



Public Health und mangelhafte empirische Grundlagen

Public Health und mangelhafte empirische Grundlagen

Replik auf das übersetzte Buchkapitel „*Il Medico Cantonale*“ betreffend **Public Health**,
SAeZ 2010;91:636-644 von Dr. med. Ignazio Cassis, Nationalrat und Vizepräsident der FMH

Für den Verein Ethik und Medizin Schweiz



Dr. med. Michel Romanens
Innere Medizin und Kardiologie FMH
Ziegelfeldstr. 1
4600 Olten

Dr. med. Franz Ackermann
Innere Medizin FMH
Bergstr. 46
4600 Olten

Dr. med. Bernhard Hofmeier
Allg. Medizin FMH
Gartenstrasse 6
4632 Trimbach

Korrespondenz:

info@kardiolab.ch

Tel: 062 212 44 10

Fax: 062 212 44 30

Olten, 28.04.2010

Wortzahl: 1899

Public Health und mangelhafte empirische Grundlagen

Public Health wird in der Referenz 13 des Artikels (im Buch über Public Health von Ivo Giulietti und Ignazio Cassis) als wissenschaftliche Methode folgendermassen beschrieben: „Die wissenschaftliche Methode ist die typische Vorgehensweise in der Wissenschaft, um zu einer objektiven, zuverlässigen, nachprüfaren und allgemein nutzbaren Erkenntnis der Realität zu gelangen. Hier ist gemeint: «Die Gesamtheit der Regeln und Instrumente, die es gestatten, mit einem hohen Grad an Sicherheit festzustellen, dass das beobachtete Ereignis nicht dem Zufall unterliegt»“.

Die Public Health Institute sind alle den Universitäten in der Schweiz angegliedert und produzieren öffentlich relevante Publikationen zu Fragen der Gesundheit und unter anderem zu Fragen der gerechten Ressourcen-Allokation. Trotzdem hat keines dieser Institute die Charta der öffentlichen Statistik des Ethikrates unterzeichnet¹ oder ist dem Ethikrat beigetreten². Damit besteht die Gefahr einer universitären Inzucht in der Statistik ohne Kontrolle durch unabhängige öffentliche Stellen.

Die Public Health Institute beobachten, wie eingangs erwähnt, Unterschiede und stellen allenfalls *statistisch relevante Unterschiede* fest, z.B. betreffend Hüftoperationen in der Schweiz.

Das Problem: es findet keine Beschränkung auf die Aussage statt, dass Unterschiede festgestellt wurden. Im Gegenteil: aus den Unterschieden werden Schlüsse gezogen, welche wissenschaftlich teils tendenziös oder unzulässig sind, da die empirischen Grundlagen fehlen bzw. nicht dokumentiert oder nicht erarbeitet wurden. Damit setzt sich die Mittelwert-Statistik der Public Health Institute der Gefahr von Confounding und Bias aus und entspricht etwas dem Niveau der *santésuisse* Statistiken zur Erfassung von Überarztung (auch dort werden Unterschiede auf falsche Ursachen zurückgeführt, weil der Confounder, die Morbidität des Patienten, in völlig unzureichender Weise berücksichtigt wird).

Anhand eines Beispiels unter vielen soll die Problematik illustriert werden. In der NZZ vom 11.04.2010 bezog sich die Journalistin Irène Dietschi auf eine wissenschaftliche Public Health Arbeit von Herrn Prof. Dr. med. vet. André Busato³. Im Kasten des NZZ-Artikels mit dem Titel „Tausende Operationen pro Jahr“ stand zu lesen: „Tausende von Schweizerinnen und Schweizern legen sich jedes Jahr unters Messer, um schmerzende Knie, Schultern oder Hüften durch künstliche Gelenke ersetzen zu lassen – Tendenz steigend, wie eine Studie des Berner Epidemiologen André Busato zeigte. Was Busato und sein Team viel mehr erstaunte als die Zunahme der Eingriffe, waren die regionalen Unterschiede: Die Zahlen für Gelenkoperationen variieren insgesamt zwischen 1 und 3 Eingriffen pro 1000 Einwohner. Beispiel Hüftprothesen: In manchen Regionen der Waadt, des Tessins oder im Aargau werden jährlich nur 1 bis 2 Hüftprothesen pro 1000 Einwohner eingesetzt, während in den Kantonen Bern, Thurgau oder Obwalden mehr als 3 solcher Eingriffe gezählt werden. Spitzenreiter ist der Kanton Schwyz, wo pro Jahr fast 4 von 1000 Bewohnern ein künstliches Hüftgelenk bekommen. «Bei einem Standardeingriff wie dem Hüftgelenkersatz, der gut validiert ist und bei dem klare Behandlungskriterien bestehen, sind diese grossen Unterschiede zwischen den Regionen nicht nachvollziehbar », sagt Busato. In anderen Worten: Die Unterschiede sind nicht medizinisch bedingt. Für Busato sind die Befunde ein Beweis, dass die Qualitätssicherung unseres Gesundheitssystems ein bruchstückhaftes Gebilde sei. «Es braucht allgemein gültige klinische Richtlinien für solche Eingriffe», fordert Busato. Nicht nur Indikationen und Operationsmethoden, sondern auch das Kosten-Nutzen-Verhältnis müssten von einer unabhängigen Instanz überprüft werden.“

Im Originalartikel von Prof. Busato steht dazu: „Although the underlying causes of geographic variations in health care *are not fully understood*, there is agreement in the literature that these variations *can be conceptualized* as those justified by the effective needs of patients and those unwarranted, which are not directly related to the illness of patients.“

Dies ist eine der wesentlichsten, jedoch nie genügend bewiesenen Unterstellungen in dieser Art von Wissenschaft: es wird insinuiert, dass Unterschiede teils durch unnötige Eingriffe bedingt sein *müssen*. Typischerweise fehlt bei solchen Meinungsäusserungen – welche zudem wie die santésuisse Wirtschaftlichkeitsprüfungen einen eigentlich rufschädigenden Aspekt aufweisen – eine Literaturangabe noch werden die Behauptungen empirisch begründet. Somit fehlt in der Argumentationskette – obwohl anders Wirtschaftlichkeitsprüfungen behauptet – schlicht der Beweis. Es werden unter dem Deckmantel der Wissenschaft Vermutungen zu Beweisen umgedeutet.

Natürlich können solche Konzepte durchaus gedacht werden, der Beweis für unnötige Hüftoperationen liefert der Artikel jedoch aus mehreren Gründen nicht:

- Es wurden nur Raten (Mittelwerte) miteinander verglichen, ohne die Ursachen der Unterschiede zu erforschen. Um die Ursachen zu erforschen benötigt man einen detaillierten Fragebogen, welcher am Ort des Geschehens die Gründe für medizinische Eingriffe quantifiziert. Diese Forschung ist relativ aufwendig und

kostet entsprechend viel Geld, weswegen in vielen Studien des Public Health darauf verzichtet wird. Das bedeutet somit, dass die empirischen Grundlagen für die beobachteten Unterschiede fehlen.

- In der Arbeit von Prof. Busato wurden u.a. interessante Unterschiede betreffend Perzentilenwerte festgestellt, jedoch auch hier ohne Angabe von statistisch signifikanten Unterschieden.
- Es wurden bei den festgestellten Unterschieden generell keine Signifikanz-Niveaus berechnet, das heisst, die Unterschiede, zumal im Promille-Bereich (statistisches Problem der kleinen Zahlen) sind nicht einmal statistisch signifikant.
- Die Berechnungsgrundlagen für die Unterschiede, insbesondere die getroffenen Gruppierungen auf Provider-Ebene. (Zoneneinteilungen) sind wissenschaftlich nicht validiert und empirisch nicht überprüft.
- Die zugrunde gelegten Zahlenmaterialien unterliegen einem möglichen Coding- oder Misclassification Bias, welcher nicht quantifiziert wurde.
- Es fehlen Artikel Berechnungen im Sinne einer Kosten-Nutzen Analyse.

Folgende Confounding Variablen können bei einer niedrigen Inzidenz in der Bevölkerung (1-6/1000) durchaus gewichtige Folgen auf die Distribution haben und müssten im besagten Artikel ebenfalls diskutiert werden:

- lokale Expertise der operierenden Zentren (Berühmtheits- und Attraktivitätsgrad bedeutet keineswegs unnötige Operationen).
- lokale "genetische" und Umwelt-Variablen, welche das degenerative Hüftleiden favorisieren.
- Unterversorgung im Ausland und speziellen Regionen in der Schweiz aus Kostengründen.
- Unterschiede in der Toleranz gegenüber Schmerzen („die Urner sind härtere Burschen als die Schwyzer“).
- Sparfaktoren wie etwas durch Überarztungsprozesse der *santésuisse* bedrohte Orthopäden.

Die Frage ist besonders wichtig: sind Unterschiede in der Häufigkeit beim Problem der niedrigen Fallzahlen (bezogen auf die Bevölkerung!) statistisch signifikant? So waren beispielsweise mit einer Ausnahme die vom BAG publizierte institutionsbasierten Mortalitätsraten für bestimmte Krankheiten, welche ja auch das "Problem" der niedrigen Fallzahlen aufweisen, mit einer Ausnahme *statistisch nicht signifikant*. In der Tagespresse wurde dieser feine Unterschied oder Mangel skotomisiert und Schuldzuweisung ohne jegliche Grundlage betreiben („schlechtere Versorgungsqualität kleinerer Spitäler“). Dazu kommt noch die Unsicherheit bei den CHOP Daten (besteht eine interne Validierung? Mislabelingrate?) und anderer confounders, so dass wir hier sicher eine Statistik mit 95%CI fordern müssen.

Der besagte Unterschied betrug in der Arbeit von Prof. Busato: drei *Hüftoperationen auf 1000 Personen*. Der Chi-Quadrat Test ist für den Unterschied 3/1000 versus 6/1000 mit 0.45 auf einem Signifikanzniveau, der rein zufällig ist (Münzenwerfen). Der Unterschied müsste 5x grösser sein (15/1000 versus 30/1000), damit der Chi-Quadrat Test mit 4.46 einen knapp signifikanten p-Wert von < 0.05 erreichen würde⁴. Die Nichtbeachtung einfachster statistischer Grundsätze widerspricht nachgerade diametral dem von Ignazio Cassis

formulierten Anspruch an die Sicherheit der Interpretation, wie das Zitat aus dem Buchkapitel belegt: «Die Gesamtheit der Regeln und Instrumente, die es gestatten, mit einem hohen Grad an Sicherheit festzustellen, dass das beobachtete Ereignis nicht dem Zufall unterliegt».

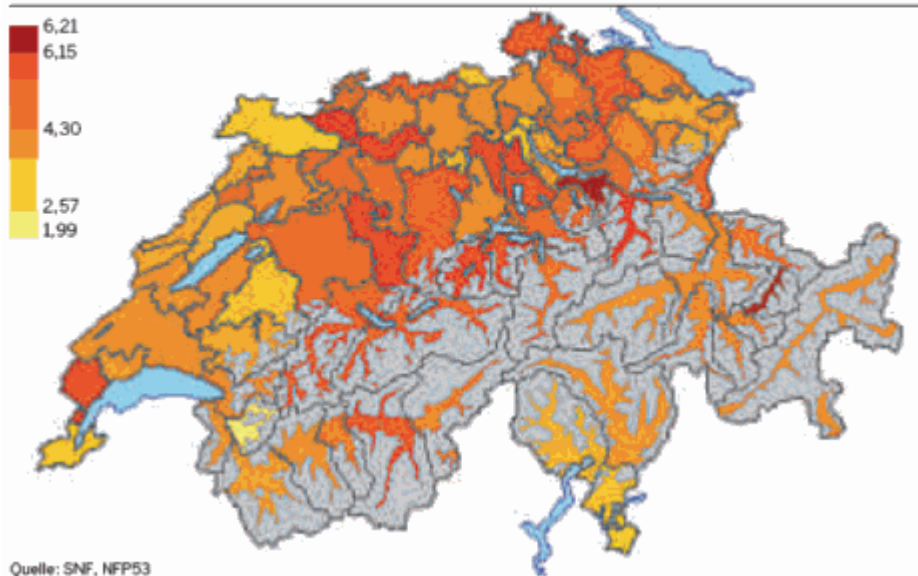
Da in der Arbeit von Prof. Busato im Diskussionsteils die Ursachen für die beobachteten Unterschiede nicht beleuchtet werden, weil die dazu notwendige empirische Grundlagenforschung nicht durchgeführt wurde, sind die von ihm geäußerten Schlussfolgerungen tendenziös. Um die Aussagen zu untermauern, müsste im Sinne eines Audit-Prozesses zufällig ausgewählte Operierte und die Leistungserbringer eingehend mittels mindestens 50 relevanten Fragen beobachtet werden, bevor die Konklusion, dass Qualitätsmängel oder gar Bereicherungssucht, wie aus dem NZZ Artikel leicht gefolgert werden kann, auf Beweisebene postuliert oder weisgemacht werden dürfen.

Ein weiteres Problem betrifft folgendes: Internationale Konsensuspapiere in Gesundheitsökonomie und Epidemiologie fordern: „Bereits bei der Planung einer Studie sollten Maßnahmen zur Abwehr von Biases ergriffen werden, die durch Selektion, Confounding etc. entstehen können. Dazu zählen zum Beispiel das Matching oder eine Einschränkung der Variabilität von Störfaktoren oder aber die Erfassung von Informationen, die zu der Kontrolle von Confounding erforderlich ist. Zur Abschätzung der Auswirkungen von Messfehlern auf das Studienergebnis können zusätzliche Erhebungen zur Durchführung von Sensitivitätsanalysen geplant werden.“ Die Leitlinien für die gute epidemiologische Praxis (GEP) sind anderswo publiziert⁵. Diese Leitlinien fanden im besagten Artikel weitgehend keine Beachtung, wenn man die nicht belegten, aber provozierenden Schlussfolgerungen der Autoren berücksichtigt.

Aus unserer praktischen Erfahrung im medizinischen Alltag sind zudem rein ökonomische Überlegungen welche zu höheren Operationsraten führen würden, völlig fremd. Im Gegenteil: bevor es zur Operation kommt, besteht eine jahrelange Geschichte mit Schmerzen, Abwarten, Leiden und Angst vor dem Eingriff.

Regionale Unterschiede

Gelenkersatzoperationen im Jahr 2005 (pro 1000 Einwohner)



In den Zeiten der DRG wird aber sicher das passieren, wovor im besagten Artikel gewarnt wird: die Ökonomisierung der Medizin wird unnötigen Operationen erzwingen, um die Institutionen überlebensfähig bzw. finanzierbar zu halten, wie Beispiele aus Deutschland belegen⁶.

Im Buchkapitel zum Public Health, welches der Auslöser dieser Stellungnahme ist, findet sich eine Art Heiligsprechung des Public Health. Zitat aus dem Buchkapitel: „Das Konzept der Public Health entspringt einer strengen Denkweise, die auf wissenschaftlichen Methoden gründet. Das Konzept der «Evidence-based Public Health» – also der «auf Wirksamkeitsnachweisen beruhenden Public Health» – **lässt keinen Raum für Improvisationen oder Gefühlsentscheidungen** und umfasst die Entwicklung, Umsetzung und Bewertung wirksamer Programme und Politiken durch Anwendung wissenschaftlicher Überlegungen und systematische Nutzung von Daten und Programmiermodellen.“

Das Paradebeispiel für den präliminären Charakter von Beobachtungen aus Public Health Studien ist die HERS Studie (JAMA 1998;280:605-13):

Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B, Vittinghoff E.).

Entgegen der Mehrzahl von Public Health Studien, welchen einen positiven Effekt des Hormonersatzes bei der Frau postulierten, zeigte die kontrollierte, randomisierte HERS Studie (CRT) in der Sekundärprävention das Gegenteil.

Aus meiner Sicht ist diese glorifizierende Beurteilung betreffend Public Health Studien als „unbestechlicher“ Forschung nicht zutreffend und gefährlich (Siehe Kasten). Die meiste Public Health Forschung generiert lediglich Hypothesen, welcher näher überprüft oder im Rahmen von kontrollierten Studien (controlled randomized trials, CRT) auf ihren Wahrheitsgehalt empirisch überprüft werden müssen. Wenn wir die zahlreichen Mittelwertvergleiche ohne dazugehörige unabdingbare empirische Grundlagenforschung und die häufig in keiner Weise wissenschaftlich bzw. statistisch belegten Schlussfolgerungen in vielen Public Health Papers unkritisch übernehmen, wird *die politische Lenkung der angeblich „gerechten“ Ressourcen-Allokation fehlgeleitet.*

Die statistischen Qualitäts-Standards des Public Health müssen kritischer betrachtet und öffentlich diskutiert werden. Public Health Institute sollten sich einer öffentlichen Kontrolle ihrer Aussagen und Statistiken durch den Ethikrat für Statistik unterziehen, vor allem und gerade dort, wo sie im öffentlich-rechtlichen Raum operieren und potentiell Ressourcen-Allokationen über politische Meinungsbildung bei Politikern und in der Öffentlichkeit beeinflussen wollen.

¹ <https://docs.google.com/viewer?url=http://www.stat.ch/docs/ethics/charter-subscribers-g.pdf>

² <http://www.stat.ch/index.php/de/stat/ethikrat.html>

³ Widmer M, Matter P, Staub L, Schoeni-Affolter F, Busato A. Regional variation in orthopedic surgery in Switzerland. Health & Place 15 (2009) 761–768.

⁴ http://www.kardiolab.ch/riskcalc_JSI.html

⁵ www.dgepi.de/doc/Empfehlungen.doc

⁶ <http://www.epochtimes.de/articles/2010/04/18/567377.html>