



© Vilevi - Fotolia.com

# Effizienz und Qualität im Spitalswesen – eine Frage der Größe?

Immer wieder flammt in Fragen der Gesundheitsreform die Diskussion um Spitalsstandorte auf. Tatsache ist, dass Österreich ein spitalslastiges System aufweist und so in seinen Bemühungen um die Gesundheit der Bevölkerung dann ansetzt, wenn Krankheiten schon weiter fortgeschritten sind. Dies dürfte einer der wichtigsten Gründe dafür sein, weshalb wir bei im internationalen Vergleich hohen Kosten höchstens mittelmäßige Ergebnisse in der Bevölkerungsgesundheit aufweisen.

Egal, ob es um Bad Aussee, Hartberg/Oberwart oder Baden/Mödling geht, das Thema erhitzt die Gemüter oft derart, dass Sachargumente kaum noch wahrgenommen werden können. Den Höhepunkt der Unsachlichkeit erreichte die Diskussion, als im Juni 2010 der Rechnungshof den Problemaufriss des Themenfeldes Gesundheit an die Mitglieder der Arbeitsgruppe der Verwaltungsreform, deren Steuerungsgremium, übermittelte. Durch gezielte Indiskretion einiger Mitglieder wurden Journalisten mit Halbwahrheiten versorgt, die in einer seltenen (und seltsamen) Eigendynamik absurde Stilblüten trieben. So schrieb z. B. eine Zeitung:

„Verwaltungsreform: SPÖ prüft Schließung von 61 % der Spitäler“, und veröffentlichte in der folgenden Ausgabe eine „Rote Liste“. Tags darauf hatte jeder Landeshauptmann bereits Standortgarantien für seine Spitäler abgegeben und damit jede sachliche und ruhige Diskussion im Rahmen der Verwaltungsreform obsolet gemacht. Wieder war wertvolle Zeit für den Umbau des österreichischen Gesundheitswesens vergeudet worden. Dabei bestand der Auftrag an die Arbeitsgruppe nur darin, Probleme des österreichischen Gesundheitswesens aufzulisten, Maßnahmen im Sinne von Lösungsvorschlägen sollten erst im entsprechenden Vorbereitungsgremium erarbeitet werden. Von Schließungsabsichten oder dergleichen konnte also keine Rede sein.

Mit dem nunmehrigen Bekenntnis, im Spitalswesen in den nächsten fünf Jahren 1,8 Mrd. Euro „einsparen“ zu wollen, erhält dieses Themenfeld neue Bedeutung. Denn während dieses Ziel mittlerweile außer Streit steht, wird es ab nun darauf ankommen, die dafür notwendigen Maßnahmen zu entwickeln. Daher lohnt sich auch ein Blick auf die



Dr. Thomas Cypionka, ist Senior Researcher und verantwortlich für den Bereich Gesundheitsökonomie und Gesundheitspolitik am Institut für Höhere Studien (IHS HealthEcon).

Tatsachen und den Erkenntnisstand, über den wir bezüglich des Zusammenhangs zwischen Größe und Effizienz bzw. Größe und Qualität verfügen.

### Größe und Effizienz von Spitälern

Bereits die Grundzüge der Produktionstheorie legen einige sehr allgemein anwendbare Gesetzmäßigkeiten fest. Während in der kurzen Frist nicht alle Produktionsfaktoren variabel sind und daher die Ausweitung dieser variablen Faktoren zwar den Output weiter erhöhen kann, aber dies bei fortlaufender Ausweitung mit immer geringerem Erfolg, können in der langen Frist auch fixe Produktionsfaktoren erweitert werden. In diesem Fall spricht man von Skalenerträgen: Verdoppelt man alle Inputfaktoren, so kann sich der Output ebenfalls verdoppeln (konstante Skalenerträge), weniger als verdoppeln (abnehmende Skalenerträge) oder mehr als verdoppeln (zunehmende Skalenerträge). In der Praxis beobachtet man im Allgemeinen zunächst zunehmende, dann konstante und schließlich fallende Skalenerträge. Während eine Standortvergrößerung zunächst viele Synergien, Lerneffekte und Einkaufsmacht erzeugt, sind diese Vorteile irgendwann aufgezehrt und es kehren sich diese Einflüsse um: Die Wege in einem Spital werden weiter, die Kommunikation schwieriger, die Logistik immer komplexer. Das Spiegelbild dieser Entwicklung ist in der langfristigen Durchschnittskostenkurve zu finden, die typischerweise U-Form hat und bei der das Kostenminimum der tiefste Punkt und gleichzeitig der Wechsel zu abnehmenden Skalenerträgen (mit dann wieder steigenden Durchschnittskosten) ist.

Die Literatur zur Frage von Größe und Effizienz im Spitalswesen ist sehr umfangreich; im Folgenden soll ein historischer Abriss der wissenschaftlichen Arbeiten gegeben werden, die sich mit dem Thema beschäftigt haben. Die ersten Studien zum Thema zogen zunächst Kostenfunktionen zur Schätzung des optimalen Verhältnisses zwischen Kosten und Output heran. Die optimale Größe einer Krankenanstalt ergibt sich dabei aus dem Verhältnis der langfristigen Durchschnittskosten und der Anzahl der erbrachten Leistungen, die sich typischerweise im Minimum der langfristigen Durchschnittskosten, dem Trade-off-Punkt zwischen zunehmen-

den und abnehmenden Skalenerträgen, befindet. Zur Überprüfung von Skaleneffekten in Spitälern wurden im Laufe der Zeit verschiedene Methoden angewandt. Anfangs wurden vor allem Kostenfunktionen gegebener Krankenanstalten geschätzt, aus deren Form auf Skaleneffekte geschlossen wurden. Seit den 1990er Jahren wurde die Frage der Effekte von Mergern zwischen Spitälern immer wichtiger, woraus tatsächliche oder hypothetische Vorher-nachher-Studien zur Kostenstruktur in den Vordergrund traten. In der jüngeren Vergangenheit fand eine Art Methodenwechsel statt. Im Vordergrund standen nun eher die Stochastic Frontier Analyse bzw. die Data Envelopment Analyse, welche Produktionseffizienz besser messen können. Durch geeignete mathematische Formulierungen oder sekundäre Analysen lassen sich dabei auch Skaleneffekte berechnen.

### Was sagt die Literatur?

Eine der ersten Studien, die die Existenz positiver Skaleneffekte im Spitalsbereich nachwies, kam von P. Feldstein (1961), der die marginalen Kosten von Leistungen im Spitalsbereich empirisch untersuchte. In einer späteren Studie kritisiert Ralph E. Berry (1967) dieses Ergebnis. Feldstein habe wichtige Faktoren, wie etwa die Qualität und Bandbreite der unterschiedlichen Krankenhausdienstleistungen, nicht berücksichtigt. Diese Aspekte müssten jedoch in eine Kosten-Output-Analyse mit einfließen, um aussagekräftige Ergebnisse zu erzielen und eine Überschätzung tatsächlich vorhandener Skaleneffekte zu vermeiden. Berry versucht diese Faktoren in seine Analyse der Skaleneffekte von Krankenanstalten mit einzuschließen.

Ein großes Problem bei der Messung der optimalen Spitalsgröße ist das hohe Ausmaß an Produktdifferenzierung der Krankenanstalten und die Variation der angebotenen Leistungen. Für eine aussagekräftige Untersuchung der Skaleneffekte von Krankenanstalten bedarf es daher einer Schätzung der Produktionskosten, die frei von Einflüssen der Produktdifferenzierung ist.<sup>1</sup> Berry versucht diesen optimalen Zustand einer homogenen Produktpalette an Dienstleistungen zu erhalten, indem er Krankenanstalten mit ähnlicher Ausstattung und Bandbreite an



**Johanna Kemper**  
ist Studentin der Volkswirtschaftslehre und Stipendiatin am Institut für Höhere Studien (IHS), Bereich IHS HealthEcon.



**Eva Schweiger**  
ist Studentin der Internationalen Entwicklung (Universität Wien) und Stipendiatin am Institut für Höhere Studien (IHS), Bereich IHS HealthEcon.

<sup>1</sup> In seinem Paper nimmt Berry Bezug auf Studienergebnisse von Ingbar und Taylor (1967), die, unter Berücksichtigung der angebotenen Bandbreite an Leistungen, aus einem Sample von 72 Gemeindekrankenanstalten in Massachusetts mit einer Größe zwischen 30 und 330 Betten eine langfristige Durchschnittskostenkurve in Form eines „invertierten U“ schätzten. Sie fanden heraus, dass Dienstleistungen bei einer Krankenanstaltsgröße von unter 200 Betten unter Bedingungen abnehmender, bei einer Größe von über 200 Betten jedoch unter Bedingungen zunehmender Skalenerträge produziert wurden. Berry argumentiert, dass diese spezielle Form der Kostenkurve Ingbar und Taylors vermutlich auf die Interdependenz von Kosten, Größe, aber auch auf die Bandbreite der angebotenen Leistungen zurückzuführen ist. Krankenanstalten sollten folglich nicht nur bezüglich ihrer Größe und ihrer Kosten, sondern auch hinsichtlich der Bandbreite der angebotenen Leistungen untersucht werden. Eine Obergrenze für die optimale Anzahl der angebotenen Leistungen sollte zu einem Output-Niveau korrespondieren, das mit zunehmenden Skaleneffekten einhergeht. Zusätzlich lässt sich nach heutiger Kenntnis sagen, dass die Durchschnittskostenkurve nur scheinbar diesen atypischen Verlauf hatte, da sich ihr eigentlicher Wendepunkt jenseits der Spitalsgrößen befindet, die Ingbar und Taylor untersuchten.



Dienstleistungen in Gruppen zusammenfasst. So kann ein weitgehend unverzerrtes Kosten-Output-Verhältnis geschätzt und die Frage nach der Existenz positiver Skaleneffekte beantwortet werden. Für seine Analyse benutzt er Daten der American Hospital Association von 5.293 Akutspitalern. Krankenanstalten mit ähnlicher Ausstattung und ähnlichem Angebot an Dienstleistungen wurden in Gruppen zusammengefasst. Für die Analyse wurde das Verhältnis zwischen Durchschnittskosten und dem Output-Level für 40 Gruppen von Krankenanstalten untersucht.<sup>2</sup> Das Ergebnis der Untersuchung zeigt, dass Akutkrankenanstalten der Basisversorgung mit homogener Ausstattung und Auswahl an Dienstleistungen positive Skalenerträge aufweisen. Dabei sind Einflussfaktoren wie Größe, Kosten und Bandbreite an Dienstleistungen eng miteinander verknüpft. Für eine gegebene Anzahl von Dienstleistungen sind die Durchschnittskosten der Produktion für große Krankenanstalten geringer als für kleinere. Politikmaßnahmen sollten laut Berry darauf abzielen, die positiven Effekte zunehmender Skalenerträge im Krankenanstaltswesen zu nutzen, jedoch unter Berücksichtigung der Bandbreite der angebotenen Dienstleistungen, um Ineffizienzen (sinkende Skalenerträge) zu vermeiden.

William J. Lynk (1995) geht in seiner Arbeit der Frage nach, ob Fusionen von Krankenanstalten empirisch nachweisbare Effizienzpotentiale schaffen. In seiner Analyse untersucht er, ob durch eine Reduzierung der Variabilität zufälliger Patientennachfrage und technische Zusammenlegung klinischer Dienstleistungen Kosteneinsparungen vorgenommen und somit steigende Skalenerträge erzielt werden können. Die Patientennachfrage in klinischen Fachbereichen ist hochgradig dem Zufall unterworfen. Je größer der Fachbereich, ob durch Ausweitung der Bandbreite oder des Volumens, desto kleiner die Variabilität der Patientennachfrage. Mit der Zusammenlegung von Fachbereichen könnten folglich Effizienzgewinne erzielt werden. Für seine Untersuchung schätzt Lynk diese Variabilität auf Grundlage der Daten zweier fusionierender Krankenanstalten mit vier Standorten in einem Ballungsgebiet.

Die Daten geben Aufschluss über alle aufgenommenen und entlassenen Krankenhauspatienten der vier Standorte in den Jahren 1991 und 1992, sowie über die Zugehörigkeit jedes Patienten zu einer Gruppe mit ähnlicher Diagnose, sodass die genaue tägliche Patientennachfrage für einen speziellen Fachbereich errechnet werden kann. Es wurden sieben klinische Fachbereiche definiert, für die die Variabilität der Patientennachfrage an allen vier Stand-



© iiro - Fotolia.com

orten gemessen werden kann. Zur Analyse der Variabilität der Patientennachfrage führt er eine Schätzung der Standardabweichung für die speziellen Fachbereiche durch. Die Analyse der Daten ergibt ein erhebliches Potential zur Reduzierung der relativen Variabilität der Patientennachfrage durch die Zusammenlegung spezieller klinischer Dienstleistungen an einen Standort innerhalb eines gemeinsamen Verbundsystems (wie etwa einer Holding). Die relative Variabilität der Patientennachfrage und damit das Ausmaß der personellen Überbesetzung (Vorhalteleistung) fallen mit ansteigender Größe des Fachbereiches und erreichen ein Minimum bei deren vollständiger Zusammenlegung. Die Kosten aller sieben untersuchten klinischen Fachbereiche fallen bei vollständiger Zusammenlegung wenigstens um 4,3 % und bis zu 28,2 %. Die durchschnittliche, gewichtete Kostenreduzierung durch Zusammenlegung beträgt 8,8 %. Des Weiteren findet Lynk Evidenz für steigende Skalenerträge und somit für Effizienzgewinne, auch wenn eine Zusammenlegung der Fachbereiche nur teilweise erfolgt. Ferner stützen die Ergebnisse Lynks die sogenannte „Excess-Capacity-Hypothese“, der zufolge Vorhaltekosten mit zunehmender Größe der Krankenanstalt/des Fachbereiches abnehmen sollten.

Infolge der Fusion von Krankenanstalten werden bevorzugt Hilfskostenstellen (Verwaltung, Technik, Wäscherei, ...) zusammengelegt, da hier hohe Skalenerträge lukriert werden können. Dranove (1998) analysiert die Skaleneffekte solcher Zusammenlegungen und versucht dabei, die Skalenerträge möglichst exakt zu messen. Durch die ausschließliche Betrachtung der Hilfskostenstellen umgeht er potentielle Verzerrungen durch nicht berücksichtigte

<sup>2</sup> Durch Schätzung einer linearen Regression mit den Durchschnittskosten als abhängiger Variable und der Gesamtmenge aller Patiententage als unabhängiger Variable.


© Franz Pfluegl - Fotolia.com

Faktoren wie Qualität oder Unterschiede im Leistungsangebot. Für die semiparametrische Schätzung seiner Kostenkurve benutzt Dranove als Outputs nicht, wie in der bisherigen Literatur üblich, die Anzahl der Betten einer Krankenhaus, sondern den tatsächlichen Output, die Anzahl der entlassenen Patienten. Anhand von Finanzdaten aus dem Jahr 1992 untersucht er die Kosten von 14 verschiedenen Hilfskostenstellen in 302 privaten Krankenhäusern mit 100 bis 430 Betten, die in jeder typischen Krankenhaus vorhanden sind und rund 15 % der Gesamtausgaben ausmachen. Steigende Skalenerträge können, bis zu einem gewissen Grad, in fast allen untersuchten Abteilungen nachgewiesen werden. Abteilungen, deren Beanspruchung durch andere Abteilungen korreliert, wiesen besonders hohe positive Skalenerträge auf. Die Aggregation aller 14 Abteilungen zeigt substantielle positive Skaleneffekte für Krankenhäuser mit bis zu 7.500 Entlassungen (etwa 200 Betten). Eine Krankenhaus dieser Größe hat im Durchschnitt um 6 % geringere Kosten pro entlassenen Patienten als eine Krankenhaus mit 5.000 Entlassungen (etwa 136 Betten) und rund 32 % geringere Kosten als Krankenhäuser mit 2.500 Entlassungen (etwa 100 Betten). Hingegen schienen für größere Krankenhäuser nur noch geringere (aber immer noch positive) Skaleneffekte vorzuliegen. Dranove kommt hier also in seinem Sample in den Bereich vor dem Wendepunkt der Durch-

schnittskostenkurve, an dem die Skalenerträge wieder abnehmen würden. Allerdings bleibt aufgrund des Ranges der Krankenhausgrößen unklar, wann die Skalenerträge wieder abnehmend sind. Zudem muss bedacht werden, dass seit 1992 die höhere Kapitalintensität der Medizin einerseits und Entwicklungen in der EDV andererseits heute eher eine Rechtsverschiebung des langfristigen Durchschnittskostenminimums bewirkt haben dürften.

Preyra und Pink (2006) führten eine Untersuchung zum Krankenhaussektor in der Provinz Ontario in Kanada durch, der Ende der 1990er Jahre eine massive Umstrukturierung erlebte, im Zuge derer viele Krankenhäuser geschlossen oder zusammengelegt wurden. Für den Umstrukturierungsprozess wurde ein auf Kostenrechnungsergebnissen basierender Analyserahmen benutzt, um davon Politikmaßnahmen und Schätzungen daraus resultierender Kosteneinsparungen abzuleiten.

Ziel der Untersuchung von Preyra und Pink ist es herauszufinden, welche Effekte durch die Zusammenlegung von Krankenhäusern entstehen konnten. Für ihre Analyse benutzen Preyra und Pink Daten für 210 Krankenhäuser aus den letzten zwei Jahren vor der Restrukturierung, 1994/95 und 1995/96, und schließen somit beinahe alle Akut- und Nicht-Akutkrankenhäuser Ontarios in ihr Sample ein. Die Krankenhäuser wurden, je nach Anzahl der Betten, in acht Vergleichsgruppen aufgeteilt.

Preyra und Pink schätzen die Skaleneffekte verschiedener Kombinationen an Krankenhäusern proportional zum Mittelwert des Outputvektors des gesamten Samples. Ändern sich die Outputs der zusammengelegten Krankenhäuser, so ändert sich auch der optimale aggregierte Kapitalstock. Würden zwei (theoretische, identische) Krankenhäuser, deren Output jeweils den Mittelwert des Outputvektors ausmacht, zu einer Krankenhaus zusammengelegt werden, so würde das neue Spital nur 78 % des zusammengelegten, gemeinsamen Kapitalstocks benötigen und gemessen an den Gesamtkosten eine Gesamtkostenreduktion von 10,5 % lukrieren können. Positive Skaleneffekte stellen die Autoren für alle Größen fest.

Ein weiterer wichtiger Beitrag der Autoren ist ihre Abhandlung von Economies of Scope, bei der sie einen Outputvektor aus vier verschiedenen Outputs verwenden. Sie modellieren verschiedene Kombinationen von Outputs, indem sie einen Spezialisierungsparameter definieren, der angibt, inwieweit ein Spitalverbund seine Standorte auf nur einen der Outputs spezialisiert bzw. Kombinationen bevorzugt. Auf diese Weise erhält die Kostenfunktion eine dreidimensionale Form (Kosten bei unterschiedlichen Spezialisierungsgraden bei unter-



schiedlichen Größen). Die Untersuchung bleibt an diesem Punkt zwar eher theoretisch, weist jedoch bereits darauf hin, dass auch die Betrachtung von Spezialisierungen und der Aufgabenverteilung zwischen Häusern in die Diskussion einfließen sollte. Als Reaktion auf die Schwierigkeiten bei der Messung des Effizienzpotentials von Spitälern anhand von Kostenfunktionen verwendeten immer mehr Studien das Verfahren der Data Envelopment Analyse (DEA). Auch auf diesem Gebiet wurden mit der Zeit methodische Verbesserungen vorgenommen. Für viele Untersuchungen zur Effizienz von Spitälern wurden ausschließlich Outputeinheiten verwendet. Erste Studien haben begonnen, auch den Outcome einzubeziehen.

Harris/Ozgen/Oczan (2000) untersuchen 20 US-Spitäler zwischen 120 und 694 Betten, die im Jahre 1992 aus 41 Spitälern fusioniert wurden und verwenden dazu Daten von 1991 bis 1993. Sie wenden auf das Sample folgerichtig DEA-Modelle mit konstanten und mit variablen Skalenerträgen an. Die Skaleneffizienz lässt sich nämlich aus den Effizienzergebnissen der beiden unterschiedlichen mathematischen Formulierungen erschließen. Ihr Resultat ist eindeutig: Die Merger haben die Gesamteffizienz der Spitäler steigen lassen, und dies zu einem Großteil durch Skaleneffekte.

Ferrier und Valdmanis (2004) führen eine ähnliche Untersuchung für Spitalsfusionen des Jahres 1997 mit Daten von 1996 bis 1998 durch und verwenden dabei zusätzlich eine Kontrollgruppe ähnlicher, nicht fusionierter Spitäler. Die Ergebnisse dieser Untersuchung sind nicht eindeutig. Merger hatten nicht zwingend einen Effizienzgewinn zur Folge. Tiemann und Schreyögg (2009) gehen der Frage nach, wie sich die Effizienzleistung öffentlicher gegenüber privaten gewinnorientierten und privaten nichtgewinnorientierten Krankenanstalten verhält und untersuchen dabei auch den Einfluss von Größe auf die Effizienzwerte. Ihre Untersuchung erfolgt anhand einer DEA-Analyse mit anschließender linearer Regression.

Die Daten für die Analyse schließen 1.046 öffentliche, privat-gewinnorientierte und privat-gemeinnützige Krankenanstalten in Deutschland für den Zeitraum von 2002 bis 2006 ein. Tiemann und Schreyögg benutzen ein Input-orientiertes DEA-Modell mit variablen Skalenerträgen (BCC-Modell) mit der Mortalität als zusätzlicher Outputvariable. In einem zweiten Schritt benutzen Tiemann und Schreyögg die Punktwerte der DEA-Analyse als abhängige Variable einer linearen Regression, in

die sie weitere Kontrollvariablen einschließen, darunter eine zur Messung des Wettbewerbs im Spitalssektor. Wird die Anzahl der Betten als Maßzahl für die Größe in die Analyse mit einbezogen, ergibt sich daraus, dass große Krankenanstalten signifikant effizienter sind. Mit Vorsicht ist dieses Ergebnis aufgrund der BCC-Modellierung zu interpretieren: Es stellt eigentlich nur die technische Effizienz dar. Die statistische Analyse sagt also genau genommen nur, dass größere Spitäler technisch effizienter sind, also bei jeweils gegebener Größe effizient sind. Eine Berechnung der Skaleneffizienz (= Gesamteffizienz/technische Effizienz) wäre interessant gewesen.

Wie bereits in vielen Ländern, so wurde auch in Dänemark in den letzten Jahren der Spitalssektor massiv umstrukturiert. Kristensen, Bogetoft und Pedersen (2010) beschäftigen sich mit der Frage von Effizienzpotentialen von Spitälern durch Fusion. Ihre Studie bezieht sich auf die damals noch bevorstehenden Fusionen und Zusammenlegungen der dänischen Krankenanstalten, die im Zuge der Restrukturierung vorgenommen werden sollten. Dabei sollten durch Skaleneffekte, Verbundeffekte und erhöhte Produktivität Einsparungen vorgenommen werden. Genauer Ziel der Studie ist es, potentielle Effizienzgewinne der geplanten Fusionen zu untersuchen und diese in technische, Größen- und Verbundeffekte zu unterteilen, um herauszufinden, ob diese Fusionen Kosteneinsparungen generieren können, die derartige Pläne rechtfertigen. Motivation für diese Aufteilung ist, dass offenbare Fusionsgewinne bereits mit einer teilweisen Zusammenlegung der Standorte zu erreichen sein können.



© Perry - Fotolia.com

Infolge der Fusion von Krankenanstalten werden bevorzugt Hilfskostenstellen zusammengelegt, da hier hohe Skalenerträge lukriert werden können.


© Andrey Rakhatulin - Fotolia.com

Für die Studie werden Daten von 36 öffentlichen dänischen Krankenanstalten für das Jahr 2004 analysiert, die zwischen 31 und 1.137 Betten haben. Kristensen/Bogetoft/Pedersen verwenden DEA-Modelle mit konstanten, variablen und nichtsinkenden Skalenerträgen, um die effiziente Grenze zu schätzen. Basierend auf den Effizienzgrenzen berechnen sie die Effekte von verschiedenen angelegten Fusionen. Nicht überraschend scheiden manche Fusionen aus, weil aufgrund der resultierenden Größe abnehmende Skaleneffekte resultieren. Wichtig erscheint das Ergebnis der Untersuchung, dass nämlich in diesem dänischen Sample technische Ineffizienz überwiegt und ebenfalls bedeutende Economies of Scope erzielt werden könnten. Dies hat große praktische Bedeutung: Technische Effizienz kann auch ohne Standortzusammenlegungen gewonnen werden, indem z. B. best-Practice-Abläufe optimiert werden. Verbundeffekte lassen sich auch durch Spezialisierung und optimale Verteilung der Aufgaben erreichen, sodass Zusammenlegungen nicht zwingend notwendig sind.

### Größe und Qualität

Die Diskussion um den Zusammenhang von Größe und Qualität brachte der Artikel „Should Operations be Regionalized?“ von Luft et al. (1979) ins Rollen. Die Studie untersucht Mortalitätsraten für zwölf Operationen in 1.498 Krankenanstalten auf ihren Zusammenhang mit dem Operationsvolumen sowohl des Krankenhauses als auch der Chirurgen.

Die Autoren stellen fest, dass die Mortalitätsraten mit zunehmender Anzahl an Operationen abnehmen. Luft et al. berechnen, dass die Mortalitätsraten<sup>3</sup> um 25 bis 41 % niedriger sind in Krankenhäusern, die jährlich mehr als 200 Operationen durchführen. Für andere Interventionen werden niedrigere Schwellwerte erfasst. Parallel dazu gibt es Eingriffe (z. B. Cholecystektomie), für die kein Zusammenhang berechnet werden konnte. Die Autoren argumentieren, dass ihre Ergebnisse die Konzentration gewisser Operationen in größeren Einheiten rechtfertigen.

Einen weiteren Beitrag zur Debatte um Größe und Qualität lieferten Hughes et al. (1987). Sie analysieren die kausale Beziehung zwischen großem Patientenvolumen und besseren Behandlungsergebnissen. Dazu werden Daten von 900 US-Spitälern für 17 ausgewählte Diagnosen und Behandlungsmethoden verwendet, wobei die Autoren für Charakteristika von Patienten und Spitalern kontrollieren. In ihrem Artikel stellen die Autoren zwei Hypothesen vor, die auch heute noch viel diskutiert sind:

Die „Practice-makes-perfect-Hypothese“ misst vor allem Lerneffekten die größte Bedeutung bei und impliziert, dass durch höhere Volumina das Personal eingespielter und erfahrener ist und daher auch spezialisiertere Gerätschaften zum Einsatz kommen können.

Die „Selective-referral-Pattern-Hypothese“ besagt, dass ein größeres Patientenvolumen deshalb vorliegt, weil viele wissen, wo die Qualität in einem Spital besser ist. Eine gute Reputation zieht also Patienten an, sodass dann typischerweise hohe Qualität in Häusern beobachtet wird, die viele entsprechende Patienten betreuen. Hughes et al. merken an, dass diese Reputation einerseits auf messbaren Ergebnissen, aber auch auf allgemeiner Einschätzung beruhen kann.

Die Studie zeigt, dass die Selective-referral-Pattern-Hypothese bei komplexeren Patientenkategorien zutrifft. Die beiden Hypothesen haben unterschiedliche Implikationen für die Politik. Hughes et al. wollen durch ihre Analyse aufzeigen, dass für gewisse Diagnosen und Behandlungsmethoden unterschiedliche Hypothesen eine Rolle spielen und stellen fest, dass die dominierende Hypothese für die abzuleitenden Politikmaßnahmen eine große Rolle spielt. Trifft die Selective-referral-Pattern-Hypothese zu, erhöht sich die Qualität bei Vergrößerung des Patientenvolumens nicht zwingend. Wenn in diesem Fall die Practice-makes-perfect-Hypothese zusätzlich nicht zutrifft, könnte sich das

3 Operation am offenen Herzen; Gefäßoperationen; transurethrale Resektion der Prostata; Koronarbypass.



Die „Practice-makes-perfect-Hypothese“ misst vor allem Lerneffekten die größte Bedeutung bei und impliziert, dass durch höhere Volumina das Personal eingespielter und erfahrener ist und daher auch spezialisiertere Gerätschaften zum Einsatz kommen können.

Gesamtergebnis sogar verschlechtern, sofern Patienten aus den Spitälern besserer Qualität abgezogen werden. Weiters könnten Ärzte, welche Überweisungen auf Qualitätsunterschieden begründen, ihre Patienten weiterhin an bestimmte Gesundheitseinrichtungen überweisen. Damit könnten sie sich auch möglichen einschränkenden Vorgaben bezüglich der Spitalwahl widersetzen. Bestehende Zuweisungsmuster – basierend auf Reputation – könnten unbeobachtbare Faktoren bereits berücksichtigen. Zentral erscheint, dass bei der Auswahl von Spitälern zu beachten ist, dass die Bedeutung von Erfahrung und Training, Zuweisungsverhalten und anderen Faktoren sich je nach Diagnose und Behandlungsmethode unterscheiden.

Eine breite systematische Literaturrecherche von Halm et al. (2002) beleuchtet die Studien zum Zusammenhang zwischen Größe und Qualität, die bis 2000 veröffentlicht wurden, kritisch. Die Literaturrecherche erstreckt sich über 272 gesichtete Artikel und 137 inkludierte Studien für 27 Behandlungen und Krankheitsbilder, die im Zeitraum von 1980 bis 2000 publiziert wurden. Die bewerteten Studien untersuchten den Zusammenhang zwischen der Größe der Krankenanstalt bzw. der Fallzahl des Arztes und der Ergebnisqualität. 71 % aller Studien zu Spitalgröße und 69 % der Studien zu Fallzahl pro Arzt zeigten einen statistisch signifikanten Zusammenhang mit besserer Ergebnisqualität. Allerdings zeigt sich eine große Variation in der Stärke des Zusammenhangs für die unterschiedlichen Behandlungen und Krankheitsbilder. Demnach wurde ein starker Zusammenhang z. B. für die Behandlung von AIDS sowie für Pankreatektomie und Operationen bei Ösophaguskarzinom, Bauchaortenaneurysmen und kindlichen Herzfehlern festgestellt, aber nur ein schwacher für Koronarbypass, Koronarangioplastie, Karotisendarteriektomie etc. Eine mögliche Erklärung für den Zusammenhang sehen die Autoren in einem unterschiedlichen Case-Mix und unterschiedlichen Prozessen. Gerade risikoadjustierte Studien lieferten nicht immer signifikante Ergebnisse für den Zusammenhang. Die Autoren heben in ihrem Review auch hervor, dass viele Studien methodologische Defizite aufweisen. Auch Gandjour et al. (2003) untersuchen in einer systematischen Recherche Literatur von 1990 bis Ende 2000 auf den Zusammenhang zwischen Fallzahl pro Arzt bzw. Spital und Mortalität. Gandjour et al. analysierten Studien zu 34 unterschiedlichen

Diagnosen und Interventionen und stellen einen allgemeinen Zusammenhang zwischen Spitals- sowie Ärztevolumen und Ergebnisqualität fest.

Kritik an bisherigen Studien wird auch in dem Artikel „Reliability of Provider Volume and Outcome Associations for Healthcare Policy“ von Sheikh (2003) zum Ausdruck gebracht. Der Autor beleuchtet die gegenwärtige Literatur über den Zusammenhang zwischen Größe und Mortalität kritisch und warnt vor leichtfertig gezogenen Schlüssen bezüglich Richtlinien für die Patientenzuweisungen. Er unterstreicht, dass Studien, die für Zuweisungen von Patienten sprechen und als Grundlage für jede politische Entscheidung herangezogen werden, auf ihre Zuverlässigkeit und ihren wissenschaftlichen Gehalt geprüft werden müssen. Die analysierten Studien bewertet er als nicht widerspruchsfrei. Methodologisch merkt er an, dass nur prospektive, kontrollierte Studien Beweise für einen kausalen Zusammenhang liefern können. In Bezug auf die Datenqualität für Studien sei eine angemessene Risikoadjustierung klinischer Daten nötig und nicht die bloße Verwendung administrativer Daten, um die Validität zu erhöhen.

Sheikh weist zudem auch auf die Existenz von Ausreißern in Studien hin: So kann ein kleines Spital ähnliche Ergebnisse liefern wie große Krankenhäuser. Umgekehrt gibt es auch große Spitäler, die schlechtere Ergebnisse aufweisen als kleine. Dieses Erkenntnis ist aus unserer Sicht sehr wichtig: Das konkrete Spital kann sich in der Realität eben nicht wie der Durchschnitt verhalten, und eine Bewertung des Einzelfalls ist bei Entscheidungen notwendig.

Auch bezüglich der Bewertung der Fähigkeiten des Arztes stellt Sheikh Schwierigkeiten fest. So kann z. B. ein Arzt in mehreren Spitälern tätig sein und seine Erfahrung korrespondiert demnach nicht mehr mit der Größe der Einheit. Sheikh kritisiert auch die Anwendung von dichotomen Maßen für die Größe der Einheiten. In Studien, welche nicht nur zwischen „kleinen“ und „großen“ Einheiten unterscheiden, konnte kein linearer Zusammenhang zwischen Größe und Ergebnis festgestellt werden. Die Maße für die Größe eines Spitals innerhalb der Literatur sind sehr heterogen. Er kritisiert, dass Wissenschaftler oftmals einen willkürlichen Trennpunkt zwischen „groß“ und „klein“ festlegen, der zu den Ergebnissen ihrer Studie passt. Im Ergebnis kommt Sheikh zum Schluss, dass die Größe des


© Helder Almeida - Fotolia.com

Spitals ein ungenaues, indirektes und unzuverlässiges Maß für Qualität ist, da der Zusammenhang weder nachvollziehbar noch konsistent und oft nur schwach ausgeprägt ist.

In Hinblick auf eine Politik der absoluten Konzentration von Spezialisierungen weist Sheikh schließlich auf konkrete Probleme hin, die unter anderem aus einer Einschränkung des Wettbewerbs resultieren. Durch eine höhere Auslastung größerer Einheiten könnte es zudem auch zu einer Überbeanspruchung der gegebenen Ressourcen kommen und letztlich zur Verschlechterung der Ergebnisse für die Patienten. Es muss auch berücksichtigt werden, dass kleine Spitäler höhere Erreichbarkeit bieten. Schließlich bestehen auch andere Möglichkeiten, die Qualität von Spitälern zu erhöhen. Hierbei sind Telemedizin, die Zusammenarbeit von erfahrenen mit weniger erfahrenen Ärzten und Projekte zur Qualitätssteigerung durch Wissenstransfer zu nennen. In diesem Sinne können auch Kooperationen von kleineren und größeren Spitälern die gewünschten Spill-overs erzeugen.

Eine weitere interessante Studie über den Zusammenhang zwischen Operationsvolumina und Mortalität kommt von Urbach und Baxter (2004). Die Autoren werfen die Frage auf, ob die höhere Qualität großer Spitäler hinsichtlich einer Behandlung auch von der Fallzahl anderer Behandlungen beeinflusst wird (Spill-over-Effekt). Die kanadische Untersuchung basiert auf sekundären Daten für fünf unterschiedliche Operationen, die zwischen 1994 und 1999 erhoben wurden. Die Spitalsgröße wurde in zwei Kategorien geteilt. Analysiert wurde die Ergebnisqualität anhand von Mortalitätsraten 30 Tage nach der Operation. In der Analyse wurden Daten zu 31.632 Patienten untersucht und für Geschlecht, Alter und Begleitkrankheiten adjustiert.

Die Analyse ergibt, dass die kurzfristigen Mortalitätsraten bei komplexen Operationen in größeren Spitälern, die eine höhere Anzahl für diese spezifische Operation aufweisen, niedriger sind. Die Ergebnisse sind aber auch in vielen Fällen besser für Spitäler, die eine große Fallzahl anderer Operationen aufweisen. Die Anzahl der jeweiligen Operationen eines Spitals, aber auch die Anzahl an anderen Operationen wirken sich negativ auf die Mortalität (30 Tage nach der Operation) aus. Urbach und Baxter zeigen auch, dass der Zusammenhang zwischen der Anzahl einer anderen Operation in manchen Fällen stärker sein kann als jener mit der gleichen.<sup>4</sup> Die Autoren schließen daraus, dass die bessere Qualität großer Spitäler eventuell auf bestimmten strukturellen Charakteristika beruht, wie dem Status als Universitätsklinik, dem Standort oder der Möglichkeit, auf gewisse Spezialisierungen zurückgreifen zu können. Gemeinsam genutzte Strukturen und Abläufe in großen Spitälern könnten laut Urbach und Baxter der Grund für die besseren Ergebnisse großer Spitäler sein.

Eine breite Übersicht über Forschungsergebnisse bis 2005 zu unterschiedlichen Bereichen wurde von Diavoli et al. (2005) erstellt. Die Autoren geben in ihrem Artikel einen weiten Überblick über verschiedene Fachgebiete, Indikatoren und unterschiedliche Behandlungsmethoden, die in Zusammenhang mit der Größe-Qualitätshypothese untersucht wurden. Die Autoren beziehen 11 von 21 unterschiedlichen systematischen Literaturrecherchen aus dem Zeitraum 1995 bis Februar 2005 in ihre Untersuchung mit ein, die insgesamt 46 Bereiche umfasst. Davoli et al. zeigen, dass für 11 Bereiche, nämlich Operationen bei Bauchortenaneurysma, PTCA, Kniearthroplastie, Bypass, Operationen von Ösophagus- und Pankreaskarzinom, Karotisendarterektomie, Herzinfarkt und neonatologische Intensivpflege ein starker Zusammenhang zwischen Größe und Ergebnisqualität (Mortalität während des Spitalsaufenthalts) festgestellt wurde. Sie weisen aber darauf hin, dass für keinen der Bereiche Schwellenwerte für die Größendefinition bestimmt werden konnten. Parallel dazu weisen die Autoren darauf hin, dass für eine Reihe anderer Bereiche keine konsistenten Ergebnisse vorliegen. Daher raten sie zu einem vorsichtigen Umgang mit den Resultaten, da die Qualität vieler der Studien Mängel aufweist. Darüber hinaus mutmaßen Davoli et al. auch über mögliche Publikationsverzerrungen, da kaum Studien mit negativen Ergebnissen existieren.

Eine weitere Literaturübersicht zu Untersuchungen über den Zusammenhang zwischen Größe und Qualität kommt von Murray und Teasdale (2006). Der

<sup>4</sup> Es lohnt sich z. B. aufgrund der anderen Operation in die technische Ausstattung zu investieren, die dann auch bei anderen eingesetzt wird.



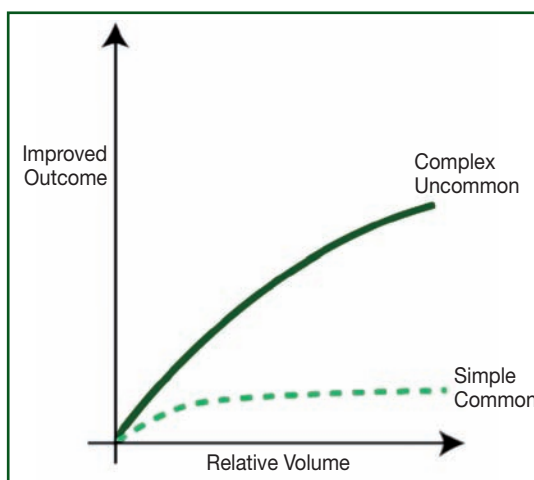


Artikel erfasst wissenschaftliche Ergebnisse von 1997 bis 2005. Der Umfang der Literaturstudie umfasst 500 Abstracts und 50 detailliert bewertete Publikationen und setzt auf einer bereits 1996 durchgeführten systematischen Literaturrecherche über den Zusammenhang von Größe und Qualität auf. Die Autoren fassen die Ergebnisse der Vorgängerstudie in drei Punkten kurz zusammen. Da die meisten Studien Unterschiede im Case-Mix zu gering berücksichtigen, werde die Wirkung von Größe auf Qualität wahrscheinlich überschätzt und auch die Kausalitätsrichtung sei (wie oben schon diskutiert) unklar. Bis 1996 konnten überdies keine Studien einen Qualitätsgewinn von bestimmten Einheiten über die Zeit (und damit bei Volumenänderung) bestätigen. Der letzte Punkt bezieht sich auf die Schwellenwerte. Die wissenschaftlichen Ergebnisse der York-Recherche deuten darauf hin, dass sofern es Qualitätsgewinne durch Größe gibt – sich diese bereits auf einem relativ niedrigen Niveau erschöpfen.

Murray und Teasdale kritisieren dass reine Mortalitätsraten für viele Behandlungen nicht mehr als adäquates Maß zur Beurteilung der Qualität herangezogen werden sollten, sondern risikoadjustierte Qualitätsindikatoren. Insgesamt stellen sie fest, dass die methodologische Qualität seit dem letzten Review zugenommen hat. Sie unterstreichen, dass Studien, die für den Case-Mix mit Hilfe von klinischen Patientendaten korrelieren, die verlässlichsten Ergebnisse hervorbringen. Das Ergebnis der Literaturstudie bestätigt den Zusammenhang zwischen Größe und Qualität insofern, als größere Fallzahlen bei Chirurgen und große Spitäler zu höherer Qualität tendieren als kleinere Einheiten. Der kausale Zusammenhang kann daraus aber noch nicht abgelesen werden. Bezüglich der Analyse von Lernkurven für Operationen kommen die bewerteten Studien zu keinem klaren Ergebnis. Murray und Teasdale verweisen auf das Problem, dass gewonnene Erfahrung auch zur Anwendung der Kenntnisse bei Patienten mit höheren Risiken führt und sich das Ergebnis demnach nicht zwingend verbessert. Aus diesem Grund sei die penible Analyse des Case-Mix für solche Studien von größter Bedeutung.

Außerdem stellen die Autoren fest, dass im Gegensatz zu Analysen, welche lediglich zwischen kleinen und großen Volumina unterscheiden, Studien, welche eine Bandbreite an Größen zulassen, zum Schluss kommen, dass mittlere Größen vergleichbare Ergebnisse wie große Einheiten liefern. Die Autoren betonen, dass es noch immer wenige Studien gibt, die den kausalen Zusammenhang zwischen Größe und Qualität untersuchen, und nennen ausgewählte Aspekte wie strukturelle Unterschiede, z. B. im Personalbestand, Unterschiede in Wei-

terbildung und Training, Kooperation zwischen Gesundheitseinrichtungen und tiefere Spezialisierung. Die Autoren kommen zu dem Schluss, dass bereits zahlreiche Studien existieren, welche den Zusammenhang zwischen Größe und Qualität für ausgewählte Behandlungen nachweisen. Der Zusammenhang ist beträchtlich bei komplexen chirurgischen Eingriffen, die mit hohem Risiko assoziiert sind. Ein mäßiger, aber klinisch relevanter Zusammenhang wurde auch für eine große Zahl von allgemeineren Behandlungen festgestellt. Bei weniger komplexen Eingriffen, so Murray und Teasdale, sei das Potenzial für die Verbesserung der Qualität durch Vergrößerung der Fallzahlen rasch erreicht.



Quelle: Murray und Teasdale (2006, S. 21)

Die Autoren stellen den Zusammenhang zwischen Größe und Qualität in der Grafik dar und merken an, dass die meisten Behandlungen zwischen diesen beiden Extremen liegen werden. Die Grafik zeigt, dass bei seltenen sowie komplexeren Eingriffen höhere Fallzahlen zu besserer Qualität führen können. Für einfachere sowie häufigere Be-

© Konstantin Sutyagin - Fotolia.com





© fuzzbones - Fotolia.com

handlungen ist der Qualitätsgewinn durch Ausweitung der Fallzahlen schnell erschöpft. Die Autoren stoßen auch in ihrer Untersuchung auf ungewöhnliche Beobachtungen bzw. Ausreißer. So gibt es kleine Spitäler, die große in ihren Ergebnissen übertreffen. Dies führt Murray und Teasdale auch zur Frage, ob die Fallzahl des Spitals oder die eines Chirurgen relevanter ist. Ihre Literaturobwohl ergibt, dass sich beide Faktoren auswirken.

Die Aussagekraft der Qualitätsmaße hat in der jüngeren Literatur zugenommen, indem z. B. postoperative Komplikationen oder spezifische klinische Indikatoren Verwendung finden. Aber auch Patientenzufriedenheit und Wahlmöglichkeiten für Patienten sollten laut Murray und Teasdale in Betracht gezogen werden. Insgesamt schließen die Autoren, dass bisherige Studien den Beweis für den kausalen Zusammenhang zwischen Größe und Qualität unzureichend erbringen. Auch eine japanische Studie von Kazui et al. (2007) bestätigt den Zusammenhang zwischen Spitalsgröße und Ergebnisqualität anhand der Analyse von 10 chirurgischen Eingriffen an Herz, Lunge und Ösophagus. Die Studie bezieht sich auf Daten zwischen 2000 und 2004. Diese wurden jedoch nicht risikoadjustiert. Mittels eines logistischen Regressionsmodells wurde der Zusammenhang zwischen kurzfristiger, operationsbezogener Mortalität und Spitalsgröße berechnet. Kazui et al. stellen einen inversen Zusammenhang für alle 10 Eingriffe fest, betonen aber, dass gerade für die Gruppe kleiner Spitäler die Mortalität eine große Streuung aufweist.

Einen Beitrag zur Analyse des Zusammenhangs bei nichtoperativen Interventionen leisten Saposnik et al. (2007) durch ihre Analyse von Behandlungen bei ischämischen Schlaganfällen. Die Autoren untersuchen kanadische Daten zu 26.676 Patienten

und 606 Krankenhäusern für den Zeitraum April 2003 bis März 2004 auf den Zusammenhang zwischen Spitalsgröße sowie Spitalstyp und Mortalität bei Schlaganfall. In der multivariaten Analyse wurde für demografische Charakteristika, Besonderheiten und Fachgebiete der Anbieter, Spitalsgröße und Clusterbildung von Beobachtungen in Einheiten adjustiert. Die Mortalität in Einheiten mit weniger als 50 Schlaganfallpatienten pro Jahr ist größer als in jenen mit großen Fallzahlen (100 bis 199 und <200). Sie zeigen, dass die 7-Tage-Mortalitätsrate in kleinen Einheiten 9,5 % im Gegensatz zu 6 % in größeren beträgt. Parallel dazu wird die Mortalität nach Entlassung mit 15,2 % bei kleinen und 12,8 % bei größeren Einheiten beziffert. Die Ergebnisse bleiben auch bei multivariater Adjustierung signifikant.

Entscheidungsträger äußern oft den Wunsch nach Schwellenwerten bezüglich der Fallzahlen, um von besserer Qualität ausgehen zu können. Die systematische Literaturrecherche von Holt et al. (2007) zu Karotisendarterektomien gehört zu den wenigen Beiträgen, die einen konkreten Schwellenwert empfehlen. Die Autoren untersuchen 25 Artikel, welche insgesamt 936.436 Fälle von Karotisendarterektomie umfassen und Daten zu postoperativer Mortalität und jährliche Fallzahlen für Karotisendarterektomien enthalten. Holt et al. berechneten den Effekt der Fallzahl auf die Mortalität mit einer Odds Ratio von 0,78. Demnach bestätigt dieses Review den inversen Zusammenhang zwischen Größe und Mortalität. Darüber hinaus errechnen Holt et al. einen Schwellenwert von 79 Karotisendarterektomien pro Jahr, welche als kritischer Wert für höhere Qualität gelten. Aus der Studie folgt, dass Krankenhäuser mindestens 79 Karotisendarterektomien durchführen sollten.

Über die Bedeutung der individuellen Leistung eines Chirurgen geben Chowdhury et al. (2007) Aufschlüsse. Die Autoren bewerten in einer systematischen Literaturrecherche 163 Artikel, die zwischen 1957 und 2002 publiziert wurden und sich auf 42 unterschiedliche Operationen beziehen. Sie untersuchen diese auf den Zusammenhang zwischen dem Operationsvolumen einer Krankenanstalt bzw. eines Chirurgen sowie seiner Spezialisierung mit der Ergebnisqualität. Die Qualität wird in den gesichteten Studien anhand der Länge des Spitalsaufenthalts, der Mortalität oder der Komplikationsrate gemessen. Chowdhury et al. betonen, dass die Hypothese, Größe führe zu Qualität, meist durch retrospektive Studien bestätigt wird. Die Autoren bemängeln jedoch die schlechte Qualität der Daten vieler (vor allem retrospektiver) Studien. Große Spitäler lieferten zwar in 74,2 % der Studien signifikant bessere Ergebnisse,



Große Spitäler lieferten zwar in 74,2 % der Studien signifikant bessere Ergebnisse, dieser Zusammenhang wird jedoch nur in 40 % der prospektiven Studien festgestellt.

dieser Zusammenhang wird jedoch nur in 40 % der prospektiven Studien festgestellt. Die Autoren merken auch an, dass die Studien unterschiedliche Resultate für bestimmte Krankenhausgrößen liefern. Bezüglich der Fallzahl eines Chirurgen stellen die Autoren fest, dass Chirurgen mit hohen Fallzahlen in 74 % der Studien bessere Ergebnisse liefern. Sie unterstreichen, dass im Gegensatz zur Spitalsgröße die Fallzahl eines Chirurgen sowohl in prospektiven als auch in retrospektiven Studien mit vergleichsweise besseren Ergebnissen zusammenhängt. 91 % der Studien bestätigen auch den Zusammenhang zwischen der Spezialisierung des Chirurgen und dem besserem Ergebnis. Allerdings ist dieser Zusammenhang nicht für alle Spezialisierungen gleich stark. Alles in allem steht eine hohe Fallzahl eines Chirurgen sowie, unabhängig davon, die Spezialisierung des Chirurgen in signifikantem Zusammenhang mit besseren Ergebnissen für ihre Patienten.

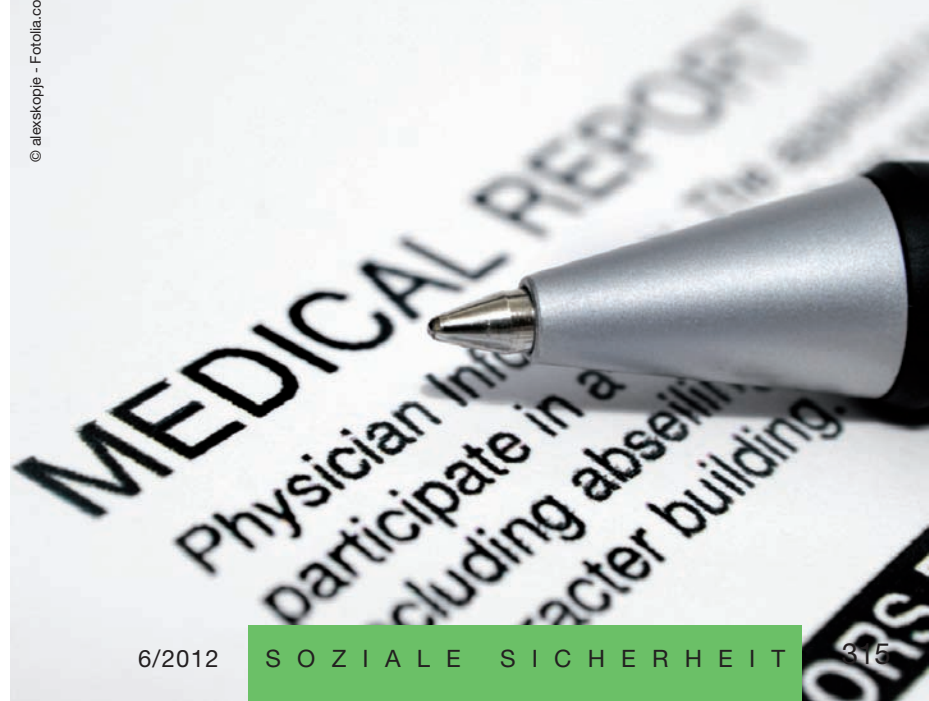
Auch Chowdhury et al. merken an, dass die Grenze zwischen großen und kleinen Fallzahlen für Operationen in den Studien sehr unterschiedlich und oft arbiträr gezogen wird. Bezüglich der Schwellwerte für eine hohe Anzahl von Operationen argumentieren die Autoren gegen einen pauschalen Wert und eher dafür, dass ein solcher für jeden Eingriff separat festgelegt werden müsse. Die Autoren merken an, dass es aufgrund von Zuweisungen von und zu Spitälern bzw. auch zwischen unterschiedlichen Abteilungen eines Spitals zu Verzerrungen der Ergebnisse kommen kann. Außerdem sehen sie in der Mortalität ein schlechtes Maß für Eingriffe mit niedrigem Risiko. Aber auch für Interventionen mit hohem Risiko müssen Überlebensraten kein Maß für Qualität darstellen, dies gilt z. B. bei orthopädischen Operationen.

Wie der Name „Hospital volume does not affect clinical outcome“ (Vetthus und Søndena 2007) schon verrät, nimmt der 2007 veröffentlichte Artikel von Vetthus und Søndena einen kritischen Blickwinkel auf die Literatur zu Größe und Qualität ein. Die Autoren rekapitulieren und erweitern Probleme der Studien zu Größe und Qualität. Primär bemängeln sie die schlechte Qualität der verwendeten Daten vieler Studien. So seien klinische Daten administrativen vorzuziehen. Sie bezweifeln auch, dass rein administrative Daten Aufschlüsse über Qualität geben können. Wie auch andere Autoren bestätigen Vetthus und Søndena, dass der Gebrauch von Mortalitätsraten für viele Behandlungen ein ungeeignetes Ergebnismaß ist. Sie er-

kennen auch Schwierigkeiten bei der Beurteilung von chirurgischen Fähigkeiten. Naturgemäß ist die Fertigkeit eines einzelnen Chirurgen von der Erfahrung, den persönlichen Fähigkeiten und dem Level an Weiterbildung abhängig. Die Veröffentlichung von Ergebnissen über die Erfahrung bzw. Qualität von Chirurgen kann sich zusätzlich als heikel erweisen, betonen die Autoren. Sie sprechen auch den Konflikt über die Verteilung von Operationen auf junge Ärzte, die Erfahrung sammeln sollen, und erfahrene qualifizierte Chirurgen an. Eine Möglichkeit, um Chirurgen zu höheren Fallzahlen zu verhelfen, sehen sie darin, Ärzte an unterschiedlichen Krankenanstalten praktizieren zu lassen. Zweifel äußern die Autoren auch an der Objektivität mancher Wissenschaftler, da sich die meisten Studien auf große Institutionen stützen und Autoren oftmals dort angestellt seien. Zusätzlich nehmen Vetthus und Søndena an, dass ein Publikationsbias bezüglich der Ergebnisse der Studien vorliegt, da negative Resultate oft nicht publiziert würden.

Hogan und Winter (2008) nehmen in ihrem Artikel eine kritische Analyse der Literatur bezüglich des Zusammenhangs zwischen Größe und klinischem Ergebnis bei Krebsbehandlungen vor. Sie kommen nach Durchsicht der Literatur zum Ergebnis, dass sich der kausale Zusammenhang zwischen Größe und Qualität nicht herauskristallisiert. So werden die Größe der Einheit, chirurgische Fähigkeiten und der Einsatz von adäquaten Begleittherapien als mögliche Gründe für die besseren Ergebnisse größerer Spitäler genannt. Die Bewertung der Erfahrung von Chirurgen sei schwierig, da Vergleichbar-

© alexskopje - Fotolia.com





© carmakoma - Fotolia.com

keit und Vollständigkeit der Daten hierzu oft nicht gegeben seien. So stellt sich auch die Frage, wie langjährige Erfahrung gegenüber der aktuellen jährlichen Operationsanzahl eines Chirurgen gewichtet werden soll? Darüber hinaus eröffnet sich die Frage, welche Implikationen das Wissen über qualitative Unterschiede der Chirurgen haben kann. Die Autoren weisen darauf hin, dass abgesehen von der individuellen Leistung des Chirurgen auch der multidisziplinäre Zugang in der Krebsbehandlung als Grund für bessere Ergebnisse größerer Einheiten genannt wird. Das Angebot von Begleittherapien scheint ebenso einen Beitrag zu guten Ergebnissen zu leisten. Hogan und Winter stellen den Zusammenhang zwischen Größe und Qualität grafisch dar. In der Grafik spiegeln Hogan und Winter links den traditionellen, linearen Zusammenhang zwischen Größe und Qualität wider. Die rechte Grafik stellt die Beziehung zwischen Größe und Qualität als Fläche unter einer Kurve dar.

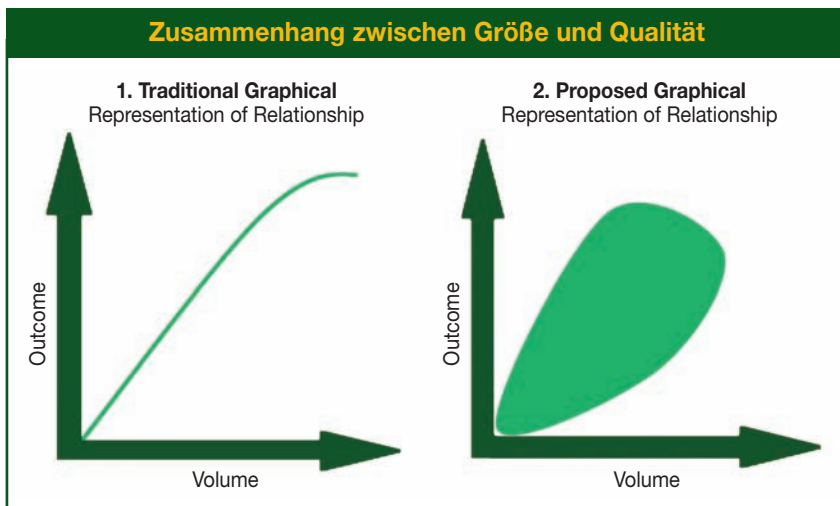
Zuletzt weisen Hogan und Winter darauf hin, dass Ergebnisse aus amerikanischen Krebs-Studien nicht als Grundlage für europäische Strategien angewendet werden können, da die Unterschiede bezüglich des Gesundheitssystems sowie der Bevölkerung zu groß seien.

Sehr viele Studien beziehen sich auf chirurgische Eingriffe. Hellinger (2008) leistet einen Beitrag, um auch nichtchirurgische Behandlungsmethoden bezüglich des Zusammenhangs zwischen Größe und Qualität zu evaluieren. In der Studie werden mittels einer multivariaten logistischen Regression die Wirkung der Spitalsgröße sowie des Behandlungsvolumens pro Arzt auf die Mortalität bei HIV-positiven Patienten untersucht. Die US-Studie bezieht sich auf Daten zu 43.325 Patienten, die im Jahr 2002 mit der Diagnose HIV in ein Krankenhaus eingewiesen wurden. Hellinger berechnet, dass nur das Behandlungsvolumen pro Arzt einen signifikanten Erklärungsbeitrag zur Mortalität der Patienten liefert.

Kulkarni et al. (2009) untersuchen in ihrem Artikel vier unterschiedliche Berechnungsmethoden für die Spitalsgröße und deren Wirkung auf die kurzfristige Ergebnisqualität. Dabei nehmen sie auch Rücksicht auf Umstrukturierungen von Spitälern während des Untersuchungszeitraums 1992 bis 2004.

Als Grundlage für die Studie wurden Kohorten von Myokardinfarktpatienten in Ontario (Kanada) herangezogen. Die Bootstrap-Technik wurde angewandt, um Odds Ratios der unterschiedlichen Größendefinitionen zu vergleichen. Kulkarni et al. zeigen, dass alle untersuchten Definitionen statistisch signifikante Ergebnisse für den Zusammenhang zwischen Größe und Qualität liefern und dass die Umstrukturierungen eines Spitals keine Auswirkungen auf diesen haben. Die Autoren weisen aber auf die signifikanten Unterschiede der Odds Ratios der einzelnen Berechnungsmethoden für die Spitalsgröße hin. Auch die Ergebnisse von Kulkarni et al. bestätigen den Zusammenhang zwischen Größe und Qualität. Die bedeutenden Unterschiede in den Odds Ratios zeigen, so Kulkarni et al., auf, dass nur sehr genaue Berechnungsmethoden für die Spitalsgröße bei Entscheidungen basierend auf den Odds Ratios angewandt werden sollen.

Gruen et al. (2009) arbeiten Ergebnisse zu chirurgischen Krebsbehandlungen in einer systematischen Literaturrecherche auf. Diese untersucht 101 Studien aus 12 Ländern auf den Zusammenhang zwischen Fallzahl und Mortalität für folgende Krebsoperationen: Ösophagus, Magen, Leber, Pankreas, Kolon, Rektum. Die Recherche bestätigt den signifikanten Zusammenhang zwischen der Fallzahl des Spitals und kurzfristiger Mortalität. Dieser



Quelle: (Hogan, Winter 2008, S.1269)

ist aber nicht sehr robust. Die Autoren geben an, dass ein Drittel der Studien keinen signifikanten Zusammenhang bei kurz- und langfristiger Mortalität feststellt. Folglich stellen Gruen et al. die Aussagekraft von Fallzahlen als Proxy für Qualität in Frage. In einer zusätzlichen Analyse für gastrointestinalen Krebs stellen die Autoren ebenfalls einen signifikanten Zusammenhang zwischen Größe und Operationszahl fest. Sie berechnen, dass bei einer Verdoppelung der Fallzahl des Spitals die Odds für perioperativen Tod um 0,1 sinken. Die Autoren führen weiters aufbauend auf der Literaturrecherche eine Meta-Studie durch, um die Stärke des Zusammenhangs zwischen Spitalsgröße und Mortalität sowie einen Schwellenwert zu berechnen. Dabei analysieren sie rohe, nicht risikoadjustierte Daten. Der Zusammenhang erweist sich als konsistent für alle Arten von Krebs, mit Ausnahme des Rektumkarzinoms. Schwellenwerte konnten jedoch nicht identifiziert werden. Die Autoren weisen darauf hin, dass risikoadjustierte Daten in der Regel einen schwächeren Zusammenhang aufzeigen.

Eine methodisch gute Studie über Qualität bei Brustkrebsoperationen in Finnland wurde von Peltoniemi et al. (2011) durchgeführt. Die Autoren untersuchten, ob die jährliche Fallzahl einer Einheit Auswirkungen auf die Anzahl von Brustkrebsoperationen, das erneute Auftreten von Brustkrebs sowie die Überlebensraten hat. Die Studie umfasst Daten zu 12.604 Frauen, bei denen in Finnland zwischen 1998 und 2001 erstmalig Brustkrebs diagnostiziert wurden und deren Fälle bis 2008 weiterverfolgt wurde. Die chirurgischen Einheiten wurden in vier Untergruppen nach der Anzahl der jährlichen Operationen (> 200, 100–200, 99–50, < 50) unterteilt. Finnland hat ein homogenes Gesundheitssystem und Brustkrebspatienten werden nach nationalen Richtlinien behandelt. Es wurde eine Risikoadjustierung für Alter und Krankheitsstadium zum Zeitpunkt der Diagnose vorgenommen und die Ergebnisse wurden anhand von Cox-Regressionsmodellen sowie logistischen Regressionsmodellen berechnet. Die Studie ergab, dass Patientinnen in Spitälern mit niedrigeren Fallzahlen häufiger einer Mastektomie unterzogen wurden. Auch bezüglich der Häufigkeit von Strahlentherapie für Patientinnen, die eine brusterhaltende Operation erhielten, wurden signifikante Unterschiede festgestellt. Die Autoren führen dies auf das geringfügig höhere Alter der Patientinnen in kleineren Spitälern zurück. Die Studie ergab, dass die Daten zu Rezidiven nicht verlässlich waren, sie wurden daher nicht weiter analysiert. Jedoch spiegelten sich größere Fallzahlen auch in höheren Überlebensraten wider. Peltoniemi et al. berechneten die Überlebensraten für die Summe aller Pa-

tientinnen und stellten fest, dass die Überlebensraten in größeren Zentren signifikant höher waren. Die Überlebensraten betragen 85,3 % in den größten Einheiten gegenüber 81,6 % in den kleinsten Spitälern ( $P = 0,031$ ). Für Patientinnen, an denen brusterhaltende Operationen durchgeführt wurden, wurden keine signifikanten Unterschiede der Überlebensraten zwischen den Einheiten beobachtet. Die von Peltoniemi et al. durchgeführte Studie bestätigt die Hypothese, dass größere Fallzahlen der Spitäler zu besseren Ergebnissen in der Krebsbehandlung führen. Die Studie zeigt höhere Überlebensraten bei Brustkrebs für größere Spitalseinheiten.

### Sollen Spitäler nun groß oder klein sein?

Eine Literaturübersicht kann in diesem Bereich notwendigerweise nicht erschöpfend sein, da mittlerweile eine sehr große Zahl von Studien vorliegt. Sie zeigt jedoch eines sehr deutlich: Schlussfolgerungen müssen sehr differenziert gezogen werden und gerade im Hinblick auf die politischen Implikationen ist eine Vereinfachung wie sie in der Diskussion immer wieder gerne vorgenommen wird gerade nicht hilfreich.

Sollen Spitäler also groß oder klein sein? Die Antwort auf diese Frage lautet: Sie sollen effizient sein und gute Behandlungsqualität ermöglichen! Denn darum geht es, und die Frage nach der Größe ist nur eine der vielen Determinanten, die es in der tatsächlichen Entscheidung zu erwägen gilt. Insbesondere muss bedacht werden, dass der statistische Zusammenhang nicht notwendigerweise für die konkrete Einheit gelten muss. Statistische Verfahren stellen in der Regel Mittelwerte dar, bei denen aber auch die Streuung Bedeutung hat. So ist keineswegs ausgeschlossen, dass ein konkretes klei-




© Bergingfoto - Fotolia.com

neres Spital effizient ist oder auf hohem Qualitätsniveau arbeitet. Im Hinblick auf diese Erkenntnis wäre es sehr wichtig, konkretes Benchmarking zu verwenden, anstatt Entscheidungen über Standorte am durchschnittlichen Zusammenhang festzumachen. Dadurch kann sogar volkswirtschaftlicher Schaden entstehen, weil effiziente und qualitätsvolle Einheiten zerstört werden. In diesem Sinne wäre die Antwort auf die Frage auch so zu interpretieren: Schließe nicht ein Spital, nur weil es klein ist, sondern schließe es höchstens dann, wenn es eben nicht effizient arbeitet oder nicht ausreichende Qualität bietet. Will ich hingegen ein neues Spital bauen, so kann es sinnvoll sein – unter anderen Kriterien – auch jenes der optimalen Größe heranzuziehen.

Die Literatur in Hinblick auf Größe und Effizienz gibt bei genauerer Betrachtung jedoch viele wichtige Informationen. Erstens sind die Zielfunktionen eines Spitals divers und müssen in Einklang gebracht werden, wobei sie leider oft Zielkonflikten bzw. Trade-offs unterliegen. So kann die Größe eines Spitalsstandortes nicht beliebig gewählt werden, wenn z. B. regionales Angebot oder Lehre eine Rolle spielen sollen. Umgekehrt kann für eine konkrete Versorgungssituation auch eine alternative Versorgungsform sinnvoller sein, weshalb die Forderung nach mehr Flexibilität in den gesetzlichen Vorgaben auch so wichtig ist.

Zweitens zeigt sich theoretisch wie empirisch, dass Spitäler nicht nur „zu klein“, sondern auch „zu groß“ sein können. Wie bei jeder wirtschaftlichen Einheit erschöpfen sich positive Skaleneffekte und werden schließlich durch Skaleneffizienz abgelöst.

Drittens zeigt sich in vielen Studien, dass Spitäler oft nicht nur an Skaleneffizienz leiden, sondern

dass das größere Potenzial in technischer Ineffizienz steckt. Mit anderen Worten: Es ist schon viel zu holen, wenn die Ressourcennutzung bei gegebener Größe verbessert wird, indem z. B. Abläufe optimiert werden. Dazu müssten aber erst Best-Practice-Beispiele gefunden, analysiert und übertragen werden. Oder um vielleicht noch grundsätzlicher anzusetzen: Es muss schon die Bereitschaft vorhanden sein, diesen Weg zu gehen.

Viertens sind die Skaleneffekte sicher auch von der Zusammensetzung eines Spitals abhängig. Bei bestimmten Abteilungen und Disziplinen können die Skaleneffekte wesentlich bedeutender sein als bei anderen. Analysen auf Abteilungsebene sind aber rar und es würde sich lohnen, die Wissensbasis diesbezüglich zu verbessern.

Fünftens sind auch Verbundeffekte von großer Bedeutung. So kann die kluge Ausrichtung einzelner Häuser in einem Spitalsverbund bloße Größeneffekte dominieren und ebenfalls viel zur Effizienzsteigerung beitragen. Auch hier besteht noch Bedarf an weiterer Forschung.

Das bloße Vorliegen von Trade-offs anderer Ziele mit der Größe entledigt einen also keineswegs der Aufgabe, trotz vielleicht gegebener Größe eine Reihe anderer effizienzsteigernder Maßnahmen zu ergreifen. Das wiederum kann aber nur gelingen, wenn der Mitteleinsatz und Prozesse im Spital auch analysiert werden und Bereitschaft zu Änderungen besteht. Was den Zusammenhang zwischen Größe und Qualität betrifft, so gilt vieles, was auch schon zur Effizienz gesagt wurde. Es ist besser, die Qualität zu messen und in einem Benchmark konkretes Continuous Quality Improvement zu betreiben, anstatt von vornherein einem statistischen Zusammenhang zu vertrauen. Nur, wenn ich „auf der grünen Wiese“ starte, werde ich ebenfalls von vorne herein ein gewisse Mindestgröße bevorzugen, weil ich dann mit höherer Wahrscheinlichkeit „richtig“ liege.

Zudem zeigt die Literatur, dass der Zusammenhang zwischen Größe und Qualität sehr stark von der Art der Intervention abhängt. Insofern erscheint es gerechtfertigt, Volumenindikatoren für bestimmte Interventionen zu verwenden. Darin darf sich aber die Qualitätsarbeit dann nicht erschöpfen.

Jedenfalls wird es sich lohnen, wenn nichts dagegen spricht, Spezialisierungen vorzunehmen und Schwerpunkte zu bilden. Dadurch müssen sich gar nicht so sehr direkte Effekte ergeben, doch ist es in diesem Fall leichter, Qualitätsmanagement ernsthaft durchzuführen, Fortbildungsmaßnahmen zu setzen und neue Prozesse nach dem Stand der medizinischen Wissenschaft zu implementieren. Auch hier ist aber noch nicht klar, ob dies zwingend einen einheitlichen Standort voraussetzt oder auch



in Zusammenarbeit geschehen kann. Die Frage von Größe und Qualität bzw. Effizienz im Spitalwesen ist ein Paradebeispiel dafür, wie komplex die Beantwortung konkreter gesundheitspolitischer Fragen aus der wissenschaftlichen Literatur ist und wie kontraproduktiv Polemik oder aus Halbwissen generierte Ratschläge sein können. Insgesamt müssen in der Anwendung wissenschaftlicher Erkenntnisse gerade auf konkrete politische Entscheidungen auch die konkreten Einschränkungen in der Wissensge-

nerierung einerseits, aber auch die realen Verhältnisse und ihre Vorgaben andererseits beachtet werden. Die Zeit für aktives Handeln in Hinblick auf die Umstrukturierung unseres Gesundheitswesens für die Zukunft wird gerade im Hinblick auf die mäßigen Wirtschaftsaussichten immer knapper. Eine sachliche und evidenzgestützte Diskussion ist vonnöten, insbesondere, wenn nun die Suche nach den 1,8 Mrd. Euro an Kostendämpfung im Spitalwesen beginnt.

## LITERATUR

- Berry, R. E., Jr. (1967): Returns to Scale in the Production of Hospital Services. *Health Services Research*, 2(2), S. 123–139.
- Chowdhury, M., Dagash, H., Pierro, A. (2007): A systematic review of the impact of volume of surgery and specialization on patient outcome. *British Journal of Surgery*, 94, S. 145–161.
- Davoli, M., Amato, L., Minozzi, S., Bargagli, A., Vecchi, S., Perucci, C. (2005): Volume and health outcomes: an overview of systematic reviews. *Epidemiol Prev.*, 29 (3-4), S. 3–63.
- Dranove, D. (1998): Economies of Scale in Non-Revenue Producing Cost Centers: Implications for Hospital Mergers. *Journal of Health Economics*, 17 (1), S. 69–83.
- Ferrier G. D., Valdmanis, V. G. (2004): Do mergers improve hospital productivity? *Journal of the Operational Research Society* 55 (10). S. 1071–1080.
- Gandjour, A., Bannenberg, A., Lauterbach, K. (2003): Threshold volumes associated with higher survival in health care: a systematic review. *Medical Care*, 41 (10), S. 1129–1141.
- Gruen, R., Pitt, V., Green, S. et al. (2009): The Effect of Provider Case Volume on Cancer Mortality: Systematic Review and Meta-analysis. *Cancer Journal for Clinicians*, 59 (3), S. 192–211.
- Halm, E., Lee, C., Chassin, M. R. (2002): Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Annals of Internal Medicine*, 137 (6), S. 511–520.
- Harris, J., Ozgen, H., Ozcan, Y. (2000): Do Mergers Enhance the Performance of Hospital Efficiency? *The Journal of the Operational Research Society*, 51 (7). S. 801–811.
- Hellinger, F. (2008): Practice makes perfect: a volume-outcome study of hospital patients with HIV disease. *Journal of Acquired Immune Deficiency Syndromes*, 47 (2), S. 226–233.
- Hogan, A., Winter, D. (2008): Does Practice Make Perfect? *Annals of Surgical Oncology*, 15 (5), S. 1267–1270.
- Holt, P., Poloniecki, J., Loftus, I., Thompson, M. (2007): Meta-analysis and systematic review of the relationship between hospital volume and outcome following carotid endarterectomy. *European Journal of Vascular and Endovascular Surgery*, 33 (6), S. 645–51.
- Hughes, R., Hunt, S., Luft, H. (1987): Effects of Surgeon Volume and Hospital Volume on Quality of Care in Hospitals. *Medical Care*, 25 (6), S. 489–503.
- Kazui, T., Osada, H., Fujita, H. (2007): An attempt to analyze the relation between hospital surgical volume and clinical outcome. *General Thoracic and Cardiovascular Surgery*, 55 (12), S. 483–492.
- Kristensen, T., Bogetoft, P., Moeller, Pedersen K. (2010): Potential gains from hospital mergers in Denmark. *Health Care Management Science*, 13, S. 334–345.
- Kulkarni, G., Laupacis, A., Urbach, D., Fleshner, N., Austin, P. (2009): Varied definitions of hospital volume did not alter the conclusions of volume-outcome analyses. *Journal of Clinical Epidemiology*, 62 (4), S. 400–407.
- Luft, H., Bunker, J., Enthoven, A. (1979): Should Operations be Regionalized? The Empirical Relation between Surgical Volume and Mortality. *New England Journal of Medicine*, 301 (25), S. 1364–1369.
- Lynk, W. J. (1995): The Creation of Economic Efficiencies in Hospital Mergers. *Journal of Health Economics*, 14, S. 507–530.
- Murray, G. D., Teasdale, G. M. (2006): The relationship between Volume and Health Outcomes: A Review. *Scottish Medical Journal*, 50 (1), S. 17–22.
- Tiemann, O., Schreyögg, N., (2009): Effects of Ownership on Hospital Efficiency in Germany. *Business Research Journal, Academy of Management*, 2 (2), S. 115–145.
- Peltoniemi, P., Peltola, M., Hakulinen, T., Häkkinen, U., Pylkkänen, L., Holli, K. (2011): The Effect of Hospital Volume on the Outcome of Breast Cancer Surgery. *Annals of Surgical Oncology*, 18 (6), o. S.
- Preyra, C., Pink, G. (2006): Scale and scope efficiencies through hospital consolidations. *Journal of Health Economics, Elsevier*, 25 (6), S. 1049–1068.
- Saposnik, G., Baibergenova, A., O'Donnell, M., Hill, M., Kapral, M., Hachinski, V. (2007): Hospital volume and stroke outcome: does it matter? *Neurology*, 69 (11), S. 1142–51.
- Sheikh, K. (2003): Reliability of Provider Volume and Outcome Associations for Healthcare Policy. *Medical Care*, 41 (10), S. 1111–1117.
- Urbach, D., Baxter, N. (2004): Does it matter what a hospital is “high volume” for? Specificity of hospital volume-outcome associations for surgical procedures: analysis of administrative data. *BMJ*, 328, S. 737–740.
- Vetthus, M., Søndena, K. (2007): Hospital volume does not affect clinical outcome. *European Journal of Surgical Oncology*, 33 (9), S. 1049–1051.