebnisse zeigen mit zunehmender Anzahl von Würfeln immer weniger Unterschiede. Diese Erkenntnis hat auch bei der Beurteilung von Daten medizinischer Versorgungsprozesse eine elementare Bedeutung. Solche Daten sollten – in Analogie zu der Zahl der Würfel – immer unter Berücksichtigung der Fallzahlen untersucht werden. Dimick et al. haben 2004 in einer grundlegenden Arbeit gezeigt, dass die Berücksichtigung der Fallzahlen bei der Beurteilung von Ergebnissen elementar ist. Tu & Gilthorpe weisen in ihrer Arbeit aus dem Jahr 2007 darauf hin, dass der französische Mathematiker de Moivre schon im Jahr 1730 die mathematische Grundlage zu dieser Erkenntnis formuliert hat. Dies ist zwar in jedem Standardlehrbuch der Statistik nachzulesen, wird aber – wie die Autoren in einem Beispiel aus England zeigen – dennoch nicht immer berücksichtigt. Daher stellen Tu & Gilthorpe die kritische Frage, ob es sich in dem untersuchten Krankenhausranking wirklich um das „gefährlichste Krankenhaus oder die gefährlichste Berechnungsmethode“ handelt. Die Bedeutung der Fallzahlen bestätigen van Dishoeck et al. (2011) im Zusammenhang mit öffentlicher Berichterstattung und favorisieren den sogenannten *Funnel Plot* zur Unterstützung bei der Interpretation von Ergebnissen. Es spielt also keine Rolle, ob es sich um nicht risikoadjustierte oder risikoadjustierte Daten handelt: Die Fallzahlen müssen bei der Interpretation immer berücksichtigt werden.

## Präzise Analyse mit einem „Funnel Plot“

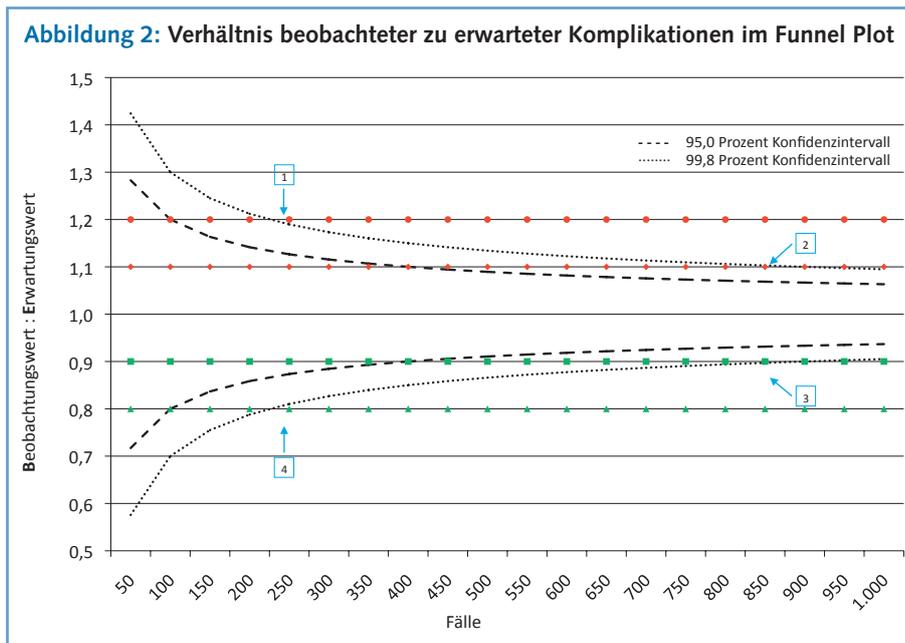
Um dies näher zu erläutern, soll eine bestimmte Komplikation betrachtet werden, die im Rahmen einer konservativen oder operativen Behandlung bzw. einem diagnostisch-therapeutischen Eingriff eintreten kann.

- Zu diesem Zweck betrachten wir die fiktiven Daten von 20 Kliniken: Fallzahl, beobachtete und erwartete Anzahl der Komplikation. Letztere wurde mittels eines risikoadjustierten Modells berechnet. Die Fallzahlen sind in unserer Stichprobe deutlich unterschiedlich und liegen in einem Bereich von 50 bis 1 000 Patienten.
- Wir unterstellen zunächst, dass die Anzahl der beobachteten Komplikationen bei allen Kliniken um 20 Prozent größer ist als die Anzahl der erwarteten Komplikationen, die mittels risikoadjustiertem Verfahren berechnet wurde. Das berechnete Verhältnis der beobachteten zu erwarteten Werten (B:E-Wert) beträgt somit für alle Kliniken 1,2 und ist in **► Abbildung 2** abgebildet (rote Kreise). Liegen 10 Prozent mehr Komplikationen als erwartet vor, beträgt dieser Wert 1,1 (rote Raute).
- Analog werden nun die B:E-Werte berechnet, wenn die beobachtete Anzahl der Komplikationen um 10 bzw. 20 Prozent unter den Erwartungswerten liegt. Sie betragen folglich 0,9 (grünes Quadrat) bzw. 0,8 (grünes Dreieck).
- Der Wert 1,0 dient als Referenzwert, denn hier ist die Anzahl beobachteter Komplikationen gleich der Anzahl erwarteter.
- Weiterhin sind in Abbildung 2 sogenannte Konfidenzintervalle abgebildet. Unter der Annahme, dass vom Referenzwert (1,0) abweichende B:E-Werte nur zufallsbedingt sind, besteht eine 95-prozentige Wahrscheinlichkeit innerhalb der inneren Grenzen und eine 99,8-prozentige Wahrscheinlichkeit, innerhalb der äußeren Grenzen zu liegen.

Eine solche Darstellung, bei der Werte gemeinsam mit den dazugehörigen Fallzahlen und den beschriebenen Grenzen abgebildet werden, nennt man *Funnel Plot*. Er leistet einen wichtigen Beitrag zur Präsentations- und Interpretationsqualität, denn er ermöglicht dem Betrachter die schnelle und fallzahlbezogene Identifizierung von Werten, die aus statistischer Sicht als auffällig bezeichnet werden können, da sie sich außerhalb der Grenzen befinden.

Bei einem B:E-Wert von 1,2 mag man geneigt sein, von *schlechter Ergebnisqualität* zu sprechen. Umgekehrt wird ein B:E-Wert von 0,8 schnell als *gute Ergebnisqualität* anerkannt. Und der direkte Vergleich der B:E-Werte 1,2 versus 0,8 würde für viele Betrachter natürlich klar aufdecken, dass zwischen den beiden dazugehörenden Kliniken deutliche Unterschiede in der Ergebnisqualität bestehen. Würde man die B:E-Werte (1,2/1,1/0,9/0,8) in Form eines Klinikrankings absteigend sortiert auflisten, so würden diese Rankings ohne Berücksichtigung der Fallzahlen zu falschen Schlussfolgerungen führen. So können Rangplätze durch eine intuitive Bewertung ohne Berücksichtigung der Fallzahlen schnell in ihrer Bedeutung überschätzt werden.

Der Funnel Plot in Abbildung 2 ermöglicht dagegen eine präzise Analyse. Zunächst fällt auf, dass die Grenzen mit zunehmender Fallzahl enger werden – mit steigender Fallzahl verengen sich die Bereiche, innerhalb derer die B:E-Werte liegen können, ohne die Grenzen (jeweilige obere und untere Grenze) zu überschreiten. Je geringer die Fallzahl, desto größer ist der *Toleranzbereich*; nur deutliche Abweichungen vom Referenzwert 1,0 führen zu einem Überschreiten der Grenzen. Bei hohen Fallzahlen werden die Grenzen schon bei deutlich geringeren Abweichungen erreicht oder sogar überschritten. Dies zeigt sich bei den beiden Kliniken mit geringster und höchster Fallzahl in Abbildung 2. Bei der Klinik mit 50 Fällen würden erst B:E-Werte



von 1,42 bzw. 0,58 die äußere obere bzw. untere Grenze erreichen (99,8 Prozent Konfidenzintervall) und als *statistisch auffällig* bezeichnet werden. Anders sieht es bei der Klinik mit 1 000 Fällen aus: Hier lägen bereits die B:E-Werte 1,1 bzw. 0,9 außerhalb der äußeren oberen bzw. unteren Grenze (99,8 Prozent Konfidenzintervall) und werden daher *statistisch auffällig*.

Eine andere Betrachtungsweise lässt – bezogen auf die äußere Grenze (99,8 Prozent Konfidenzintervall) – folgende Schlüsse zu: Der B:E-Wert 1,2 überschreitet erst bei 250 Fällen die Grenze (Pfeil 1) und der B:E-Wert 1,1 erst bei 850 Fällen (Pfeil 2). Zwischen den beiden B:E-Werten liegen 10 Prozent, aber 600 Fälle! Auch die Unterschreitung der äußeren Grenzen bestätigt die Bedeutung der Fallzahlen: Während der B:E-Wert 0,9 bei 850 Fällen die äußere Grenze erreicht (Pfeil 3), ist dies bei dem B:E-Wert 0,8 bereits bei 250 Fällen festzustellen (Pfeil 4).

Anders ausgedrückt: Innerhalb der Grenzen kann eine Klassifizierung als *statistisch auffällig* selbst bei deutlich unterschiedlichen B:E-Werten nicht stattfinden, man kann auch von einer „neutralen Zone der Nichtinterpretation“ sprechen (in Anlehnung an Aelvoet et al. 2008). Diese Erkenntnis schützt vor vorschnellen Urteilen und fördert einen sachlichen Umgang mit solchen Daten. Mit Hilfe des Funnel Plots kann also festgestellt werden, ob ein bestimmter B:E-Wert in Bezug auf die innere oder äußere Grenze eines Referenzwertes als *statistisch auffällig* klassifiziert werden kann.

Zu beachten ist an dieser Stelle, dass diese Einstufung nicht mit „guter oder schlechter, besserer oder schlechterer Versorgungsqualität“ gleichzusetzen ist. Die Risikoadjustierung ist an dieser Stelle nur prognostisch zu verstehen, sie ist keine exakte Diagnose der hinter einem B:E-Wert liegenden Versorgungsprozesse! Risikoadjustierung ist der Anfang und nicht das Ende der Qualitätsbewertung, da sie nur das Erkennen statistischer Auffälligkeiten erlaubt, die dann eine nähere Analyse der Strukturen und Prozesse erfordert. Nur so kann

festgestellt werden, ob tatsächlich Defizite in der Versorgung vorliegen.

Wenn die vom Gesetzgeber geplanten risikoadjustierten Indikatoren und auch die geplante Darstellung der Versorgungsqualität dies berücksichtigt, wären sie konstruktive Beiträge zur Verbesserung der Versorgungsqualität. Undifferenzierte Krankenhausrankings ohne Berücksichtigung der Fallzahlen führen schnell zu falschen Schlussfolgerungen. Eine solche trügerische Anordnung suggeriert klinisch relevante Unterschiede, die häufig gar nicht vorhanden sind und verhindert so eine Akzeptanz in den Versorgungssektoren.

### Zufall

Der Zufall sorgt für ein *statistisches Rauschen* und kann kaum eingeschätzt werden. Die vorstehenden Ausführungen zum Münzwurf und den Fallzahlen zeigen aber, dass sich zufällige Ereignisse oder zufälliges Schwanken des B:E-Wertes in Abhängigkeit von der Fallzahl sehr unterschiedlich auswirken können.

### Referenzbereiche

Referenzbereiche sollen als Grenzwerte auffällige von unauffälligen Ergebnissen unterscheiden können, man unterscheidet „Ziel- und Toleranzbereiche“.

Man spricht von einem *Zielbereich*, wenn aufgrund wissenschaftlicher Untersuchungen definiert werden kann, welches Ergebnis als gute Qualität anzusehen ist, ob es also im Zielbereich liegt. Für diese Indikatoren wird ein fester Wert als Referenzbereich festgelegt. Ein Beispiel für einen Zielbereich ist die Rate der Patienten mit ambulant erworbener Pneumonie aus der externen Qualitätssicherung (EQS 2014), bei denen die erste Blutgasanalyse oder Pulsoxymetrie innerhalb von acht Stunden nach der Krankenhausaufnahme durchgeführt wur-

**Abbildung 3: Anforderungen an sinnvolle und faire Vergleiche**

- Individuelle Risiken der behandelten Patienten bzw. Patientengruppen werden angemessen berücksichtigt.
  - Wissenschaftlich fundierte Vorgehensweise bei der Entwicklung von Modellen zur Risikoadjustierung.
  - Informationen zu den verwendeten Daten, deren Qualität und Definitionen (Grundgesamtheit, Zielgröße, Abgrenzung »Merkmal bei Aufnahmen vorhanden ja/nein«).
  - Bei Verwendung administrativer Routinedaten: Auswirkung von relevanten Vorgaben aus den Kodierrichtlinien und den Klassifikationen (ICD, OPS) auf die Daten und ihre Qualität werden berichtet.
  - Transparenz der Modelle: Entwicklung, Validierung, Variablen und Koeffizienten.
  - Transparenz statistischer Gütekriterien: Sensitivität, Spezifität, prädiktive Werte, Kalibrierung, Diskriminierung
  - Standard bei der Präsentation der Ergebnisse, zum Beispiel durch Grafiken und Tabellen ohne Verfälschung der Ergebnisse.
  - Werden Ergebnisse als »Ranking« abgebildet, so wird die gewählte Methode angemessen erläutert und begründet.
  - Ergebnisse werden nach definierten Standards unter Berücksichtigung der Fallzahlen interpretiert.
  - Referenzbereiche werden unter Berücksichtigung von evidenzbasierten Daten festgelegt.
- Grundlegendes Verständnis: Risikoadjustierung ist nur prognostisch, sie ist keine exakte Diagnose der Versorgungsprozesse! Risikoadjustierung ist der Anfang und nicht das Ende der Qualitätsbewertung, da sie nur das Erkennen statistischer Auffälligkeiten erlaubt.

renzbereichs liegt ... Wird ein solcher Referenzbereich im Einzelfall verfehlt, so liegt damit eine Auffälligkeit vor, die im weiteren Verlauf geklärt werden muss. Die Verwendung von relativen Referenzbereichen zu diesem Zweck lässt das absolute Qualitätsniveau der einzelnen erbrachten Leistung außer Acht und ist daher nicht zielführend. Dies wäre etwa dann der Fall, wenn immer die 15 Prozent der Leistungserbringer mit den schlechtesten Ergebnissen als ‚auffällig‘ definiert würden ...“ (DKG 2014).

Einen interessanten Ansatz zur Ermittlung eines Wertes als *Benchmark* stellten Weissmann et al. im Jahr 1999 vor. Hier wird der Referenzbereich

de. Dieser Zielbereich wird in der Bundesauswertung für das Jahr 2013 mit  $\geq 95$  Prozent angegeben.

Dem gegenüber steht der *Toleranzbereich*: Für einige Qualitätsindikatoren gibt es keine eindeutige feste Grenze. So kann man die Häufigkeit von Komplikationen, die nach Operationen auftreten können, als Indikatoren für Ergebnisqualität verwenden. Da diese Komplikationen aber auch bei bester Behandlung nicht hundertprozentig vermeidbar sind, kann man zwar das Ziel formulieren, dass die Komplikationen möglichst selten auftreten sollen, man kann aber keine sichere Grenze festlegen, die erreichbar gute Qualität kennzeichnet. Bei diesen Qualitätsindikatoren lässt sich folglich kein Referenzbereich angeben, der erreichbar gute Qualität klar beschreibt. Mit Hilfe von Vergleichsergebnissen können aber besonders auffällige Ergebnisse erkannt werden. Häufig werden in diesen Fällen Perzentil-Referenzbereiche verwendet. Ergebnisse, die innerhalb dieses Referenzbereichs liegen, können toleriert werden. Ergebnisse, die außerhalb des Toleranzbereichs liegen, müssen genauer analysiert werden (BQS-Institut 2014).

Ein Perzentil (lateinisch *Hundertstelwert*) ist ein Prozentrang, wie ein weiterer Indikator aus der externen Qualitätssicherung zur ambulant erworbenen Pneumonie aus dem Jahr 2013 zeigt: Als Toleranzbereich für die Sterblichkeit im Krankenhaus wird das 90. Perzentil mit dem Wert  $\leq 13,8$  Prozent angegeben. Dies bedeutet, dass 90 Prozent der ausgewerteten 1 123 Kliniken eine Krankenhaussterblichkeit von 13,8 Prozent oder niedriger ausgewiesen haben.

Die Festlegung von Referenzbereichen ist eine komplexe Aufgabe und hat Auswirkungen auf die Position und Bewertung eines Krankenhauses bei einer öffentlichen Berichterstattung bzw. einer zukünftigen qualitätsorientierten Vergütung. Daher formuliert die Deutsche Krankenhausgesellschaft folgerichtig: „Für die Ergebnisse der öffentlich berichteten Qualitätsindikatoren sind vorzugsweise absolute Referenzbereiche zu verwenden, die sich an vorhandenen evidenzbasierten Daten orientieren. Abwertungen als Folge der Qualitätsdarstellung sind auf Leistungserbringer zu beschränken, deren Bewertung statistisch signifikant außerhalb eines solchen Referenzbereichs liegt ...“

zwar auch unter Anwendung der einzelnen Krankenhauswerte ermittelt, die vorgestellte Methode limitiert jedoch den Einfluss von Extremwerten, die durch geringe Fallzahlen bedingt sein können.

## Zusammenfassung

*Public Reporting* soll die verschiedenen Interessenten bei ihren Entscheidungen durch Transparenz der Daten, Methoden und Ergebnisse unterstützen und bei den Teilnehmern an der Versorgung die Motivation zur Veränderung erhöhen (RAND 2011, Smith et al. 2009). Transparenz und Objektivität sind dabei legitimes Interesse aller Beteiligten, insbesondere der Patienten. In ► **Abbildung 3** werden die Anforderungen an sinnvolle und faire Vergleiche aus diesem Artikel zusammenfassend dargestellt.

Vornehmliches Ziel der Messung und Darstellung der Versorgungsqualität sollte immer die Qualitätsverbesserung sein – Messung, Beurteilung und Reporting erfüllen keinen Selbstzweck. Es handelt sich hierbei um eine komplexe Unternehmung (Glance et al. 2011) und es gilt dabei zu beachten, dass *Komplexität* keine lineare Steigerung von *einfach* ist.

Die Qualität der Definitionen, Daten, die Modellentwicklung, Referenzbereiche, Interpretation und Präsentation bestimmen die Qualität und somit zusammen mit der Transparenz die Akzeptanz und Rechtssicherheit der Qualitätsvergleiche. Daher sollten die von Daley (2013) formulierten Fragen bei jedem Verfahren und jedem Indikator gestellt und beantwortet werden:

- Wie können wir wissen, dass die zugrunde liegenden Daten das Outcome vertrauenswürdig, korrekt und valide messen?
- Wie können wir wissen, ob beobachtete Unterschiede zwischen den Ergebnissen nicht durch patientenbezogene Faktoren bedingt sind, die wir nicht beeinflussen können?
- Wie können wir wissen, ob beobachtete Unterschiede nicht das Ergebnis zufälliger Variation sind und nicht Folge beachteter Interventionen?

Um die Bedeutung der Fallzahlen intuitiv erfassbar zu machen, sollte die grafische Darstellung von Ergebnissen durch Funnel Plots unterstützt werden.

Eine fachlich qualifizierte und unabhängige Institution sollte alle Leistungserbringer nach einheitlichen, wissenschaftlich fundierten und transparenten Verfahren beurteilen und dabei auch die gewünschte Objektivität zukünftig als Qualitätskriterium der Messungen und Berichterstattung definieren. Das GKV-FQWG bietet mit dem neuen § 137 a SGB V die dazu erforderliche Rechtsgrundlage.

## Literatur

Aelvoet, W.; Windey, F.; Molenberghs, G.; Verstraelen, H.; Van Reempts, P.; Foidart, J.M. (2008): Screening for inter-hospital differences in cesarean section rates in low-risk deliveries using administrative data: an initiative to improve the quality of care. *BMC Health Serv Res.* 2008; 8. PubMed-ID: 18177493

Becker, A. (2012): Present-On-Admission-Kennzeichen (POA) für administrative Routinedaten in Krankenhäusern. Literaturübersicht und Handlungsempfehlungen. Eine Ausarbeitung für die Gesellschaft für Qualitätsmanagement in der Gesundheitsversorgung e.V. (GQMG). Quelle: <http://www.i-pdb.de/publikationen/wissenschaftliche-ausarbeitungen-fuer-fachgesellschaften.html> (letzte Einsicht 22. Juli 2014)

Becker, A.; Eissler, U. (2014): Die standardisierte primäre Sectorate (SPSR) und ihre Anwendung im Qualitätsmanagement und für Krankenhausvergleiche. *das Krankenhaus* 2014; 106 (2): 115–130. Quelle: <http://www.i-pdb.de/publikationen/publikationen-mit-erstauteurschaft.html> (letzte Einsicht 22. Juli 2014)

Benchimol, E.I.; Manuel, D.G.; To, T.; Griffiths, A.M.; Rabeneck, L.; Guttmann, A. (2011). Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol.* 2011; 64(8): 821–829. PubMed-ID: 21194889

BQS-Institut (2014), Quelle: <http://bqs-institut.de> (letzte Einsicht 22. Juli 2014)

Cram, P.; Bozic, K.J.; Callaghan, J.J.; Lu, X.; Li, Y. (2014): Use of present-on-admission indicators for complications after total knee arthroplasty: an analysis of Medicare administrative data. *J Arthroplasty.* 2014; 29(5): 923–928.e2. PubMed-ID: 24530205

Daley, J. (2013): The role of risk adjustment in managing health care organizations. In: *Risk adjustment for measuring health care outcomes.* Editor: Lisa I. Iezzoni. 2013 (4th edition). Health Administrative Press, Chicago, Illinois

Dimick, J.B.; Welch, H.G.; Birkmeyer, J.D. (2004): Surgical mortality as an indicator of hospital quality: the problem with small sample size. *JAMA.* 2004; 292(7): 847–851. PubMed-ID: 15315999

DKG (2014): Positionen der Deutschen Krankenhausgesellschaft zur Weiterentwicklung der Qualitätssicherung und der Patientensicherheit. Quelle: [http://www.dkgev.de/dkg.php/cat/216/aid/2610/title/Positionen\\_der\\_DKG](http://www.dkgev.de/dkg.php/cat/216/aid/2610/title/Positionen_der_DKG) (letzte Einsicht 22. Juli 2014)

EQS (2014): Ambulant erworbene Pneumonie. Bundesauswertung zum Erfassungsjahr 2013. Quelle: <https://www.sqg.de/ergebnisse/leistungsbereiche/ambulant-erworbene-pneumonie.html> (letzte Einsicht 22. Juli 2014)

Fonarow, G.C.; Pan, W.; Saver, J.L.; Smith, E.E.; Reeves, M.J.; Broderick, J.P.; Kleindorfer, D.O.; Sacco, R.L.; Olson, D.M.; Hernandez, A.F.; Peterson, E.D.; Schwamm, L.H. (2012): Comparison of 30-day mortality models for profiling hospital performance in acute ischemic stroke with vs without adjustment for stroke severity. *JAMA.* 2012; 308(3): 257–264. PubMed-ID: 22797643

Gaieski, D.F.; Edwards, J.M.; Kallan, M.J.; Carr, B.G. (2013): Benchmarking the incidence and mortality of severe sepsis in the United States. *Crit Care Med.* 2013; 41(5): 1167–1174. PubMed-ID: 23442987

GKV-FQWG (2014): GKV-Finanzstruktur- und Qualitäts-Weiterentwicklungsgesetz. Quelle: <http://www.bmg.bund.de/ministerium/presse/pressemitteilungen/2014-02/gkv-fqwg-beschlossen.html> (letzte Einsicht 22. Juli 2014)

Glance, L.G.; Dick, A.; Osler, T.M.; Li, Y.; Mukamel, D.B. (2006): Impact of changing the statistical methodology on hospital and surgeon ranking: the case of the New York State cardiac surgery report card. *Med Care.* 2006; 44(4): 311–319. PubMed-ID: 16565631

Glance, L.G.; Neuman, M.; Martinez, E.A.; Pauker, K.Y.; Dutton, R.P. (2011): Performance measurement at a "tipping point". *Anesth Analg.* 2011; 112(4): 958–966. PubMed-ID: 21385976

Goto, M.; Ohl, M.E.; Schweizer, M.L.; Perencevich EN (2014): Accuracy of administrative code data for the surveillance of healthcare-associated infections: a systematic review and meta-analysis. *Clin Infect Dis.* 2014; 58(5): 688–696. PubMed-ID: 24218103

Harrison, D.A.; Brady, A.R.; Parry, G.J.; Carpenter, J.R.; Rowan, K. (2006): Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med.* 2006; 34(5): 1378–1388. PubMed-ID: 16557153

Iezzoni, L.I. (2013): Reasons for risk adjustment. In: *Risk adjustment for measuring health care outcomes.* Editor: Lisa I. Iezzoni. 2013 (4th edition). Health Administrative Press, Chicago, Illinois

Jakob, J.; Marendt, D.; Sold, M.; Schlüter, M.; Post, S.; Kienle, P. (2014): Dokumentationsqualität intra- und postoperativer Komplikationen. Verbesserung der Dokumentation für die verpflichtende Qualitätssicherung und Vergleich mit DRG-Routinedaten. *Chirurg.* 2014; DOI 10.1007/s00104-013-2696-4. PubMed-ID: 24499996

Katzan, I.L.; Spertus, J.; Bettger, J.P.; Bravata, D.M.; Reeves, M.J.; Smith, E.E.; Bushnell, C.; Higashida, R.T.; Hinchey, J.A.; Holloway, R.G.; Howard, G.; King, R.B.; Krumholz, H.M.; Lutz, B.J.; Yeh, R.W.: American Heart Association Stroke Council, Council on Quality of Care and Outcomes Research, Council on Cardiovascular and Stroke Nursing, Council on Cardiovascular Radiology and Intervention, Council on Cardiovascular Surgery and Anesthesia, Council on Clinical Cardiology (2014). Risk adjustment of ischemic stroke outcomes for comparing hospital performance: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* 2014; 45(3): 918–944. PubMed-ID: 24457296

Kozar, R.A.; Holcomb JB, Xiong W, Nathens AB (2014). Are all deaths recorded equally? The impact of hospice care on risk-adjusted mortality. *J Trauma Acute Care Surg.* 2014; 76(3): 634–641. PubMed-ID: 24553529

Krumholz, H.M.; Brindis, R.G.; Brush, J.E.; Cohen, D.J.; Epstein, A.J.; Furie, K.; Howard, G.; Peterson, E.D.; Rathore, S.S.; Smith, S.C.; Spertus, J.A.; Wang, Y.; Normand, S.L.: American Heart Association, Quality of Care and Outcomes Research Interdisciplinary Writing Group, Council on Epidemiology and Prevention, Stroke Council, American College of Cardiology Foundation (2006). Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: cosponsored by the Council on Epidemiology and Prevention and the Stroke Council. Endorsed by the American College of Cardiology Foundation. *Circulation.* 2006; 113(3): 456–462. PubMed-ID: 16365198

Lienert, G.A.; Raatz, U. (1989): Testaufbau und Testanalyse. 1989 (5. Auflage). Beltz Verlag

Lilford, R.; Mohammed, M.A.; Spiegelhalter, D.; Thomson, R. (2004): Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet.* 2004; 363(9415): 1147–1154. PubMed-ID: 15064036

Meddings, J.A.; Reichert, H.; Hofer, T.; McMahon, L.F. (2013): Hospital report cards for hospital-acquired pressure ulcers: how good are the grades? *Ann Intern Med.* 2013; 159(8): 505–513. PubMed-ID: 24126644

Mull, H.J.; Chen, Q.; O'Brien, W.J.; Shwartz, M.; Borzecki, A.M.; Hanchate, A.; Rosen, A.K. (2013): Comparing 2 methods of assessing 30-day readmissions: what is the impact on hospital profiling in the veterans health administration? *Med Care.* 2013; 51(7): 589–596. PubMed-ID: 23604016

RAND (2011): How health systems make available information on service providers. Experience in seven countries. RAND Corporation. Quelle: [http://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCAQFJAA&url=http%3A%2F%2Fwww.rand.org%2Fcontent%2Fdam%2Frand%2Fpubs%2Ftechnical\\_reports%2F2011%2FRAND\\_TR887.pdf&ei=48bOU4KtjOG\\_OdaZgdg&usq=AFOjCNHsU9LXIAXns1PP3arRdE0lpdrew&bvm=bv.71198958,d.ZWU](http://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCAQFJAA&url=http%3A%2F%2Fwww.rand.org%2Fcontent%2Fdam%2Frand%2Fpubs%2Ftechnical_reports%2F2011%2FRAND_TR887.pdf&ei=48bOU4KtjOG_OdaZgdg&usq=AFOjCNHsU9LXIAXns1PP3arRdE0lpdrew&bvm=bv.71198958,d.ZWU) (letzte Einsicht 22. Juli 2014)

Reiter, A.; Fischer, B.; Köttling, J.; Geraedts, M.; Jäckel, W.H.; Barlag, H.; Döbler, K. (2007): QUALIFY: Ein Instrument zur Bewertung von Qualitätsindikatoren. Quelle: <http://bqs-institut.de/innovationen/qualify-instrument.html> (letzte Einsicht 22. Juli 2014)

Siregar, S.; Pouw, M.E.; Moons, K.G.; Versteegh, M.I.; Bots, M.L.; van der Graaf, Y.; Kalkman, C.J.; van Herwerden, L.A.; Groenwold, R.H. (2014): The Dutch hospital standardised mortality ratio (HSMR) method and cardiac surgery: benchmarking in a national cohort using hospital administration data versus a clinical database. *Heart.* 2014; 100(9): 702–710. PubMed-ID: 24334377

Smith, P.C.; Mossialos, E.; Papanicolas, I.; Leatherman, S. (2009): Principles of performance measurement. Part 1.1 Introduction. In: *Performance Measurement for Health System Improvement. Experiences, Challenges and Prospects.* Edited by Peter C. Smith, Elias Mossialos, Irene Papanicolas and Sheila Leatherman. 2009. Cambridge University Press

Steyerberg, E.W.; Vergouwe, Y. (2014): Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J.* 2014; PubMed-ID: 24898551

Talbot, T.R.; Bratzler, D.W.; Carrico, R.M.; Diekema, D.J.; Hayden, M.K.; Huang, S.S.; Yokoe, D.S.; Fishman, N.O.: Healthcare Infection Control Practices Advisory Committee (2013). Public reporting of health care-associated surveillance data: recommendations from the healthcare infection control practices advisorycommittee. *Ann Intern Med.* 2013; 159(9): 631–635. PubMed-ID: 24189597

Tu, Y.K.; Gilthorpe, M.S. (2007): The most dangerous hospital or the most dangerous equation? *BMC Health Serv Res.* 2007; 7. PubMed-ID: 18005428

van Dishoeck, A.M.; Looman, C.W.; van der Wilden-van Lier, E.C.; Mackenbach, J.P.; Steyerberg, E.W. (2011): Displaying random variation in comparing hospital performance. *BMJ Qual Saf.* 2011; 20(8): 651–657. PubMed-ID: 21228432

Weissman, N.W.; Allison, J.J.; Kiefe, C.I.; Farmer, R.M.; Weaver, M.T.; Williams, O.D.; Child, I.G.; Pemberton, J.H.; Brown, K.C.; Baker, C.S. (1999): Achievable benchmarks of care: the ABCs of benchmarking. *J Eval Clin Pract.* 1999; 5(3): 269–281. PubMed-ID: 10461579

Ziegler, A.; König, I.R. (2011): Vollständiges und präzises Berichten von Studien zur diagnostischen Genauigkeit: Die STARD-Initiative. *Dtsch Med Wochenschr.* 2011; 136: e16–e19

---

Der Autor bedankt sich bei Herrn Burkhard Fischer (Referatsleiter Qualitätsmanagement, IT und Datenanalyse, Krankenhausgesellschaft Nordrhein-Westfalen e.V.) für seine wertvollen Hinweise zu diesem Artikel.

---

## Anschrift des Verfassers

---

*Prof. Dr. med. Andreas Becker, Institut Prof. Dr. Becker, Nonnenweg 120 a, 51503 Rösrath, becker@i-pdb.de, www.i-pdb.de* ■