

Ein Kommentar zu einer Studie eines Internetanbieters

Behandlungszufriedenheit in öffentlichen Spitälern

Ranglisten sind äusserst beliebt. Sogenannte «Studien», die ohne kritische Würdigung («peer review») in die Tagespresse gelangen, werden von der Leserschaft als wissenschaftlich wahrgenommen. Der folgende Artikel beleuchtet aus statistischer Optik die vor kurzem «publizierte» Comparis-Studie zur Patientenzufriedenheit in Schweizer Spitälern und nimmt eine Bewertung für die Ärzteschaft vor.

T. Kaufmann, M. Rossi,
G. Schüpfer

Umfragen gehören zum politischen Alltag. Sie entscheiden über unser Fernsehprogramm und werden sogar von Standesorganisationen durchgeführt. Umfragen sind wichtige, aber auch herausfordernde Instrumente. Ranglisten sind äusserst beliebt, egal ob sie die Städte mit der höchsten Lebensqualität, die Korruption in verschiedenen Ländern oder, neuerdings immer häufiger, die Qualität der Spitälern betreffen. Kürzlich hat ein Internetvergleichsdienst eine Studie über die «Patientenzufriedenheit in Schweizer Spitälern» publiziert. Während die Ergebnisse der Studie breit diskutiert werden, wird die Methode kaum je hinterfragt.

Im Auftrag des Internetanbieters kontaktierte ein Marktforschungsinstitut telefonisch 15 000 Schweizer, 5800 Telefoninterviews fanden dann Eingang in die Studie, das sind knapp 3,8% aller angerufenen Personen. Der Kanton Luzern hatte 2006 eine mittlere Bevölkerung von 359 715 Einwohnern und 41 768 Austritte aus einem der öffentlichen Spitälern, ohne Psychiatrie (11,6%). Auch wenn eine Person mehrere Hospitalisationen hatte, liegt die Wahrscheinlichkeit sicher über 4%, im letzten Jahr stationär hospitalisiert gewesen zu sein. Der Internetanbieter und Auftraggeber publizierte folgende Einschlusskriterium: Patienten im Alter zwischen 15 und 85 Jahren, die mindestens eine Nacht in einem der 53 definierten Spitälern verbracht haben. Keine Aussagen finden sich zu den von der Untersuchung ausgeschlossenen Personen.

Wie die in der Befragung verwandten 19 Fragen validiert wurden, bleibt ein Geheimnis. Jede Frage hatte 5 Antwortmöglichkeiten, und die Antwortkategorien wurden mit 1 oder 0 bewertet, wobei Comparis bestimmte, wo die Grenze zu liegen hatte:

Beispiel

- Frage: Wurden Sie im Spital als Patient/Patientin respektvoll behandelt?
Antwortmöglichkeiten (Bewertung in Klammern): «immer» (1), «meistens» (0), «manchmal» (0), «selten» (0), «nie» (0)
- Frage: Hat Ihnen der Arzt/die Ärztin oder das Pflegepersonal im Spital die Medikamente für zu Hause verständlich erklärt?
Antwortmöglichkeiten (Bewertung in Klammern): «sehr verständlich» (1), «verständlich» (1), «gerade so» (0), «weniger verständlich» (0), «überhaupt nicht verständlich» (0)

Diese Methode der Stratifizierung ist doch etwas exotisch und öffnet Raum für Willkür.

Die grösste Herausforderung jeder Studie ist, eine wirklich vergleichbare Referenzgruppe zu finden. Schliesslich sind nicht alle das Ergebnis beeinflussenden Faktoren bekannt – also müssen Confounder gleich verteilt sein. Es ist üblich, dass die Geschlechter- und die Altersverteilung sich nahezu entsprechen. Wird diesem Umstand nicht Rechnung getragen, können Subgruppenfehler und sogenannte Confounderfehler auftreten. Der Subgruppenfehler bezeichnet das Vorkommen signifikanter Resultate in Untergruppen bei jedoch nicht signifikantem Resultat über das gesamte Kollektiv. Ist die Grösse der Subgruppen adäquat, ist dies meist der Ausdruck einer (indirekt) abhängigen Variablen. Ein bekanntes Beispiel aus der Kardiologie zeigt, wie sich der sozioökonomische Status und das Rauchen beeinflussen.

Bias bedeutet strenggenommen Vorurteil oder Voreingenommenheit. In der klinischen Forschung meint man damit eine systematische Abweichung, einen Drift, eine Verfälschung etc.

Korrespondenz:
Dr. med. Thomas Kaufmann
Qualitätsmanagement
Infektiologie und Spitalhygiene
Kantonsspital Luzern
CH-6000 Luzern

Man unterscheidet zudem zwischen einem selektiven Bias und einem Informationsbias. Nicht alle Abhängigkeiten sind so offensichtlich. Dies führt dann zu einem Bias, einem systemischen Fehler der Studie. Werden Populationen verschiedener Regionen miteinander verglichen, besteht die Gefahr eines Confounderfehlers.

Laut publizierter Interpretation der Daten der von Comparis beauftragten Studie sei die Westschweizer Bevölkerung grundsätzlich unzufriedener, erkennbar am Rating der Spitäler. Diese Aussage ist untenstehender Checkliste zu unterwerfen, und die Resultate sind bezüglich interner und externer Validität zu hinterfragen (Tabelle).

Eine Studie sollte möglichst transparent und damit nachvollziehbar, sogar nachrechenbar sein. Leider werden zur Patientenstruktur mit Ausnahme des Geschlechterverhältnisses (57% Frauen) keine Angaben geliefert.

Neben den rein technischen Aspekten ist bei allen Zufriedenheitsmessungen zu fragen, ob wirklich das gemessen wird, was man messen will. So hat der Zeitpunkt bzw. die Zeit zwischen Austritt und Befragung möglicherweise einen wesentlichen Einfluss auf das Ergebnis. Sogar wenn die Fragen auf den ersten Blick die Wahrnehmung zur Zufriedenheit aufnehmen, bleibt offen, was der Befragte genau ausdrücken will. Wird damit vielleicht vielmehr die Identifizierung mit «seinem» (von Schliessung) bedrohten Spital gemessen?

Zufriedenheitsmessungen sind äusserst delikat, vor allem wenn die Ergebnisse auch Massnahmen nach sich ziehen sollten. Im Bereich Patientenzufriedenheit existieren bereits validierte Methoden. Eine seriöse Untersuchung ist aufwendig und kostet demzufolge viel Geld. Innovationen sind wie überall auch auf diesem Gebiet zu begrüssen, dürfen – ebenfalls wie üblich – den Vergleich mit dem Goldstandard aber nicht scheuen.

Bei kritischer Durchsicht dieser «Studie» stellen wir fest, dass wohl eher die Identifikation mit dem Spital in der eigenen Region als die Zufriedenheit mit der Behandlung gemessen wurde, und sehen erhebliche methodische Probleme der Studie. Wird ein kleines Regionalspital mit einem universitären Zentrumsversorger in den gleichen Topf geschmissen und werden bezüglich der Behandlungszufriedenheit Schlussfolgerungen gezogen wie in der zur Frage stehenden Studie, so ist wohl die wichtigste Schlussfolgerung, dass die erhobene Patientenzufriedenheit wohl in Zusammenhang mit der Spitalgrösse steht. Für einen Zentrumsversorger ist die Studie nicht hilfreich, ist doch seine chirurgische Klinik möglicherweise grösser als das zum Vergleich herangezogene Regionalspital als Ganzes. Welche Schlussfolgerungen sollen also gezogen werden, auf welche Klinik des Zentrumversorgers ist das Ergebnis zu übertragen, und was heisst das nun für einen potentiellen Patienten? Bei der Be-

Tabelle 1

Können alle diese Fragen negativ beantwortet werden, könnten die Ergebnisse korrekt sein.

Beurteilung von Studien

<i>Interne Validität</i>	Wird das gemessen, was behauptet wird? Observationsstudie (Zensus usw.) – Biasanfälligkeit – hohe Repräsentativität
<i>Externe Validität</i>	Können die Ergebnisse generalisiert/extrapoliert werden? Kontrolliert randomisierte Studie

Checkliste für Bias

Handelt es sich um einen selektiven Bias?	Kohortenstudie: Sind alle wichtigen Einflussfaktoren sowohl bei der Interventions- wie auch bei der Vergleichsgruppe gleich verteilt mit Ausnahme der zu untersuchenden Variable? Fall-Kontrollstudie: Sind sowohl bei der Fallgruppe wie bei der Vergleichsgruppe alle wichtigen Einflussfaktoren ähnlich verteilt, mit Ausnahme der Krankheit in der Fragestellung?
Handelt es sich um einen Informationsbias?	Kohortenstudie: Sind die Endpunkte für die Interventions- und Vergleichsgruppe identisch? Fall-Kontrollstudie: Wurde bei beiden Gruppen auf dieselbe Weise die gleiche Information beschafft?
Ist eine Störvariable vorhanden («confounding»)?	Könnte das Resultat durch die Anwesenheit eines beeinflussenden Faktors, wie Alter, Rauchen, Gewicht, Sexualverhalten etc., der bei der Interventionsgruppe und Vergleichsgruppe vorhanden ist, aber keinen direkten Kausalzusammenhang aufweist, erklärt werden?
Handelt es sich um keinen dieser systemischen Fehler, ist es ein Zufallsergebnis ohne Signifikanz?	Hierzu sind das relative Risiko und das 95%-Vertrauensintervall zu berechnen. Ist das Resultat signifikant? War die Stichprobengrösse genügend hoch, um einen klinisch bedeutenden Unterschied feststellen zu können?

urteilung eines Spitals oder besser einer Klinik sind neben der Patientenzufriedenheit das medizinische Gesamtangebot, die Qualität, die Wartezeiten und die Kosten zudem mitentscheidend.

Es handelt sich um eine Observationsstudie, die typischerweise biasanfällig ist. Die externe Validität ist eher gering.

Umfragen müssen den Stichprobenbereich klar definieren. Fragen müssen zuverlässig und validiert sein. Dies kann eine Pilotstudie erforderlich machen. Abhängig von der gewünschten Präzision ist der Stichprobenumfang zu definieren. Für statistische Analysen sind grössere Stichproben erforderlich. Wenn die untersuchten Charakteristika mit dem Alter, dem Geschlecht, der Erfahrung usw. variieren, ist eine Stratifizierung der Population erforderlich. Fehlerhafte Resultate sind auch auf nicht Antwortende zurückzuführen. Um diesen Bias zu vermeiden, müssen

Wege gefunden werden, von allen Befragten auch Antworten zu erhalten.

Die Ärzteschaft wird aber in Zukunft von derartigen Publikationen nicht verschont bleiben, und wir tun gut daran, methodisch saubere Versorgungsforschung selbst in die Hand zu nehmen.

Literatur

- Burmeister LF. Principles of Successful Sample Surveys. *Anesthesiology*. 2003;99:1251-2.
- Katz MH. Multivariable analysis: Primer for readers of medical research. *Ann Int Med*. 2003;138:644-50.
- Vandembroucke JP. When are observational studies as credible as randomised controlled trials? *Lancet*. 2004;363:1728-31.
- Grimmes DA, Schulz KF. Observational studies: What they can and can not do. *Lancet*. 2002; 359:145-9.