

9 Fallzahl als Parameter in öffentlichen Qualitätsvergleichen

Ilona Köster-Steinebach

Bringt öffentliche Qualitätsberichterstattung, das sog. *public reporting*, positive Effekte für die Versorgungsqualität? Das zu beantworten ist methodisch überaus schwierig, schließlich kann man nicht wie bei einer normalen randomisiert-kontrollierten Studie vorgehen und ein Gesundheitssystem oder zumindest einen Leistungsbereich mehrfach duplizieren, um Interventions- und Kontrollgruppe zu vergleichen. Welche Verbesserungen der Versorgungsqualität (oder deren Ausbleiben) auf die öffentliche Berichterstattung zurückgehen, ist bei der Vielzahl der parallel auftretenden Veränderungen in der Realität nicht zu isolieren, und auch internationale Vergleiche bringen hier wenig Aufschluss. Entsprechend scheiden sich in dieser Frage die Geister – unter dem Blickwinkel der Patientensicherheit durchaus erstaunlicherweise, gilt doch die Maxime „Miss es oder vergiss es!“ Unbestreitbar ist aber, dass *public reporting* eine zentrale Rolle darin spielt, Patientinnen und Patienten die Grundlage für die Ausübung qualitätsorientierter Wahlentscheidungen (in diesem Fall bei der Wahl des Krankenhauses) zur Verfügung zu stellen. Patientensouveränität bleibt ohne Qualitätstransparenz immer unvollständig, auch wenn die aktuellen Nutzerdaten z.B. der Krankenhaus-Navigatoren auf Basis der Weissen Liste nahelegen, dass derzeit nur ein vergleichsweise kleiner, wenn auch wachsender, Teil der Patienten diese Möglichkeit nutzt. Allerdings ist hier schwierig zu isolieren, ob diese geringe Nutzung auf der Erwartung bzw. dem Anspruch der Patienten beruht, dass jede Versorgungseinrichtung eine mindestens angemessene Qualität aufweist oder ob die geringe Aussagekraft und Nutzbarkeit der Informationen (vgl. unten) für das eher verhaltene Interesse verantwortlich zeichnet. Umso wichtiger sind verlässliche, aussagekräftige Informationen für öffentliche Qualitätsvergleiche. Im Folgenden soll untersucht werden, welche Rolle die Fallzahl im Zusammenhang mit der Qualitätsmessung und -darstellung in Deutschland spielt. Dazu ist zunächst ein Blick auf den Status quo der Qualitätssicherung und -berichterstattung erforderlich, um darauf aufbauend abschätzen zu können, welche Möglichkeiten, aber auch Grenzen die Nutzung der Fallzahl bietet.

9.1 Qualitätsvergleiche in Deutschland: Anspruch und Wirklichkeit

Keine Frage, die beste Grundlage für Qualitätsvergleiche von Krankenhäusern ist ein intelligent aggregierter, aus verschiedenen Indikatorentypen zusammengesetzter Qualitätsindex mindestens für jeden einzelnen Leistungsbe-
reich, besser noch für jede einzelne Leistung, die in nennenswertem Umfang erbracht wird. Bei den zu berücksichtigenden Werten sind selbstverständlich gleichermaßen Struktur-, Prozess- und Ergebnisdaten, letztere natürlich sorgfältig risikoadjustiert, aufzugreifen, idealerweise aber auch die Indikationsqualität als vierte Säule. Und bezüglich der Datenquellen, aus denen diese Angaben gewonnen werden, sollten je nach Fragestellung Dokumentationsdaten der Leistungserbringer (LE), Auswertungen aus Routinedaten sowie Befragungen der betroffenen Patientinnen und Patienten zu *patient reported experience measures* (PREM) und *patient reported outcome measures* (PROM) kombiniert werden. Das Ganze muss natürlich so gestaltet sein, dass klar zwischen sehr guter, ausreichender/erwartbarer und unzureichender Qualität unterschieden wird. Solche Indices wären auch für gelegentliche Nutzer einfach verständlich, aussagekräftig und an den praktischen Fragestellungen der Patienten orientiert. Gleichzeitig enthielten sie die für die Patienten relevanten Faktoren insbesondere der patientenbezogenen Ergebnisqualität, unterlägen wegen der Risikoadjustierung und der Vielzahl der einbezogenen Aspekte keinen bedeutenden methodisch bedingten Verzerrungen und wären wegen der Elemente der Struktur- und Prozessqualität unmittelbar handlungsleitend für die Leistungserbringer bei der kontinuierlichen Verbesserung ihrer Leistungen. Sie würden also Transparenz mit kontinuierlichen Verbesserungsanreizen verbinden.

9.1.1 Unzureichende Qualitätsdifferenzierung

Geht man nach dem soeben entworfenen Idealbild für Qualitätsmessung und -vergleiche, dann gibt es in Deutschland keine einzige hinreichend aussagekräftige Qualitätssicherungsmaßnahme. Auch die (wenigen) Verfahren der Qualitätssicherung nach §§ 136ff. SGB V werden diesem Anspruch nicht gerecht. Zentraler Problempunkt hierbei ist die mangelnde Differenzierung zwischen den Leistungserbringern. Als Ziele der externen stationären Qualitätssicherung des Gemeinsamen Bundesausschusses (G-BA) werden zwar u. a. die Unterstützung des einrichtungsinternen Qualitätsmanagements und die „Vergleichbarkeit von Behandlungsergebnissen“ genannt (G-BA 2018a: 4). Dennoch findet sich bis zum Erlass des Krankenhausstrukturgesetzes (KHSG) im Dezember 2015 in der datengestützten gesetzlichen Qualitätssicherung (QS) praktisch umgesetzt ausschließlich die Zielsetzung, unzureichende Qualität zu identifizieren und durch geeignete Maßnahmen zur Angleichung mindestens auf dem Niveau der erwartbaren Qualität zu sorgen. Folglich wurden keine Daten erhoben, die auf die Differenzierung zwischen jenen Krankenhäusern

abstellten, die oberhalb unzureichender und deshalb abzustellender Qualität angesiedelt waren.

Ausdruck der Ausrichtung der gesetzlichen Qualitätssicherung am „*bad apple picking*“ ist, dass 2016 220 von insgesamt 266 Qualitätsindikatoren, mithin 83% der gesamten Indikatoren der gesetzlichen QS, einen „normalen Handlungsbedarf“ hatten. Das bedeutet in diesem Zusammenhang nicht nur, dass „der Handlungsbedarf mit den Mitteln der gesetzlichen QS adressiert werden kann“ (IQTIG 2017: 4), sondern auch, dass es kaum Ausreißer bei den Krankenhäusern gibt – also de facto keine Grundlage für den Ausweis von Qualitätsunterschieden. Geht man im Rahmen einer groben Schätzung einmal davon aus, dass die 1.887 Krankenhausstandorte, die im Jahr 2016 Datensätze in der gesetzlichen QS geliefert haben (IQTIG 2017: 3), im Durchschnitt von zehn der insgesamt 25 Verfahren betroffen waren, in denen durchschnittlich zehn Indikatoren bewertet wurden, ergibt das ca. 189.000 krankenhausindividuelle Ergebnisse. Davon waren im selben Jahr nach Auskunft des Instituts für Qualität und Transparenz im Gesundheitswesen (IQTIG) 15.858 rechnerisch auffällig. Nach dem Strukturierten Dialog wurden noch für 1.761 Kombinationen aus Krankenhausstandort und Qualitätsindikator qualitative Auffälligkeiten ausgewiesen (IQTIG 2017: 5). Das sind 0,93%. Mit anderen Worten: Mehr als 99% der Ergebnisse der gesetzlichen Qualitätssicherung differenzieren nicht zwischen den Krankenhäusern und sind deshalb für echte Qualitätsvergleiche im Rahmen des *public reporting* ungeeignet.

! Hieraus wird deutlich, dass dringend ergänzende Informationen benötigt werden, um dem Informationsanspruch der Patientinnen und Patienten gerecht zu werden. Diese Ergebnisse stehen auch nicht im Verhältnis mit dem enormen Ressourcenaufwand, der in den Krankenhäusern mit der Dokumentation von Millionen von patientenindividuellen Datensätzen verbunden ist.

Zu erwähnen ist noch, dass möglicherweise das KHSG Bewegung in diese festgefahrene Situation gebracht hat, indem es in § 136b SGB V die qualitätsorientierte Vergütung vorgab. Für die Ausschüttung von Qualitätszuschlägen muss neben der unzureichenden auch die außerordentlich gute Versorgungsqualität identifiziert werden. Wenn also die Berichterstattung auch diejenigen Standorte mit Zuschlägen nennt, erfolgt zumindest eine gewisse Differenzierung. Da allerdings die bisherigen Indikatoren der QS, die zunächst ausschließlich als Grundlage für die Qualitätsdifferenzierung dienen können, hierfür nicht konstruiert wurden (übrigens ebenso wenig wie für die qualitätsorientierte Planung, die mit dem KHSG ebenfalls eingeführt wurde), erweist sich die Umsetzung bis zur Entwicklung neuer Indikatoren oder gar ganzer QS-Maßnahmen als (nicht nur) methodisch schwierig.

9.1.2 Einschränkungen der Datengrundlagen und Qualitätsdimensionen

Wenig ermutigend bezüglich der Eignung für *public reporting* ist auch der Blick auf die Datengrundlagen und die Vielfalt der Indikatoren hinsichtlich der einbezogenen Qualitätsdimensionen. Absolute Schwerpunkte der vorhandenen Qualitätsindikatoren sind einerseits die durch Leistungserbringer dokumentierte Prozessqualität und andererseits „Ergebnisqualität“, wobei unter letzterer vor allem die Vermeidung einzelner, schwerwiegender unerwünschter Ereignisse zu verstehen ist. Beispiele hierfür sind z.B. Mortalität, Infektionsraten oder Implantatfehlagen. Die Erreichung von Behandlungszielen wird kaum erfasst. Patientenbefragungen, obwohl schon länger gesetzlich möglich, wenn nicht gar gefordert, sind bis dato als Datenquelle in der gesetzlichen QS noch nicht praktisch umgesetzt. Entsprechend fehlen bisher auch PREMs und PROMs komplett, also die durch Patienten erlebte und berichtete Prozess- und Ergebnisqualität. Das hat zur Folge, dass weder die erreichte Lebensqualität bzw. Symptomlinderung erfasst werden können, noch Vergleiche zwischen dem Ausgangsniveau einer Beeinträchtigung und dem erzielten Behandlungsergebnis (z.B. beim künstlichen Gelenkersatz) möglich sind. Derartige Erhebungen würden auch Rückschlüsse auf die Indikationsqualität erlauben, die ebenfalls bisher keinen methodisch fundierten Niederschlag in der gesetzlichen QS gefunden hat.

Bei der Strukturqualität wiederum ging man seitens des Gemeinsamen Bundesausschusses überwiegend den Weg, entweder keine Vorgaben zu machen oder durch sog. Strukturrichtlinien in einzelnen Bereichen wie z.B. bei der Versorgung von Bauchortenaneurysmen, Frühgeborenen oder in der Kinderherzchirurgie. Dabei gilt das Alles-oder-Nichts-Prinzip: Eigentlich müssen alle Strukturvorgaben erfüllt sein, um die jeweiligen Leistungen erbringen zu dürfen, auch wenn – das zeichnet sich leider ab – die Kontrolle der Einhaltung unter dem Aspekt der Patientensicherheit zu wünschen übriglässt. Aus Sicht der Patientensicherheit potenziell ausgesprochen sinnvoll, bieten derartige Richtlinien und Vorgaben aber durch dieses Prinzip ebenfalls keine Grundlage für Qualitätsvergleiche.

9.1.3 Das Problem der Indexbildung

Nachdem, wie oben hergeleitet, wesentliche Bedingungen für öffentliche Qualitätsvergleiche durch die derzeitigen Methoden und Ergebnisse der gesetzlichen QS nicht erfüllt werden, ist es wenig erstaunlich, dass auch die dritte Voraussetzung unerfüllt bleibt, die Indexbildung. Für das interne Qualitätsmanagement und für wissenschaftliche Auswertungen ist es unerlässlich, dass die Einzelergebnisse auf Indikatorebene ausgewiesen werden. Hieraus können vor allem krankenhausindividuelle Anknüpfungspunkte für Verbesserungen der Patientensicherheit abgeleitet werden, aber auch die Versorgungsforschung kann aus diesen Daten gezielten Interventionsbedarf ableiten.

Für (potenzielle) Patienten, die sich über die Qualität einzelner Häuser in einem Leistungsbereich, vielleicht sogar bezüglich einer Leistung informieren wollen, sind solche Angaben hingegen in der Regel zu komplex, um nutzenstiftend zu sein. Patienten benötigen eine einzige aggregierte Angabe oder allenfalls eine sehr begrenzte Zahl von Qualitätsangaben - vorausgesetzt, dass trotz der Aggregation relevante Qualitätsunterschiede sichtbar werden.

Dass Nutzer von Dienstleistungen an intelligent aggregierten Index-Bewertungen auch komplexer Dienstleistungsprodukte interessiert sind, zeigt der Erfolg der Stiftung Warentest beispielsweise mit ihrem Finanztest. Am Beispiel der Bewertung eines so komplexen Produktes wie geschlossener Immobilienfonds (Stiftung Warentest 2014) kann man auch die Erfolgsfaktoren für eine gute Verständlichkeit durch Laien ableiten:

- Die einzelnen Faktoren, aus denen der Index zusammengesetzt wird, sind eindeutig und verständlich beschrieben und sehr relevant für die Nutzer.
- Die Index-Ergebnisse differenzieren zwischen Leistungsanbietern, und zwar über die ganze Bandbreite der getesteten Angebote hinweg.
- Es werden normative Grenzen für die einzelnen Qualitätsstufen gesetzt, auch wenn das zur Folge hat, dass wenige oder kein Anbieter die Kriterien zu einem hohen Grad erfüllen. Auch wird eine gewisse Willkür an den Grenzen der Qualitätsstufen in Kauf genommen.
- Es gibt keine „Bonuspunkte“ für die Einhaltung grundlegender Funktionalitäts- oder Sicherheitsaspekte, aber deren Nichteinhaltung führt zum Ausschluss oder gravierenden Abwertungen.

Bisher erfolgt in der Berichterstattung der gesetzlichen QS keine Indexbildung. Die einzelnen Indikatorergebnisse werden unkommentiert und ungewichtet nebeneinander gestellt. In der Regel erfolgt kein Versuch, die Güte einer Leistung (oberhalb von unzureichender Qualität) zu benennen oder zu quantifizieren, sondern es werden Werte genannt, bei denen es dem Nutzer überlassen bleibt, herauszufinden, welche Ausprägung guter oder sehr guter Versorgung entspricht.

Grundlegende Informationsbedürfnisse der Patienten werden auf diese Weise nicht erfüllt: Es besteht also vor allem angesichts des komplexen Sachverhalts hoher Bedarf an fokussierten und aussagekräftigen Angaben.

9.1.4 Ergänzungsbedarf der derzeitigen Qualitätstransparenz

Zusammengefasst kann man sagen, dass die deutsche QS bisher den Qualitätsvergleich, also die öffentliche Transparenz über Qualitätsunterschiede, weder ermöglicht noch für positive Qualitätsentwicklungen genutzt hat. Das

ist höchst bedauerlich, da der vielzitierte Wettbewerb um Qualität voraussetzt, dass Qualität vergleichend sichtbar ist. Wenn diese Funktion nicht von den Qualitätsangaben der gesetzlichen QS erfüllt wird, gibt es dann zumindest Näherungen auf der Grundlage anderer Angaben? Vor diesem Hintergrund stellt sich die Frage, ob bzw. welche Rolle die Fallzahl für öffentliche Qualitätsvergleiche spielen kann. Die Nutzung von Fallzahlen als „Universalindikator“ für Qualität kann jedoch bestenfalls ein Substitut bzw. Surrogat sein. Der Auftrag an den Gemeinsamen Bundesausschuss, QS-Verfahren für weitere Leistungsbereiche zu etablieren und die bestehenden Verfahren so weiterzuentwickeln, dass sie sich für Qualitätsvergleiche eignen, bleibt im Interesse der Patientensicherheit und der kontinuierlichen Weiterentwicklung der Patientenversorgung bestehen und ist höchst dringend.

9.2 Fallzahl als relevanter Faktor innerhalb der gesetzlichen QS

Es gibt kaum einen Bereich der QS, der so umstritten ist wie Ansätze rund um die Fallzahl. Das hat auch Auswirkungen auf die Studienlage. So scheint es immer wieder notwendig, den Zusammenhang von Fallzahl und Outcome neu herzuleiten und zu belegen (z.B. Nimptsch u. Mansky 2018: 66–67). Hintergrund mag auch sein, dass bis zum KHSG für die Festlegung einer Mindestmenge ein einfacher Zusammenhang zwischen Fallzahl und Ergebnisqualität nicht ausreichte, sondern ein „besonderer Zusammenhang“ gefordert war, der hohe Hürden für die beizubringende Evidenz mit sich brachte.

9.2.1 Transparenz über Mindestfallzahlen und Durchsetzung von Mindestmengen

Schon die Alltagserfahrung besagt, dass Erfahrung bei bestimmten Tätigkeiten unabdingbar dafür ist, bessere Ergebnisse zu haben. Im Zusammenhang mit medizinischen Leistungen darf dabei aber nicht nur die Erfahrung einer Einzelperson, z.B. des verantwortlichen Operateurs, betrachtet werden. Medizinische Ergebnisse sind Teamleistungen, die weit über die isolierte Performance einer einzelnen Person hinausgehen. So kann selbst der begnadetste Transplantateur keine guten Ergebnisse erreichen, wenn nicht das gesamte Team, von der Betreuung der Patienten auf der Warteliste über die eigentliche Operation bis hin zur Nachsorge, ebenso erfahren und kompetent ist. Das kann aber nur gegeben sein, wenn alle Beteiligten, ärztlich, pflegerisch und therapeutisch, in den Prozessen und insbesondere in der Zusammenarbeit untereinander geübt sind. Ein weiterer wichtiger Faktor ist die Personalausstattung: Um beispielsweise einen durchgängigen Facharztstandard für die Betreuung der Patienten vorhalten zu können, müssen mindestens zwei bis drei Ärzte vorhanden sein. Dass diese auch durchgängig mit entsprechenden Behandlungen beschäftigt sein sollten, versteht sich von selbst. Da kleine Einheiten diese Personalausstattung nicht auslasten können, halten sie sie in der Regel auch

nicht vor, was eben zu Patientengefährdungen führt (Hoffmann et al. 2018: 114). Und auch die apparative Ausstattung darf nicht übersehen werden. Wenn spezielle, vor allem teure Geräte nur selten und für wenige Patienten genutzt würden, ist der Anreiz verständlicherweise hoch, die nötigen Investitionen nicht zu tätigen. Für mehrere Indikationen, darunter Schlaganfall und Herzinfarkt, weisen Mansky et al. (2017) folgerichtig nach, dass in einem erheblichen Anteil der versorgenden Krankenhäuser, typischerweise jenen mit geringen Fallzahlen, die notwendige apparative Ausstattung für eine leitliniengerechte Versorgung nicht vorliegt. Ohne entsprechende apparative Ausstattung hingegen sind bestimmte diagnostische und therapeutische Verfahren schlicht nicht durchführbar, auch wenn sie medizinisch dringend geboten sind – mit Folgen für die betroffenen Patienten. Das alles macht deutlich, dass es aller Wahrscheinlichkeit nach für die allermeisten Leistungen im Krankenhaus einen Fallzahlbereich gibt, in dem qualitativ hochwertige Versorgung mangels Erfahrung des Teams und oft auch apparativer wie personeller Ausstattung nicht möglich ist. Derartige „Gelegenheitsversorgung“ muss also dringend identifiziert und nach Möglichkeit ausgeschlossen werden.

Die Überlegung, dass unterhalb gewisser Fallzahlen Gefährdungen für die Patientensicherheit zwangsläufig und strukturell zunehmen, liegt der gesetzlichen Mindestmengenregelung in § 136b Abs. 1 Nr. 2 SGB V zugrunde. Als Folge der bis 2016 bestehenden hohen Evidenzhürden für die Etablierung von Mindestmengen hat der G-BA jedoch bis dato nur für acht Leistungsbereiche Mindestmengen etabliert (G-BA 2018b). Diese Zahl scheint bei weitem nicht ausreichend, da der Zusammenhang von niedriger Menge und strukturell schlechteren Ergebnissen bzw. höherer vermeidbarer Patientengefährdung für die allermeisten Leistungen gegeben sein dürfte. Aber auch in Bezug auf die bestehenden Mindestmengen gibt es noch Nachholbedarf: Obwohl eigentlich Krankenhäuser unterhalb der Mindestmengen die Leistungen nicht erbringen durften, war die Durchsetzung der Regelung bis zum KHSG im besten Fall als lückenhaft zu bezeichnen. In der Folge gab es eine nicht unerhebliche Anzahl von Krankenhausstandorten, die die Mindestmengen nicht erreichten. Diese Standorte hätten in der vergleichenden Qualitätsberichterstattung unmissverständlich markiert werden müssen, da von Patienten nicht zu erwarten ist, dass sie die Leistungsbereiche und Fallzahlen der Mindestmengenregelungen des G-BA bei ihren Suchvorgängen kennen und einbeziehen. Ob die neuen Regelungen zur prospektiven Feststellung der Erlaubnis zur Leistungserbringung in § 136b Abs. 4 SGB V dafür sorgen, dass zumindest die Gelegenheitsversorgung unterhalb definierter Mindestmengen unterbleibt, wird die Zukunft zeigen. Für Patienten wäre es im Rahmen der Qualitätsberichterstattung jedenfalls von zentraler Bedeutung, wenn für möglichst viele Leistungen eine Mindestfallzahl angegeben wäre, unterhalb derer von potenziell verstärkt patientengefährdender Gelegenheitsversorgung auszugehen ist. Die hierzu nötigen Überlegungen und statistischen Erhebungen, wie sie z.B. Hoffmann et al. (2018) für die Thoraxchirurgie aufführen, fehlen für viele Leistungsbereiche.

Und selbst dort, wo in der Fachliteratur derartige Angaben vorliegen, werden diese in der vergleichenden Qualitätsberichterstattung in der Regel nicht verwendet, da bei den Betreibern der Internetportale die (berechtigte!) Befürchtung besteht, in diesem Fall von Krankenhäusern mit Fallzahlen unterhalb dieser Marke in Rechtsstreitigkeiten verwickelt zu werden. In der Summe führt das dazu, dass die wichtigen Aussagen darüber, wo aufgrund niedriger Fallzahlen mit erhöhter Patientengefährdung zu rechnen ist, die Patienten (und die einweisenden Ärzte) nicht erreichen und folglich auch nicht versorgungssteuernd wirken können. Die vergleichende Qualitätstransparenz für stationäre Einrichtungen bleibt damit weit hinter dem zurück, was z.B. durch die Stiftung Warentest für andere Dienstleistungsbereiche praktiziert wird und im Interesse der Patientensicherheit geboten wäre.

9.2.2 Fallzahlen und statistische Aussagekraft innerhalb der datengestützten QS

Eingang dieses Artikels wurde ausgeführt, dass datengestützte QS-Ergebnisse der einfachen Betrachtung von Fallzahlen in ihrer Aussagekraft überlegen sind. Entfällt also der Einbezug von Fallzahlen in Qualitätsvergleiche für all jene Leistungsbereiche, in denen datengestützte QS-Maßnahmen etabliert sind?

Zur Beantwortung dieser Frage lohnt ein Blick in eine Ausarbeitung des AQUA-Instituts aus der Zeit, in der es nach § 137a SGB V für den G-BA die gesetzliche QS betreute. Ziel der damaligen Ausarbeitung war, die Unterscheidungsfähigkeit der Indikatoren der datengestützten gesetzlichen Qualitätssicherung zu ermitteln. Diese war nach den damals vom AQUA-Institut getroffenen Festlegungen dann gegeben, wenn eine Verdoppelung der Ergebnisse eines Qualitätsindikators im Vergleich zum Bundesdurchschnitt (also z.B. eine doppelt so hohe Komplikationsrate als im Durchschnitt) mit einer gewissen Wahrscheinlichkeit statistisch nachweisbar war. Hierzu wurde eine erforderliche minimale Fallzahl abgeleitet (AQUA 2015: 12). Mit anderen Worten: Wann immer die Fallzahl eines Standortes unter dieser mit statistischen Methoden ermittelten Zahl liegt, ist selbst die Verdoppelung von Raten schwerer Komplikationen nicht sicher an den Qualitätsindikatoren ablesbar, mithin die Aussagekraft der vermeintlich zuverlässigen QS-Ergebnisse hinsichtlich Vergleichen zur Patientensicherheit stark eingeschränkt. Bleibt die Frage, wie viele Einrichtungen unter diesen „erforderlichen“ Fallzahlen bleiben. Einen Rückschluss darauf, dass ein Großteil der in der gesetzlichen QS eingesetzten Indikatoren bei den Fallzahlen, welche die typischen Leistungserbringer aufweisen, nicht verlässlich zwischen guter und schlechter Qualität unterscheiden, gibt die vom AQUA-Institut getroffene Festlegung. Demnach liegt bereits dann eine „gute Diskriminationsfähigkeit“ eines Indikators vor, wenn mehr als 10% (!) der Krankenhausstandorte die minimal erforderliche Fallzahl erreichen (AQUA 2015: 12). Diese Interpretation, die nach allen Gesetzen der Logik nicht mit dem Ziel der Patientensicherheit in

Einklang zu bringen ist, dürfte nur aus der Not geboren gewesen sein, da ansonsten offenbar geworden wäre, dass für einen Großteil der deutschen Krankenhausstandorte ein Großteil der Indikatoren der gesetzlichen Qualitätssicherung nur bei ausgesprochen großen Qualitätsdifferenzen verlässliche Ergebnisse erbringt: Bei niedrigen Fallzahlen versagen die Instrumente der gesetzlichen Qualitätssicherung aufgrund statistisch-mathematischer Gegebenheiten. Das liegt auch daran, dass eine ganze Reihe der Indikatoren vergleichsweise seltene, dafür aber schwerwiegende negative Ereignisse in den Blick nimmt. Für statistisch sichere Aussagen sind dann besonders große Fallzahlen erforderlich. Ergänzend wäre die Entwicklung von Ergebnisindikatoren sinnvoll, die auf das Ausmaß der Erreichung von Behandlungszielen abstellen.

Angesichts der soeben dargelegten Zusammenhänge sind mit Blick auf öffentliche Qualitätsvergleiche und Patientensicherheit einige Schlussfolgerungen zu ziehen:

- Unterhalb der für die Qualitätsmessung statistisch notwendigen Fallzahlen können die Krankenhausstandorte weder gute Versorgungsqualität nachweisen, noch gezielte Qualitätsentwicklung etwa auf der Grundlage von Benchmarks mit anderen Einrichtungen steuern.
- In der vergleichenden Qualitätsberichterstattung auf der Grundlage datengestützter Indikatoren sollte die Mindestfallzahl, wie sie das AQUA-Institut für die Prüfung der statistischen Diskriminationsfähigkeit ermittelt hat, unbedingt angegeben werden. Alle Indikatorergebnisse, die auf niedrigeren Fallzahlen beruhen, sollten mit ausdrücklichen Warnhinweisen gekennzeichnet sein, damit Patienten und einweisende Ärzte nicht in falscher Sicherheit gewiegt werden.
- Es sollten möglichst robustere Indikatoren entwickelt werden, die bereits bei geringeren Fallzahlen und damit anteilig mehr Einrichtungen belastbare Ergebnisse liefern.

Zusammenfassend kann man also folgern, dass mitnichten die Erhebung datengestützter Qualitätsindikatoren die Information, die sich aus (geringer) Fallzahl ergibt, obsolet macht. Im Gegenteil: Gerade aus der datengestützten QS ergeben sich wichtige Hinweise auf einen Schwellenwert für die Beurteilung der Fallzahl und mithin für die Festlegung inhaltlich begründeter Mindestmengen.

! Wenn die Fallzahl eines Hauses unterhalb des statistisch notwendigen Wertes liegt, um relevante Qualitätsprobleme anzuzeigen, bedeutet das, dass die Patientensicherheit in der betreffenden Einrichtung mit den herkömmlichen Instrumenten nicht überprüft werden kann und damit die Gefahr besteht, dass vermeidbare Patientengefährdungen vorliegen. Diese Information sollte sowohl für Patienten als auch für einweisende Ärzte verfügbar gemacht werden.

9.3 Fallzahl als universeller Ansatzpunkt für Qualitätsvergleiche

Bisher wurde hergeleitet, dass sich die Qualitätsberichterstattung auf der Grundlage der Informationen aus der gesetzlichen Qualitätssicherung, sowohl was die datengestützten Ansätze angeht als auch mit Blick auf die Strukturqualität, praktisch nicht für Vergleiche zwischen Krankenhausstandorten eignet. Davon ganz abgesehen, wird auch nur ein Bruchteil der Leistungen überhaupt von QS-Maßnahmen erfasst. In der Summe besteht also hoher Bedarf daran, Informationen über weitere Leistungsbereiche verfügbar zu machen. Vor diesem Hintergrund liegt nahe, die Fallzahl, die als Datum vergleichsweise vollständig und zuverlässig vorliegt, als eine Art „universellen Indikator“ für Qualitätsvergleiche heranzuziehen.

9.3.1 Limitationen

Bei aller Euphorie über die gute Verfügbarkeit des „Indikators Fallzahl“ für die Qualitätsberichterstattung sollten dennoch nicht die Limitationen einer solchen Vorgehensweise außer Acht gelassen werden. Hierbei sind zwei Problemkreise relevant. Der erste besteht darin, dass die Fallzahl wahrscheinlich nicht über die gesamte Bandbreite der LE hinweg differenziert. Typische Verläufe von Lernkurven (unabhängig davon, ob Individuen oder Systeme betrachtet werden) weisen am Anfang einen steilen Anstieg auf, während in oberen Bereichen nur noch marginale Zuwächse erzielt werden. Übersetzt bedeutet das, dass die Patientensicherheit und die Ergebnisqualität wahrscheinlich im unteren Fallzahlbereich in deutlich höherem Ausmaß mit der Fallzahl korrelieren als im oberen. Gerade mit Blick auf kritische Ressourcen wie Personal und Geräte kann sich bei ausgesprochen hohen Fallzahlen sogar ein negativer Effekt ergeben, wenn diese überbeansprucht, also überlastet werden. Besonders offensichtlich wird das derzeit mit Bezug auf das Pflegepersonal: Die RN4CAST-Studie hat ergeben, dass für jeden Patienten, den eine Pflegekraft pro Schicht mehr betreuen muss, die 30-Tage-Sterblichkeit aller betroffenen Patienten um sieben Prozent ansteigt (zitiert nach Busse u. Zander 2015). Und aus übergroßer Routine kann sich, ohne entsprechendes Qualitätsmanagement, auch Sorglosigkeit oder Flüchtigkeit ergeben. Damit ist ein „Universalindikator Fallzahl“ nur dazu geeignet, Leistungserbringer zu identifizieren, bei denen sich erhöhte Risiken für die Patientensicherheit ergeben, nicht aber dazu, zwischen guten und sehr guten Versorgungsangeboten zu unterscheiden.

Die zweite Einschränkung des „Universalindikators Fallzahl“ besteht darin, dass er keinen Umkehrschluss zulässt: Auch Standorte, die deutlich oberhalb von kritischen Fallzahlen arbeiten, müssen nicht zwangsläufig gute Qualität erbringen. Wenn dort beispielsweise kein gutes internes Qualitätsmanagement etabliert ist oder ausgeprägter Mangel beim Pflegepersonal herrscht, können die positiven Effekte hoher Fallzahlen von den negativen Auswirkungen

gen anderer Faktoren überkompensiert werden. Eine ausreichende Fallzahl ist damit nur eine notwendige, aber keine hinreichende Bedingung für gute Versorgungsqualität.

9.3.2 Ausgestaltung

Die Limitationen, denen der „Universalindikator Fallzahl“ unterliegt, sind kein Grund, auf diese wichtige Angabe in der Qualitätsberichterstattung zu verzichten. Stattdessen müssen die Grenzen, denen dieser Ansatz unterliegt, klar kommuniziert und durch entsprechende Ausgestaltung bestmöglich kompensiert werden. Das bedeutet vor allem, dass deutlich gemacht wird, dass Angaben zur Fallzahl insbesondere im niedrigen Fallzahlbereich eine wesentliche Orientierung bieten, nicht aber notwendigerweise im hohen.

Derzeit wird beispielsweise in der Weissen Liste bei Suchanfragen bereits eine geobasierte Darstellung von Krankenhäusern vorgenommen, bei der drei Angaben für die Einordnung eines Standortes herangezogen werden: die Bewertung im Rahmen von Patientenbefragungen, ein Index über die Erfüllung von Qualitätsvorgaben wie z.B. behandlungsrelevante Ausstattung und Hygieneergebnisse und eben die Fallzahl. Aus Sicht der patientenverständlichen und -relevanten Berichterstattung ist diese Vorgehensweise äußerst begrüßenswert. In Ermangelung inhaltlich determinierter und durch z.B. den G-BA festgelegte Grenzwerte kann hier allerdings nur zwischen unterdurchschnittlicher, durchschnittlicher und überdurchschnittlicher Fallzahl unterschieden werden. Diese Vorgehensweise birgt die Gefahr, dass sehr viele kleine Standorte den Durchschnitt so weit senken, dass dieser deutlich unter einem für die Patientensicherheit angemessenen Niveau liegt. Wesentlich besser wäre eine inhaltliche Festsetzung von Mindestfallzahlen im Interesse der Patientensicherheit, mithin eine normative Entscheidung auf der Basis der vorhandenen Studien und methodischen Ansätze, wie auch die Stiftung Warentest in den Feldern, in denen sie tätig wird, vorgeht. Hierzu sollten die in diesem Artikel angesprochenen Erkenntnisse herangezogen werden:

- **Erkenntnisse aus der Lerntheorie und der Organisationsforschung:** Es sollte ermittelt werden, wie hoch typischerweise Fallzahlen liegen müssen, um Einzelpersonen und Versorgungsteams zu ermöglichen, die Lernkurve bis zu einem Punkt zu durchlaufen (und sich dort zu halten!), an dem nur noch unwesentliche Verbesserungen durch mehr Erfahrung und Übung erzielt werden können.
- **Ökonomische Überlegungen zur Auslastung:** Da unzureichende Auslastung von personellen und materiellen Ressourcen immer den Anreiz bietet, auch qualitätskritische Ausstattung „einzusparen“, sollte die Orientierungsfallzahl immer so berechnet sein, dass eine sinnvolle Auslastung möglich ist. (Ergänzend ist wichtig, zu überprüfen, dass Versorgung nur dort passiert, wo diese kritischen Ressourcen nachweislich vorhanden sind.)

- **Statistische Aussagekraft von Qualitätsmessungen:** Die Fallzahl sollte mindestens so hoch liegen, dass Qualitätsmessungen auch bei eher seltenen patientenrelevanten Komplikationen mathematisch-statistisch aussagekräftig sind und damit die Grundlage für internes Qualitätsmanagement wie auch externe Benchmarks im Zusammenhang mit der Patientensicherheit bieten.

Auch wenn hier weiterer Forschungsbedarf bezüglich der Frage besteht, wo diese jeweiligen Schwellenwerte liegen und welche Erkenntnisse sich daraus für die Versorgungsqualität ableiten lassen, sollte das nicht als Argument dafür genutzt werden, auf Mindestmengen bzw. weitere Ansätze mit Bezug zur Fallzahl zu verzichten, während man an der Methodik der idealen Festlegung der Mindestfallzahlen arbeitet. Hier sollte das Vorsichtsprinzip im Interesse der Patientensicherheit überwiegen.

9.4 Zusammenfassung und Ausblick

Ein „Indikator Fallzahl“ muss mit Blick auf seine Limitationen gestaltet und berichtet werden, denn er kann nur notwendige, aber keine hinreichenden Bedingungen für gute Versorgungsqualität anzeigen. Hierbei wäre wünschenswert, zu einer multidisziplinären Ermittlung von Schwellenwerten bezüglich der Fallzahlen zu gelangen. Aber angesichts der Tatsache, dass der Ressourcenaufwand für eine umfassende Qualitätsmessung in der Realität nicht für alle Leistungsarten zu erbringen ist, stellt die Fallzahl einen ausgesprochen sinnvollen und tragfähigen „Surrogatindikator“ dar, der dringend umfassender als bisher genutzt werden sollte. Festlegungen von Mindestfallzahlen nach den hier vorgestellten Mustern wären auch eine überaus wichtige Information für die Qualitätsberichterstattung und könnten darüber hinaus Verbesserungen im Bereich der Patientensicherheit nach sich ziehen. Hier sind Gesetzgeber und Krankenkassen aufgerufen, entsprechende Impulse zu setzen. Auf diesem Weg besteht noch viel Raum, ja im Patienteninteresse dringender Bedarf nicht nur an weiterer Forschung, sondern insbesondere an konkretem Handeln!



Take home messages

- *Qualitätsmessung und nachfolgende -berichterstattung entsprechen derzeit nicht den Anforderungen, die an gute Informationen für Patienten und einweisende Ärzte zu richten sind.*
- *Der Aspekt der strukturellen potenziellen Patientengefährdung bei geringen Fallzahlen wird zu wenig beachtet.*
- *Die Mindestmengenregelungen des G-BA ändern diese Situation bisher nicht, da zu wenige Leistungen erfasst sind und auch die Fallzahlen zu niedrig erscheinen.*
- *Auch wenn eine umfassende Qualitätsmessung mittels Indikatoren zu Struktur-, Prozess-, Ergebnis- und Indikationsqualität sowie PREMs und*

PROMs, deren Ausweitung dringend geboten ist, erfolgen würde, verlieren ergänzende Ansätze rund um die Fallzahl nicht an Bedeutung, da der Aufwand für diesen umfassenden Ansatz nicht für alle Leistungsarten zu erbringen ist.

Literatur

- AQUA-Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen (AQUA) (2015): „Bericht zur Prüfung und Bewertung der Indikatoren der externen stationären Qualitätssicherung hinsichtlich ihrer Eignung für die einrichtungsbezogene öffentliche Berichterstattung, Stand: 17. April 2015“. URL: https://www.aqua-institut.de/fileadmin/aqua_de/Projekte/578_QI_Pruefung/Bericht_QI_Pruefung_2014.pdf (abgerufen am 20.07.2018).
- Busse, R.; Zander, B. (2015): „Welchen Einfluss haben qualitative und quantitative Parameter der Pflege in Akutkrankenhäusern auf Personal- und Patienten-Outcomes?“ URL: https://www.mig.tu-berlin.de/fileadmin/a38331600/2013.lectures/Berlin_2013.11.11.rb_RN4Cast-Web.pdf (abgerufen am 30.08.2018).
- Gemeinsamer Bundesausschuss (G-BA) (2018a): „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 (Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern/QSKH-RL) in der Fassung vom 15. August 2006, zuletzt geändert am 16. März 2018“. URL: https://www.g-ba.de/downloads/62-492-1569/QSKH-RL_2018-03-16_iK-2018-04-27.pdf (abgerufen am 11.06.2018).
- Gemeinsamer Bundesausschuss (G-BA) (2018b): „Regelungen des Gemeinsamen Bundesausschusses gemäß § 136b Absatz 1 Satz 1 Nummer 2 SGB V für nach § 108 SGB V zugelassene Krankenhäuser (Mindestmengenregelungen, Mm-R) in der Fassung vom 20. Dezember 2005, zuletzt geändert am 17. Mai 2018“. URL: https://www.g-ba.de/downloads/62-492-1601/Mm-R_2018-05-17_iK-2018-06-08.pdf (abgerufen am 19.07.2018).
- Hoffmann, H.; Passlick, B.; Ukena, D.; Wesselmann, S. (2018): „Mindestmengen in der Thoraxchirurgie: Argumente aus der deutschen DRG-Statistik“, in: Dormann, F. et al.: Qualitätsmonitor 2018, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin, S. 103–120.
- Institut für Qualität und Transparenz im Gesundheitswesen (IQTIG) (2017): „Qualitätsreport 2016 – Zusammenfassung“. URL: https://iqtig.org/downloads/berichte/2016/IQTIG_Qualitaetsreport-2016_Zusammenfassung.pdf (abgerufen am 15.06.2018).
- Mansky, T.; Drogan, D.; Nimptsch, U.; Günster, C. (2017): „Eckdaten stationärer Versorgungsstrukturen für ausgewählte Krankheitsbilder in Deutschland“, in: Dormann, F. et al.: Qualitätsmonitor 2017, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin. URL: https://aok-bv.de/imperia/md/aokbv/presse/pressemitteilungen/archiv/2016/download_qualitaetsmonitor_themenblock2_final.pdf (abgerufen am 19.07.2018).
- Nimptsch, U.; Mansky, T. (2018): „Volume-Outcome-Zusammenhänge in Deutschland“, in: Dormann, F. et al.: Qualitätsmonitor 2018, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin, S. 55–69.
- Stiftung Warentest (2014): „Geschlossene Immobilienfonds: Neun Angebote im Test – So haben wir getestet“. URL: <https://www.test.de/Geschlossene-Immobilienfonds-Neun-Angebote-im-Test-4703450-4703453/> (abgerufen am 17.06.2018).



Dr. rer. pol. Ilona Köster-Steinebach

Studium der Japanologie (M.A.) und Volkswirtschaftslehre (Promotion) in Marburg. 2004 Fachreferentin, Senior-Projekt- sowie Teamleiterin in strategischen Stabsstellen der KV Bayerns, München. 2010 Referentin im Verbraucherzentrale Bundesverband, Berlin, und Patientenvertretung im Gemeinsamen Bundesausschuss. Seit 2018 Geschäftsführerin im Aktionsbündnis Patientensicherheit.