Freiwilliges Benchmarking der Psychiatrischen Universitätskliniken Basel, Bern und Zürich

Methodisches Vorgehen und Erfahrungsbericht

H.-C. Kuhl, D. Krneta, I. Warnke, J. Herdt, C. Cassidy, U. von Allmen, W. Rössler

Das massgebliche Ziel von Klinikvergleichen ist, vor dem Hintergrund der Referenz anderer Kliniken Verbesserungspotenziale aufzuzeigen. Damit ein Klinikvergleich das leisten kann, ist die Wahl der Indikatoren entscheidend. Diese müssen messmethodischen Kriterien genügen und so differenzierungsstark sein, dass bestehende Unterschiede auch tatsächlich aufgezeigt werden können. Aus Sicht der Qualitätssicherung wären Indikatoren wünschenswert, die die Ableitung von ganz konkreten Veränderungsmassnahmen erlauben. Bei allen Klinikvergleichen stellt sich im Anschluss die Frage, ob und welche Konsequenzen sich daraus ergeben beziehungsweise inwiefern ein Handlungsbedarf besteht.

Einführung

In den letzten 20 Jahren ist die Zahl der externen Qualitätssicherungsprogramme und der internen Qualitätsmanagementinitiativen in allen Teilbereichen der Medizin weltweit deutlich angestiegen. Bei den externen Qualitätssicherungsprogrammen handelt es sich im Wesentlichen um Klinikvergleiche bezüglich spezifischer Qualitätsindikatoren. Das Ziel des Vergleichs mit anderen Einrichtungen ist, wichtige Hinweise für Verbesserungsmöglichkeiten für das interne Qualitätsmanagement zu gewinnen. Dahinter steht unter anderem der Kostendruck im Gesundheitswesen, der dazu zwingt, mit annähernd gleichen Ressourcen eine bessere Versorgung zu erreichen, also die Effizienz des Gesundheitswesens zu steigern. So ist in den letzten Jahren eine Vielzahl von Projekten und Programmen für Klinikvergleiche in verschiedenen medizinischen Teilgebieten entwickelt und durchgeführt worden. In Deutschland beispielsweise besteht für die Leistungserbringer die gesetzliche Verpflichtung zu externer Qualitätssicherung (vgl. auch Glattacker et al., 2007).

Auch in der Schweiz werden zunehmend Benchmarkings von Kliniken geplant beziehungsweise durchgeführt. Massgeblich ist dabei die nationale Koordinations- und Informationsstelle für Qualitätsförderung der Schweiz (KIQ, 2007). Die KIQ ist eine durch H+ und Santésuisse gegründete Gesellschaft mit dem Ziel der aktiven Förderung einer schrittweisen Einführung von Ergebnismessungen in der ganzen Schweiz, und dies in den Fachbereichen Akutsomatik, Rehabilitation und Psychiatrie. Nach dem Beginn eines gesamtschweizerischen Pilotprojekts in der Rehabilitation plant die KIQ zurzeit ein ähnliches Pilotprojekt für Klinikvergleiche in der Psychiatrie.

Beim Benchmarking von Kliniken kann bereits auf einige Erfahrungen aus dem somatischen Bereich, aber auch auf psychiatriespezifische Projekte zurückgegriffen werden. Zu nennen ist hier zum Beispiel das in Deutschland vom Bundesministerium für Gesundheit geförderte Projekt «Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung in der Psychiatrie» (Frick et al. 2003). In jüngerer Zeit liegen jedoch auch in der Schweizer Psychiatrie Erfahrungen mit Klinikvergleichen vor.

Aufgrund der Initiative der Direktionen sowie der Qualitätsbeauftragen der drei deutschschweizerischen psychiatrischen Universitätskliniken Basel, Bern und Zürich wurde die Durchführung eines Benchmarkings beschlossen. Ziel ist zum einen, Verbesserungspotenziale für jede Einrichtung aufzuzeigen, und zum anderen, insgesamt voneinander zu lernen. Das freiwillige Benchmarking wird bereits seit dem Jahre 2005 durchgeführt und enthält neben den patientenbezogenen auch wissenschaftliche und finanzbezogene Indikatoren, die vergleichend gegenübergestellt werden.

source: https://doi.org/10.7892/boris.27576 | downloaded: 12.10.2024

Es wurde vereinbart, die Ergebnisse zunächst für die internen Qualitätsverbesserungen zu verwenden und erst zu einem späteren Zeitpunkt auch extern transparent zu machen. Neben dem Benchmarking wurde die Einführung eines gemeinsamen *Critical Incident Reporting System* (CIRS) vereinbart. Ein Erfahrungsbericht zum seit 2007 laufenden CIRS wird der erste Schritt der Veröffentlichung von gemeinsamen Ergebnissen sein und ist für dieses Jahr geplant.

Der vorliegende Artikel soll vorab die bisherigen Erfahrungen mit dem psychiatrischen Klinikbenchmarking darstellen und enthält aus genannten Gründen keine konkreten Ergebnisse. Der Erfahrungsbericht soll die Konzeption, das Vorgehen sowie auch Schwierigkeiten bei der Durchführung diskutieren, da dies wichtige Hinweise für schweizweite Klinikvergleiche liefern kann und es teilweise Parallelen zu dem von der KIQ vorgestellten Konzept gibt.

Methodisches Vorgehen

Die festgelegten Indikatoren dieses freiwilligen Benchmarkings beziehen sich auf finanzielle, wissenschafts- und patientenbezogene Indikatoren. Im Folgenden werden nur die patientenbezogenen Indikatoren beschrieben, da die wissenschaftsbezogenen Kriterien eher spezifisch für Universitätskliniken sind und sich die der finanziellen Konzeption noch im Aufbau befindet.

Eine grosse Herausforderung stellt die gemeinsame Festlegung der qualitätsrelevanten Indikatoren dar. Bis anhin gibt es auch international keinen Standard für die Wahl der Indikatoren, sodass eine Vielzahl von Variablen diskutiert und verwendet wird (z.B. Hermann et al., 2006). Für aussagekräftige Ergebnisse zum Outcome sind hier die messmethodischen Gütekriterien zu beachten, die jedoch bei vielen Benchmarkings nur unzureichend untersucht beziehungsweise nicht ausreichend gesichert sind. Häufig spielen bei der Auswahl der Indikatoren auch Fragen des Dokumentations- und Erhebungsaufwands für die behandelnden Personen eine wichtige Rolle.

Das Erhebungs- und Auswertungskonzept des hier beschriebenen Klinikbenchmarkings wurde von einer multizentrischen Arbeitsgruppe entwickelt. Die drei psychiatrischen Universitätskliniken Basel, Bern und Zürich haben sich bei den patientenbezogenen Qualitätsindikatoren auf die Messung von zunächst drei Indikatoren geeinigt:

- Häufigkeit von Wiedereintritten
- Outcomemessung über die CGI (Clinical Global Impressions)
- Anteil an Fürsorgerischen Freiheitsentzügen (FFE) bezogen auf die Einwohner im Versorgungsgebiet.

Inzwischen liegt die zweite Auswertungsrunde dieses Benchmarkings vor, sodass sowohl ein Vergleich zwischen den Kliniken als auch ein zeitlicher Vergleich möglich ist, was eine Abschätzung von normalen Schwankungen erlaubt.

Datengrundlage

Um den Aufwand für die Datenerhebungen auf ein Mindestmass zu beschränken und ausserdem Doppelerhebungen in den drei Kliniken zu vermeiden, wurde das freiwillige Benchmarking so konzipiert, dass im Wesentlichen auf bereits routinemässig erhobene Daten zurückgegriffen werden konnte. Nicht zuletzt aus diesem Grund wurde für die Outcomemessung die CGI verwendet, die bereits in allen Kliniken elektronisch erfasst wurde. Der Vorteil ist, dass eine Vollerhebung aller Behandlungsfälle in das Benchmark eingeht. Alle Angaben für die Berücksichtigung der Patientenstruktur (s.u. Risikoadjustierung) wurden den standardmässig vorliegenden Daten der medizinischen Statistik (Beschreibung: Bundesamt für Statistik) entnommen. Auch die Angaben zu den zwei anderen Indikatoren – Wiedereintritten und FFE – konnten aus der medizinischen Statistik gewonnen werden. Die Auswertungen wurden aufgrund der zum Teil unterschiedlichen Strukturen der drei Kliniken zunächst auf stationäre Behandlungen begrenzt.

Beschreibung der Risikoadjustierung

Das unkritische Nebeneinanderstellen von Indikatoren für verschiedene Kliniken ohne die Berücksichtigung der möglicherweise unterschiedlichen Patientenstruktur wäre ein Vergleich von Äpfeln mit Birnen. Hätte eine Klinik zum Beispiel aufgrund von geografischen Besonderheiten eine Sturktur mit deutlich älteren Patienten, so wäre unter Umständen ein schlechteres Outcome kein Zeichen für eine geringe Qualität, sondern eine Folge der Patientenstruktur. Ziel muss es daher sein, Klinikvergleiche insofern korrekt zu gestalten, dass Unterschiede bei den Ergebnissen nicht auf der verschiedenen Patientenstruktur beruhen, sondern auch tatsächlich eine unterschiedliche und von der Klinik beeinflussbare Behandlungsqualität widerspiegeln.

In einigen somatischen Gebieten wird dies dadurch gelöst, dass Klinikvergleiche nur für relativ homogene Erkrankungsbilder, sogenannte Tracerdiagnosen, durchgeführt werden. Dies bedeutet jedoch, dass nur ein kleiner Teil der Patienten in einen Klinikvergleich eingeht. Aufgrund des weit gefächerten Spektrums psychiatrischer Störungsbilder kann ausserdem die Generalisierbarkeit der Behandlungsqualität von einzelnen Störungsbildern auf die gesamte Klinik nicht sicher vorausgesetzt werden.

Für die Auswertung des hier beschriebenen Benchmarkings der drei Universitätskliniken wurde eine regressionsanalytische Risikoadjustierung durchgeführt, die den Vorteil hat, dass alle Diagnosegruppen in den Klinikvergleich mit einbezogen werden können (vgl. dazu Iezzoni et al., 2003; Farin et al., 2005). Die gemeinsame Auswertung wurde vom Arbeitsbereich Versorgungsforschung und Qualitätssicherung der UPK Basel durchgeführt. Folgende Patientenmerkmale wurden bei der Risikoadjustierung statistisch kontrolliert: Alter, Geschlecht,

Schweregrad bei Aufnahme, Diagnosegruppe, Ausbildungsstand, Erwerbstätigkeit, Nationalität und Aufenthaltsdauer (aufgrund der schiefen Verteilung logarithmiert). Da alle diese Variablen für alle Behandlungen aus den drei Kliniken vorliegen, wurden auch diese Parameter klinikvergleichend ausgewertet.

Bewertung der Indikatoren

Outcome: Clinical Global Impressions (CGI)

Der grosse Vorteil der CGI ist die sehr hohe Effizienz dieser arztseitigen Outcomebeurteilung. Beim Eintritt muss der Schweregrad der psychischen Erkrankung lediglich mit einem siebenstufigen Item erfasst werden, beim Austritt wird eine Gesamtbeurteilung der Zustandsänderung sowie eine Einschätzung des Schweregrads zum Austritt mit jeweils einem siebenstufigen Item erfragt. Diese hohe Effizienz der Outcomemessung und der geringe Dokumentationsaufwand dürften hauptverantwortlich sein für die hohe Akzeptanz bei den behandelnden Ärzten. Die CGI ist ein Teil der Empfehlung zur psychiatrischen Basisdokumentation der DGPPN (Cording et al., 1995).

Der geringe Umfang der CGI bringt jedoch aus messmethodischer Sicht einen gravierenden Nachteil mit sich. Die Einstufung des psychischen Gesundheitszustands mit einem Item lässt kaum wirkliche Differenzierungen zu. Dabei wirkt sich besonders ungünstig aus, dass von den sieben Antwortkategorien nur drei bis vier im stationären psychiatrischen Routinebetrieb relevant sind. Bei den Auswertungen des Benchmarkings verteilen sich 94,2 Prozent der Beurteilungen zum Eintritt auf nur drei Kategorien der siebenstufigen Skala (Patient ist «mässig», «deutlich» und «schwer» krank). Auch bei der direkten Beurteilung der Veränderung zum Austritt verteilen sich 90 Prozent der Einschätzungen auf drei Antwortkategorien (Zustand ist «sehr viel besser», «viel besser» und «nur wenig besser»). Drei Kategorien zu Zustandsverschlechterungen sind bei stationären psychiatrischen Behandlungen kaum sinnvoll.

Die CGI zeigt in mehreren Studien eine erfreulich hohe Beurteilerübereinstimmung (z.B. Haro et al., 2003). Dies kann jedoch allein aus methodischen Gründen mit grosser Wahrscheinlichkeit auf diese geringe Differenzierungsfähigkeit der CGI zurückgeführt werden. Bestätigt wird dies dadurch, dass bei der Beurteilung des Schweregrads der Erkrankung beim Eintritt und Austritt die Mittelwerte der drei Universitätskliniken äusserst eng zusammen liegen. Aus messmethodischer Sicht ist jedoch schwerwiegender, dass es bei den Beurteilungen kaum Varianz gibt. Das bedeutet, dass auch die Outcomebestimmung als Differenzwert von Eintritt und Austritt nur geringe Varianz aufweist. Dies führt unter anderem zu unrealistisch hohen Effektstärken der CGI (SES: 1,11) und damit zu einer zu optimistischen Einschätzung der Veränderungssensitivität.

Wie wirken sich diese messmethodischen Probleme beim Einsatz in einem Benchmarking aus? Insgesamt kann man sagen, dass sich schon allein aufgrund dieser schlechten Skaleneigenschaften der CGI kaum relevante Unterschiede zwischen Kliniken aufzeigen lassen. Für eine solche Zielsetzung wäre ein wirklich quantitatives Mass notwendig und nicht eine nur ordinal skalierte Beurteilung.

Die Risikoadjustierung beziehungsweise das Einbeziehen des Patienten-Case-Mix kann regressionsanalytisch 14 Prozent der Varianz der Outcomemessung aufklären. Es zeigt sich also, dass es durchaus notwendig ist, diese oben aufgelisteten Patientenvariablen mit in die Analyse einzubeziehen. Der Einbezug dieser Kontrollvariablen führt zu einer höheren Differenzierung der patientenbezogenen Werte, kann jedoch die oben beschriebenen Schwachstellen der CGI nicht ausgleichen.

Wiedereintritte

Stationäre Wiedereintritte werden häufig als Qualitätsindikator in dem Sinne diskutiert, dass bei einem sinnvollen Nachsorgekonzept ein (rascher) erneuter stationärer Aufenthalt potenziell zu verhindern wäre. Es ist jedoch gerade bei psychiatrischen Störungen bekannt, dass diese häufig chronisch und teilweise auch schubweise verlaufen können. Im Sinne einer integrierten Versorgung kann ein häufiger Wechsel zwischen ambulanten und (kurzen) stationären Behandlungen sogar therapeutisch sinnvoll sein (Kuhl et al., 2006). Es stellt sich somit die Frage, inwiefern die Zahl der Wiedereintritte ein Qualitätsindikator sein kann.

Der Abstand zwischen Wiedereintritt und der vorausgegangenen stationären Behandlung spielt dabei sicher eine wesentliche Rolle. Vom OECD-Projekt *Health Care Quality Indicators* wurde als Kriterium Wiedereintritte zwischen 7 und 28 Tagen vorgeschlagen. Für die multizentrische Benchmarkingarbeitsgruppe der drei psychiatrischen Universitätskliniken schien das Kriterium Wiedereintritte innerhalb von 14 Tagen als Indikator jedoch sinnvoller, da hier bei einer Grosszahl von Fällen davon ausgegangen werden kann, dass eine adäquate Patientenversorgung möglich gewesen wäre.

Die Wiedereintritte wurden diagnosespezifisch ausgewertet, wobei auch hier der Patienten-Case-Mix regressionsanalytisch kontrolliert wurde. Für die Klinikvergleiche wurden so die diagnosespezifischen Wahrscheinlichkeiten eines Wiedereintritts gegenübergestellt. Daneben wurde auch der Abstand zwischen den Wiedereintritten im Klinikvergleich ausgewertet. Über alle Diagnosen hinweg zeigen sich nur geringe Unterschiede zwischen den Kliniken. Diagnosespezifisch sind jedoch einige Unterschiede zu erkennen, die wichtige Hinweise für das interne Qualitätsmanagement liefern können.

Zwangsmassnahmen: Fürsorgerischer Freiheitsentzug (FFE)

Auch bei dem Indikatorenkonzept der KIQ werden Zwangsmassnahmen als Indikator für Klinikvergleiche

vorgeschlagen. Bei dem hier beschriebenen Benchmarking wurde die Anzahl der Behandlungsmassnahmen mit Fürsorgerischen Freiheitsentzug (FFE) bezogen auf 100 000 Einwohner des Versorgungsgebiets ausgewertet, um eine standardisierte Grösse zu erhalten. Die Ergebnisse zeigen jedoch, dass dieser Indikator aufgrund von ungleichen gesetzlichen Rahmenbedingungen der beteiligten Kantone und damit zusammenhängenden unterschiedlich hohen FFE-Raten nur bedingt für einen Klinikvergleich geeignet ist.

Fazit

Die Outcomemessung ist in der Psychiatrie ungleich problematischer als in somatischen Bereichen, da Patienteneinschätzungen hier schwierig zu erheben sind und häufig nur für einen Teil der Patienten vorliegen. Gleichzeitig stellt aber analog zur somatischen Medizin die Patientenbeurteilung eine wichtige Säule der Outcomebeurteilung dar und sollte daher auch in einem Benchmarking berücksichtigt werden. Für die arztseitige Outcomebeurteilung gilt es abzuwägen, wie viel Dokumentationsaufwand akzeptabel und welcher Differenzierungsgrad notwendig ist. Die vorliegenden Ergebnisse zeigen deutlich, dass die Erhebung mit einem siebenstufigen 1-Item-Mass für ein sinnvolles Benchmarking nicht ausreichend ist, sondern um differenziertere Schweregradindikatoren ergänzt werden sollte.

Die Erfahrungen aus dem beschriebenen Benchmarking zeigen weiterhin, dass es bei den patientenbezogenen Indikatoren notwendig ist, eine Risikoadjustierung vorzunehmen, um die Vergleiche korrekt zu gestalten. Bei einer grösseren Zahl von Kliniken wären Multi-Level-Verfahren dafür die Methode der Wahl.

Für eine sinnvolle Interpretation sind ausserdem mehrere Auswertungsrunden in einem ausreichenden Zeitintervall nötig. Während halbjährliche Auswertungen in dem dargestellten Benchmarkingprojekt noch einige zeitliche Unterschiede bei einzelnen Kliniken ergaben, lieferten jährliche Auswertungen deutlich stabilere Ergebnisse. Erst durch die Berücksichtigung zeitlicher Schwankungen innerhalb einzelner Kliniken lassen sich Unterschiede zwischen Kliniken sinnvoll interpretieren.

Insgesamt zeigen die Erfahrungen der bisherigen zwei Jahre aber auch, dass neben den konkreten Ergebnissen im Klinikvergleich auch der Austausch über organisatorische und prozessbezogene Klinikunterschiede sowie Unterschiede der Rahmenbedingungen wichtige Anregungen für das Qualitätsmanagement liefern können. Auch das auf allen Ebenen durch Besprechungen und Diskussionen wachsende Qualitätsbewusstsein ist ein hervorzuhebender Effekt bei der Durchführung eines Benchmarkings. Aus der Sicht der Qualitätssicherung kann man trotz einiger problematischer Punkte insgesamt eine positive Bilanz ziehen. Für gesamtschweizerische Klinikvergleiche in der Psychiatrie sollte die Wahl der Indikatoren unter den genannten Gesichtspunkten in jedem Fall gut überlegt werden. ■

Korrespondenzadresse:
Dipl.-Psych. Hans-Christian Kuhl
Universitäre Psychiatrische Kliniken Basel (UPK)
Psychologischer Dienst
Arbeitsbereich Versorgungsforschung & Qualitätssicherung
Wilhelm-Klein-Strasse 27
4025 Basel

Interessenkonflikte: keine

Literatur:

1. *Cording C, Gaebel W, Spengler A, Stieglitz RD, Geiselhart H, John U, Netzold DW, Schönell H, Spindler P, Krischker S (1995). Die neue psychiatrische Basis-Dokumentation. Eine Empfehlung der DGPPN zur Qualitätssicherung im (teil-)stationären Bereich. Spektrum der Psychiatrie und Nervenheilkunde 24: 3–41.*
2. *Farin E, Glattacker M, Follert P, Kuhl HC, Klein K, Jackel WH (2005). Einrichtungsvergleiche in der medizinischen Rehabilitation. Z Arztl Fortbild Qualitatssich 98(8): 655–662.*
3. *Frick U, Krischker S, Cording C (2003). Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung in der Psychiatrie. Bericht an den Bundesminister für Gesundheit.*
4. *Glattacker M, Jäckel WH (2007). Evaluation der Qualitätssicherung – aktuelle Datenlage und Konsequenzen für die Forschung. Gesundheitswesen 69: 277–283.*
5. *Haro et al. (2003). The Clinical Global Impression – Schizophrenia Scale: a simple instrument to measure the diversity of symptoms present in schizophrenia. Acta Psychiatr Scand 107 (Suppl. 416): 16–23.*
6. *Hermann RC, Mattke S, Somekh D, Silfverhielm H, Goldner E, Glover G, Pirkis J, Mainz J, Chan JA (2006). Quality indicators for international benchmarking of mental health care. Int J Qual Health Care 18 Suppl 1: 31–38.*
7. *Iezzoni, LI (2003). Risk adjustment for Measuring Healthcare Outcomes (3rd ed.). Chicago: Health Administration Press.*
8. *Kuhl HC, Müller-Spahn, F (2006). Psychiatrische Rehospitalisierungen in Basel-Stadt: Häufigkeit, Abstand und Einflussfaktoren. Schweizer Zeitschrift für Psychiatrie und Neurologie (4), 27–32.*
9. *Nationale Koordinations- und Informationsstelle für Qualitätssicherung (KIQ). (2007). Q-Info 16. www.kiq.ch.*