

2 Messen und Zählen: Voraussetzung für Transparenz und Anreizbildung

2.1 Über Daten und blinde Flecken

Sind Daten erstmal vorhanden, ist ihr Charme unwiderstehlich. Man kann vergleichen, man kann sie interpretieren, man kann damit Verbesserungen und evtl. Verschlechterungen nachweisen, und wer die Datenhoheit hat, ist fein raus. Noch interessanter ist allerdings die Frage, für welche Bereiche Daten vorliegen, und insbesondere **in welchen Bereichen keine Daten vorliegen**. Die Identifikation der Themenbereiche, für die Daten erhoben werden sollen, ist also der erste Schritt, Daten an sich haben keinen Wert. Daten gewinnen nur einen Wert auf der Basis einer vorangegangenen Problemanalyse. Da die Problemanalyse stark vom jeweiligen Entwicklungszustand eines Gesundheitswesens abhängt, ergibt sich jeweils ein typisches Bild: Zeig' mir Deine Daten, dann sage ich Dir wer Du bist. So stellen Gesundheitssysteme in der Phase der Kostenkontrolle (in Deutschland vor 1993) nur Daten zu ausgewählten Krankheitsbildern zur Verfügung (in Deutschland z.B. die Perinatalerhebung), Gesundheitssysteme in der DRG-Ära dagegen sektorale Daten vor allem aus dem Krankenhausbereich, und höhergradig integrierte Systeme (z.B. *Managed Care*) dann schon eher populationsbezogene Qualitätsdaten.

Jedes Gesundheitssystem, jede Entwicklungsphase, die ein Gesundheitssystem durchläuft, hat also seine typischen Daten zu den aktuellen Problemreichen – aber genauso auch seine typischen „blinden Flecken“. Die Ersteren kann man als „**bekanntes Wissen**“ bezeichnen (s. Abb. 4), es handelt sich um

		Daten	
		vorhanden	nicht vorhanden
Problem	bekannt	bekanntes Wissen elektive Akuterkrankungen, z.B. Cholezystektomie	bekanntes Unwissen konservative Fachgebiete, Stürze
	unbekannt	unbekanntes Wissen stationäre Aufnahme wegen Exsikkose Patienten >65 J	unbekanntes Unwissen Verhinderung der Pflegebedürftigkeit

Abb. 4 Datenauswahl und -verfügbarkeit. „Problem unbekannt“ bezieht sich auf das Gesundheitssystem, wissenschaftlich ist die Exsikkoseproblematik natürlich bekannt.

Daten über relevante Problemstellungen, die auch als Problem anerkannt sind. Als in Deutschland mit dem Gesundheitsstrukturgesetz von 1992 für einige operative Indikationen die Sonderentgelte und Fallpauschalen eingeführt wurden (im Gegensatz zu den heutigen DRGs noch nicht komorbiditäts- und risikoadjustiert), hatte man die Befürchtung, dass es bei pauschalierter Vergütung z.B. über eine zu starke Verkürzung der Verweildauer zu Qualitätsmängeln kommen könnte. Daher führte man Qualitätssicherungsprogramme ein, organisiert auf Länderebene. Heute, unter DRG-Bedingungen, ist die externe Qualitätssicherung entsprechend den Bestimmungen der §§ SGB V 135a, 137 und 137a weitgehend auf Bundesebene organisiert, von 2002 bis 2008 getragen durch die Bundesgeschäftsstelle Qualitätssicherung (heute BQS-Institut GmbH) und seit 2009 getragen durch das AQUA-Institut GmbH.

Man erkennt auf den ersten Blick, dass es sich bei den Qualitätssicherungsdaten nach § 137 und § 137a nicht um solche handelt, die mehrere der oben beschriebenen Qualitätsperspektiven oder -dimensionen abbilden (s. Kap. 1.3 und 1.4). Es handelt sich um rein **diagnose- und prozedurenbezogene Qualitätsparameter**, und selbst hier fast ausschließlich nur um solche, die die stationäre, operative Medizin betreffen. Bei der Analyse der 31 Krankheitsbilder bzw. Prozeduren, für die wir unsere Qualitätssicherung vorhalten, wird klar, dass diese nur einen sehr kleinen Teil unserer Gesundheitsversorgung bzw. sogar unserer stationären Versorgung abbilden. Es handelt sich im Einzelnen um

- **Transplantationschirurgie:** Herz, Leber, Lunge, Niere, Pankreas-Niere, weiterhin Leber- und Nierenlebendspende (7 Krankheitsbilder)
- **Kardiologie/Herzchirurgie:** Schrittmacher, Defibrillatoren, Stenteinlage Koronararterien durch Katheter (PCI), Aortenklappen isoliert chirurgisch/per Katheter, Bypass isoliert, Kombination Bypass/Aorten-chirurgie (11)

- **Hüftgelenk und Endoprothetik:** Hüftgelenksnahe Femurfraktur, Hüft-TEP/Wechsel, Knie-TEP/Wechsel (5)
- Geburtshilfe und Neonatologie (2)
- Operative Gynäkologie und Mamma-Chirurgie (2)
- Weitere Operationen: Cholezystektomie und Carotis-Revaskularisation (2)
- Konservative Fächer: Ambulant erworbene Pneumonie (1)
- Pflege: Dekubitus (1)

Es besteht also ein massives Übergewicht akuter, operativer Krankheitsbilder bzw. Prozeduren mit Schwerpunkt in der Maximalmedizin (Transplantationschirurgie), der Herzerkrankungen (Schrittmacher, Herzchirurgie) und Endoprothetik; die konservativen Fächer fehlen bis auf die ambulant erworbene Pneumonie vollständig, chronische konservative Krankheitsbilder sind nicht berücksichtigt (s. Abb. 5).

Das Bild wird noch weitaus deutlicher, wenn man die **Indikatoren bzw. Indikatorengruppen** entsprechend der Spezifizierungen für das Jahr 2013 (www.zqe.de) zugrundelegt (s. Abb. 6). Unter Berücksichtigung aller 464 einzelnen Indikatoren sind 105 Indikatoren Mortalitätsindikatoren (23%), davon kommen ganze 45 Indikatoren aus der Transplantationschirurgie (43%) und 35 Indika-

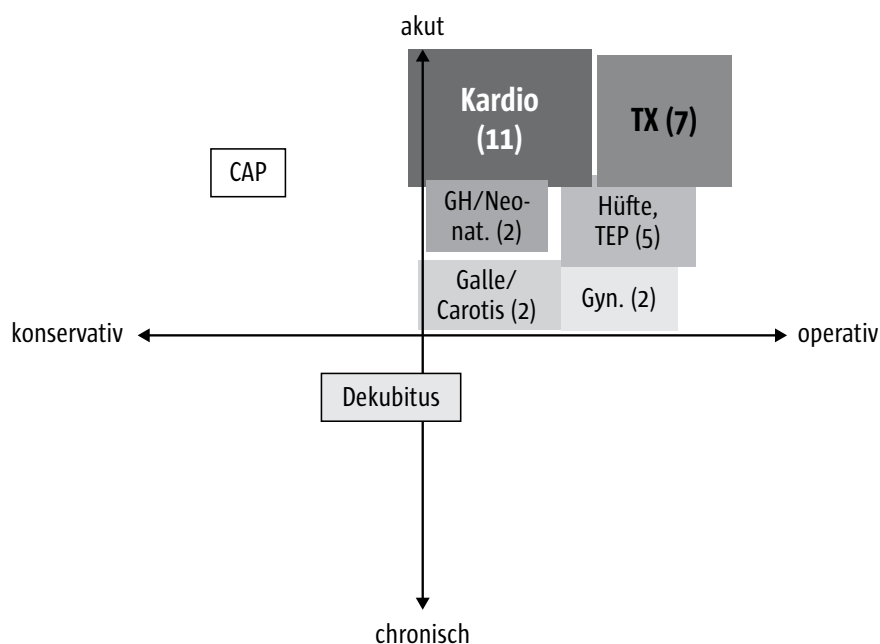


Abb. 5 Schematische Darstellung der Krankheitsbilder/Prozeduren der Ext. Qualitätssicherung nach § 137a SGB V für 2013 entsprechend der Merkmale akut vs. chronisch, operativ vs. konservativ. TX Transplantation, GH Geburtshilfe, TEP Totalendoprothetik, CAP Ambulant erworbene Pneumonie, in Klammern: Zahl der Krankheitsbilder (www.zqe.de)

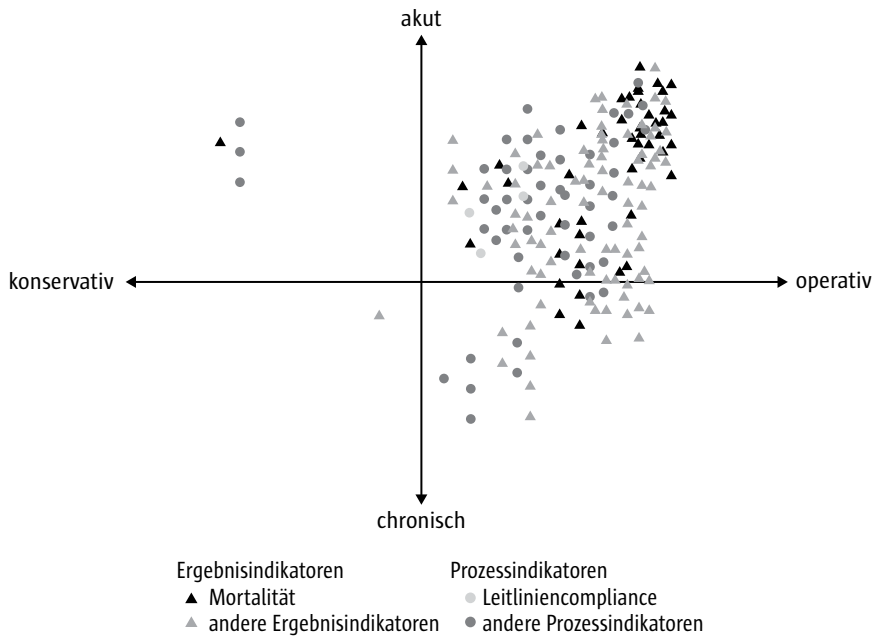


Abb. 6 Indikatoren und Indikatorengruppen der Ext. Qualitätssicherung nach § 137a SGB V entsprechend der Merkmale akut vs. chronisch und operativ vs. konservativ

toren (33%) aus der interventionellen Kardiologie/Herzchirurgie (exkl. Herztransplantation). Wie oben bereits ausgeführt, sind weite Bereiche der Gesundheitsversorgung wie die konservativen Fächer und chronische Erkrankungen kaum abgebildet, während die operative Maximalversorgung außerordentlich dominant erscheint.

Wie ist dieser Befund umfangreicher Areale des „**bekannten Unwissens**“ zu interpretieren? Drei Punkte sind hier zu nennen:

1. Die externe Qualitätssicherung nach § 137a SGB V in der heutigen Form ist dringend überarbeitungsbedürftig, weil selbst die stationäre Gesundheitsversorgung nicht adäquat abgebildet wird.
2. Wenn man von der Hypothese ausgeht, dass die mit der Qualitätssicherung verbundene Transparenzsteigerung eine Verbesserung der Versorgungsqualität zur Folge hat (zu den Grundlagen dieser Annahme und den impliziten Rahmenbedingungen s. Kap. 5.2), dann sind weite Bereiche der Gesundheitsversorgung von diesen Verbesserungsoptionen ausgeschlossen.
3. Die ausgeschlossenen Bereiche betreffen – und das ist besonders schwerwiegend – genau die Bereiche, die für die zukünftige Gestaltung des Gesundheitssystems von besonderer Bedeutung sind (s. „fünf Ziele“ in Kap. 1.4): die chronischen Erkrankungen bei älteren Patienten, die kon-

servativ zu behandelnden Erkrankungen wie Diabetes mellitus, Herzinsuffizienz, Hypertonie – hier herrscht völlige Fehlanzeige.

An dieser Stelle muss man sich aber nochmals der Abbildung 4 zuwenden, denn neben dem „bekannten Wissen“ und „bekannten Unwissen“ gibt es ja noch zwei weitere Bereiche. Interessant ist zunächst das „**unbekannte Wissen**“, also Qualitätsprobleme in Bereichen, wo Daten zur Genüge vorliegen, aber sich entweder niemand daran interessiert oder man die Problematik überhaupt nicht im Blick hat. Ein Beispiel könnte hier z.B. die stationäre Aufnahme wegen unerwünschten Arzneimittelwirkungen sein, grundsätzlich durch die Aufnahmediagnose dokumentiert, die bei jeder Krankenhausaufnahme erhoben wird. Dieses Phänomen ist international (Pirmohammed et al. 2004, Pouyanne et al. 2000) und in Deutschland (Schneeweiss et al. 2002) hervorragend untersucht, bei 2 bis deutlich über 5% der Patienten, die ins Krankenhaus aufgenommen werden, ist dies wegen unerwünschter Arzneimittelwirkungen der Fall. Dies entspricht der eindrucksvollen Zahl von mindestens 380.000 Krankenhausaufnahmen im Jahr in Deutschland. Wie gesagt, hier handelt es sich um Probleme, wo Daten zur Verfügung stehen, „man müsste nur die Augen auf machen“.

Weitaus schwieriger gestaltet sich die Situation, wenn Probleme nicht bekannt sind oder nicht realisiert werden und außerdem keine Daten vorliegen (die „**blinden Flecken**“, „**unbekanntes Unwissen**“). So war im Jahr 1992 bei der Einführung des Gesundheitsstrukturgesetzes die Notwendigkeit der Qualitätssicherung wegen der Fallpauschalen vielleicht noch klar, aber eine zweite Entscheidung, nämlich die Festschreibung sektoraler Budgets, wurde hinsichtlich ihrer Qualitätsauswirkungen ausgeblendet – ein blinder Fleck, der bis heute Bestand hat. Trotz mannigfaltiger Bemühungen der transsektoralen Qualitätssicherung wissen wir heute über die Qualitätsprobleme an den Schnittstellen ambulant/stationär/Pflege/Rehabilitation und über Qualitätsprobleme in der regionalen Versorgung (*area*-Indikatoren) so gut wie nichts – obwohl hier die relevantesten Qualitätsprobleme vorliegen, die wir für die Versorgung der Bevölkerung gerade in den nächsten Jahrzehnten zu vergegenwärtigen haben. Ältere Menschen mit chronischen und Mehrfacherkrankungen brauchen vor allem eine gute Koordination über die Sektorgrenzen hinweg. Als Beispiel sei hier die Verhinderung der Pflegebedürftigkeit von gesetzlich krankenversicherten Patienten genannt, wo immer wieder Berichte auftauchen, dass die „Entlassung“ in den Nachbarsektor (Pflegeversicherung) einer aktiven Mobilisierung vorgezogen wird.

Zusammenfassend kann also Folgendes festgehalten werden:

- Die externe Qualitätssicherung nach § 137/137a SGB V bezieht sich lediglich auf diagnose- und prozedurenbezogene Qualitätsaspekte der stationären Versorgung und lässt andere Perspektiven (z.B. Nutzen, Bedarf, Patientenperspektive) außer Betracht.

- Selbst die stationäre Gesundheitsversorgung wird durch die externe Qualitätssicherung nach §137a SGBV nur sehr unvollständig abgedeckt und bedarf daher – vor Diskussion anderer Aspekte – einer energischen Neuorientierung.
- Bereiche, in denen Daten zwar vorliegen, diese aber nicht aktiviert werden, sind vorhanden (z.B. medikationsbedingte Notfallaufnahme ins Krankenhaus) und müssen im Weiteren aktiv identifiziert werden.
- Wenn man die heutige Qualitätssicherung unter dem Aspekt der Entwicklungsnotwendigkeiten im deutschen Gesundheitssystem betrachtet („fünf Ziele“), dann kann nur festgehalten werden, dass keines dieser Ziele angesprochen wird (weder Qualitätsorientierung [statt Menge] noch chronische Mehrfacherkrankungen, Integration, Prävention oder Patientenbezug). Eine zukunftsfähige Qualitätssicherung muss dringend auf diese „blinde Flecken“ Bezug nehmen.

Worin liegt diese Situation begründet? Die Antwort ist eindeutig: in der Vergütung. Die Vergütungssysteme begünstigen allein die sektorale Optimierung hinsichtlich Preis und Menge, und Qualitätsstrategien haben gegen die Vergütungsanreize keinerlei Chance. Wie später noch auszuführen sein wird: Qualität hat nur eine Chance, wenn gleichzeitig Vergütungsanreize verändert werden, das Gesundheitssystem also weiterentwickelt wird. Somit ist die erste Phase der Entwicklung des Themas Qualität in Deutschland, die 1993 mit sektoraler Perspektive begonnen hat, endgültig zu ihrem Ende gekommen. Diese Aussage gilt ausdrücklich auch angesichts der vom Gesetzgeber stark betonten transsektoralen Ansätze zur Qualitätssicherung (§ 137a), denn diese versuchen die Sektoren durch Qualitätsvereinbarungen zu verbinden, ohne die Sektorengrenzen anzutasten (s. Kap. 2.4.8 und 2.4.9). Ohne analoge Änderungen der Vergütungsstrukturen wird sich keine wirkliche Änderung an der jetzigen, auf operativ zu behandelnde Akuterkrankungen zugeschnittenen Qualitätssicherung ergeben. Wir stehen erst am Anfang einer 2. Phase, die durch Koordination, netzwerkartige Zusammenarbeit, die Versorgung von Populationen und die Demographiefrage bestimmt ist.

2.2 Qualität: welcher Blick auf die Zahlen?

Um nur Eines vorwegzunehmen: ein Problem haben wir in Deutschland wirklich nicht, nämlich einen Mangel an belastbaren Daten. Wir erheben Qualitätssicherungsdaten *en masse*, wir forschen mit feingeschliffenen klinisch-epidemiologischen Instrumenten, wir kümmern uns um hochspezialisierte Teilbereiche, wir vergleichen mit anderen Ländern, wir befragen die Patienten als die Nutzer unseres Gesundheitssystems, wir kombinieren das Wissen aus Studien aus der ganzen Welt. Und trotzdem hat es des Öfteren den Anschein, als glaubten wir unseren eigenen Zahlen nicht, als sähen wir den Wald vor Bäumen nicht, so zuletzt bei Frage der der Größenordnung von unerwünschten Ereignissen im Krankenhaus. Was sind die Gründe?

Mit dem Zählen und den Zahlen ist es so eine Sache. Meist will man es gar nicht so genau wissen, man weiß doch sowieso wo man dran ist. Und wenn man gezählt hat, dann – ist da plötzlich eine Zahl. 18.800 Todesfälle. 5mal zu spät gekommen in den letzten drei Monaten. Zahlen lassen sich nicht mehr wegdiskutieren, sie üben eine geheimnisvolle Anziehungskraft aus. Doch was sagen Zahlen eigentlich aus?

Das hängt maßgeblich damit zusammen, **mit welcher Absicht**, d.h. mit welchem Erkenntnisinteresse man sich ans Zählen macht. Manch einer hat vielleicht schon mal auf dem Weg zur Arbeit genau die Minuten im Stau gezählt. Er wollte es genau wissen. Aber dann ist es auch schon zu spät, im Stau ist nun mal im Stau. Eine Autofahrerin neben ihm macht es anders, sie sagt: was sie sich noch ärgern, wenn man schon im Stau steckt. Sie zählt die Ampelphasen an der Tankstelle an der Ausfallstraße, kurz vor der Autobahneinfahrt. Wenn es mehr als dreimal Rot wird, bevor sie über die Ampel kommt, dann – ruft sie bei der Arbeit an, dass sie später kommt. Die Zahl der Ampelphasen ist ihr eigentlich egal, aber sie weiß jetzt, wo sie dran ist, und macht sich vielleicht gute Musik im Auto an, dauert heute ja sowieso länger. Oder sie nimmt einen Umweg, um den Stau zu vermeiden. Als sie dann schon fast bei der Arbeit angekommen ist, fährt sie über eine Brücke, es ist Winter muss man dazu wissen. Hier fällt ihr immer ein, jedesmal wenn sie diese Brücke überquert, dass ihr ein Arbeitskollege erzählt hat, dass hier an dieser Stelle im Winter einmal Glatteis war, dass er ins Schleudern gekommen sei. Und das hat sich ihr eingeprägt, sie passt unwillkürlich auf, obwohl – so kalt ist es heute gar nicht. Aber besser ist besser, aufpassen kann man ja.

Genauso machen wir es auch im Gesundheitswesen, aber oft achten wir nicht darauf, dass wir verschiedenen gewonnene Zahlen gar nicht miteinander vergleichen können. Richtig, manchmal müssen wir es so machen wie der Autofahrer im Stau, wir zählen die *bad apples*, die schlechten Äpfel: es handelt sich dann um sog. **quantitative Verfahren**. Man muss halt manchmal wirklich genau wissen, wo man steht, auch wenn es unangenehm ist. Hier kommt es dann darauf an, dass man sich vorher überlegt, was man zählt, denn man braucht ja einen Bezugspunkt. 5 nosokomiale Infektionen auf 100 Patienten, das trifft es besser, als 5 Infektionen – ja in Bezug auf was denn eigentlich. Vermeiden lässt sich das Zählen von *bad apples* nicht, denn irgendwie muss man ja feststellen, wie groß der Handlungsbedarf ist. Es ist nicht gerade beliebt, darüber zu sprechen, aber vor allem ist das Kind ja schon in den Brunnen gefallen, es ist schon ein Schaden eingetreten. Es geht bei den *bad apples* also um Komplikationen oder **unerwünschte Ereignisse (UE)**, es geht aber auch um Erfahrungen, vor allem solche, die die Patienten gemacht haben. Ein Beispiel für eine Komplikation bzw. UE ist eine postoperative Wundinfektion, ein Beispiel für die zweite Form der **Patientenerfahrung** wäre die Tatsache, dass der Hausarzt nach der Entlassung aus dem Krankenhaus keinen Arztbrief erhält und der Patient sich unsicher ist, ob sein Arzt richtig informiert ist für das kommende Wochenende. Patientenerfahrungen spielen in der Qualitäts-

berichtserstattung eine immer wichtigere Rolle, so z.B. im *Value-Based Purchasing* (VBP) Programm in den USA, wo 20% der qualitätsbezogenen Vergütung auf der Basis von Patientenerfahrungen bzw. -befragungen verteilt wird (Ryan und Blustein 2012).

Man kann aber auch die Zahl der falsch beschrifteten Laborröhrchen auf der internistischen Station zählen, vielleicht fehlt nur das Geburtsdatum und passiert ist deswegen gar nichts, trotzdem man weiß: die Gefahr, dass ein Röhrchen verwechselt wird (eine Form der Patientenverwechslung), ist hier höher als bei einer anderen Station, wo das nicht vorkommt. Man nennt das einen **Indikator**, da kommt es weniger auf das Ereignis selbst an (wie beim Unerwünschten Ereignis), sondern darauf, dass dieses etwas Unangenehmes vorhersagt, so wie die Autofahrerin es mit ihrer Ampel macht (man sagt auch, der Indikator hat eine „Ampelfunktion“).

Und die dritte Antriebskraft des Zählens kommt in der neueren Zeit sogar immer häufiger vor. Wir zählen zum Beispiel in einem *Critical Incident Reporting System* (CIRS), einem freiwilligen Fehlerberichtssystem, dass irgendwo im Krankenhaus etwas schief gelaufen ist, dass ein Patient vor dem Röntgen fast aus dem Bett gefallen war. Das hat allen Beteiligten, dem Patienten natürlich am meisten, gehörigen Schreck eingejagt, aber dem Patienten ist nichts passiert. Jedoch hat man deswegen die Übergabe vom Transportdienst zur Röntgenabteilung komplett anders organisiert. Hier kommt es gar nicht darauf an, ob das einmal oder zweimal passiert ist, sondern nur dass man auf ein Problem aufmerksam gemacht wird, damit man etwas dagegen tun kann: **Analyse und Prävention**. Man kann auch sagen: diese Methoden beheben den blinden Fleck, sie **generieren** „bekanntes Wissen“ aus sog. *Unknown Unknowns*. Eine ähnliche Funktion haben die Morbiditäts- und Morbiditätskonferenzen der Abteilungen, auch die Analyse der Patientenbefragungen und die Berichte der Patientenbeauftragten, und deswegen ist es auch so wichtig, dass die Patienten „sich beschweren“, eine kostenlose Beratung sagt man immer. Ja, das stimmt, wenn man sich nicht beschwert, merkt niemand, dass sich etwas ändern muss. Auf der Ebene des Gesundheitssystems ist das Pharmakovigilanzsystem des Bundesinstitutes für Arzneimittel und Medizinprodukte ein Beispiel für ein ähnliches System. Wichtig ist: mit diesen Daten kann man nicht zwischen „viel“ und „wenig“ unterscheiden. Der Bürgermeister, der sagt, im Krankenhaus der Stadt sei die Patientensicherheit schon viel „besser geworden“, weil die Meldungen im CIRS zurückgegangen ist, hat wahrscheinlich nicht recht, denn meist ist so ein Rückgang auf eine zunehmend schlechte Motivationslage der Mitarbeiter zurückzuführen, die nicht erkennen, dass sich aufgrund der Meldungen etwas tut, und es daher nicht recht einsehen, warum sie sich die Mühe machen sollen. Wäre die Patientensicherheit wirklich besser geworden, hätte man eine Zunahme der CIRS-Meldungen erwartet.

Tab. 1 Abgrenzung von Analyseinstrumenten, quantitativer Erfassung und Indikatoren. Sens. Sensitivität, Spez. Spezifität

	Erkenntnisinteresse	Zielereignis	Statistik	Besonderheiten
Generierende Methoden (z.B. CIRS)	Generierung: Identifikation unbekannter Ereignisse	Unbekannt	Sens. < 1%, Spez. gering	Nicht zur Quantifizierung geeignet, Exploration
Quantitative Erfassung	Quantitative Analyse definierter Ereignisse	Unerwünschtes Ereignis	Spez. und Sens. hoch	Kann nur erfassen, was man kennt; Umgebungsvariablen erlauben Risikofaktoranalyse
Indikatoren	Monitoring und Risikobewertung	Qualität und Sicherheit	Sens. hoch, Spez. sekundär	Zielereignisse sollten beeinflussbar sein

Die **drei geschilderten Zugänge** zu Daten und Messergebnissen, nämlich

- quantitative Erfassung,
- Indikatoren und
- generierende Methoden (z.B. CIRS)

weisen große Unterschiede auf, werden aber häufig miteinander vermengt (s. Tab. 1 und Abb. 7). Zum einen, wie oben genannt, spielt das Erkenntnisinteresse eine große Rolle: mit den Analyseinstrumenten wie CIRS möchte man auf Sachverhalte aufmerksam gemacht werden, über die man vorher keine Kenntnis hatte (Generierung), die quantitativen Methoden dienen der genauen Erfassung und die Indikatoren bewerkstelligen das anschließende Monitoring (s. Abb. 8). Bei Analyseinstrumenten wie CIRS ist das Zielereignis also nicht bekannt; dagegen kann die quantitative Erfassung nur bei Ereignissen eingesetzt werden, die vor Beginn des Messvorganges bekannt und

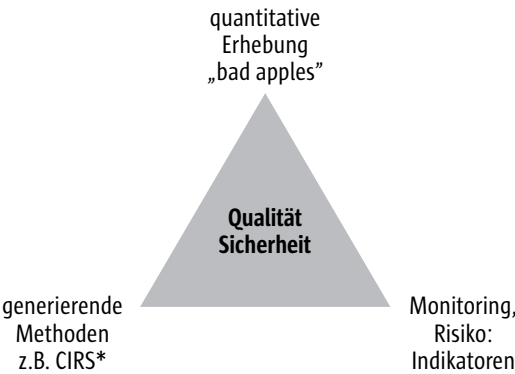


Abb. 7 Messen und Zählen: drei verschiedene Zugänge. * CIRS Critical Incident Reporting-System

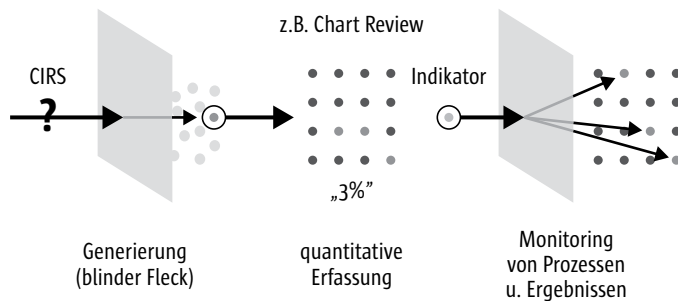


Abb. 8 Drei Ziele der Datenerfassung: CIRS zur Generierung vorher nicht bekannter Ereignisse und als Voraussetzung einer quantitativen Erfassung. Die Identifikation eines Indikators erlaubt das darauf folgende Monitoring.

genau spezifiziert werden konnten. Generierende Analyseinstrumente wie CIRS sind nicht sensitiv und müssen auch nicht spezifisch sein, während die quantitative Erfassung nur mit Verfahren Sinn macht, in denen Sensitivität und Spezifität in einem ausgewogenen Verhältnis stehen (mit leichten Vorteilen der Spezifität). Eine besondere Rolle spielt die Sensitivität bei Indikatoren: diese müssen hochsensitiv sein, sie sollen alle interessierende Ereignisse vorhersagen (u.U. auf Kosten der Spezifität, man wird „lieber einmal zu viel gerufen“).

Nur auf den ersten Blick mag diese Unterscheidung lediglich theoretischer Natur erscheinen, ist aber **in der Praxis und der Interpretation** der qualitätsrelevanten Daten von grundlegender Bedeutung (s. Abb. 9). Die quantitativen Verfahren statten uns mit grundlegenden Informationen aus (Mortalität, Komplikationen), die entweder einer Einzelfallanalyse oder – wenn als Rate dargestellt – einer kumulierten Analyse zugeführt werden. Indikatoren weisen auf Auffälligkeiten hin oder nicht – in ersterem Fall muss es auch zu einer Analyse kommen. Man bedient sich hier entweder epidemiologischer Verfahren (z.B. Regressionsanalyse) oder einem Vorgehen wie dem Ishikawa-Diagramm bzw. im Risikomanagement der *Root-Cause-Analyse* oder ähnlichen Instrumenten. Allen diesen Instrumenten ist Eines gemeinsam: sie können sich nur auf Ursachen beziehen, die als Faktor bekannt sind; im Bereich der „Unknown Unknowns“ (blinde Flecken) sind sie machtlos, denn der nicht bekannte Faktor kann nicht in die Analyse einfließen. Hier helfen die analytisch generierenden Verfahren wie CIRS weiter, denn sie können unbekannte Faktoren aus dem Bereich des unbekannten Unwissens in den Bereich der bekannten Faktoren überführen, so dass sie in die Ursachenanalyse jedweder Provenienz einfließen können.

Im **Risikomanagement** sind diese unterschiedlichen Zugänge Bestandteil der täglichen Praxis (s. Kap. 4.2). Unerwünschte Ereignisse, über die man genau Bescheid wissen muss, werden mit quantitativen Methoden erhoben, das Monitoring wird mit Patientensicherheitsindikatoren bewerkstelligt, und mittels

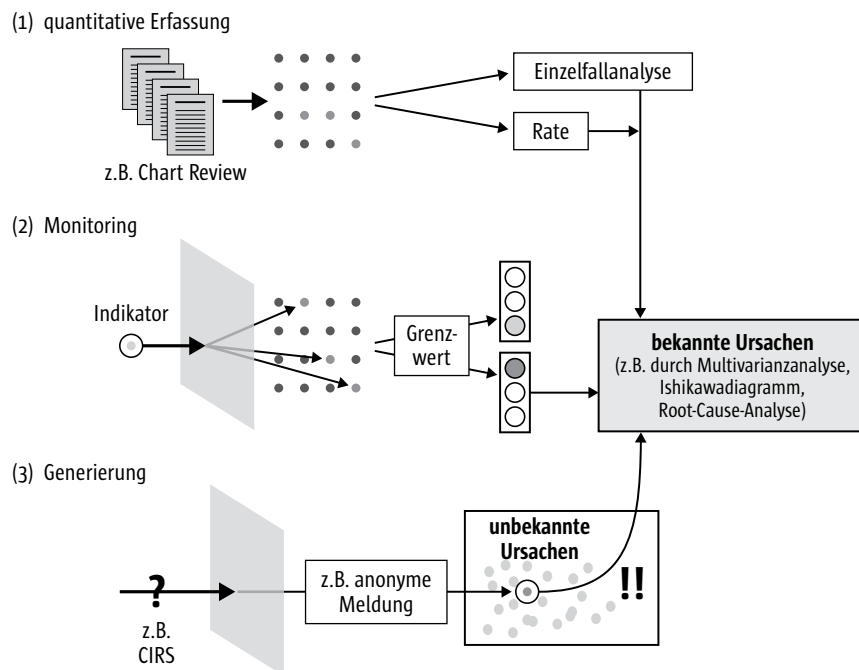


Abb. 9 Die Kombination der drei Zugangswege kann auch unbekannte Faktoren mit einbeziehen, soweit sie von den generierenden Verfahren (z.B. CIRS) identifiziert werden.

CIRS und verwandten generierenden Methoden (z.B. *Morbidity-Mortality-Konferenz*, *Obduktion*, Analyse von Patientenbefragungen und Beschwerdemanagement) werden Ursachen identifiziert, über deren Existenz man vorher keine Kenntnis hatte, und die dann in die Ursachenanalyse einfließen.

Auch (und das ist erstaunlich) das Sozialgesetzbuch V kennt alle drei hier genannten Zugangsmöglichkeiten; der Schwerpunkt wird dabei auf den Begriff „Qualität“ (§ 2; Überschrift und Text des neunten Abschnitts: „Sicherung der Qualität der Leistungserbringung“) bzw. „Verfahren zur Messung und Darstellung der Versorgungsqualität“ (§ 137a, Abs. 1) und „Indikatoren“ (§ 137a Abs. 2) gelegt. Aber auch wichtige quantitativ erhobene Daten werden genannt, z.B. in § 137 Abs. 1a unter Bezugnahme auf das Infektionsschutzgesetz. In § 137 Abs. 1d (in Verbindung mit § 135a Abs. 3 sowie der Qualitätsmanagementrichtlinie Krankenhäuser [KQM-RL] in der Fassung vom 23.1.2014) werden wichtige Details zu einrichtungsinternen und -externen CIRS-Systemen zur Generierung von Daten genannt (ausführliche Schilderung der gesetzlichen Regelungen in Kapitel 7).

In der Folge bzw. im nächsten Kapitel wird daher noch genauer auf die Unterscheidung und die unterschiedliche Anwendung von quantitativer Erfassung und Indikatoren eingegangen; CIRS und andere generierende Verfahren werden hier nur genannt, sie sind nicht primär Gegenstand dieses Gutachtens.

Zusammenfassung:

Je nach Zugang und Erkenntnisinteresse sind quantitative Verfahren, Monitoringverfahren (Indikatoren) und generierende Verfahren (zur Identifikation „blinder Flecken“ [z.B. CIRS]) zu unterscheiden. Quantitative Verfahren ergeben valide Informationen z.B. über Komplikationen oder Patientenerfahrungen, Sensitivität und Spezifität müssen in einem ausgewogenen Verhältnis stehen. Bei den Indikatoren (Monitoring) stehen die Anforderungen an die Sensitivität ganz oben auf der Prioritätenliste, denn man will in der Vorhersage unerwünschter Ereignisse immer gewarnt sein. Sowohl quantitative Verfahren als auch Indikatoren können jedoch nur messen, was als Parameter bekannt ist (gleiches gilt für Verfahren wie die Regressionsanalyse). Gerade im Rahmen der Diskussion um die Patientensicherheit sind daher explorative bzw. generierende Verfahren in den Vordergrund gerückt, die Informationen über Ereignisse liefern, die vorher unbekannt waren. Neben dem CIRS (*Critical Incident Reporting System*) sind hier *Morbidity-Mortality*-Konferenzen und z.B. die Analyse von Patientenbeschwerden zu nennen. Gemeinsam ist diesen Verfahren, dass sie keine nennenswerte Sensitivität aufweisen, also zum Vergleich über die Zeit oder interinstitutionell nicht zu verwenden sind. Bemerkenswerterweise führt das Sozialgesetzbuch V alle drei existierenden Datentypen für Qualitätssicherung und -management auf.

2.3 Quantitative Erfassung

Die quantitative Erhebung von Qualitätsparametern hat in vielerlei Hinsicht eine hohe Priorität, weil auf die entsprechende Information nicht verzichtet werden kann. Wichtige Beispiele sind

- nosokomiale Infektionen,
- unerwünschte Arzneimittelereignisse (populationsbasiert) und
- Mortalität.

Die Erfahrung mit der quantitativen Erfassung von Ereignissen ist mit Abstand am größten in der Infektiologie bzw. im Bereich *Infection Control* (**Infektions-epidemiologie**; s. Kap. 3.2.5). Ausgehend von den Falldefinitionen der *Centers of Disease Control* (CDC) in Atlanta aus den 70er-Jahren sind hier Kriterien für die epidemiologische Erfassung erarbeitet worden, die international große Akzeptanz gefunden haben und auch in Deutschland angewandt werden (z.B. Gaynes et al. 1996). Daten über unerwünschte Arzneimittelereignisse, nicht zu verwechseln mit Daten aus der Pharmakovigilanz, die auf bislang unbekannte Wirkungen ausgerichtet ist, weisen einen Populationsbezug auf und stammen aus Phase III- oder Versorgungsforschungsstudien. Natürlich ist auch die Mortalität z.B. im Krankenhaus, in seinen unterschiedlichen Ausführungen (s.u.), ein wichtiger Parameter, aber ein schlechter Qualitätsindikator (s.u.).

Auf die klinisch epidemiologischen Details (Datenformat, prospektive und retrospektive Designs, Risikofaktoranalyse etc.) soll hier nicht weiter einge-

gangen werden. Von Wichtigkeit ist aber die Wahl der verwendeten **Erhebungsmethode**, denn die Ergebnisse sind teilweise stark von der verwendeten Methode abhängig. Dies wurde insbesondere bei der Thematik Patientensicherheit deutlich, hier kommen Methoden, die sich auf administrative Daten stützen, zu deutlich niedrigeren Daten als die Untersuchung der Krankenakten oder gar die klinische *surveillance* (in Kapitel 2.4.6 wird ausführlich darauf eingegangen). Der *Chart Review* (Analyse von Krankenakten) durch interne oder externe Experten ist das am häufigsten angewandte Verfahren, führt aber ebenfalls zu einem *underreporting*, wenn man die Ergebnisse mit der klinischen *surveillance* oder der direkten Beobachtung durch trainierte Untersucher vergleicht (Andrews et al. 1997). Der Goldstandard auf institutioneller Ebene ist das in zahlreichen Studien eingesetzte, zweistufige *Harvard Medical Practice Study Design*, das eine Vorsichtung der Akten durch speziell ausgebildete Pflegekräfte mit einer endgültigen Festlegung durch ärztliche Experten kombiniert (Leape et al. 1991). Die Unterstützung durch die Krankenhausinformationssysteme mit Anwendung definierter Trigger kann die Ergebnisse deutlich verbessern (z.B. Anforderung eines Antidots als Hinweis auf ein unerwünschtes Arzneimittelereignis; s. Bates et al. 1994). Die geringe Sensitivität von administrativen Abrechnungsdaten zur Erfassung von UE tut ihrer zunehmenden Beliebtheit keinen Abbruch und haben auch für bestimmte Fragestellungen ihren Wert (AOK Bundesverband 2007). Von besonderem Wert wäre auch die Obduktion, die allerdings von der Obduktionsrate abhängig ist, die in Deutschland nicht sehr hoch ist (Madea 2008).

Im **überinstitutionellen bzw. nationalen Kontext** sind in Deutschland natürlich in erster Linie die Daten der externen Qualitätssicherung nach § 137 Abs. 1 bzw. § 137a SGB V für Erhebungen von Qualitätsdaten und von Daten zur Patientensicherheit zu nutzen. Wichtige Informationsquellen sind außerdem die Daten der Schiedsstellen bei den Landesärztekammern (Schaffartzik und Neu 2008), die im optimalen Fall mit den Datenbanken der Haftpflichtversicherer und der Krankenkassen abgeglichen werden, die aber alle keinen klaren Populationsbezug aufweisen und daher für Vergleiche oder zeitliche Verläufe nur eingeschränkt geeignet sind. Allerdings können sie auf Fall- und Patientengruppen sehr wichtige Hinweise geben (z.B. Analyse von Seitenverwechslungen). Im internationalen Bereich werden zunehmend Patientenangaben eingesetzt (sog. *Patient-Reported Outcome Measures*, PROMs, s. Kap. 2.4.7).

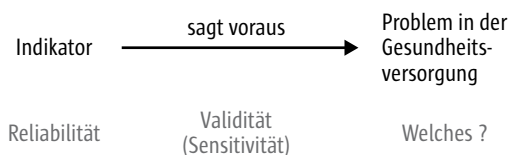
Zusammenfassung:

Quantitative Erfassungen sind für Vergleichszwecke und zur Bestimmung des Status quo unerlässlich. Die breiteste Erfahrung mit diesem Datentyp wurde in der Infektionsepidemiologie gemacht. Daten ohne exakten Populationsbezug (z.B. Schiedsstellen) sind bei aller Bedeutung für die Analyse von Fallgruppen für Vergleiche nicht geeignet.

2.4 Methode der Qualitätsmessung – das Indikatorenkonzept

2.4.1 Was ist ein Indikator?

Anders als der Begriff Patientensicherheit (Abwesenheit von unerwünschten Ereignissen) ist der Begriff Qualität nicht normativer Natur, sondern bezieht sich nach allgemein gültiger Definition (Deutsches Institut für Normung 2001) auf die Erfüllung von Anforderungen (s. Kap. 1.3). Der Begriff „Qualität“ hat also ohne die vorherige Einigung auf das „wesensmäßige“ Merkmal, an dem man die Qualität eines Produkts oder im Fall des Gesundheitswesens einer Dienstleistung bestimmt, keine Existenz. An dieser Zielorientiertheit knüpft das Indikatorenkonzept an: Indikatoren oder Kennzahlen unterstützen den Nutzer dabei zu verstehen, ob die Anforderungen (mit einer bestimmten Sicherheit) erfüllt sind oder nicht. Indikatoren sagen Qualität bzw. Qualitätsprobleme voraus (Ampelfunktion), so wie z.B. die Rückverlegungsrate auf Intensivstation für die Qualität der intensivmedizinischen Versorgung steht (s. Abb. 10).



Und was kann man tun ?

Abb. 10 Begriff des Indikators. Neben Reliabilität und Validität muss das Problem definiert sein, das der Indikator beobachten soll. Dieses muss veränderbar sein (aus Schrappe 2014, Kap. 1.4).

Die Frage „**Was ist ein Indikator**“ gehört zu den größten Problemen in der gegenwärtigen Qualitätsdiskussion, weil durch den ungenauen Gebrauch dieses Konzeptes die Diskussion ungemein erschwert wird. Insbesondere kommt es zur Verwechslung mit Parametern, die der quantitativen Erfassung dienen (s. Kap. 2.3). Mehrere Punkte sind zu berücksichtigen:

- Wie oben bereits ausgeführt, stellen Indikatoren nicht zwingend selbst Qualitätsprobleme dar, sondern sagen solche voraus (vgl. auch AQUA 2013B). Die JCAHO stellte dies schon vor mehr als 20 Jahren heraus:

„Ein Indikator ist ein Maß, welches zum Monitoring und zur Bewertung der Qualität wichtiger Leitungs-, Management-, klinischer und unterstützender Funktionen genutzt werden kann, die sich auf das Behandlungsergebnis des Patienten auswirken. Ein Indikator ist kein direktes Maß der Qualität. Es ist mehr ein Werkzeug, das zur Leistungsbewertung benutzt werden kann, das Aufmerksamkeit auf potenzielle Problembereiche lenken kann, die einer intensiven Überprüfung innerhalb einer Organisation bedürfen können.“ (JCAHO 1991, in der Übers. v. Sens und Fischer 2003)

Wie das Beispiel *readmission-rate* zeigt (s.u. „Beispiel *readmission-rate*“), ist diese Vorhersagefunktion das zentrale Kriterium eines Indikators (sog. **Abstraktion**, s. Schrappe 2010A). Es gibt aber durchaus (gute) Indikatoren, die selbst qualitätsrelevant sind (aber das ist die Ausnahme!), z.B. nosokomiale intravenöse Katheterinfektionen, die selbst ein Qualitätsproblem darstellen, gleichzeitig und in erster Linie aber die Qualität der Versorgung einer ganzen (Intensiv-)Station abbilden.

Beispiel *readmission-rate*:

Es geht in der Funktion als Indikator bei diesem Parameter nicht darum, ob er „zu hoch“ oder „o.k.“ ist (kein direktes Qualitätsmaß), sondern darum, ob er valide die Qualität der intensivmedizinischen Behandlung in Zusammenarbeit mit den Normalpflegestationen vorhersagt (Reliabilität der Erfassung vorausgesetzt).

- Eine damit verwandte Eigenschaft von Indikatoren ist die sog. **Selektivität** (Schrappe 2010A), denn Indikatoren müssen aus einer großen Anzahl potenziell qualitätsrelevanter Parameter ausgewählt werden und „sprechen“ in der Folge für einen größeren Sektor der Versorgung. Eines der häufigsten Missverständnisse besteht darin, dass Parameter aller Art sogleich als Indikatoren bezeichnet werden, obwohl solche Messwerte nur einen kleinen Teilaspekt eines Prozesses beschreiben, die Gesamtheit des Prozesses jedoch nicht erfassen. Der Wert der Blutsauerstoffsättigung ist bei der Pneumonie-Behandlung zwar ein wichtiger Einzelparameter, jedoch nicht in jedem Fall aussagekräftig für die Qualität der Behandlung, sondern hängt genauso von der Krankheitsschwere oder vom Zeitpunkt der Bestimmung ab.
- Auch muss, wie von Szescenyi und Mitarbeitern immer wieder hervorgehoben wird, eine hohe **Relevanz und Umsetzbarkeit** von Indikatoren gegeben sein (Willms et al. 2013). Es hat wenig Sinn, einen Indikator zu implementieren, der auf Qualitätsdefizite verweist, die nicht abänderbar sind (z.B. außerhalb des institutionellen Kompetenzrahmens liegen), sondern es sollten „Qualitätspotenziale“ existieren. Dies ist z.B. nicht gegeben, wenn Indikatoren zu lange in Gebrauch sind, so dass das Verbesserungspotenzial ausgeschöpft ist (sog. *ceiling*-Effekt).
- Weiterhin können Indikatoren, vor allem wenn sie unachtsam implementiert werden, **unerwünschte negative Effekte** zur Folge haben (zur zusammenfassenden Darstellung s. AQUA 2013B). Die wichtigste dieser unerwünschten Wirkungen besteht in der Risikoselektion, um die Ergebnisse besser erscheinen zu lassen; diese ist insbesondere bei Ergebnisindikatoren möglich. Auch Datenverfälschungen (sog. *gaming*) sind zu beachten, außerdem das *teaching to the test*-Phänomen, das dann auftritt, wenn alle Anstrengungen auf die durch Indikatoren gemessenen Versorgungsbereiche konzentriert und andere Ziele bzw. Patienten ver-

nachlässigt werden (Lester et al. 2011). Soweit – und das sollte die Regel sein – Indikatoren dazu verwendet werden, Qualitätsverbesserungsprozesse einzuleiten und anzuregen, bedient man sich eines *Feedback-Mechanismus*. Wie weiter unter P4P diskutiert werden wird, sind hier alternative Verhaltensformen kritisch zu prüfen, z.B. die Mengenausweitung („ein Fall geht noch“) statt der Qualitätsinvestition bei Vergütungssystemen mit starkem Mengenreiz (z.B. DRG-Finanzierung, in die P4P-Vergütungsbestandteile integriert werden) (s. Kap. 8.6).

Zusammenfassung:

Indikatoren sind ein indirektes Maß für Qualität, und hinsichtlich ihrer Vorhersagefunktion sollten sie möglichst sensitiv sein. Sie werden aus mehreren Parametern anhand ihrer Sensitivität ausgewählt (Selektion), sollten veränderbare Sachverhalte beschreiben und können (vor allem bei falscher Anwendung) durchaus auch unerwünschte Folgen haben, z.B. *teaching to the test*, also die Konzentrierung aller Anstrengung auf die Bereiche, die mittels Indikatoren beobachtet werden, unter Vernachlässigung anderer Bereiche.

2.4.2 Was ist ein guter Indikator? Die zentrale Rolle der Sensitivität

Um Parameter, die als Indikatoren in Frage kommen, weiter zu charakterisieren, stehen

- die Zuverlässigkeit (Reliabilität:): wird der Parameter richtig und technisch zuverlässig bestimmt? und
- die Validität: werden die Qualitätsprobleme identifiziert, die vorliegen, wird also das gemessen, was gemessen werden soll?

im Vordergrund. Auf die Methodik der Überprüfung von Reliabilität und Validität soll hier nicht genauer eingegangen werden (zusammenfassende Darstellung der Anforderungen s. z.B. Schrappe 2010A), es ist jedoch auch für die aktuelle deutsche Diskussion von größter Wichtigkeit, beide Begriffe voneinander abzugrenzen (s. Kap. 2.4.3).

„Gute“ Indikatoren sind sinngemäß solche, die ihre Vorhersagefunktion richtig erfüllen. Technisch unzuverlässige, nicht reliable Indikatoren sind verständlicherweise nicht von Nutzen, ebenso wenig wie solche, die zwar exakt zu bestimmen sind, aber keinen nachweisbaren Bezug zum jeweiligen Qualitätsaspekt aufweisen (mangelnde Validität). Bei der Validität wird der Sensitivität der Vorzug gegenüber der Spezifität gegeben, denn man will alle Qualitätsprobleme bzw. alle Einrichtungen mit Qualitätsproblemen erkennen. Bei der Spezifität (falsch-positive Befunde) macht man zur Not Abstriche, eher will man „umsonst gerufen worden sein“, als Qualitätsprobleme zu übersehen, wie es bei mangelnder Sensitivität der Fall wäre (s. Abb. 11). Die **Sensitivität ist daher das Schlüsselmerkmal von Indikatoren**.

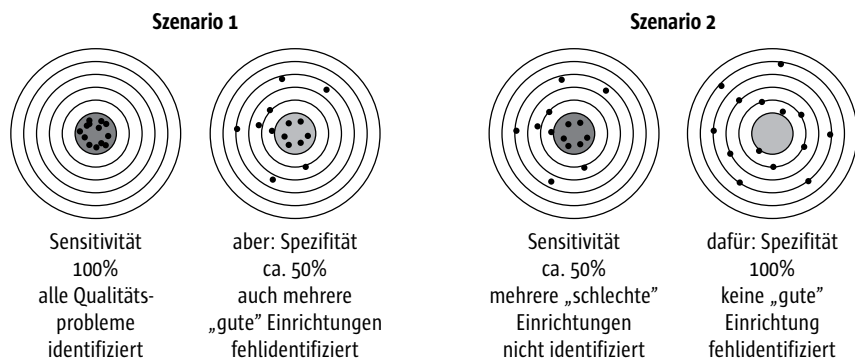


Abb. 11 Zwei Szenarien – welches würden Sie wählen? (s. Text)

Die große Bedeutung der Sensitivität von Indikatoren spielt in mehrerlei Hinsicht eine große Rolle in der gegenwärtigen Diskussion um Qualitätsindikatoren:

- die **„intrinsische Ungerechtigkeit“** von Indikatoren beschreibt den Tatbestand, dass aus statistischen Gründen eine hohe Sensitivität zu Lasten der Spezifität geht, ein sensibler Indikator folglich immer einige solcher Einrichtungen als „auffällig“ identifiziert, die gar kein Qualitätsproblem haben (s.u.),
- die **schwere Verständlichkeit** von Indikatoren für Ärzte und die Angehörigen der anderen Gesundheitsberufe, denn das Indikatorenkonzept läuft ihrem Verständnis diagnostischer Verfahren zuwider (s. Kap. 2.4.4) und
- die **geringe Eignung von administrativen und Abrechnungsdaten** als Qualitätsindikatoren, denn diese haben ein gravierendes Sensitivitätsproblem (s. Kap. 2.4.6).

Die **„intrinsische Ungerechtigkeit“** von Indikatoren mag aus dem unten aufgeführten Beispiel hervorgehen: man wird als Patient oder Angehöriger im Zweifel immer einen Indikator vorziehen, der alle problematische Einrichtungen identifiziert und dafür in Kauf nehmen, dass auch einige Einrichtungen als auffällig identifiziert werden, die keine Qualitätsprobleme haben, während man mit einem Indikator, der dies zwar vermeidet, dafür aber die problematischen Einrichtungen nicht valide identifiziert, kaum etwas anfangen kann (s. Abb. 12).

Beispiel: Ihr Vater, Sie selbst oder ein Angehöriger hat ein Prostata-Ca. Angenommen, es gäbe einen Indikator, der die Qualität der Behandlungszentren bei der Behandlung dieser Erkrankung beschreibt und mit dessen Sensitivität qualitativ problematische Einrichtungen identifiziert werden („Treffer“ im dunkelgrauen Zentrum der Zielscheiben) und mittels dessen Spezifität beschreiben wird, ob „Treffer“ bei den unproblematischen Einrichtungen ausbleiben (hellgrau Zentrum der Zielscheibe bleibt leer) (s.a. Abb. 11).

Welches Szenario würden Sie wählen? Einen Indikator, der alle problematischen Einrichtungen als solche identifiziert (hohe Sensitivität) und dafür allerdings einige Einrichtungen als „auffällig“ klassifiziert, die gar kein Problem haben (Spezifität nicht bei 100%) (**Szenario 1**), oder einen Indikator, der nur einen Teil der qualitativ problematischen Einrichtungen identifiziert (geringere Sensitivität), allerdings auch keine Einrichtungen fälschlicherweise als auffällig einstuft (Spezifität 100%) (**Szenario 2**).

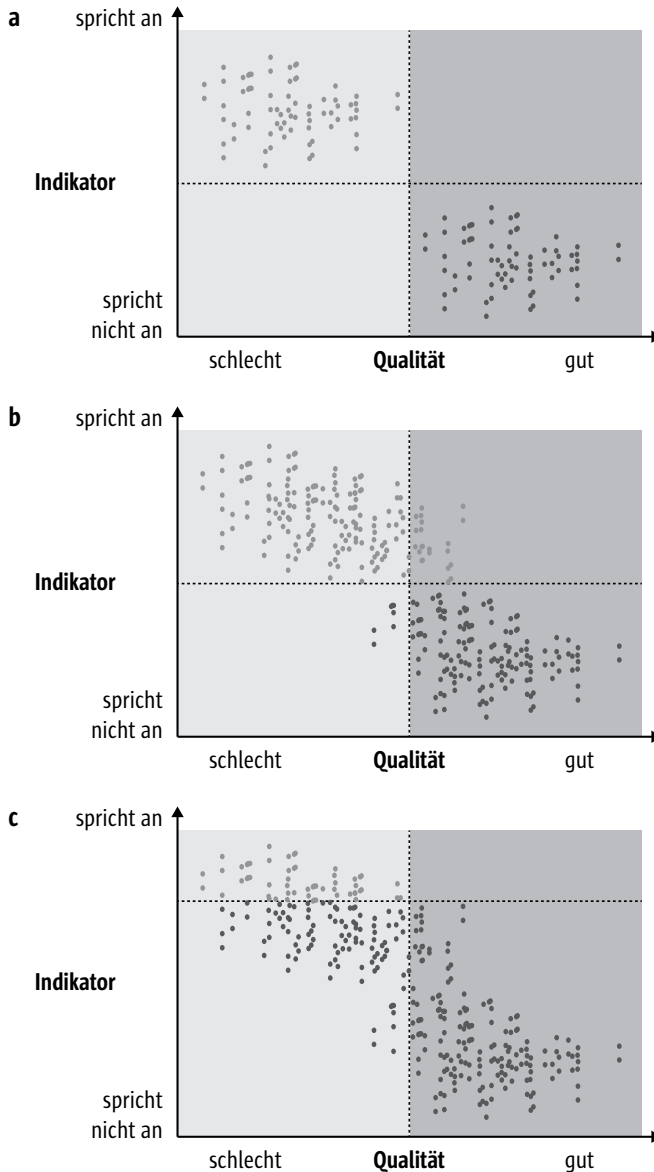


Abb. 12 Zur intrinsischen Ungerechtigkeit von Indikatoren. Fall b ist der Normalfall.

Es folgt also daraus: **Ein Indikator kann niemals gleichzeitig eine Sensitivität von 100% sowie eine Spezifität von 100% aufweisen.** Ein Szenario A wie in Abbildung 12 ist nicht realistisch. Indikatoren ergeben immer ein Bild wie in Szenario B, es gibt also neben den richtig als problematisch identifizierten Einrichtungen (Quadrant oben links) und den richtig als unauffällig identifizierten Einrichtungen (unten rechts) solche, die „übersehen“ werden (schlechte Qualität bei fehlenden Indikatoransprechen, unten links) und vor allem solche die eine gute Qualität haben, aber vom Indikator als problematisch eingestuft werden (oben rechts). Man kann zwar letzterem Problem entgehen, indem man den Indikator so „einstellt“ (d.h. die Grenzwerte verschiebt), bis der Quadrant oben rechts leer bleibt (keine „ungerecht“ als auffällig eingestufte „gute“ Einrichtungen, Spezifität ist jetzt 100%), aber zu einem hohen Preis, denn äußerst zahlreiche problematische Einrichtungen werden vom Indikator nicht mehr angezeigt (unten links).

Ein aktuelles Beispiel aus Deutschland ist das der **Mindestmengenregelung** (s. § 137 Abs. 3 Satz 1 Nr. 2 SGB V). Wenn der Mindestmengenindikator bei weniger als 50 Operationen anspricht, dann werden mehrere Einrichtungen als auffällig markiert, die trotzdem gute Qualität „liefern“ (Szenario B), obwohl sie z.B. nur 30 Operationen pro Jahr machen. Setzt man den Indikatorgrenzwert jetzt aber auf 20 Operationen fest, dann ist dies nicht mehr der Fall, der Indikator ist nicht mehr „ungerecht“. Allerdings gibt es bei einem so verwässerten Grenzwert viele Einrichtungen, bei denen der Indikator nicht anspricht, obwohl hier wirklich Handlungsbedarf gegeben wäre.

Ein weiteres Beispiel ist der **Händedesinfektionsmittelverbrauch** pro Krankenhausmitarbeiter. Wenn dieser niedrig ist, so dass auf Station um 9 Uhr morgens schon kein Händedesinfektionsmittel mehr vorhanden sein dürfte (s. Kap. 3.3.2), dann kann die Händedesinfektion auf keinen Fall adäquat durchgeführt worden sein, ein entscheidender Risikofaktor für das Auftreten nosokomialer postoperativer Wundinfektionen (oben links). Häuser mit einem hohen Desinfektionsmittelverbrauch werden im Allgemeinen niedrigere Raten nosokomialer Wundinfektionen haben (unten rechts). Wegen des komplexen Zustandekommens von Wundinfektionen ist es allerdings weder ausgeschlossen, dass diese Rate trotz niedrigem Desinfektionsmittelverbrauch niedrig ist (z.B. weil die „übrige Hygiene“ hervorragend gemacht wird, Quadrant oben rechts in Szenario B), noch dass Wundinfektionen häufig sind, obwohl sehr viel Desinfektionsmittel verbraucht wird (unten links), z.B. weil die anderen Hygienemaßnahmen nachlässig gehandhabt werden, die Händedesinfektion trotz hohen Verbrauchs falsch (z.B. zu kurz) durchgeführt wird oder (was immer wieder vorgebracht wird) Händedesinfektionsmittel verworfen werden. Hier handelt es sich also um einen Indikator, der zwar nicht sensitiv die nosokomialen Infektionen beschreibt, wohl aber die mangelnde Handdesinfektion als einen der wichtigsten Risikofaktoren für Wundinfektionen.

Es soll an dieser Stelle kurz noch darauf hingewiesen werden, dass die Forderung einer hohen Sensitivität besonders für **Patientensicherheitsindikatoren** gilt (vgl. SVR 2008, Nr. 649ff.), da sicherheitsrelevante Ereignisse natürlich besonders dringlich und vollständig erkannt werden sollten. Bei diesen Indikatoren handelt es sich oft um Beinahe-Schäden, also Fehler, die noch ohne Folgen geblieben sind, die in der Fehlerkette möglichst vorne stehen, damit es die Möglichkeit gibt, die Fehlerkette durch Präventionsmaßnahmen zu unterbrechen. Durch diesen Tatbestand wird die Vorhersagefunktion der Indikatoren betont und außerdem theoretisch unterlegt, dass Risiko- und Qualitätsmanagement gemeinsame Wurzeln aufweisen.

Bei der **Validierung** von Indikatoren, es sei nur dieser eine Punkt ausgeführt, gibt es grundsätzlich zwei Herangehensweisen. Zum einen kann man die Vorhersagefunktion des Indikators dadurch untersuchen, dass man parallel einen „Goldstandard“ untersucht und die Übereinstimmung als Sensitivität bzw. Spezifität ausdrückt. Die andere Methode, in Deutschland durch die langjährige Beschäftigung mit Leitlinien und der *Evidence-Based Medicine* gut begründet, besteht darin, dass man die wissenschaftlichen Belege für den Zusammenhang zwischen Indikator und Zielereignis systematisch darstellt und daran einen Konsensprozess anschließt, der Expertenmeinungen genauso einschließt wie die Überprüfung der Praxistauglichkeit des Indikators. Die Bundesgeschäftsstelle Qualitätssicherung und das AQUA-Institut haben hier wichtige und grundlegende Arbeiten geleistet (vgl. Reiter et al. 2007).

Zusammenfassung:

Gute Indikatoren sind solche, die ihre Vorhersagefunktion richtig erfüllen. Die Sensitivität ist daher das Schlüsselmerkmal von Indikatoren. Die „intrinsische Ungerechtigkeit“ von Indikatoren ist der Tatsache geschuldet, dass aus statistischen Gründen eine hohe Sensitivität immer zu Lasten der Spezifität geht, ein sensitiver Indikator folglich immer einige solcher Einrichtungen als „auffällig“ identifiziert, die gar kein Qualitätsproblem haben. Für Ärzte sind Indikatoren deswegen oft schwer verständlich, denn das Indikatorenkonzept läuft ihrem Verständnis diagnostischer Verfahren zuwider (ihnen fehlt die Spezifität). Administrative Daten sind wegen ihres Sensitivitätsproblems grundsätzlich als Qualitätsindikatoren schlecht geeignet. Patientensicherheitsindikatoren haben besonders hohe Anforderungen an die Sensitivität. Die Validierung von Indikatoren kann durch empirische Untersuchungen (Vergleich mit Goldstandard) oder durch einen EBM-gestützten Konsensprozess erfolgen.

2.4.3 Abgrenzung quantitative Erfassung vs. Indikatoren

Die Unterscheidung und das richtige Verständnis von Reliabilität und Validität spielen außerdem eine große Rolle in der Differenzierung von quantitativer Erfassung und den eigentlichen Indikatoren (s. Kap. 2.2 und 2.3). In der fol-

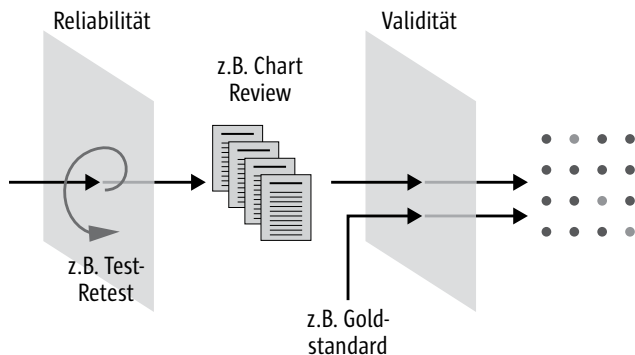
genden kurzen Analyse lässt sich daher feststellen, dass in Deutschland häufig eine quantitative Erfassung von Parametern durchgeführt wird, die zwar von großem Interesse sind (Paradebeispiel Mortalität), die jedoch Defizite in ihrer postulierten Eigenschaft als Qualitätsindikatoren haben. Entsprechend bestehen **Ungenauigkeiten im Umgang mit den Begriffen Reliabilität und Validität** fort.

Grundsätzlich ist davon auszugehen, dass man im Qualitätsmanagement und bei sinnvollen Qualitätsstrategien auf nationaler Ebene eher mit Indikatoren arbeitet, die Probleme vorhersagen *bevor* sie aufgetreten sind, als mit Parametern, die unerwünschte Ereignisse und *bad apples* aus der *ex post*-Perspektive beschreiben. So ist es zum Beispiel bei den beiden größten *Pay for Performance*-Projekte weltweit der Fall, dem *Value-Based Purchasing* Programm (VBP) von Medicare in den USA, bei dem verpflichtend alle Krankenhäuser der USA mitwirken (CMS 2013), und dem *Quality and Outcome Framework* (QOF) in Großbritannien, das für den gesamten ambulanten Bereich zuständig ist (NHS 2013). Es gibt aber Situationen, bei denen es trotzdem sinnvoll ist, sich über die *bad apples* ein genaues Bild zu machen: wenn es auf einer chirurgischen Station ein Hygieneproblem gibt, muss man nosokomiale Wundinfektionen über einen gewissen Zeitraum genau quantitativ erfassen, und auf nationaler Ebene kann es angebracht erscheinen, z.B. die Mortalität eines bestimmten Eingriffes exakt zu dokumentieren. Ganz allgemein: um sich ein zutreffendes Bild von der Patientensicherheit zu machen, ist es unabdingbar, quantitative Erhebungen z.B. über die (insbesondere vermeidbaren) unerwünschten Ereignisse zu machen.

Dieses differenzierte Plädoyer für die differenzierte Bedeutung von Indikatoren und quantitativen Parametern darf aber nicht aus dem Blickfeld geraten lassen, dass **Reliabilität und Validität** hinsichtlich der quantitativen Erhebungen und der Indikatoren eine unterschiedliche Anwendung erfahren müssen (s. Abb. 13). In einer quantitativen Erhebung wird z.B. per Analyse der Abrechnungsdaten die Zahl von nosokomialen Infektionen erhoben. Um die Reliabilität dieser Erhebung zu prüfen, wird vorab z.B. in einem Test-Retest-Verfahren die Stabilität dieser Erhebung über die Zeit untersucht, oder man untersucht, ob alle mit der Erhebung betrauten Personen zum gleichen Ergebnis kommen (und das ist auch bei administrativen Daten dringend notwendig, wie die Erfahrung zeigt, denn auch diese müssen interpretiert werden). Die Validität dieses Verfahrens wird dann z.B. mittels eines zeitweise miterhobenen „Goldstandards“, also eines anderen Verfahrens, das als Reverenz angesehen wird, überprüft, z.B. eines *Chart-Review* (Aktenstudium). Hier wird nachgesehen, ob die administrativen Daten wirklich alle nosokomialen Infektionen identifizieren (tun sie nicht, nicht einmal die Hälfte, s.u.). Dieser Fall ist in Abbildung 13 im oberen Teil dargestellt.

Bei einem Indikator „**Rate nosokomialer Infektionen**“ besteht die Reliabilität darin, ob diese Rate zuverlässig erhoben wird (auch hier Test-Retest oder an-

a quantitative Erhebung



b Indikator

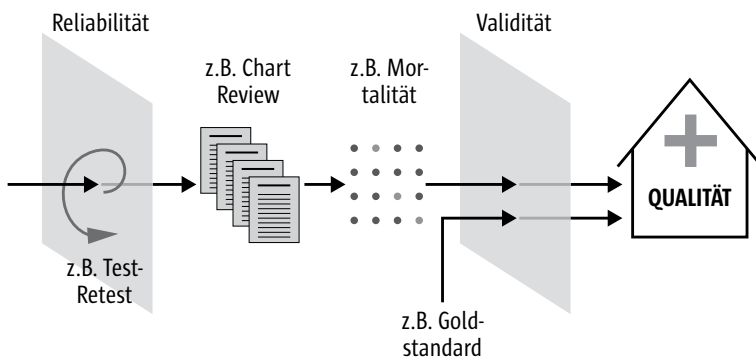


Abb. 13 Unterschied von Reliabilität und Validität bei quantitativen Erfassungsmethoden (a) und bei Indikatoren (b)

dere Verfahren), die Validität besteht allerdings darin, ob diese Rate in irgend-einer Art und Weise die Qualität einer Krankenstation oder eines Kranken-hauses anzeigt, verglichen mit einem anderen Goldstandard, der diese Quali-tät beschreibt (z.B. Prozessparameter wie Rückverlegungsraten). Dieser zwei-te Fall ist im unteren Teil der Abbildung 13 dargestellt. Man erkennt deutlich, dass die quantitative Erhebung und das Indikatorenkonzept beim gleichen Parameter, nämlich der Rate nosokomialer Infektionen, unterschiedliche An-wendungen aufweisen, obwohl sie natürlich prinzipiell gleich definiert sind.

Aus unserer **aktuellen deutschen Diskussion** sei ein Beispiel aus der externen Qualitätssicherung nach § 137 SGB V herangezogen. Wenn etwa bemängelt wird, Indikatoren seien deswegen nicht „valide“ (erfüllen also nicht ihre Vorhersagefunktion), weil sie wegen Dokumentationsmängeln falsche Ergebnis-se erbringen, dann ist hier gar keine Aussage zur Validität, sondern nur zur

Reliabilität gemacht – zur Validität gibt diese Aussage nur dann Auskunft, wenn man die Erhebung (und das dürfte der Fall sein) nicht als Indikatorerhebung, sondern als quantitative Erhebung versteht (z.B. Petzold et al. 2013). Natürlich kann ein nicht reliabler Indikator kaum valide sein, aber die Konsequenzen sind andere: Reliabilitätsprobleme kann man meist beheben (z.B. die Spezifikationen überarbeiten, besseres Training der erhebenden Personen, bessere Reliabilitätskontrolle), und danach kann durchaus eine hervorragende Validität vorliegen. Wirkliche Validitätsprobleme lassen sich jedoch nicht durch „Verbesserungen“ der Reliabilität beeinflussen. Interessant wäre es aber in der zitierten Arbeit gewesen nachzuschauen, wie hoch der Anteil von (*a priori*) bekannten Qualitätsdefiziten ist, der von den Indikatoren entdeckt wird (Sensitivität). Bei niedriger Sensitivität wären die Indikatoren wirklich nutzlos, denn sie machen nicht valide auf Probleme aufmerksam, die vorliegen. Wenn sie dagegen (mit vertretbarem Aufwand) alle Qualitätsprobleme identifizieren, ist die Validität hoch, auch wenn es zusätzlich einige falsch-positive Ergebnisse geben mag.

Zusammenfassung:

Von zentraler Bedeutung für die heutige Qualitätssicherung ist die Unterscheidung von der quantitativen Erfassung unerwünschter Ereignisse und dem Einsatz von Indikatoren. Die Validität von quantitativen Methoden bezieht sich auf die Messung des unerwünschten Ereignisses, die Validität von Indikatoren darauf, ob das gemessene Ereignis andere Ereignisse bzw. die Qualität vorhersagt. In der deutschen Qualitätssicherung nach § 137 SGB V werden sehr oft nur quantitative Erfassungen durchgeführt (Qualitätsparameter erhoben), die Indikatorenfunktion ist oft nicht belegt.

2.4.4 Abgrenzung Indikatoren – diagnostische Verfahren

Wegen der großen Bedeutung, die der Sensitivität der Indikatoren zugeiwiesen ist, ist der Unterschied zwischen Indikatoren und diagnostischen Parametern in der ärztlichen (Differential-)Diagnostik besonders wichtig (s. Abb. 14). Eine gute diagnostische Methode muss in erster Linie eine therapeutische Entscheidung tragen können, daher darf sie keine Erkrankung vorspiegeln, obwohl sie gar nicht vorliegt. Sie sollte also wenig falsch-positive Ergebnisse erbringen (hohe „Spezifität“ wird dann gesagt, es handelt sich aber im Grunde um einen hohen Positiven Prädiktiven Wert [PPW], da man allein von der Kenntnis des Testergebnisses ausgeht). Anderenfalls würde eine falsche Therapie durchgeführt bzw. durch Unterlassen weiterer Untersuchungen die wirklich vorliegende Erkrankung nicht erkannt.

Aus diesem Grund wird von ärztlicher Seite immer wieder die mangelnde „Spezifität“ von Indikatoren ins Feld geführt („war doch gar nicht so“) (z.B. Albrecht et al. 2013). Dabei wird aber der grundlegende Unterschied zu diagnos-

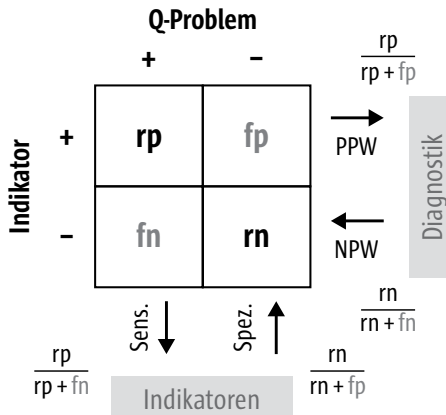


Abb. 14 Gebrauch von Indikatoren in Abgrenzung zu diagnostischen Tests. Bei Indikatoren sind vor allem die falsch-negativen Befunde wichtig (Sensitivität), während bei der Diagnostik die Rate von falsch-positiven Ergebnissen niedrig sein muss (PPW, pos. prädiktiver Wert) (aus Schrappe 2014, Kap. 1.4).

tischen Test verkannt; anders als in der Diagnostik dürfen Indikatoren durchaus in einem gewissen Umfang falsch-positive Ergebnisse erbringen, wenn sie „dafür“ aber alle Qualitätsdefizite erkennen, die tatsächlich vorliegen (keine falsch-negativen Ergebnisse).

An dieser Stelle kann vielleicht angemerkt werden, dass die Übertragung der EBM-Anforderungen für diagnostische Methoden auf die Entwicklung von Indikatoren aus dieser Sicht diskussionswürdig ist (Schmitt et al. 2013).

Zusammenfassung:

Diagnostische Tests bedürfen eines hohen Positiven Prädiktiven Wertes (PPW), Indikatoren einer hohen Sensitivität.

2.4.5 Prozess- versus Ergebnisindikatoren

Die Frage, welcher Indikatortyp eher Verwendung finden sollte, Ergebnis- oder Prozessindikatoren, erhitzt die Gemüter gerade in Deutschland ganz besonders. In anderen Ländern ist man pragmatisch vorgegangen, indem man zunächst mit Prozessindikatoren startete. Sie sind nicht nur in der Vergangenheit z.B. im HQIP-Programm von Premier und CMS in den USA eingesetzt wurden (Lindenauer et al. 2007), sondern stehen auch im aktuellen *Value-Based Purchasing*-Programm (VBP) von Medicare (Tompkins et al. 2009, Ryan et al. 2012) und in Großbritannien (*Quality and Outcome Framework* [QOL], s. NHS 2013A) ganz im Mittelpunkt. Im Verlauf wurden Ergebnisindikatoren jeweils nach Klärung der methodischen Probleme schrittweise integriert (z.B. Maryland, s. Calikoclu et al. 2012).

Natürlich sind **Prozessindikatoren** *eo ipso* weniger relevant, und Ergebnisindikatoren (wie der Name schon sagt) sind dem *Outcome* deutlich näher. Die Kontroverse wird noch dadurch erschwert, dass Prozess- und Ergebnisindikatoren nicht immer klar voneinander abzugrenzen sind (eines der Kritikpunkte an dem alten Donabedian'schen Konzept; s. Abb. 15). Eine nosokomiale, postoperative Pneumonie ist natürlich outcomerelevant, streng genommen aber ein Prozessproblem im Behandlungsablauf (z.B. einer Hüft-OP). Gleiches gilt für Rückverlegungsraten oder postoperative Thrombosen.

Will man die Auseinandersetzung nun kurz zusammenfassen, lässt sich konstatieren, dass Prozess- und Ergebnisindikatoren beide große Vor- und Nachteile haben, aber man kann es vielleicht so formulieren: im Elfmeterschießen gehen die Prozessindikatoren dann doch als Sieger vom Platz (s. Tab. 2). Dies liegt insbesondere am Anreiz zur Risikoselektion bei Ergebnisindikatoren und der aus der deswegen notwendigen Risikoadjustierung resultierenden Bildung von kleinen Gruppen, die kleine Einrichtungen benachteiligen.

Prozessindikatoren, die den Behandlungsablauf in organisatorischer, aber auch medizinisch pflegerischer Sicht beschreiben, haben zunächst den unschlagbaren Vorteil, dass sie bereits ansprechen, bevor das Kind in den Brunnen gefallen ist, sich also negative Ergebnisse zeigen. Dies entspricht dem präventiven Charakter sowohl des Qualitäts- als auch des Risikomanagements (s. z.B. Cannon 2006). Die vermehrt genutzten auf Patientenerfahrungen zurückgehenden Indikatoren (s. Kap. 2.4.7) beziehen sich meist auf die Prozesssicht (Kommunikation, Koordination). Weiterhin liegen Prozessindikatoren im Allgemeinen tatsächlich im Verantwortungsbereich des jeweiligen Leistungserbringers (z.B. Arztbriefschreibung), während Ergebnisindikatoren



Abb. 15 Die Abgrenzung zwischen Ergebnis- und Prozessindikatoren ist nicht immer unproblematisch.

Tab. 2 Abgrenzung von Prozess- und Ergebnisindikatoren (vgl. Text)

Indikatortyp	Vorteile	Nachteile
Prozess	<ul style="list-style-type: none"> ■ präventiver Charakter („Kind noch nicht im Brunnen“) entspricht QM ■ meist keine Risikoadjustierung notwendig ■ Verantwortlichkeit klar ■ international große Erfahrung ■ entsprechen der Patientensicht 	<ul style="list-style-type: none"> ■ Outcome-Relevanz gering (trotz EBM-Absicherung) ■ Mengenanreiz möglich
Ergebnis	<ul style="list-style-type: none"> ■ hohe Outcome-Relevanz ■ selbsterklärend ■ gegenüber Ärzten und Politik gut zu vertreten 	<ul style="list-style-type: none"> ■ Risikoadjustierung ■ bei mangelhafter Risikoadjustierung Anreiz zur Risikoselektion ■ Benachteiligung kleiner Einrichtungen ■ ex-post Perspektive: „Kind schon im Brunnen“ ■ „bad apples“ fördern nicht gut die Motivation ■ Verantwortlichkeit nicht zuzuordnen

stark von der Komorbidität sowie der Vor- und Nachbehandlung in den benachbarten Versorgungssektoren beeinflusst sind (s.u.).

Allerdings folgt dem Licht auch Schatten, vor allem weil Prozessindikatoren aus wissenschaftlicher Perspektive nicht so stark mit den Ergebnissen korreliert sind, wie man es meinen sollte, gerade wenn man in Rechnung stellt, dass es sich teilweise um stark EBM-abgesicherte Parameter handelt (Beispiel: Zeitpunkt der ersten Antibiotikagabe bei der ambulant erworbenen Pneumonie, vgl. Horn 2006). Es ist zwar durchaus ein (hoch-signifikanter) Effekt von Prozessindikatoren auf *Outcomes* nachweisbar, wie Werner et al. (2006) bei immerhin 3657 Krankenhäusern in den USA anhand der Daten aus dem *Hospital Compare* Programm, einem Vorläufer von HQIP, nachweisen konnten (ähnlich wie auch Jha et al. 2007, weiterhin Haller et al. 2005, Higashi et al. 2005, Houck et al. 2005, Peterson et al. 2006 u.a.), aber quantitativ enttäuschen die Ergebnisse dennoch. Evtl. sind hierfür auch nicht die „schwächelnden“ Prozessindikatoren verantwortlich, sondern die Studiendesigns, die zu ihrer Evaluation verwendet werden und den komplexen Bedingungen, in denen solche Prozessparameter wirken, nicht gerecht werden (Horn 2006).

Ergebnisindikatoren sind dagegen bereits „selbsterklärend“ *outcomerelevant*, so dass sie gegenüber der Ärzteschaft und auch der politischen Ebene gut vertreten werden können. Diese Eingängigkeit hat dazu geführt, dass in den USA und stärker noch in Deutschland schon länger die Forderung nach einer stärkeren Berücksichtigung erhoben wird (Rosenthal et al. 2007, Shekelle et al. 2001). Aber auch Ergebnisindikatoren haben ihre Probleme, in erster Linie wegen der unumgänglichen Risikoadjustierung, die durch die Berücksichti-

gung begleitender Daten durchgeführt wird. Dieser statistische Prozess führt oft zu kleinen statistischen Gruppen, und hierdurch werden kleinere Einrichtungen ohne eigenes Zutun benachteiligt, weil sie ein einziges negatives Ereignis schlecht ausgleichen können (Cannon 2006, Davidson et al. 2007, Nicholson et al. 2008, Scott und Ward 2006). Bei nur mangelhafter Risikoadjustierung würden dagegen die Anbieter zur Risikoselektion veranlasst, was ebenfalls nicht wünschenswert erscheint (s. Nicholson et al. 2008). Die statistische Aussagekraft ist daher bei Prozessindikatoren, die nicht oder in geringerem Maße risikoadjustiert werden müssen (s. allerdings Mehta et al. 2008), deutlich besser.

Weiterhin ist zu berücksichtigen, dass die Leistungsanbieter – aus ihrer Sicht – Prozesse und evtl. auch die Struktur gut beeinflussen können, wohingegen dies wegen der ihrem Zugriff entzogenen Risikoverteilung nicht für die Ergebnisse gilt, mit der Konsequenz dass sie daher sie weniger auf Ergebnisindikatoren reagieren (Frolich et al. 2007). Wird der Beobachtungszeitraum auf die poststationäre Periode ausgedehnt (z.B. postoperative Thrombose nach Kniegelenksoperation über 90 Tage), verschärft sich diese Zuständigkeitsproblematik noch, denn der Großteil der Versorgung ist der Verantwortlichkeit des Krankenhauses entzogen, zumindest solange, bis die sektorale Finanzierung nicht aufgehoben wird (s. Abb. 16). Wichtig ist es hervorzuheben, dass innerhalb eines Sektors diese Problematik nicht argumentiert werden kann: der Operateur kann nicht die Verantwortung an der Ausschleusung abgeben, denn die medizinische Verantwortung besteht in Kongruenz mit der finanziellen Verantwortung für die gesamte Dauer der stationären Versorgung.

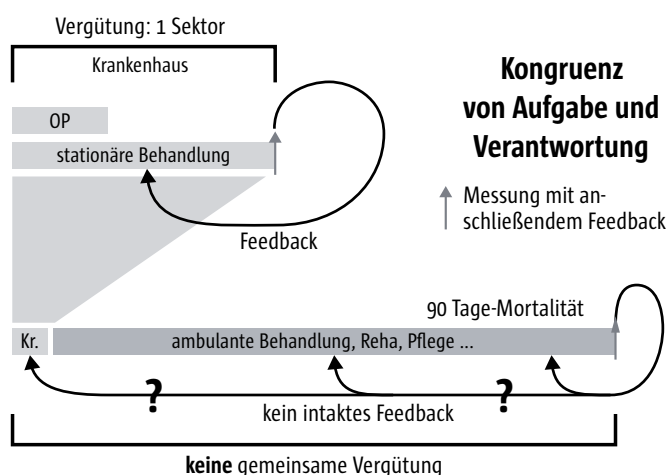


Abb. 16 Verantwortung und Finanzierung bei sektorübergreifender Versorgung: Feedback und somit Lernzyklus sind unterbrochen.

Ungeachtet der entsprechenden Beiträge der *principal-agent* Theorie (s. Kap. 8.5.4), kann man zusammenfassen, dass zumindest für den Beginn der Schwerpunkt auf den Prozessindikatoren liegen sollte, häufig wechselnd und Patientensicherheit (unerwünschte Ereignisse, Komplikationen, nosokomiale Infektionen) sowie koordinative Prozesse betonend (z.B. in der Behandlung von chronischen Erkrankungen, transsektorale Koordination). Für die weitere Entwicklung ist der Einsatz von Ergebnisindikatoren nicht ausgeschlossen, wenn deren methodische Probleme erfolgreich bearbeitet wurden.

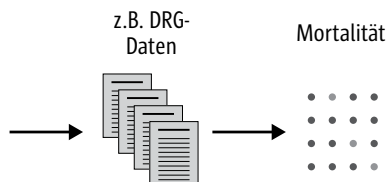
Von besonderem Interesse ist der klassische Ergebnisindikator **Mortalität** in seinen verschiedenen Ausführungen (*in house*, 30-Tage, 90-Tage, 1-Jahr), dem typischerweise in Gesundheitssystemen in den Jahren nach der DRG-Einführung große Bedeutung zukommt. Diese Bedeutung speist sich einerseits aus der Notwendigkeit, einen umfassenden Ergebnisparameter zur Verfügung zu haben, der über eine etwaige Verschlechterung der Gesundheitsversorgung unter einem fallpauschalierten Vergütungssystem mit seinen spezifischen Anreizen (z.B. Verkürzung der Liegezeit, Mengenanreiz) Zeugnis ablegt, andererseits aus der stark verbesserten Verfügbarkeit von Daten über die Sterblichkeit und für die notwendige Risikoadjustierung. Die Mortalität gilt daher als eine Art Aggregatparameter, der über die Leistungsfähigkeit einer Einrichtung zusammenfassend Auskunft gibt. Allerdings gibt es große Unterschiede zwischen verschiedenen Krankheitsgruppen und Versorgungsstrukturen; so ist die Mortalität bei Erkrankungen mit niedriger Mortalität bzw. Letalität die anders einzuschätzen als bei hoher Sterblichkeit. In Anlehnung an Abbildung 13 und 17 sei hier noch einmal darauf verwiesen, dass es eine Sache ist, Daten über die Mortalität zu erheben und eventuell zu vergleichen, und eine andere Angelegenheit, die institutionenbezogene Mortalität als Indikator für die Qualität zu nutzen.

In der Analyse werden in diesem Zusammenhang drei Hauptfaktoren unterschieden, die die Mortalität beeinflussen:

- die Qualität der Versorgung
- die Versorgungsstruktur und
- Zufallseffekte.

Diese drei Faktoren führen dazu, dass die Mortalität zwar als Qualitätsindikator eingesetzt wird, aber eine schlechte Validität für das Indikatorziel „Qualität der Versorgung einer Institution“ aufweist. Die Versorgungsstruktur spielt gerade bei sich verkürzender Verweildauer im Krankenhaus eine große Rolle, denn um so stärker ist der Einfluss der anderen an der Behandlung beteiligten Partner (z.B. ambulante oder rehabilitative Versorgung) und um so schlechter ist die Zuordnung auf einzelne Einrichtungen (Leyland und Boddy 1998, Seagroatt und Goldacre 2004). Es ist für die Entwicklung der Mortalität daher außerordentlich wichtig, ob z.B. ein Pflegeheim oder ein Hospiz in der Nähe vorhanden ist oder wie sich die Verlegungspraxis zwischen den Kran-

a Erhebung der Mortalität



b Indikator Mortalität

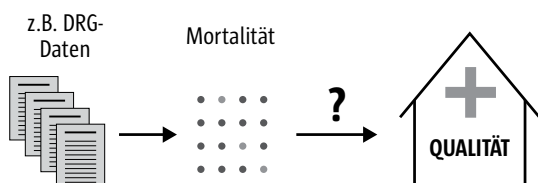


Abb. 17 Mortalität als Parameter und als Indikator (vgl. auch Abb. 13)

kenhäusern darstellt. Einfache longitudinale Verläufe der Mortalität ohne Aussage über derartige Faktoren (Nimptsch und Mansky 2013) sind daher nicht als valide anzusehen.

Noch entscheidender ist die **Abgrenzung von Zufallseffekten** und der Versorgungsqualität (z.B. Thomas und Hofer 1999), denn speziell kleinere Einrichtungen sind Zufallseffekten stark ausgesetzt und dadurch benachteiligt (Davidson et al. 2007, Scott und Ward 2006). Da die Risikoadjustierung nicht vollständig sein kann, gibt es für die Leistungserbringer starke Anreize zur Risikoselektion (Abweisen von Patienten mit hohem Risiko) und Indikationsausweitung bei Patienten mit niedrigem Risiko (Werner und Asch 2005). Andererseits ergeben sich für große Krankenhausanbieter gerade bei adjustierten Mortalitätsraten (*Standardized Mortality Rates*, SMR) Möglichkeiten des *gaming*, denn sie können einerseits Risikoselektion und durch Marketing eine Indikationsausweitung betreiben, andererseits die Überweisungs- bzw. Entlassungs- und Rückverlegungspraxis beeinflussen und vor allen Dingen durch eine „bessere“ Kodierung die Komorbiditätsadjustierung derart anpassen, so dass eine erhöhte SMR im Rahmen einer höheren Komorbidität plötzlich normal oder sogar niedrig erscheint.

Die vorangehenden Ausführungen sollen jedoch nicht heißen, dass man der Mortalität bzw. Erkrankungs-bezogenen Letalität keine Beachtung schenken sollte, insbesondere wenn die Integration des Gesundheitssystems fortschreitet und es möglich ist, unabhängig von der reinen *Provider*-Perspektive einzelner Einrichtungen die Leistungsfähigkeit von integrierten Systemen über längere Zeiträume zu beschreiben; einzelne Versuche hat es in Deutschland bereits gegeben (AOK-Bundesverband 2007). Es gibt auch Hinweise im Sinne

einer Konstruktvalidität, dass die Mortalität zumindest zusammen mit anderen Indikatoren beobachtet werden sollte. So wurde die Mortalität z.B. vom *Cardiac Surgery Reporting System* (CSRS) in New York auf der Basis klinischer Daten genutzt (Jha und Epstein 2006). Dieses Programm startete 1990 mit Daten aus 1989 und berichtete die krankenhausbezogene risikoadjustierte Letalität, ab 1992 auch für einzelne Chirurgen. Im Jahr 1991 mussten aufgrund einer Klage auf der Basis des *Freedom of Information Act* die Reports öffentlich zugänglich gemacht werden. Die Evaluationsstudien zunächst im Longitudinalvergleich (Hannan et al. 1994), dann im Vergleich mit anderen Bundesstaaten (Peterson et al. 1998, Hannan et al. 2003c) gehören zu den besten Hinweisen auf eine Wirksamkeit der Veröffentlichung von Qualitätsdaten. Die Mortalität ist außerdem ein sinnvoller Parameter für die Untersuchung umfassender, das ganze Gesundheitssystem betreffender Phänomene. In Deutschland konnten in einer Untersuchung anhand des Endpunktes risikoadjustierter Mortalität strukturelle Unterschiede von Krankenhäusergruppen identifiziert werden (Heller 2010). In Massachusetts wurde nach Einführung der 2006 Health Care Reform eine deutliche Abnahme der Mortalität (alle Ursachen, unabhängig von Leistungserbringern) nachgewiesen (Sommers et al. 2014).

Zusammenfassung:

Im Endeffekt ist es bei neu gestarteten Qualitätsprogrammen (z.B. der qualitätsorientierten Vergütung) sinnvoll, mit Prozessindikatoren zu beginnen und schrittweise Outcome-Indikatoren hinzuzufügen, so wie es international gehandhabt wurde. Mortalität ist kein sehr valider Ergebnisindikator, da sie von zu vielen nicht kontrollierbaren Effekten beeinflusst wird. Zum Vergleich auf Systemebene (z.B. von Krankenhausgruppen) und als Parameter zur begleitenden Information ist sie aber sinnvoll. Vor *gaming*-Strategien (Risiko-selektion verbunden mit Alterationen der Verlegungspraxis und *Upcoding* der Komorbidität) ist zu warnen.

Eine Präzisierung des differenzierten Gebrauchs von Prozessindikatoren bei *Pay for Performance* Programmen ist nur unter Berücksichtigung der *Principal-Agent*-Theorie möglich (s. Kap. 8.5.4, dort Abb. 40).

2.4.6 Bedeutung Routine- vs. klinischer Daten

Über die Vor- und Nachteile der Nutzung administrativer Daten gegenüber klinischen *surveillance*-Daten wird ähnlich erbittert gestritten wie über die Abgrenzung von Ergebnis- und Prozessindikatoren. Ähnlich wie in der Diskussion um die Mortalität ist festzustellen, dass administrative Daten regelmäßig in den Jahren nach Einführung des DRG-Systemes an Aktualität gewinnen. Einerseits sind Daten plötzlich in einem ungeahnten Ausmaß vorhanden (und harren ihrer Verwertung), andererseits gibt es die Notwendigkeit zur Qualitätssicherung wegen der vermuteten Fehlanreize der Pauschalierung. Da die

Einführung des DRG-Systems ja gerade die Zunahme der Transparenz im stationären Bereich als Ziel hat, nimmt der Dokumentationsaufwand schon für die Vergütung zu, daher wird die konsekutive Dokumentation für die Qualitätssicherung als übermäßig belastend abgetan und die Nutzung der Routinedaten erscheint in einem sehr günstigen Licht.

In Deutschland handelt es sich insbesondere um folgende **Datenquellen**:

- Fall- und krankenhausbezogene Daten nach § 21 KHEntgGes vom 15.7.2013, insbes. Abs. 3a;
- Daten nach § 299 Abs. 1 Satz 1 SGB V, der alle Leistungserbinger verpflichtet, Daten für Qualitätssicherung und Risikoadjustierung bereitzustellen, wobei der GBA beauftragt wird, „insbesondere diejenigen Daten zu bestimmen, die für die Ermittlung der Qualität von Diagnostik oder Behandlung mit Hilfe geeigneter Qualitätsindikatoren, für die Erfassung möglicher Begleiterkrankungen und Komplikationen, für die Feststellung der Sterblichkeit sowie für eine geeignete Validierung oder Risikoadjustierung bei der Auswertung der Daten medizinisch oder methodisch notwendig sind“ (Satz 3);
- Daten nach §§ 301 SGB V, die die nach § 108 zugelassenen Krankenhäuser in elektronischer Form bereitstellen müssen, und die u.a. die Aufnahme- und folgende Diagnosen und die durchgeführten Operationen umfassen;
- Daten nach §§ 303a-f SGB V, die den Einrichtungen des Gesundheitswesens nach § 303e Abs. 1 Satz 1 in anonymisierter oder bedarfswesie pseudoanonymisierter Form genutzt werden können, soweit sie Steuerungsaufgaben wahrnehmen und Reformen, Planungsaufgaben sowie politische Entscheidungsprozesse unterstützen und zur Verbesserung der Qualität der Versorgung beitragen.

Für administrative Daten wird immer wieder ins Feld geführt, dass sie leicht zu erheben bzw. „sowieso schon da“ seien, auch sei die Reliabilität (Zuverlässigkeit der Erhebung) relativ hoch. Beides ist auf der Erfahrungsebene so nicht ganz richtig, jeder der mit administrativen Daten gearbeitet hat, weiß wie anfällig diese sind, und wie aufwendig es ist, die Datenabfragen zu spezifizieren und zu verproben. Trotzdem muss natürlich festgehalten werden, dass die Erhebung von klinischen Daten noch weitaus aufwendiger ist.

Das eigentliche Problem administrativer Daten besteht jedoch in ihrer zu **geringen Sensitivität**, Sie sind primär für Vergütungszwecke entwickelt worden und werden auch nur zu diesem Zweck erhoben. Nach der dritten Komplikation (z.B. nosokomialen Infektion) wird keine Steigerung des relativen Fallgewichts mehr erreicht, und die Motivation zur Dokumentation aus Vergütungsperspektive nimmt rapide ab. Die in der Konsequenz auftretenden Defizite in der Sensitivität, der alles entscheidenden Eigenschaft von Indikatoren, sind beträchtlich. In deutschen Untersuchungen liegt die Sensitivität z.B. in einer Untersuchung zu nosokomialen Pneumonien, einer schwerwie-

genden und daher für die vergütungsbezogene Dokumentation attraktiven Komplikation, bei nur 40% (Azaouagh und Stausberg 2008), diese Daten entsprechen den internationalen Erfahrungen (Calderwood et al. 2014, Miller et al. 2001, Pawlson et al. 2007, Powell et al. 2001; vgl. auch SVR 2008 Nr. 654ff)..

In der letzten Zeit werden auf Auftrag des GBA durch das AQUA-Institut Untersuchungen durchgeführt, die das Ziel haben zu prüfen, ob man klinische Qualitätssicherungsdaten mit administrativen Daten validieren kann. So werden § 21-Daten zur Validierung der **Neonatalerhebung** der Perinatalzentren eingesetzt, Ergebnisse liegen hier noch nicht vor (Heller 2014). Interessant ist die Untersuchung des AQUA-Institutes zur Häufigkeit der **Dekubitalulzera**, einem Bestandteil der Qualitätssicherung nach § 137a SGB V. Eine Prävalenzuntersuchung der Berliner Charité ergab über alle Risikoklassen für die Stadien II–IV eine Häufigkeit von 4,7% der Patienten (Dassen 2004), eine Zahl, die bezogen auf 18,8 Mill. Krankenhauspatienten (im Jahr 2013) 883.600 Dekubitusfällen Grad II–IV entspricht, wobei allerdings nicht in neu während der stationären Versorgung neu aufgetretenen und bei Aufnahme vorbestehenden Dekubitalulzera (*present on admission*, POA) unterschieden wird (meist wird 50:50 angenommen). Eine Untersuchung an der Essener Universitätsklinik zeigte auf der Basis von über 41.000 Patienten eine Punktprävalenz von 5,4% für alle Schweregrade (Stausberg et al. 2005). Es gibt hier also ein beträchtliches Problem (s. a. Kap. 3). Da die Ergebnisse der Externen Qualitätssicherung nach § 137a SGB V bei zuletzt 0,6% neu (!) aufgetretener Dekubitalulzera Grad II bis IV bei Patienten älter als 75 Jahre lagen, war der Verdacht auf ein *underreporting* nicht von der Hand zu weisen, denn die Beschränkung auf ältere Patienten wählt natürlich eine Patientengruppe mit besonderes hohem Risiko aus. Weiterhin ergab sich die Schwierigkeit, dass mehrere Dekubitalulzera nicht getrennt dokumentiert werden konnten. Die Routinedaten nach § 21 KHEntgGes erbrachten bei den gleichen Auslösekriterien noch niedrigere Daten. Erstaunlicherweise konnten aber nach Einbeziehung auch jüngerer Patienten (ab 20 Jahre) rund 72.000 Fälle durch die Routinedaten identifiziert werden (Grad II–IV), mehr als 10mal so viel wie es auf der Basis der klinischen Daten zu erwarten gewesen wäre. Bei isolierter Betrachtung der Grad IV-Fälle lag die Rate mit mehr als 2.000 Fällen sogar um den Faktor 20 über den § 137a-Daten.

Diese Diskrepanz macht die **ganze Problematik der administrativen Daten** deutlich. Einerseits besteht ein Sensitivitätsproblem, insbesondere wenn Ereignisse wie Komplikationen etc. gezählt werden, da diese bei den besonders anfälligen Langliegern und Patienten mit einer hohen Komplikationsträchtigkeit und Komorbidität in den Vergütungsdaten nur bis zu einem gewissen Grad dokumentiert werden. Wenn allerdings jüngere Patienten mit einbezogen werden und Daten auf dieser Datenbasis berichtet werden, steigt die Wahrscheinlichkeit an, dass gerade aus Gründen der „ergebnisorientierten Dokumentation“ Komplikationen wie ein Dekubitus gerne dokumentiert wird, so dass höhere Zahlen entstehen. Administrative und Routinedaten sind also nur

mit äußerster Vorsicht zu Qualitätssicherungszwecken zu verwenden, was aber nicht heißen soll, dass solche Untersuchungen, wie sie zur Zeit vom AQUA-Institut oder vom Wissenschaftlichen Institut der Ortskrankenkassen (WIdO-Institut: Qualitätssicherung durch Routinedaten, sog. QSR-Projekt [AOK 2007]) vorgenommen werden, nicht zur Mehrung unserer Kenntnisse über die Qualitätssicherung sehr sinnvoll sind. Ein gutes Beispiel für die sinnvolle Nutzung von Routinedaten sind die Daten über die Behandlungshäufigkeit bei bestimmten Eingriffen bzw. Erkrankungen, die nachweisen, dass ein höheres Behandlungsvolumen regelmäßig mit einer besseren Qualität und einem besseren Behandlungserfolg vergesellschaftet ist (zu Mindestmengen s. Kap. 3.2.3).

Zusammenfassung:

Routinedaten sind vorhanden und sollten daher differenziert, insbesondere zur Validierung von klinischen Daten und zur Analyse von klärungsbedürftigen Sachverhalten (z.B. Dekubitus) eingesetzt werden. Die Sensitivität ist aber sehr niedrig, daher ist die wichtigste Eigenschaft, die einen guten Indikator auszeichnet, nicht erfüllt. Zu warnen ist vor der Annahme, die Reliabilität von Routinedaten sei *per se* hoch; vergütungsrelevante Daten sind immer unter dem Blickwinkel der ökonomischen Anreizbildung zu betrachten. Deswegen wird eine aussagekräftige Qualitätssicherung nicht überwiegend auf Routinedaten aufzubauen sein.

2.4.7 Daten auf der Basis von Patientenerfahrungen

Die großen aktuellen P4P-Projekte in den USA (*Value-Based Purchasing* [VBP], s. Ryan et al. 2012C) und in Großbritannien (*Quality and Outcome Framework* [QOL], s. NHS 2013A) legen einen klaren Schwerpunkt auf Indikatoren aus Patienten-Erfahrungen bzw. *Patient-Reported Outcome Measures* (PROM) (Varagunam et al. 2014). Im VBP-Programm machten diese Indikatoren zunächst 30%, später 20% der der Zahlungen (!) aus. Im VBP-Programm werden

- *communication with nurses*
- *communication with physicians*
- *responsiveness of hospital staff*
- *pain management; communication about medicines*
- *hospital cleanliness and quietness*
- *discharge instructions und*
- *overall rating of hospital*

erfragt, dokumentiert und als Basis der qualitätsbezogenen Vergütung bewertet (Ryan et al. 2012C). Im QOF-Programm in Großbritannien wird ein hoch bewerteter *Composite*-Indikator zu Patientenerfahrungen verwendet (NHS 2013A).

Besonders im Hinblick auf die **Thematik Patientensicherheit** wird schon länger diskutiert, dass Patienten nicht nur über differenzierte Meinungen zu

ihrer Behandlungsablauf verfügen und diese adäquat äußern können (Schwappach und Wernly 2010), sondern dass die von Patienten geäußerten Aspekte auch förderlich für ihr gesundheitliches *Outcome* sind (Lyons 2007, Hall et al. 2009). Patienten haben z.B. eine sehr realistische Einschätzung der Häufigkeit von unerwünschten Ereignissen und Schäden im Gesundheitswesen (Blendon et al. 2002). Patientenbefragungen kann man somit als Grundlage für Untersuchungen über die Qualität der Versorgung auf Populationsebene verwenden (McGlynn et al. 2003), und es ist möglich, Daten für bestimmte strukturelle Defizite eines Gesundheitssystems zu gewinnen (z.B. Schoen et al. 2005, 2007, 2009 und 2011 für den Commonwealth-Fund, Steel et al. 2008). *Patient-Reported Outcome Measures* sind aber natürlich nicht in jedem Fall aussagekräftig (Varagunam et al. 2014). Soziale Gradienten sind zu beachten (Davis und Vincent 2008).

Die letztgenannten **Daten des Commonwealth-Funds** liegen auch für Deutschland vor und werden in Kapitel 3.4 genauer dargestellt. In den letzten Gesundheitsreformen in Deutschland hat man schrittweise die Vertretung der Patienten in den Gremien der Selbstverwaltung ausgebaut (s. Kap. 7.4). Auch im Rahmen der externen Qualitätssicherung nach § 137a SGB V wird entsprechend eines Auftrages des GBA über die Rolle von Patientenbefragungen und *Patient-Related Outcomes* in der Qualitätsberichterstattung diskutiert (s. AQUA-Tagung 14.5.2014). Es ist zu erwarten, dass Patienten verstärkt auf Indikatoren aus dem Bereich der Prozessqualität Bezug nehmen (z.B. Information, Kommunikation, Organisation). Im Finanzstruktur- und Qualitätsweiterentwicklungsgesetz vom 5.6.2014 (FQWG) sind als Aufgabe für das neu zu gründende „Institut für Qualitätssicherung und Transparenz im Gesundheitswesen“ im Zusammenhang mit der Erfassung der Versorgungsqualität auch „Module für ergänzende Patientenbefragungen“ gefordert (§ 137a [neu], Abs. 3 Satz 2 Nr. 1).

Zusammenfassung:

Indikatoren auf der Basis von Patientenerfahrungen und *Patient-Reported Outcome Measures* sind international gebräuchlich und bieten valide und relevante Informationen über die Qualität der Gesundheitsversorgung. Sie betreffen vornehmlich den Bereich der Prozessindikatoren (Koordination, Information, Kommunikation). Entsprechende Entwicklungen durch das „Institut für Qualitätssicherung und Transparenz im Gesundheitswesen“ sind daher von großer Wichtigkeit.

2.4.8 Transsektorale Indikatoren

Neben der akutmedizinischen Ausrichtung und der fehlenden Präventionsorientierung ist die starke Sektorisierung der Gesundheitsversorgung eines der tragenden Charakteristika des deutschen Gesundheitssystems. Die Anreizwirkungen der sektoralen Budgets, eingeführt im Gesundheitsstrukturgesetz

(1993) zeitgleich mit den Sonderentgelten und Fallpauschalen, werden durch die ebenso sektoral orientierte DRG-Finanzierung noch deutlich verstärkt. Der viel diskutierte Mengenanreiz ist eine Folge davon, aber auch die Qualität der Versorgung wird letztlich nur unter sektoralen Gesichtspunkten optimiert. Von den 31 Indikatorengruppen aus der externen Qualitätssicherung des AQUA-Institutes verfügt kein einziger über eine transsektorale Perspektive (AQUA 2013). Für sektorübergreifende Aktivitäten werden zusätzliche finanzielle Mittel eingefordert (z.B. MRSA), obwohl sich die Frage stellt, inwiefern es sich nicht um selbstverständliche und durch die Finanzierung einer zweckmäßigen Versorgung schon abgegoltene Leistungen handelt.

Es ist also folgerichtig, dass der Gesetzgeber seit geraumer Zeit fordert, dass die Qualitätssicherung und das Qualitätsmanagement sich einer **transsektoralen Perspektive** bedient. Neben der „einrichtungsübergreifenden Maßnahmen“ aus § 135a wird in § 137 Abs. 2 nochmals gesondert auf die sektorübergreifende Perspektive, die für die Richtlinien des GBA gelten sollen, hingewiesen, weiterhin wird in § 137a die „unabhängige Institution“ bzw. das „Institut für Qualitätssicherung und Transparenz im Gesundheitswesen“ (Finanzstruktur- und Qualitätsweiterentwicklungsgesetz vom 5.6.2014 [FQWG]) aufgefordert, „für die Messung und Darstellung der Versorgungsqualität möglichst sektorenübergreifend abgestimmte Indikatoren und Instrumente zu entwickeln“ (alte Fassung). Der Gemeinsame Bundesausschuss hat daher in seiner Richtlinienkompetenz reagiert und zunächst vier Diagnosen für die transsektorale Qualitätssicherung identifiziert (Konisation, Katarakt, kolorektales Karzinom und Perkutane Koronarangiographie).

Die zweite Entwicklungslinie basiert auf den Daten des WIdO-Institutes (Wissenschaftliches Institut der Ortskrankenkassen, Projekt **Qualitätssicherung durch Routinedaten**, QSR-Projekt) und verknüpft stationäre und ambulante Daten der AOK-Versicherten, beruht folglich auf Routinedaten. Für mehrere Diagnosen konnte hier die stationäre Qualitätssicherung um eine Nachbeobachtungszeit von 30 oder 90 Tagen erweitert werden, was einer erheblichen Ausweitung der Perspektive bedeutet. In einer Untersuchung zur elektiven Cholezystektomie trat bei 10% der knapp 150.000 Patienten innerhalb von 90 Tagen mindestens eine postoperative Komplikation auf, 3,1% davon nach der stationären Entlassung (Heller und Jeschke 2012). Bei einer nicht-risikoadjustierten Sterblichkeit von 0,6% (die Zahlen des AQUA-Qualitätsberichtes 2013 mit einer Sterblichkeit von 1,0% beinhalten auch Patienten mit akuten Krankheitsbildern) hatten 145 von 1.051 Krankenhäusern eine standardisierte Mortalitätsrate (SMR) von > 2 und 30 Häuser von > 4. Ein aggregierter Qualitätsindex zeigte eine ähnliche Streuung. Bei einer weiteren Untersuchung zur Koronarangiographie mit bzw. ohne Intervention traten bei über 330.000 Patienten in 9,62% Komplikationen auf, 5,86% aber erst nach dem stationären Aufenthalt (Jeschke und Günster 2013).

Beide Vorgehensweisen, sowohl die transsektorale Qualitätssicherung des AQUA-Institutes als auch die routinedatengestützten Projekte des WIdO-In-

stitutes, sind wichtig und verdienstvoll. Allerdings darf nicht vergessen werden, dass diese „trans-sektoralen“ Analysen nur für einzelne Diagnosen bereitgestellt werden können, schon aus Gründen des Arbeitsaufwandes kann eine Gesamtsicht der Versorgung so nur scheitern. Es ist abzusehen, dass nur eine gewisse Zahl solcher „transsektoralen Tunnel“ (Schrappe 2010B) gefertigt werden können, eine integrierte Sicht auf unsere Gesundheitsversorgung ist dadurch nicht oder nur in kleinen Ausschnitten zu erreichen. Insofern ist durch diesen Ansatz die sektorale Logik des Systems nicht zu überwinden. Dies gilt ebenso für den WIdO-Ansatz, bei dem noch die in Kapitel 2.7 diskutierte Problematik hinzukommt, dass der *Feedback-Mechanismus* nicht funktionieren kann, da die Krankenhäuser teilweise aufgrund des ambulanten Behandlungsablaufes bewertet werden, für den sie aber fachlich nicht verantwortlich sind. Sie erhalten auch auf Nachfragen die Daten nicht, da dem datenschutzrechtliche Gründe entgegenstehen, daher ist ein Lernen aus diesen Daten nicht möglich, und selbst die Gutwilligen werden vom Gedanken der transsektoralen Qualitätssicherung auf diese Weise abgebracht.

Zusammenfassung:

Die transsektorale Qualitätssicherung kann die sektorale Perspektive nicht überwinden. Keiner der 31 in der externen Qualitätssicherung nach § 137a SGB V aufgenommen Krankheitsbilder bzw. Prozeduren weisen eine sektorübergreifende Perspektive auf, das AQUA-Institut hat aber gemäß dem gesetzlichen Auftrag eine kleine Zahl von transsektoralen Indikatoren entwickelt. Da die Entwicklung sehr aufwändig ist, bleibt es naturgemäß bei einzelnen Diagnosen, für die eine transsektorale Qualitätssicherung umgesetzt werden kann („transsektorale Tunnel“). Die Sektorlogik wird daher nicht aufgehoben, es ergibt sich kein integriertes Bild für die Versorgung einer Population. Gleiches gilt für die routinedatenbasierten Projekte des WIdO-Institutes, die für einzelne Diagnosen stationäre und ambulante Daten koppeln. Außerdem kann es hier nicht zu einem sinnvollen Feedback kommen, denn den Krankenhäusern obliegt nicht die Verantwortung für die poststationäre ambulante oder rehabilitative Behandlung. Die Krankenhäuser können weiterhin nicht auf die deanonymisierten Daten zugreifen und die Fälle nicht analysieren, so dass ein Lerneffekt nicht möglich ist.

2.4.9 area-Indikatoren

Von leistungserbringerbezogenen und transsektoralen Indikatoren sind *area-Indikatoren* abzugrenzen (Schrappe 2011C). Die Versorgung ist nicht mehr sektoral, sondern regional organisiert, es geht um die Behandlung (und evtl. auch die Versicherung, s. *Managed Care*) von Bevölkerungsgruppen (SVR 2009, Nr. 501). Während die Qualitätssicherung in Deutschland mit sektoralen Indikatoren und einigen transsektoralen Ansätzen arbeitet, schreitet die Integ-

ration im Rahmen einzelner Konzepte (z.B. Disease Management nach § 137f, g und Integrierte Versorgung nach § 140a ff.) fort. Die wichtigsten gesetzlich in Deutschland vorgesehenen Konzepte, die diese fortschreitende Integration tragen, sind in Abbildung 18 tabellarisch dargestellt, wobei die Integrations-tiefe (Zahl der beteiligten Sektoren) gegen die Integrationsbreite (Zahl der abgedeckten Indikationen) aufgetragen ist.

Zunächst erscheint folgender Gesichtspunkt von großer Wichtigkeit: während wir einerseits vor der Situation stehen, dass wir in unserer Qualitätssicherung zu sektoral denken, ist andererseits die reale Entwicklung (vom Gesetzgeber so gewollt) bereits dabei, die transsektorale Perspektive zu überspringen und **populationsbezogene Versorgungsmodelle** auf die Beine zu stellen. Es ist daher dringend notwendig, die Diskussion um Indikatoren zu führen, die die Versorgung von Populationen abbilden.

Natürlich sind alle in Abbildung 18 dargestellten Versorgungsformen gesetzlich verpflichtet, sich an der Qualitätssicherung zu beteiligen (Richtlinienkompetenz des GBA). Die Entwicklung von *area*-Indikatoren muss jedoch viel stärker forciert werden, denn wenn integrierte Projekte mit einem Integrationsgrad, der bis hin zur Übernahme der Versicherungsfunktion geht (*Managed Care*-ähnlich), kann es bereits zu spät sein, um unerwünschten Entwicklungen Einhalt zu gebieten. Bei aller Befürwortung einer fortschreitenden Integration der Versorgung, durch die die sektorbezogene Optimierung zugunsten einer Populationsperspektive abgelöst wird, darf nicht aus dem Blickfeld geraten, dass auch solche Konzepte spezifische Gefahren in sich bergen. Der Sachverständigenrat hat diese Gefahren in seiner Analyse zur Qualität der *Managed Care*-Versorgung ausführlich dargestellt (SVR 2009, Nr. 988). Die Problemstellungen, mit denen sich die *area*-Indikatoren auseinandersetzen müssen, sind kürzlich zusammenfassend dargestellt worden (Schrappe 2011C):

Integrations- tiefe \ breite	Anzahl Indikationen			
	eine	zwei	mehrere	alle
Prävention				
amb. Versorgung			HZV §73b/BAV §73c	
stat. Versorgung	IV §140a	ASV §116b		
Rehabilitation				
Pflege				Managed Care
alle Sektoren				

Abb. 18 Integration der Versorgung in Deutschland. IV Integrierte Versorgung, DMP Disease Management Programme, ASV Ambulante Spezialärztliche Versorgung, HZV hausarztzentrierte Versorgung, BAV Besondere ambulante Versorgung. Modifiziert und aktualisiert nach SVR 2009, Nr. 961.

- Vorenthaltung von Leistungen,
- Qualität der Versorgung chronisch und multipel erkrankter Personen und in Abhängig von Geschlecht und ethnischer Zugehörigkeit,
- Qualität der Versorgung älterer Patienten und von Kindern bzw. Jugendlichen und der Übergang zwischen den Altersgruppen (*transitional care*),
- Auswirkungen in Abhängigkeit von der Träger- bzw. Eigentümerstruktur (*for profit* vs. gemeinnütziger Träger)
- negative Risikoselektion durch Ausschluss von Patienten mit ressourcenintensiven Erkrankungen,
- Verstärkung der sozialen Ungleichheit der Versorgung, da sozial benachteiligte Patientengruppen für die *Managed Care*-typischen Ansätze wie Verstärkung der Prävention etc. nicht erreichbar sind,
- Verbesserung der Prävention in Abhängigkeit vom Zeithorizont der Patientenbindung an die *Managed Care*-Struktur (bei zu kurzer Bindungszeit nimmt das Interesse an Prävention ab),
- Versorgung in wenig besiedelten Regionen und Auswirkungen von Monopolstrukturen,
- Auswirkung auf die Organisationsstruktur der ambulanten Versorgung (in dicht und wenig besiedelten Regionen),
- Auswirkung auf das Zusammenwirken von primärfachärztlicher (Hausarzt-) und sekundärfachärztlicher Versorgung,
- Auswirkung auf das Zusammenwirken von ambulanter und stationärer sekundärfachärztlicher Versorgung,
- Entwicklung der ambulanten und stationären Pflege,
- Effizienz und Kostenentwicklung auf Systemebene.

Zusammenfassung:

Area-Indikatoren beschreiben die Versorgung von Populationen und müssen wegen der fortschreitenden Integration rechtzeitig entwickelt werden: Die Alternative zur sektoralen Perspektive ist nicht die transsektorale Perspektive, sondern die Versorgung von Populationen. Die Entwicklung in Deutschland kann sich auf eine relativ große Anzahl von Versorgungsformen stützen (zuletzt ambulante spezialärztliche Versorgung), so dass die Tendenz zur Integration langsam aber stetig zunimmt. Die Qualitätsdiskussion muss diese Entwicklungen antizipieren, damit rechtzeitig Indikatoren vorliegen, die die unerwünschten Effekte hochgradig integrierter Konzepte beschreiben können. Es ist auch die Übernahme der Versicherungsfunktion durch die kooperierenden Leistungserbringer mit einzubeziehen (Aufhebung des *provider-payer splits* in *Managed Care*), die zwar eine wünschenswerte Tiefe und Breite der Integration gewährleistet, aber auch Nebeneffekte wie Vorenthaltung von Leistungen mit sich bringen kann.