

# Leistungsvergleiche in der Medizin

## Bedarf, Anforderungen und Wege zur Akzeptanz

S. Schneeweiss<sup>1</sup>, O. Sangha<sup>2</sup>

<sup>1</sup> Department of Medicine, Brigham and Women's Hospital, Harvard Medical School und Department of Epidemiology, Harvard School of Public Health, Boston

<sup>2</sup> Bayerischer Forschungsverbund Public Health und Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Ludwigs-Maximilians Universität, München

Der Zugang zu zuverlässigen Vergleichsdaten von Qualität und Wirtschaftlichkeit medizinischer Leistungserbringer ist Voraussetzung für die Wahl einer effektiven und wirtschaftlichen Versorgung (10, 34).

Im deutschen Sprachraum gibt es bislang nur wenige Erfahrungen mit umfassenden und regelmäßigen Vergleichen von medizinischen Leistungsanbietern (46, 19, 38). Einzelne Vorstöße durch Massenmedien haben ebensoviel Aufmerksamkeit wie Widerspruch hervorgerufen (52). Diese haben eindrucksvoll den Bedarf an vergleichenden Informationen zur Qualität der gesundheitlichen Versorgung demonstriert. So berichtet der »Focus«, dass 67% einer Patientenbefragung gern mehr Informationen über Kenntnisse und Erfahrungen ihrer Ärzte hätte, und 72% würden eine Aufhebung des existierenden Werbeverbots befürworten (14). Eine Veröffentlichung von Leistungsvergleichen wurde kürzlich auch von Mitgliedern des Sachverständigenrates gefordert (30, 45). Daneben wurde gleichzeitig verdeutlicht, dass mit der Veröffentlichung von derartigen Ranglisten mit z.T. beschränkter Aussagekraft erhebliche Implikationen verbunden sein können.

Es ist dringend notwendig, die Diskussion um Leistungsvergleiche im deutschen Sprachraum einen Schritt weiter zu bringen und zu erörtern, wie Leistungsvergleiche publiziert werden können, so dass alle Beteiligten profitieren und ein möglicher Missbrauch minimiert wird. Nachdem in Deutschland derzeit mehrere bundesweite Projekte zur Qualitäts- und Wirtschaftlichkeitsmessung – die jeweils umfangreiche Leistungsvergleiche erlauben – fertiggestellt sind oder kurz vor ihrer Fertigstellung stehen, sollten die in der Gesundheitsversorgung beteiligten Partner möglichst zügig eine Einigung über das Vorgehen bei der Veröffentlichung von Leistungsvergleichen anstreben.

### Anforderungen an Leistungsvergleiche

#### Leistungsvergleiche in der Medizin

Leistungsvergleiche in der Medizin (*benchmarking*) haben potenziell denselben Nutzen wie Produktvergleiche, die uns aus dem täglichen Leben als Konsumenten bekannt sind. Sie helfen, gute Leistungen zu erkennen oder umgekehrt Optimierungspotenziale zu identifizieren und den Effekt von Verbesserungsmaßnahmen im Verlauf der Zeit transparent zu machen. Dabei ist der Begriff Leistungsvergleich sehr breit zu sehen. Im Wesentlichen lassen sich die zu vergleichenden Leistungen in die Qualität der Versorgung und die Wirtschaftlich-

**Tab.1** Beispiele von Leistungen der medizinischen Versorgung, die für mögliche Leistungsvergleiche von Interesse sind.

Dimensionen	Beispiele von Leistungen	Literatur
Versorgungsqualität	Angemessenheit von Leistungen (»Fehlbelegung«)	AEP(40)
	Struktur- und Prozessqualität	KTQ (29), HEDIS, HCFA
	Ergebnisqualität	QMK (43), ORYX, HCFA
Wirtschaftlichkeit	Ausgaben	WldO-Klip (32)
	Kosten	TSI (51)

keit der Versorgung unterteilen (Tab.1). Ersteres ist insbesondere für Patienten und überweisende Ärzte von Interesse, letzteres für Kostenträger und das Krankenhausmanagement. Leistungen können gegen fixe Sollgrößen verglichen werden, so genannte »benchmarks«, oder gegen die durchschnittliche Leistung aller Teilnehmer. Die Untersuchungseinheit eines Vergleichs können einzelne Ärzte, Abteilungen, ganze Krankenhäuser oder Regionen sein (36). Aus den unterschiedlichen Interessenlagen und der Variabilität in der Patientenzusammensetzung ergeben sich eine Reihe von Argumenten, warum derartige Vergleiche zu verfälschten Ergebnissen führen können.

### Messung und Messbarkeit von Leistung

Die Messung von Leistung in der Medizin ist in vielen Fällen komplex, da Leistungen in der Medizin oft schwer fassbare Konstrukte sind. So ist es noch relativ leicht, Vergleiche der Inanspruchnahme von Mammographie-Screenings oder Kinderimpfungen mit epidemiologischen Methoden durchzuführen. Es ist jedoch deutlich schwieriger, den medizinischen Erfolg in volkswirtschaftlich so bedeutenden Bereichen wie der Behandlung von chronischen Erkrankungen zu messen (2).

Die beiden grundlegenden Anforderungen an die Messung von Leistung sind Zuverlässigkeit einer Messung, d.h. die Reproduzierbarkeit der Messung durch dieselben und verschiedene Untersucher und die Validität einer Messung, d.h. das Ausmaß, zu dem eine Messung das Konstrukt erfasst, welches sie vorgibt zu messen (49). Weitere Eigenschaften von Leistungsindikatoren, die für eine breite Akzeptanz bedeutend sind, sind in Tab.2 zusammengefasst.

**Tab. 2** Generelle Anforderungen an Leistungsindikatoren. Nach 12, 39, 43.

Reliabilität der Messung
Validität der Messung
Sensitivität der Indikatoren
Variabilität der Leistung zwischen Leistungsanbietern (Ärzte oder Krankenhäuser)
Bedeutung der Indikatoren für Morbidität, Mortalität und Kosten
Geringer Aufwand bei der Datenerfassung
Modifizierbarkeit der Leistungsindikatoren
Zuordbarkeit von Leistung und Leistungserbringer
Repräsentativ für das Servicespektrum der Leistungsanbieter
Ausreichende Möglichkeit der Risiko-Adjustierung

Nachdem es einen nicht unerheblichen Aufwand darstellt, valide Indikatoren zu identifizieren, die gleichzeitig zuverlässig messbar sind, ist es üblich, sich auf wenige, aber sorgfältig ausgewählte Leistungsindikatoren zu beschränken (8), eine Einsicht, die auch in Deutschland aus der Frustration über eine umfangreiche Dokumentation für Fallpauschalen gewonnen wurde (31).

**Qualität der medizinischen Behandlung:** Für Patienten und Kostenträger sind letztlich nur die Ergebnisse der Behandlung (*outcomes*) bzw. die Relation von *outcomes* zu dem Preis der Leistung entscheidend (6, 9). Für Ärzte, Krankenhäuser und deren Manager ist die Darstellung von *outcomes* gemeinsam mit den Prozessen die zu diesen *outcomes* führen hilfreicher, da sie Ansätze für eine Leistungsverbesserung aufzeigen können.

Letztlich ist medizinische Behandlungsqualität nur bis zu einem gewissen Grad messbar, d.h. man misst Surrogate, die möglichst eng mit dem tatsächlichen Konstrukt korrelieren (4). Nachdem diese Korrelation jedoch nicht perfekt ist, entstehen bei der Messung Fehler. Die Güte von Leistungsindikatoren der Behandlungsqualität hängt u.a. ab von der standardisierten Messbarkeit, dem Zeitrahmen der Messung, der Zuordbarkeit. Je größer die Summe der Messfehler ist, desto unpräziser ist die Messung von Leistung. Als Konsequenz verlieren Leistungsvergleiche dadurch an Trennschärfe (53), d.h. einige Leistungsanbieter werden fälschlich als unauffällig bewertet, obwohl sie tatsächlich außerhalb eines definierten Normbereichs liegen. Konservativere Ergebnisse sind im Zusammenhang mit Leistungsvergleichen jedoch eher tolerabel.

Leistungen werden zum Teil in Form von Reputation oder Renomee gemessen. Reputation scheint für kostengünstig durchzuführende Leistungsvergleiche eine attraktive Größe zu sein, da eine Vielzahl von Einzelparametern in diese Globalbewertung durch Laien und/oder Mediziner in einem Parameter gebündelt werden. Der wesentliche Nachteil ist jedoch, dass unklar bleibt, welche Faktoren eingehen, wie diese Faktoren gewichtet werden und welche inter- und intraindividuelle Variabilität diese Faktoren und Gewichtungen bei den Befragten besitzen. Nachdem die Bewertung von Renomme wenig nachvollziehbar und damit für alle Betroffenen wenig informativ ist, ist von derartigen Leistungsbewertungen abzuraten.

**Wirtschaftlichkeit der medizinischen Behandlung:** Die Messung von Wirtschaftlichkeit hängt von der jeweiligen Perspektive ab. Kostenträger sind an ihren Ausgaben pro Krankenhaus oder Patient interessiert, Ärzte und Krankenhäuser jedoch an den tatsächlich angefallenen Kosten, bzw. dem Gewinn (Erstattung minus Kosten) pro Patient. Die Erfassung der Ausgaben ist in der Regel leicht, da sie sich an Bewertungspunkten, Liegezeiten oder pauschalen Vergütungen, wie z.B. *Diagnosis-related Groups* (DRG) orientieren. Die tatsächlich anfallenden Kosten sind jedoch eine komplexe Zusammensetzung aus den eingesetzten personellen und sachlichen Ressourcen.

### Vergleichbarkeit von Patienten und die Notwendigkeit einer Risiko-Adjustierung

Patienten unterscheiden sich hinsichtlich Alter, Geschlecht, Hauptdiagnose, Schweregrad, Begleiterkrankung(en), Funktionsfähigkeit und einer Vielzahl weiterer Faktoren, die den Behandlungserfolg beeinflussen können (Tab. 3). Diese Risikoprofile von Patienten zu Beginn einer Krankheitsepisode (z.B. Krankenhausaufnahme) haben z.T. erheblichen Einfluss auf das erreichbare und tatsächlich erreichte Behandlungsergebnis und die eingesetzten Ressourcen. Gemeinsam mit Eigenschaften der Versorgungsstruktur sind Patienteneigenschaften neben der ärztlichen Leistung die wichtigsten Prädiktoren von medizinischen Ergebnissen (Abb. 1). Ziel einer Risiko-Adjustierung ist es, nur Patienten mit ähnlichem Risikoprofil, bzw. nur Krankenhäuser mit ähnlicher Patientenzusammensetzung («case mix») zu vergleichen. Dies kann mit statistischen Verfahren wie multivariaten Regressionsmodellen erreicht werden. Die Darstellung der Risiko-Adjustierung in einem verallgemeinerten Versorgungsmodell lässt sich direkt in ein einfaches Regressionsmodell mit *i* Patienteneigenschaften (PE) und *p*-i Charakteristika der Versorgungsstruktur (VS) übertragen:

$$\text{Leistungsindikator} = \alpha + \beta_0 \text{Leistungserbringer} + \beta_1 \text{PE}_1 + \dots + \beta_i \text{PE}_i + \beta_{i+1} \text{VS}_1 + \dots + \beta_p \text{VS}_p + e$$

Bei Vergleichen von Krankenhäusern, Versorgungsstufen oder Regionen müssen zusätzlich Faktoren der Krankenhausstruktur und umliegender Versorgungsstrukturen in komplexen hierarchischen Modellen berücksichtigt werden (47, 13). Einer guten Risiko-Adjustierung kommt in Leistungsvergleichen der gleiche Stellenwert zu wie der validen Messung von Leistungsindikatoren. Ist eines von beiden fragwürdig, so ist der gesamte Leistungsvergleich fragwürdig.

Mit der Einführung von prospektiven Vergütungssystemen im Krankenhausbereich (DRGs) wird eine Risiko-Adjustierung bezüglich der Finanzierung von Krankenhausaufhalten geschaffen. Fraglich ist derzeit jedoch noch, ob sich die DRG-Einteilung in unterschiedlich schwer kranke Patienten in Deutschland für eine Risiko-Adjustierung mit ausreichend homogenen Gruppen eignet.

Insbesondere für die Qualitätsbewertung nach der Einführung von DRGs ist der Vergleich von Strukturen und Prozessen ungeeignet. Diese lassen sich deutlich leichter als «*outcomes*» messen, können jedoch nur dokumentieren, wie ein Leistungsüberangebot langsam in eine kosten-effektive Versorgung umgewandelt wird. Patienten und Ärzte sind über mögliche Verschlechterungen der Behandlungsergebnisse be-

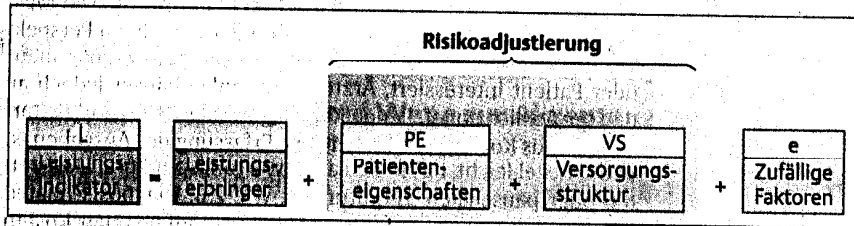


Abb.1 Leistungsvergleiche in einem verallgemeinerten Versorgungsmodell: Risiko-Adjustierung eliminiert die Effekte von Patienteneigenschaften und Versorgungsstruktur auf medizinische Leistungen (z.B. Qualität, Kosten) und erlaubt damit einen direkten Vergleich von Leistung zwischen Leistungserbringern.

Tab.3 Patientenbezogene Faktoren, die den Behandlungserfolg oder aufzuwendenden Ressourcen beeinflussen können und damit in Risiko-Adjustierungen berücksichtigt werden sollten. In Anlehnung an 24 und 23.

Dimensionen	Beispiele für Faktoren innerhalb einer Dimension
Schweregrad der Index-Diagnose	NYHA-Klassifikation einer Herzinsuffizienz
Akuter klinischer Allgemeinzustand	Aufnahmeart, Vitalzeichen, Bewusstseinszustand, Funktionseinschränkungen
Begleiterkrankungen und deren Schweregrad	Index of CoExisting Diseases (ICED), Charlson Index
Inanspruchnahme von Gesundheitsleistungen	Anzahl Arztbesuche, Anzahl Krankenhausaufenthalte, Anzahl Notfalleinweisungen
Demographische Faktoren	Geschlecht, Alter, Versicherungsstatus
Patientenzentrierter Gesundheitsstatus	Short form SF-12 oder SF-36
Lebensstil-Faktoren	Rauchen, Ernährungszustand, Alkoholkonsum
Kognitive, seelische und soziale Funktionsfähigkeit	Pflegekategorie
Soziales Umfeld	Wohnsituation

sorgf. Daher müssen auch Risiko-adjustierte outcomes in Benchmarking eingehen (9).

**Aspekte der statistischen Verfahren zur Erstellung von Leistungsvergleichen**

**Risiko-Adjustierung mit multivariaten Regressionsmodellen:**

Aussagekräftige Leistungsvergleiche stellen Risiko-adjustierte Leistungsindikatoren unterschiedlicher Anbieter gegenüber. Die Risiko-Adjustierung erfolgt durch Regressionsmodelle, die die in Abbildung 1 dargestellte Grundstruktur aufweisen. Die Parametrisierung der Modelle unterscheidet sich jedoch von Indikator zu Indikator, da unterschiedliche Faktoren als Einflussgrößen für die unterschiedlichen Leistungsindikatoren berücksichtigt werden müssen. Generalisierte lineare Modelle besitzen die notwendige Flexibilität, die statistische Modellierung ist jedoch in der Regel komplex. Trotz aller Bemühungen ist die Vorhersagekraft dieser Modelle noch limitiert (37). Ärzte innerhalb einer Praxismgemeinschaft oder eines Krankenhauses verhalten sich ähnlicher als im Vergleich zu den Kollegen außerhalb, oft aufgrund von Faktoren, die sie selber nicht beeinflussen können, z.B. das Fehlen von Nachsorgeeinrichtungen. Diese Variationen auf unterschiedlichen Ebenen werden in hierarchi-

schen Regressionsmodellen realitätsnäher abgebildet werden (5, 13, 20, 50). Selbst perfekte Modelle können nicht unterscheiden ob eine Adjustierung für Patientenalter zu recht ein altersassoziiertes Risiko für schlechtere outcomes kontrolliert, oder zu unrecht eine schlechtere Behandlung von älteren Menschen maskiert (34).

**Stichprobengröße:** Eine exakte Fallzahlab-schätzung ist in Leistungsvergleichen nur eingeschränkt möglich, da eine Vielzahl von Faktoren eingehen. Die notwendige Anzahl von Patienten pro Klinik oder Praxis hängt zum Großteil von der Sensitivität der Indikatoren sowie von der Anzahl von Parametern ab, die für eine hinreichend gute Risiko-Adjustierung notwendig sind. Untersuchungen empfehlen eine minimale Anzahl von 20-30 Patienten pro Indikator pro Klinik oder Arzt, um zu aussagekräftigen Ergebnissen zu kommen (7).

**kurzgefasst:** Benchmarking und damit Transparenz von medizinischen Leistungserbringern wird zunehmend gefordert, um Patienten, Ärzten und Kassen bei der Wahl einer effektiven und wirtschaftlichen Versorgung zu unterstützen. Wichtigste Anforderungen an Leistungsvergleiche sind die valide und reproduzierbare Messung von Leistung und eine umfassende Risiko-Adjustierung.

**Veröffentlichung von Leistungsvergleichen**

**Bedarf an öffentlich zugänglichen Leistungsvergleichen**

Der Bedarf an Leistungsvergleichen leitet sich aus dem Recht der Patienten an Informationen über die Qualität der medizinischen Versorgung ab. Patienten und überweisende Ärzte benötigen diese Information, um den geeignetesten Versorger auswählen zu können. Nur informierte Patienten können die effektivste und wirtschaftlichste medizinische Leistung wählen. Nachdem sich Patienten zum Zeitpunkt der Versorgungsentscheidung oft in einer emotional und körperlich belastenden Situation befinden, kommt den überweisenden Ärzten als Fürsprecher der Patienten eine zentrale Rolle zu (33). Dadurch, dass vergleichende Informationen zur Qualität und Wirtschaftlichkeit im Gesundheitswesen den Patienten, Überweisern und Kostenträgern, d.h. den Kunden und Zahlern des Gesundheitssystems, nicht in ausreichendem Maße zugänglich sind, erklärt sich, warum der Gesundheitsmarkt sich nicht nach den Regeln andere Märkte verhält. Durch das bestehende Informationsungleichgewicht können sich Angebot und Nachfrage nicht ausreichend nach Qualität und Wirtschaftlichkeit orientieren (35). Daraus leitet sich neben dem Bedarf der Patienten auch ein gesellschaftlicher Bedarf an Leistungsvergleichen ab.

**Tab. 4** Typische Argumente von Ärzten und Kassen im Zusammenhang mit risiko-adjustierten Leistungsvergleichen in der Medizin. In Anlehnung an 25.

**Argumente von Ärzten:**

Die Risiko-Adjustierung ist nicht aussagekräftig, d. h. wichtige Risikofaktoren für schlechte klinische *outcomes* sind nicht erfasst.

Datensammlung ist zu aufwendig und teuer und fügt nur eine weitere administrative Belastung zum klinischen Alltag hinzu.

Entscheidungssträger, die Leistungsvergleiche einfordern, verstehen nicht die Grenzen des Datenmaterials und der analytischen Technik und ziehen daraus unangemessene Schlüsse.

Diese mangelhaften Daten werden benutzt, um punitive Entscheidungen gegenüber den scheinbar weniger leistungsfähigen Anbietern zu treffen.

Die Daten haben häufig einen begrenzten Umfang (z. B. nur Mortalität) und adressieren nicht die für Ärzte und Patienten wichtigsten Fragen.

**Argumente von Krankenkassen:**

Bei den gegenwärtig stetig wachsenden Gesundheitsausgaben müssen Ärzte und Krankenhäuser für ihre Leistungen verantwortlich gemacht werden können.

Andere Industrien mit zum Teil weit geringeren Umsätzen müssen auch die Qualität und Preise ihrer Produkte rechtfertigen.

Angesichts der gesamten Gesundheitsausgaben sind die Kosten von Leistungsvergleichen trivial.

Patienten und überweisende Ärzte brauchen Leistungsvergleiche, um informierte Entscheidungen über die Wahl der Gesundheitsversorgung treffen zu können.

Angesichts der ungenügenden Informationslage muss irgendwann mit Leistungsvergleichen begonnen werden, selbst wenn die Daten und Ergebnisse zunächst nicht perfekt sind.

In der Debatte um Leistungsvergleiche werden jedoch eine Reihe von Bedenken genannt (Tab. 4), die vor einer Veröffentlichung berücksichtigt werden müssen. Die wesentlichen Punkte werden im Folgenden besprochen.

**Bedarf, einen gemeinsamen Nutzen aller Beteiligten an einem Leistungsvergleich zu identifizieren**

Alle Leistungsvergleiche benötigen vorab klar definierte Ziele, die von allen Beteiligten einverständlich geregelt werden sollten. Dabei ergeben sich verschiedene Kriterien, die an eine sinnvolle Anwendung von Leistungsvergleichen in der Medizin gestellt werden sollten (Tab. 5). Die beiden wichtigsten Kriterien für einen Erfolg sind, einen Konsensus zu erreichen über 1.) die vorrangige Bedeutung einer Leistungsverbesserung, jedoch nicht Identifizierung von »schwarzen Schafen« und 2.) eine anerkannte und zuverlässige Methodik mit begleitender Informationsarbeit und Interpretationshilfe. Diese Ziele und Kriterien lassen sich erreichen, wenn die Interessen und die damit verbundenen möglichen Nutzen aller Beteiligten offengelegt werden. In einem »Heile-Welt-Szenario« würden alle Beteiligten einen Nutzen in Leistungsvergleichen sehen. In der Realität wird der Nutzen jedoch unterschiedlich beurteilt (Tab. 6).

Am wenigsten Interesse an Leistungsvergleichen haben in der Regel Ärzte, die ultimativen Leistungserbringer. Sind Ärzte einzeln identifizierbar, so tragen diese das größte Risiko, meist ohne einen klaren Nutzen sehen zu können. Werden

**Tab. 5** Kriterien für einen sinnvollen Einsatz von Leistungsvergleichen in der Medizin.

**Leistungsvergleiche in der Medizin sollten ...**

... für möglichst viele Abnehmer von Interesse sein: Patienten, Ärzte, Krankenhausmanager, Kassen etc.

... helfen, Prozesse und Ergebnisse der Gesundheitsversorgung zu verbessern.

... eine solide wissenschaftliche Basis haben und dennoch die hohe Variabilität in der medizinischen Versorgung berücksichtigen.

... eine hinreichende Validität und Reproduzierbarkeit aufweisen.

... möglichst wenig kosten und leicht durchzuführen sein.

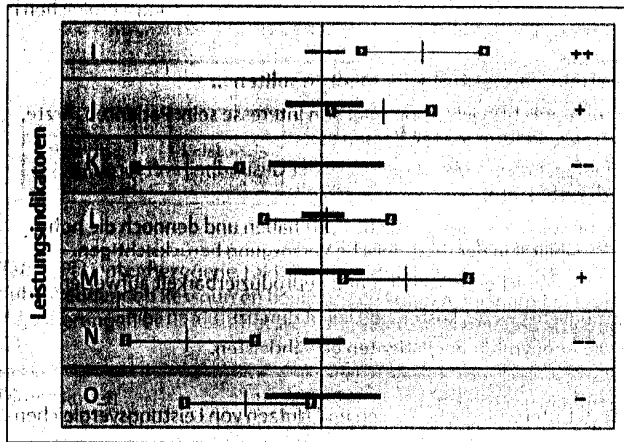
... die Anonymität der Patienten gewährleisten.

**Tab. 6** Unterschiedliche Risiken und Nutzen von Leistungsvergleichen für Patienten, Ärzte und Kassen.

	Träger eines gesundheitlichen Risikos	Mögliches finanzielles Risiko	Möglicher Nutzen
Patienten	Ja	Nein	Identifikation der besten Qualität, Vermeidung schlechter Qualität
Ärzte, Krankenhäuser	Nein	Ja	Leistungsverbesserung, Qualitätsmanagement
Kassen	Nein	Ja	Identifikation der wirtschaftlichsten Anbieter (Preis pro Leistung, pro Qualität)

Krankenhäuser miteinander verglichen, so haben Ärzte keinen direkten Anreiz sondern lediglich die Pflicht, sich für einen Leistungsvergleich positiv zu engagieren. Egal ob Krankenhäuser oder einzelne Ärzte verglichen werden: Sobald einzelne Leistungserbringer identifizierbar sind, besteht häufig die Meinung, dass dies nur Nachteile einbringen kann, insbesondere, wenn sie *a priori* nicht von ihrer eigenen Leistung überzeugt sind. Wenn jedoch Leistungsvergleiche nicht als im besten Interesse von Ärzten erscheinen, dann werden Ärzte freiwillige Programme vermeiden und versuchen, die Ergebnisse oder die Personen und Institutionen, die Leistungsvergleiche propagieren, zu diskreditieren. Damit müssen Leistungsvergleiche entweder ohne direkte Beteiligung von Ärzten (z. B. aufgrund von Krankenakten), durch eine Verpflichtung von Ärzten, oder durch finanzielle Anreize umgesetzt werden. Finanzielle Anreize haben den Vorteil einer positiven Belohnung beim Erreichen einer festgelegten Leistungssteigerung. Verschiedene Autoren meinen, dass alle Beteiligten auch ohne Verpflichtung oder finanzielle Anreize in Gruppentreffen zu einer Teilnahme motiviert werden können (15), andere sind weniger optimistisch (18, 21).

**kurzgefasst:** Die Identifizierung eines gemeinsamen Nutzens ist entscheidend für den Erfolg von Benchmarking. Die gemeinsame Anstrengung zu einer Leistungsverbesserung aller Teilnehmer muss im Vordergrund stehen, die Auslese und Bestrafung von »schwarzen Schafen« führt nicht zu längerfristigen Erfolgen. In der Realität sind meist direkte oder indirekte finanzielle Anreize notwendig.



**Abb.2** Graphische Darstellung eines Leistungsprofils für einen Leistungserbringer, z. B. ein Krankenhaus. Dargestellt sind 7 metrische Leistungsindikatoren (von I bis O). Die vertikale Linie zeigt die mittlere Leistung der Vergleichsgruppe mit den entsprechenden 95% Konfidenzintervallen. Rechts der vertikalen Linie ist mit + der überdurchschnittliche Leistungsbereich indiziert. Interpretationshilfe: Indikator I: Leistung überlappt mit dem Vergleichsmittel = liegt im Vergleichsstandard (0). Indikator M: Leistung überlappt nur mit dem Konfidenzintervall des Vergleichsmittels = Übergangsbereich, Tendenz zu einer besseren Leistung (+). Indikator N: Leistung liegt ausserhalb vom Konfidenzintervall des Vergleichsmittels = unterdurchschnittliche Leistung.

### Bedarf an einer klaren Darstellung von Ergebnissen eines Leistungsvergleichs

Graphische Leistungsprofile von wenigen aber sorgfältig ausgewählten Leistungsindikatoren haben sich bewährt (Abb.2) (28). Leistungsprofile erlauben eine ausgewogene Bewertung eines Spektrums wichtiger Leistungen, wohingegen die Betrachtung lediglich eines Indikators den Blickwinkel bis hin zu einer Verzerrung einengen kann und für Patienten meist wenig hilfreich ist. Leistungsprofile helfen Ärzten und Krankenhäusern, ihre Schwächen und Stärken zu identifizieren und entsprechend an den zugrundeliegenden Prozessen zu arbeiten.

Leistungsvergleiche werden gerne zu Ranglisten zusammengefasst. Für Ranglisten müssen jedoch die einzelnen Indikatoren mit Gewichten zusammengeführt werden, die häufig einer inhaltlichen Grundlage entbehren (17). Gewichtung ist letztlich eine Frage von Präferenzen, die für wirtschaftliche Vergleiche leichter zu finden sein mögen als für Qualitätsvergleiche (34). Größere Flexibilität bei einer Veröffentlichung erlauben interaktive Internet-Seiten. Je nach Interesse des Lesers kann eine interaktive Liste nach einzelnen Leistungsindikatoren aufsteigend oder absteigend sortiert und entsprechend grafisch dargestellt werden (16). Der begleitende gedruckte Überblick zählt hingegen alle Indikatoren in einer Tabelle auf, die regional und alphabetisch sortiert sein kann (26).

### Bedarf an allgemein anerkannten methodischen Standards

Eine breite Akzeptanz von Leistungsvergleichen wird es nur dann geben, wenn die höchstmögliche methodische Qualität gewährleistet ist und alle Beteiligten die Methodik akzeptieren. Es ist daher sinnvoll, das methodische Fundament aus

Komponenten zusammensetzen, die in der wissenschaftlichen Gemeinschaft bereits allgemeine Akzeptanz gefunden haben.

**Datenqualität und Fallzahl:** Die Datenqualität für differenzierte Bewertungen medizinischer Leistungen ist bei prospektiv erhobenen Daten mit externem »auditing« am besten. Insbesondere für wirtschaftliche Zielgrößen und Wiedereinweisungen nach operativen Eingriffen eignen sich jedoch auch Krankenakten oder Abrechnungsdaten. Bezüglich der Datenqualität sollten Validierungsstudien vorliegen, die die Reproduzierbarkeit und Validität der Rohdaten darstellen. Vor einer Veröffentlichung von Leistungsvergleichen sollten sich alle Beteiligten auf eine mindestens zu erreichende Fallzahl einigen, die aufgrund von Vorerfahrungen festgelegt wird. Erreichen Leistungsanbieter diese Fallzahl für einzelne Diagnosen nicht, so werden die entsprechenden Leistungsindikatoren nicht dargestellt.

**Risiko-Adjustierung:** Nachdem eine Risiko-Adjustierung genauso wichtig wie die Validität der Leistungsindikatoren ist, sollten die wesentlichen Aspekte der Risiko-Adjustierung und die Güte der Risiko-Adjustierung verständlich dargestellt werden. Dabei muss insbesondere für jeden Leistungsindikator getrennt dargestellt werden, welche Variablen in die Risiko-Adjustierung eingegangen sind, wie die Variablen modelliert wurden und welche prädiktive Wertigkeit jedes einzelne Indikatormodell besitzt (42, 48). Kommerzielle Produkte zur Risiko-Adjustierung gewähren häufig keinen Einblick.

**Evidenz-basierte Leistungsindikatoren:** Wie bereits oben besprochen, müssen Leistungsindikatoren eine Reihe von Kriterien erfüllen, um eine breite Akzeptanz zu erreichen (Tab.2). Soweit als möglich sollte die Erfüllung dieser Kriterien mit empirischen Angaben geprüft werden (3, 43). Begründungen für jeden verwendeten Leistungsindikator sollten verständlich aufbereitet sein und allgemein zugänglich, beispielsweise im Internet, publiziert sein (z. B. 10).

Die Bedeutung dieser Punkte für die Akzeptanz von Leistungsvergleichen wurden in einer Untersuchung zu den Konsequenzen des Pennsylvania »Consumer Report to Coronary Artery Bypass Graft Surgery« (44) deutlich gezeigt. Kardiologen als wichtigste Überweiser für Bypass-Operationen kritisierten, dass Sterblichkeit als einziger Leistungsindikator nicht aussagekräftig sei, dass klinisch wichtige Indikatoren fehlten und dass die Risiko-Adjustierung für einen fairen Vergleich nicht ausreichte. Demzufolge hatte der Report bei 87% der Zuweiser keinen oder nur einen minimalen Effekt auf das Überweisungsverhalten.

Generell wird es nie die perfekte Daten- und Analysequalität geben, die sich Forscher von Leistungsvergleichen erhoffen würden. Es ist daher eine kontinuierliche Forschungsarbeit notwendig, um ein optimales Maß an Daten- und Analysequalität bei gleichzeitig möglichst geringem finanziellen Aufwand und hoher Praktikabilität zu erreichen. »Don't let the perfect be the enemy of the good« (12) ist in den USA im Zusammenhang mit Leistungsvergleichen viel zitiert und dennoch nicht ohne Vorsicht zu interpretieren: Vergleiche von ungenügender Qualität schaden der Öffentlichkeit, den Ärzten und Krankenhäusern (27). So wurde es als Folge des oben erwähnten »Consumer Report to Coronary Artery Bypass Graft

*Surgery*« und dessen Schwäche in der methodischen Durchführung für Kardiologen zunehmend schwieriger, Operateure zu finden, die noch willens waren, schwerkranke Patienten am Herzen zu operieren, da sie ein schlechtes Abschneiden in der jährlichen Mortalitäts-Rangliste befürchteten (11).

**kurzgefasst:** Die Akzeptanz von Benchmarking-Ergebnissen hängt ab von einer klaren Darstellung, Fokussierung auf ausgewählte Indikatoren, Verwendung anerkannter Methoden und Veröffentlichung des methodischen Vorgehens. Eine kontinuierliche Weiterentwicklung von Benchmarking ist notwendig.

### Bedarf an Aufklärung über Stärken und Schwächen von Risiko-adjustierten Leistungsvergleichen

Aufklärung und Schulung über die richtige Nutzung von Leistungsvergleichen ist vermutlich die wichtigste Vorarbeit vor einer Veröffentlichung von Leistungsvergleichen. Dies beugt möglichen Überinterpretationen und selektiven Darstellung durch Interessengruppen vor. 1986 veröffentlichte die staatliche Krankenversicherung für Senioren, Medicare, eine Liste von Sterblichkeitsraten für alle Krankenhäuser der USA. (1) Der Bericht erschien ohne wesentliche Interpretationshilfen und einer ungenügenden Risiko-Adjustierung, was zu einer erheblichen Verunsicherung führte. Schließlich wurden von 361 Krankenhäusern mit stark erhöhten Sterblichkeitsraten nur 6% durch externe Gutachter als auffällig identifiziert (41). Dieses Beispiel sollte in Erinnerung bleiben, wenn aufgrund von Leistungsvergleichen weitgehende Entscheidungen getroffen werden. Dies können von Bonuszahlungen für besonders gut arbeitende Ärzte bis hin zu einer selektiven Kontrahierung von überdurchschnittlich arbeitenden Leistungsanbietern geschehen.

Jeder Ergebnisbericht sollte von umfangreichen Schulungen begleitet sein, die die besprochene Punkte erläutern. Eine Internetpräsenz oder ein Servicetelefon können hilfreich sein, um übertriebenen Interpretationen vorzubeugen und gleichzeitig Anregungen für Verbesserungen des Leistungsvergleichs zu sammeln.

### Bedarf an einverständigen Rahmenbedingungen

Entscheidend für das Gelingen von Leistungsvergleichen ist ferner eine eindeutige Festlegung, welche Ziele verfolgt werden und welche positiven wie negativen Konsequenzen sich ergeben können. Vor einem ersten Leistungsvergleich herrscht im Allgemeinen bei den meisten Beteiligten erhebliche Unsicherheit, da sie sich zunächst nicht im Vergleich zu anderen Anbietern positionieren können und damit ein erhebliches Risiko auf sich zukommen sehen. Diese Unsicherheit kann gemindert werden, indem nicht der erste Leistungsvergleich sondern nur die konsekutiven öffentlich gemacht werden. Die Beteiligten können ihre Leistungen gegebenenfalls steigern und Erfahrungen sammeln. Bis umfangreiche Erfahrungen mit Leistungsvergleichen vorliegen, sollten diese hauptsächlich zur Messung einer Qualitätsverbesserung eingesetzt werden. Sollten Ausreißer identifiziert werden, so müssen diese zunächst genauer untersucht werden und eine Möglichkeit zur Stellungnahme erhalten.

**kurzgefasst:** Nutzer von Benchmarking-Ergebnissen müssen über die Grenzen dieser Verfahren aufgeklärt werden, um übertriebenen Interpretationen vorzubeugen.

### Missbrauch und Ansätze zu deren Vermeidung

Die Möglichkeit zum Missbrauch von Leistungsvergleichen besteht bei allen oben erwähnten Punkten im Wesentlichen bei der Datenerfassung, Risiko-Adjustierung und einer einseitigen Berichterstattung.

Leistungsvergleiche basieren auf empirischen Daten. Sie stellen damit einen erheblichen Anreiz dar, Leistungsindikatoren und Risikoprofile einzelner Patienten aufzuwerten (*upcoding*) (11). Obwohl statistische Methoden Hinweise auf *upcoding* geben können, scheinen stichprobenhafte Prüfungen durch externe Untersucher am erfolgreichsten zu sein.

Die Neutralität und hochqualitative Durchführung von Leistungsvergleichen kann durch die Schaffung einer gemeinsam getragenen Benchmarking-Stelle erreicht werden, die in einer möglichst transparenten Form Leistungsvergleiche durchführt oder die Durchführung von Leistungsvergleichen supervidiert und akkreditiert. Alle an einem Leistungsvergleich beteiligten sollten die Möglichkeit haben, die Durchführung unter Wahrung datenschutzrechtlicher Bestimmungen einzusehen und sich von einer sorgfältigen Arbeit zu überzeugen. Jedoch nicht einen vorher beschlossenen Verfahrensablauf zu beeinflussen.

Eine Fehlinformation durch eine selektive Berichterstattung kann nur dadurch minimiert werden, dass alle Beteiligten, von Patienten bis hin zu Ministerien über die Möglichkeiten und Grenzen von Leistungsvergleichen aufgeklärt werden.

### Fazit

Leistungsvergleiche werden eine zunehmende Bedeutung erlangen. Nachdem diese Entwicklung nicht umkehrbar ist, muss intelligent, gemeinnützig und zügig gehandelt werden. Es ist entscheidend, frühzeitig einen Konsensus über das Vorgehen bei Durchführung und Veröffentlichung von Leistungsvergleichen zu finden und an den methodischen Herausforderungen zu arbeiten, denn die Thematik ist für Patienten und Ärzte zu wichtig, um sie einer oberflächlichen Betrachtung durch Massenmedien oder Organisationen mit einseitigen Interessen zu überlassen.

**Danksagung:** Dr. Schneeweiss wurde unterstützt durch die Deutsche Forschungsgemeinschaft (schrn 527/3 & schr 527/4), U.S. Agency for Health Care Research and Quality (RO3HSO9855, RO1HS10881) und dem AOK Bundesverband. Die Ausführungen geben ausschließlich die Meinungen der Autoren wieder. Unser besonderer Dank gilt Prof. Malcolm Maclure für hilfreiche Diskussionen.

## Literatur

- 1 Anonymous. 1986 hospital mortality information. Federal Register, August 17 1987; 52: 30741-30745
- 2 Angell M, Kassierer JP. Quality and the medical marketplace - following elephants. *N Engl J Med* 1996; 335: 883-885
- 3 Brook RH, Avery AD, Greenfield S. Assessing the quality of medical care using outcomes measures: An overview of the method. *Med Care* 1977; 15: (suppl)
- 4 Brook RH, McGlynn EA, Cleary PD. Quality of health care. Part 2: Measuring quality of care. *N Engl J Med* 1996; 335: 966-970
- 5 Cary K. A multilevel modeling approach to analysis of patient costs under managed care. *Health Economics* 2000; 9: 435-446
- 6 Clancy CM, Eisenberg JM. Outcomes research: measuring the end results of health care. *Science* 1998; 282: 245-246
- 7 Pennsylvania Health Care Cost Containment Council. Coronary artery bypass graft surgery: a technical report. Harrisburg, PA, November 1992
- 8 Corrigan JM, Rogers LS. Comparative performance measurement for health plans. Gaithersburg, MD, Aspen, In: Kazandjian VA: The epidemiology of quality, 1995
- 9 Donabedian A. Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly* 1966; 44: 166-206 (Suppl)
- 10 Eddy DM. Performance measurement: problems and solutions. *Health Affairs* 1998; 17: 7-25
- 11 Epstein A. Performance reports on quality- prototypes, problems, and prospects. *N Engl J Med* 1995; 333: 57-61
- 12 Epstein AM. The role of quality measurement in a competitive marketplace. Ann Arbor, MI: Health Administration Press, In: Reinhard U. (ed): Strategic choices for a changing health care system, 1996
- 13 Gatsonis CA, Epstein AM, Newhouse JP, Normand SL, McNeil BJ. Variations in the utilization of coronary angiography for elderly patients with acute myocardial infarction. An analysis using hierarchical logistic regression. *Med Care* 1995; 33: 625-642
- 14 GFK Umfrage, Mehr Transparenz, besserer Service. *Focus* 2000; 43: 202-203
- 15 Goldfield N. Public disclosure of case mix adjusted clinical information: practical and theoretical challenges. Gaithersburg, MD, Aspen, In: Physician profiling and risk adjustment, 2nd ed. 1999: 77-78
- 16 Goldstein H, Healy JR. The graphical presentation of a collection of means. *J R Stat Soc A* 1995; 158: 175-177
- 17 Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. *J R Stat Soc A* 1996; 159: 385-443
- 18 Goodspeed R. Public disclosure: A hospital chief executive's perspective. Gaithersburg, MD, Aspen, In: Physician profiling and risk adjustment, 2nd ed. 1999: 91-93
- 19 Gottwik M, Kretschmar R, Vogt A, Hepp A, Weber MA, Sechtem U, Hauptmann KE, Tebbe U, Grube E, Glunz HG, Neuhaus KL. 30-Tage-Sterblichkeit nach Herzoperation. Ein Modellprojekt der Arbeitsgemeinschaft Leitender Kardiologischer Krankenhausärzte. *Deutsche Medizinische Wochenschrift*. 1999; 124: 1090-1094.
- 20 Greenland S. Principles of multilevel modeling. *Int J Epidemiol* 2000; 29: 158-167
- 21 Gross R, Löffler M. *Prinzipien der Medizin*. Berlin, Springer, 1997: 366
- 22 Health Care Financing Administration quality indicators: <http://www.hcfa.gov/quality/11a.htm>.
- 23 Hornbrook MG, Goodman MJ. Chronic disease, functional Health Status, and demographics: A multi-dimensional approach to risk adjustment. *Health Services Research* 1996; 31: 283-307
- 24 Iezzoni LI (ed.). Risk Adjustment for Measuring Healthcare Outcomes, 2nd ed. Health Administration Press, Chicago, 1997
- 25 Iezzoni LI, Shwartz M, Restuccia J. The role of severity information in health policy debates: A survey of state and regional concerns. *Inquiry* 1991; 28: 117-128
- 26 Jencks SF, Cserdon T, Burwen DR et al. Quality of medical care delivered to medicare beneficiaries. A profile at state and national levels. *JAMA* 2000; 284: 1670-1676
- 27 Kassierer JP. The use and abuse of practice profiling. *N Engl J Med* 1994; 330: 634-636
- 28 Kazandjian VA. The epidemiology of quality. Gaithersburg, MD, Aspen, 1995
- 29 Kolkman FW, Scheinert IID, Schoppe C, Walger M. KTQ-Projekt zur Zertifizierung von Akutkrankenhäusern. *Z ärztl Fortbild Qual sich* (<http://www.ktq.de>) 2000; 94: 651-657
- 30 Lauterbach KW, Lungen M. Was hat die Vergütung mit der Qualität zu tun? Schattauer, Stuttgart, In: Arnold M, Litsch M, Schellschmidt H: Krankenhaus-Report 2000, 2001: 119
- 31 Leber W-D, Schlottmann N, v Stackelberg J-M. Die Einführung von DRGs in Deutschland - Erste Vereinbarungen der gemeinsamen Selbstverwaltung. Schattauer, Stuttgart, In: Arnold M, Litsch M, Schellschmidt H: Krankenhaus-Report 2000, 2001: 84
- 32 Litsch M, Sahlmüller H. Die Methodik des leistungsorientierten WIDO-Krankenhausvergleiches. Berlin: Springer Verlag, In: Sieben G, Litsch M: Krankenhausbetriebsvergleich, 2000
- 33 Marzuk PM. The right kind of paternalism. *N Engl J Med* 1985; 313: 1474-1476
- 34 McNeill BJ, Pedersen SH, Gatsonis C. Current issues in profiling quality of care. *Inquiry* 1992; 29: 298-307
- 35 Millenson ML. Demanding medical excellence: Doctors and accountability in the information age. Chicago, University of Chicago Press, 1997
- 36 Morgan KO, Morgan S, Uhlig MA. Health care state ranking 1998. Morgan Quitno, Lawrence, KA, 1998
- 37 Newhouse JP. Patients at risk: Health reform and risk adjustment. *Health Affairs* 1994; 13: 132-146
- 38 Paeger A. Benchmarking auf der Basis einer Krankenhausbewertung nach EFQM und Joint Commission am Beispiel der Asklepios Kliniken. I Lucherhand, Neuwied, n: v. Wilfried, v. Eiff: Krankenhaus Betriebsvergleich - Controlling-Instrumente für das Krankenhausmanagement, 2000
- 39 Sangha S, Schneeweiss S. Lebensqualität und Qualitätssicherung im Krankenhaus. Ecomed Verlag: Landsberg, In Ravens-Sieberer U, Cieza A, v. Steinbüchel N, Bullinger M (Hrsg): Lebensqualität und Gesundheitsökonomie in der Medizin, 2000: 413-421
- 40 Sangha O, Wildner M, Schneeweiss S, Siebert H, Hebeisen H, Hübner M, Bauer H, Witte J. Fehlbelegung im Krankenhaus. Entwicklung eines standardisierten Verfahrens zur Beurteilung der Notwendigkeit von vollstationären Krankenhausbehandlungen. *Der Chirurg* 1999; 38: 201-210.
- 41 Scheckler WE. Counterpoint: Public disclosure of process and outcome measures. *Clin Performance Quality Health Care* 1999; 7: 41-42
- 42 Schneeweiss S, Sangha O. Weiterentwicklung von Krankenhausbetriebsvergleichen: Wie wichtig ist Risiko-Adjustierung für den Krankenhausvergleich. Springer Verlag, Berlin, In: Sieben G, Litsch M (Hrsg): Krankenhausbetriebsvergleiche pp131-147, 2000
- 43 Schneeweiss S, Sangha O, Manstetten A, Schlottmann N, Liebetrau M, Hartman W, Walter B, Schultz W, Paeger A, Euler P, Eichenlaub A, Leber W-D, Rauh G. Identifikation von medizinischen Indikatoren für Ergebnisqualität in der internistischen Krankenhausversorgung: Ergebnisse der QMK Pilotstudie. *Gesundheitsökonomie und Qualitätsmanagement* 2000; 5: 173-182
- 44 Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care. *N Engl J Med* 1996; 335: 251-256
- 45 Schwarz F. Zitiert im Stern 2000; 50: 228
- 46 Selbmann HK. Qualitätskontrolle in der Perinatalogie Betrachtungen am Beispiel der Münchner Perinatal-Studie. *Münch Med Wschr* 1978; 720: 595-598
- 47 Shwartz M, Ash AS, Iezzoni LI. Comparing outcomes across providers. Health Administration Press, Chicago, In: Iezzoni LI (ed.): Risk Adjustment for Measuring Healthcare Outcomes, 2nd ed. 1997
- 48 Smith DW. Evaluating risk adjustment by partitioning variation in hospital mortality rates. *Stat Med* 1994; 13: 1001-1013
- 49 Streiner DL, Norman GR. *Health Measurement scales*, 2nd ed. Oxford University Press, Oxford, 1995
- 50 Sullivan LW, Wilensky GR. Medicare Hospital Mortality Information 1987-1989. Dept. of Health and Human Services, Health Care Financing Administration, 1991
- 51 Transition I, Transition Systems Inc., Boston, MA,
- 52 Wegweiser zum Spezialisten. Die neue Ärzte Liste Teil1. *Focus Magazin* 2000; 43: 218-232
- 53 Zelen M. Author's reply to compliance, bias, and power in clinical trials. *Biometrics* 1991; 47: 778-779

## Korrespondenz

Dr. med. Sebastian Schneeweiss, Sc.D.  
 Department of Medicine  
 Brigham and Women's Hospital and  
 Harvard Medical School  
 221 Longwood Ave (BLI-341)  
 Boston, MA U2115  
 Tel.: 617/278-0937  
 Fax: 617/232-8602  
 E-Mail: [schneeweiss@post.harvard.edu](mailto:schneeweiss@post.harvard.edu)