

Foto: kei 907 – Fotolia

Ohne Transparenz keine Rechtssicherheit

Qualität der Qualitätssicherung – ausgewählte Aspekte zu Qualitätsindikatoren

Von Prof. Dr. med. Andreas Becker

Nach den Vorgaben des Gesetzgebers spielt die Messung der Versorgungsqualität mittels der einrichtungsübergreifenden Qualitätssicherung (EQS) nach § 135a SGB V eine ganz besondere Rolle, soll sie doch die verschiedenen Interessengruppen, beispielsweise bei der Krankenhauswahl und der Krankenhausplanung, unterstützen. Sollten Ergebnisse der Qualitätsmessung für Krankenhäuser zukünftig negative Konsequenzen bei der Krankenhausplanung oder Krankenhausvergütung haben, so wird die Frage nach der Qualität der Qualitätsmessung spätestens dann relevant, wenn ein Krankenhaus einem entsprechenden Bescheid auf dem Klageweg widerspricht.

Dem Gesetzgeber ist dies durchaus bewusst, denn das Institut

für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) soll für die Messung der Versorgungsqualität risikoadjustierte Indikatoren und Instrumente entwickeln (§ 137a Abs. 3 Nr. 1 SGB V) und dabei gewährleisten, dass hierbei international anerkannte Standards der Wissenschaften erfüllt werden (§ 137a Abs. 5 SGB V). 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 Krankenhäusern) nur dann Sinn macht und fair ist, wenn die individuellen Risiken der behandelten Patientengruppen angemessen berücksichtigt werden. Die Anwendung von mathematischen

Durch einrichtungsübergreifende Qualitätssicherung (EQS) sollen die verschiedenen Interessengruppen, beispielsweise bei der Krankenhauswahl und der Krankenhausplanung, unterstützt werden. Wenn Ergebnisse der Qualitätsmessung für Krankenhäuser zukünftig negative Konsequenzen bei der Krankenhausplanung oder Krankenhausvergütung haben, so wird spätestens dann die Frage nach der Qualität der Qualitätsmessung relevant, wenn ein Krankenhaus einem entsprechenden Bescheid auf dem Klageweg widerspricht. Umso wichtiger ist es, dass wesentliche in der Literatur formulierte Anforderungen eingehalten und die statistischen Verfahren, welche die Qualität der Qualitätssicherung beschreiben, transparent dargestellt werden.

Keywords: Qualitätsmanagement, Recht, Risikomanagement

Verfahren zur Risikoadjustierung ist also für die Qualitätsbewertung von elementarer Bedeutung.

Gleichzeitig besteht aber die Gefahr, dass die Anwendung dieser Verfahren suggeriert, dass es sich hierbei um eine exakte Wissenschaft handelt, deren Ergebnisse eine exakte Unterscheidung in „außerordentlich gute“, „unzureichende“ (beide in § 136b Abs. 9 SGB V) oder „in erheblichem Maß unzureichende Qualität“ (§ 109 Abs. 3 Nr. 2 SGB V) im Sinne einer sicheren Beweisführung ermöglichen. Aus diesem Grund lohnt es sich, ausgewählte Aspekte zum Thema risikoadjustierter Qualitätsindikatoren (QI) näher zu beleuchten.

Qualitätsindikatoren sind nur Indikatoren, nicht die Versorgungsrealität

Ein QI dient der Bewertung, ob ein Qualitätsziel erreicht wird. Er ist kein direktes Maß der Qualität, sondern nur ein Werkzeug, das zur Leistungsbewertung dient und das die Aufmerksamkeit auf Problemfelder lenken kann, die einer Überprüfung innerhalb einer Organisation bedürfen. Solche Überprüfungen kennen wir im Rahmen der EQS als Strukturierten Dialog, der ausgelöst wird, wenn ein Ergebnis nicht im festgelegten Referenzbereich liegt.

Wie wichtig der Abgleich eines QI-Ergebnisses mit der Versorgungsrealität ist, soll am Beispiel des Indikators „Verhältnis der beobachteten zur erwarteten Rate an Todesfällen“ (Indikator-ID 50778) bei der ambulant erworbenen Pneumonie (CAP) aus dem Erfassungsjahr 2015 und dem dazugehörigen Strukturierten Dialog im Jahr 2016 gezeigt werden. Die erwartete Rate an Todesfällen wird bei diesem QI mittels der sogenannten „logistischen Regression“ risikoadjustiert berechnet. Der Referenzbereich für den QI wird mit $\leq 1,58$ angegeben, Werte darüber sind also auffällig. Im Bericht zum Strukturierten Dialog 2016 wird angegeben, dass bei 89 von 1.465 (rund 6%) betrachteten Krankenhausstandorten zu diesem QI eine rechneri-

sche Auffälligkeit festgestellt wurde. Nur bei 2 der 89 Standorte wurde die Bewertung nach Strukturiertem Dialog als qualitativ auffällig mit „Hinweisen auf Struktur- und Prozessmängel“ eingestuft, dies entspricht 2 %. Von den verbleibenden 87 Standorten wurden 49 in die Kategorien „Besondere klinische Situation“ beziehungsweise „abweichendes Ergebnis erklärt sich durch Einzelfälle“ eingestuft. Dieses Beispiel zeigt, wie wichtig eine differenzierte Analyse auch bei einem risikoadjustierten QI ist. Würde nur die Überschreitung des Referenzbereichs als Qualitätsaussage gewertet und mit negativen Konsequenzen für die Krankenhäuser belegt, so wäre dies bei 87 von 89 Krankenhäusern nicht gerechtfertigt.

Referenzbereich ist nicht gleich Referenzbereich

Nach Verständnis des IQTIG sind QI dadurch definiert, dass sie die Messung einer qualitätsrelevanten Größe (z. B. „Verhältnis der beobachteten zur erwarteten Rate an Todesfällen“) ermöglichen. Dazu müssen sie ein bestimmtes Qualitätsziel (hier: $\leq 1,58$) vorgeben, die Messmethodik darlegen (Rechenregeln) und eine Bewertung der Versorgungsqualität erlauben. Da Bewertungen der Qualität der Versorgung nur mittels eines Abgleichs des QI-Werts eines Leistungserbringers mit einem Referenzwert erfolgen können, erfordern die QI des IQTIG einen Referenzbereich. Ohne Referenzbereich bleibt unklar, welche QI-Werte als Hinweise für gute oder schlechte Qualität gelten können. Referenzbereiche werden danach unterschieden, auf welcher wissenschaftlichen Basis die Festlegung erfolgt, man unterscheidet sogenannte „Ziel- und Toleranzbereiche“.

Für einige Qualitätsindikatoren kann aufgrund wissenschaftlicher Untersuchungen definiert werden, welches Ergebnis als gute Qualität anzusehen ist, ob es also im Zielbereich liegt. Für diese Indikatoren wird ein fester Wert – also ein Zielbereich – als Referenzbereich festgelegt. So wurde beispielsweise für den QI

„Erste Blutgasanalyse oder Pulsoxymetrie innerhalb von 8 Stunden nach Aufnahme“ im Jahr 2015 als Zielbereich ≥ 95 % angegeben.

Für andere QI gibt es keine eindeutige feste Grenze und daher wird ein Toleranzbereich angegeben. So kann man das Verhältnis der beobachteten zur erwarteten Rate an Todesfällen bei der CAP als QI für Ergebnisqualität verwenden. Da CAP-Patienten aber auch bei besserer Behandlung versterben können, kann man zwar das Ziel formulieren, dass möglichst wenig Patienten versterben sollen. Man kann aber keine sichere Grenze festlegen, die erreichbar gute oder schlechte Qualität kennzeichnet. Bei diesen Qualitätsindikatoren lässt sich folglich kein Referenzbereich angeben, der erreichbare Qualität klar beschreibt. Mit Hilfe von Vergleichsergebnissen können aber besonders auffällige Ergebnisse erkannt werden. Häufig werden in diesen Fällen sogenannte „Perzentil-Toleranzbereiche“ verwendet. Ergebnisse, die innerhalb dieses Toleranzbereichs liegen, können toleriert werden. Ergebnisse, die außerhalb des Toleranzbereichs liegen, müssen genauer analysiert werden.

Bei dem vorliegenden QI wurde im Jahr 2015 festgelegt, dass 5 % der Krankenhäuser bzw. deren Standorte mit den höchsten Werten als auffällig gelten sollen. Damit wurde der Toleranzbereich von ≤ 95 %-Perzentil definiert; er beträgt $\leq 1,58$ und basiert auf den Ergebnissen der Krankenhausstandorte mit mindestens 20 Fällen. Dies bedeutet also: Erst wenn die beobachtete (also: tatsächliche) Rate an Todesfällen um mehr als 58 % über der erwarteten risikoadjustierten Rate liegt, wird der Toleranzbereich überschritten und ein Strukturiertes Dialog ausgelöst.

Umso mehr mag auf den ersten Blick erstaunen, dass – wie oben ausgeführt – der Strukturierte Dialog nur bei 2 von 89 Krankenhausstandorten Hinweise auf Struktur- und Prozessmängel ergeben hat. Würde man doch erwarten, dass die Überschreitung eines hohen Toleranzwertes mit einem hohen

Anteil an festgestellten Qualitätsdefiziten einhergeht. Tatsächlich entspricht das Resultat aber auch den Erkenntnissen aus der Fachliteratur: Sie zeigt insgesamt, dass der Anteil der Todesfälle, der durch Mängel in der Behandlung verursacht und somit als „vermeidbar“ einzustufen ist, nur einen geringen Teil aller Todesfälle ausmacht. Es wird also ersichtlich, dass selbst ein hoch angesetzter Toleranzbereich nicht als „Beweis“ für Qualitätsmängeln geeignet ist und daher ohne eine systematische Prüfung der Behandlungsprotokolle von einer vergleichenden Veröffentlichung von QI-Ergebnissen oder gar einer negativen Sanktionierung von Krankenhäusern abgesehen werden sollte. Die zukünftige Festlegung von Referenzbereichen wird in Anbetracht ihrer möglichen Auswirkungen eine große Herausforderung darstellen.

Regressionsmodelle – auch nur eine Schätzung

Sind viele potenzielle Einflussfaktoren bei der Messung eines Qualitätsindikators gemeinsam zu berücksichtigen, stellen multiple Regressionsmodelle eine geeignete Methode dar, um Risikofaktoren zu identifizieren. Bei dem QI „Verhältnis der beobachteten zur erwarteten Rate an Todesfällen“ wird die erwartete Rate der Todesfälle auf Basis eines sogenannten logistischen Regressionsmodells für die CAP-Fälle des gesamten Bundeskollektivs (2015: 289.302 Fälle) und jeden einzelnen Standort ermittelt. Das Modell berücksichtigt unter anderem das Patientenalter, das Vorliegen einer pneumoniebedingten Desorientiertheit, die Atemfrequenz und den systolischen Blutdruck.

Allgemein gesagt gilt dabei: Die erwartete Rate der Todesfälle bei den an einem bestimmten Krankenhausstandort behandelten CAP-Patienten steigt mit dem Anteil der Patienten in den höchsten Altersgruppen, dem Anteil mit pneumoniebedingter Desorientiertheit, einer hohen Atemfrequenz beziehungsweise einem niedrigen systolischen Blutdruck. Die erwartete Rate für einen Krankenhausstandort wird ermittelt, indem für jeden

CAP-Patienten die individuelle Wahrscheinlichkeit für das Eintreten des Todes unter Berücksichtigung seiner gemessenen Risikofaktoren bestimmt. Diese Ergebnisse werden dann im weiteren Verlauf zur Berechnung der risikoadjustierten QI zusammengefasst. So ergibt sich die für einen Standort erwartete Anzahl an CAP-Todesfällen aus der Summe aller modellbasierten Wahrscheinlichkeiten seiner Patienten für das Ereignis Tod. Ist die tatsächlich beobachtete Rate der CAP-Todesfälle an einem Standort kleiner als die erwartete Rate, so ist das Verhältnis der beiden Raten (hier also der QI) kleiner als 1. Im Gegensatz dazu zeigt ein QI-Wert über 1, dass die beobachtete Rate größer als die erwartete Rate an Todesfällen ist. Dies bedeutet im letztgenannten Fall, dass mehr Todesfälle eingetreten sind als aufgrund des Risikoprofils der Patienten zu erwarten gewesen wäre. Das so ermittelte Verhältnis wird auch „Standardisierte Mortalitätsrate (SMR)“ genannt. Jedoch sollte das Wort „Standardisiert“ mit Vorsicht interpretiert werden!

Selbst ausgefeilte und mit größter Gründlichkeit entwickelte logistische Regressionsmodelle ermöglichen für den einzelnen Patienten nur die Berechnung einer mit Unsicherheit behafteten Wahrscheinlichkeit im Hinblick auf ein Zielergebnis (hier: Tod bei CAP). Somit ist auch der für einen ganzen Standort ermittelte Erwartungswert als Summe der individuellen Wahrscheinlichkeiten aller CAP-Patienten nur als „Gesamtwahrscheinlichkeit“ anzusehen und auch die auf dieser Basis berechnete SMR ist daher kein „harter Beweis“, der zu Zwecken von Krankenhausrankings, öffentlicher Berichterstattung oder Sanktionen unkritisch eingesetzt werden sollte. Es gilt: Auch eine SMR ist nur prognostisch, nicht diagnostisch! Die näheren Gründe hierfür sind vielfältig und ihre Erläuterung würde über den Rahmen dieses Beitrages hinausgehen. Daher wird an dieser Stelle auf die einschlägige Fachliteratur verwiesen. Für die Akzeptanz der Verfahren und somit auch ihrer Ergebnisse ist von Vorteil, wenn sie mit dem

Qualitätsanspruch entwickelt und angewendet werden, der von den Krankenhäusern bei der Patientenversorgung berechtigterweise erwartet und gefordert wird. Da die Entwicklung insbesondere komplexer Modelle eine außerordentliche statistische Herausforderung darstellt, sollten die in der Fachliteratur beschriebenen Qualitätsanforderungen schon bei der Modellentwicklung und Modellvalidierung konsequent berücksichtigt werden.

Von großer Bedeutung ist dabei, dass wesentliche in der Literatur hierzu formulierte Anforderungen eingehalten und die statistischen Verfahren, die hierzu verwendeten Prozeduren und auch die Gütekriterien, welche die Qualität der Modelle zur Risikoadjustierung beziehungsweise der darauf aufbauenden QI beschreiben, transparent dargestellt werden. Eine solche Transparenz wurde in der Vergangenheit auch nicht ansatzweise in befriedigendem Umfang gepflegt. Die Akzeptanz und Rechtssicherheit risikoadjustierter QI hängt jedoch gerade davon ab und es ist daher im Interesse aller Beteiligten, wenn relevante Informationen nachvollziehbar offengelegt werden. Geheime Blackboxen erhöhen nicht die Akzeptanz und Rechtssicherheit, sie führen eher zu Misstrauen. ■

Literatur beim Verfasser

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



Prof. Dr. med. Andreas Becker