



Begleitforschung SwissDRG und das Problem
der pluralistischen Ignoranz in der Statistik, 18.10.2009



Begleitforschung SwissDRG und das Problem der pluralistischen Ignoranz in der Statistik

Viewpoint Artikel

Autor und Korrespondenz-Adresse:

Dr. med. Michel Romanens
Innere Medizin FMH, speziell Kardiologie
Präsident Verein Physician Profiling
Ziegelfeldstr. 1, 4600 Olten
☎ 062 212 44 10
Fax 062 212 44 30
E-Mail: info@kardiolab.ch
Homepage: www.physicianprofiling.ch

**Begleitforschung SwissDRG und das Problem
der pluralistischen Ignoranz in der Statistik, 18.10.2009**

Zusammenfassung

Die Forschung mit dem Ziel, negative Auswirkungen der DRG-basierten Spitalfinanzierung und neuer Benchmarks für Performance-Vergleiche zwischen Spitälern zu erfassen, sollte vorab selbst Gegenstand der Forschung und eines Nationalen Konsensus werden. Damit kann verhindert werden, dass das Monitoring Anreize schafft, welche über Risikoselektionsmechanismen die medizinische Versorgungsqualität und Versorgungssicherheit der Schweizerischen Bevölkerung vermindern. Zwei Hauptpunkte sind zu erwähnen: das weitgehende Fehlen der Diskussion um das Confounding in der Forschung; das Fehlen einer Validierung von DRG mittels pharmaceutical cost groups (PCG).

Research tools evaluating undesirable effects of hospital financing and benchmark comparison between hospitals based upon diagnosis related cost groups (DRG) should themselves become object of research in order to define a national consensus. This would help to avoid negative effects of DRG based hospital financing on medical supply quality and security for the Swiss. Two major points have to be mentioned: lack of discussion about the statistical problem of "confounding variables"; lack of validation of the variable DRG using pharmaceutical cost groups (PCG).

Glossar

CONFOUNDER: Unter einem Confounder (englisch: „Störfaktor“) versteht man innerhalb von epidemiologischen Studien einen nicht erfassen und daher nicht kontrollierbaren Störfaktor, der mit einem Faktor unter Beobachtung (Exposition) in Beziehung steht. Ein Confounder ist eine Variable, die das Auftreten eines Risikofaktors (z. B. Krankheit) und die beobachtete Exposition mitbestimmt. Untersucht man den Zusammenhang zwischen Nikotinkonsum und Leberzirrhose oder hepatozellulärem Carcinom, kann man eine signifikante Korrelation feststellen. Es besteht jedoch kein kausaler Zusammenhang, Nikotinkonsum führt nicht zur Leberzirrhose. Vielmehr sind viele Raucher auch Trinker und es besteht ein kausaler Zusammenhang zwischen Alkoholkonsum und Leberzirrhose. Der Alkoholkonsum ist also der Confounder, Nikotinkonsum ist dagegen nur ein Indikator für Leberzirrhose (statistisch ausgedrückt: Typ 1 Fehler).

R²: Das Bestimmtheitsmaß (abk. R² oder B, auch Determinationskoeffizient) ist ein Maß der Statistik für den erklärten Anteil der Variabilität (Varianz) einer abhängigen Variablen Y durch ein statistisches Modell. Beispiel: Um die Gesundheitskostenausgaben einer Person vorherzusagen, können verschiedene Variablen benutzt werden. Gemäss Beck¹ erklären 0 Variablen 0% der Kosten, die Variablen „Alter und Geschlecht“ 11% (R² = 0.11), die Variablen „Alter, Geschlecht, Hospitalisation im Vorjahr“ 21% und die Variablen „Alter, Geschlecht, Hospitalisation im Vorjahr und pharmaceutical cost groups (PCG)“ 30% der Kosten. Somit fehlen noch die Confounders (Variablen), welche die 70% der nicht erklärten Kosten beleuchten. Die weltweit durchgeführte Interheart Study² zeigte demgegenüber, dass 9 kardiovaskuläre Risikofaktoren (Rauchen, Lipide, Diabetes, Übergewicht, Diät, Bewegung, Alkoholkonsum und psychosoziale Faktoren) 90% des Risikos für Herzinfarkt erklären. Hier ist die Wahrscheinlichkeit für das Vorhandensein eines Confounders gering.

IGNORANZ: Pluralistische Ignoranz ist ein Begriff aus dem Gebiet der Sozialpsychologie: Er bedeutet, dass sich das eigene Verhalten in dem anderer widerspiegelt. Dies führt in einer Notfallsituation dazu, dass niemand einschreitet und hilft, da sich alle im Nicht-Reagieren einander anpassen. Dies kann fatale Folgen haben, wenn sich niemand aus dieser pluralistischen Ignoranz herauslöst und zum Modell wird, dem die anderen Bystander (Schaulustigen) folgen können. Beispiel: Mittelwert-Betrachtungen haben per se **keinen Aussagewert**, da Unterschiede in Mittelwerten keinen Erklärungswert haben. Die Instrumentalisierung von Mittelwerten zwecks Diskriminierung ärztlicher Leistungen und Leistungsabbau in der medizinischen Versorgung ist hingegen eine politisch gewollte und in den Medien beliebte Falsifizierung statistischer Evidenz.

¹ <http://physicianprofiling.ch/HealthEconomicsPCGArticleBeckWorkingPaper2006.pdf>

² Yusuf S, Hawken S, Ounpuu S, on behalf of the INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004;364:937-952.

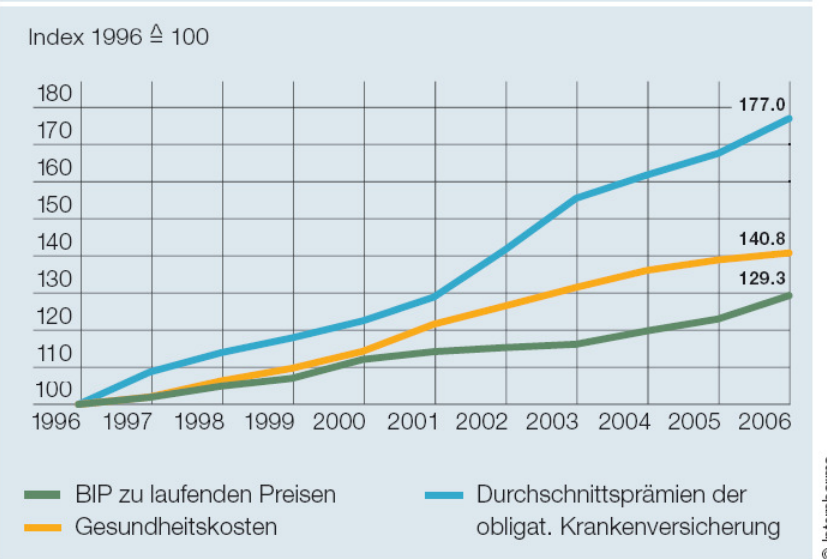
**Begleitforschung SwissDRG und das Problem
der pluralistischen Ignoranz in der Statistik, 18.10.2009**

Hintergrund

Im Jahr 2006 betragen die gesamten Ausgaben für das Gesundheitswesen rund 52.8 Milliarden Franken. Zwischen 1996 und 2006 sind die Gesundheitskosten um 41

Prozent angestiegen, während sich die Prämien in der Grundversicherung im gleichen Zeitraum viel stärker erhöht haben (+77 Prozent, Siehe Grafik 1).

Entwicklung der Indizes des BIP, der Gesundheitskosten und der monatlichen Durchschnittsprämien



Quelle: Bundesamt für Statistik, Neuchâtel; Statistik der obligatorischen Krankenversicherung, Bundesamt für Gesundheit, Bern.

Die Schweiz ist eines der reichsten Länder der Welt. Jeder 25. Einwohner nennt eine Million Franken oder mehr sein eigen. Dies animiert zu Luxus. So können medizinische Leistungen gekauft werden, einfach weil man sich das leisten kann und will, zB ein Klassenwechsel bei Hospitalisationen. Im Jahre 2002 betragen die „out of the pocket“ Zahlungen der SchweizerInnen im Gesundheitswesen 13 Milliarden Franken oder 26% der gesamten Kosten im Gesundheitswesen, kaufkraftbereinigt pro Kopf US-\$ 1085 (zum Vergleich die USA für 2002: 737 US-\$). Dazu kommen noch 2.5 Milliarden Franken für Kostenbeteiligungen.

Da diese Zahlungen zum Teil als freiwillige Luxusmedizin aufgefasst werden müssen, welche sich die Schweizer als Bewohner eines reichen Landes ja auch leisten dürfen, hinkt der Vergleich der Ausgaben für den Anteil der Gesundheitskosten am BIP (%BIP) mit anderen Ländern erheblich. Das %BIP betrug für die Schweiz im Jahre 2002 11.2 % (USA 14.6%). Gehen wir davon aus, dass rund 5 der 13 Milliarden Franken der „out of the pocket“ Zahlungen freiwillige Luxusmedizin betreffen, welche eben in anderen, ärmeren Ländern

nicht bezahlt werden wollen oder können, würde das %BIP auf 10.0% sinken, vergleichbar mit Deutschland (10.9%), Frankreich (9.7%) oder Griechenland (9.5%) (CD-ROM OECD Health Data 2004. www.oecd.org).

Gemäss Dr. med. Alex Schwank steht „den gestiegenen Kosten aber auch eine stark gestiegene Qualität in der medizinischen und pflegerischen Versorgung gegenüber. Die Lebenserwartung ist in der Schweiz auch in den letzten Jahren gestiegen, und dies deutlich stärker als in vergleichbaren Ländern wie zum Beispiel Holland. Das Geld wird im Gesundheitswesen nicht einfach verschwendet und der überwiegende Teil der Bevölkerung wünscht hier auch keine Sparprogramme und keinen Leistungsabbau.“ (Soziale Medizin, September 2009).

Vor dem Hintergrund der politisch verordneten Prämienexplosion sind die Kosten im Gesundheitswesen für viele SchweizerInnen zu einem finanziellen Problem geworden. Dies erhöht auch politisch betrachtet das Risiko, dass die Durchsetzung alter Forderungen zur Kostensenkung im Gesundheitswesen

Begleitforschung SwissDRG und das Problem der pluralistischen Ignoranz in der Statistik, 18.10.2009

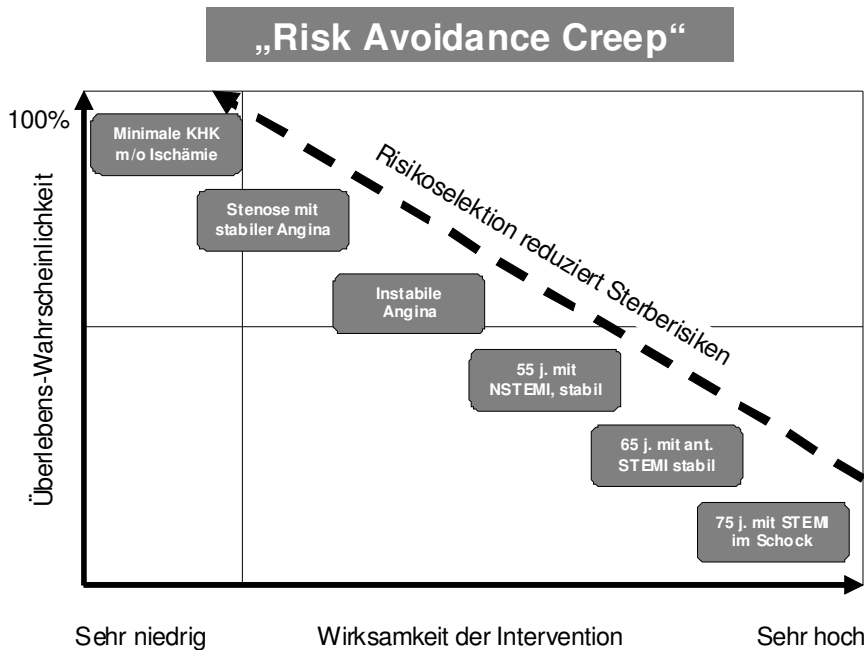
durchsetzbar geworden sind, insbesondere betreffend Kosten im Primary Care (Globalbudget) und Spitalschliessungen (politisch gewollte Reduktion der Baserate im DRG-System).

Diese einseitig auf Kostenreduktion ausgerichtete Budgetverantwortung für medizinisch begründete Interventionen entspricht einer falsch verstandenen „Ökonomisierung“ der medizinischen Versorgung der Schweizerischen Bevölkerung. Dies birgt erhebliche Risiken, insbesondere betreffend Verminderung der Versorgungsqualität, Verminderung der Patientensicherheit, vermehrte versteckte oder offene Rationierung notwendiger medizinischer Leistungen oder Behandlungsverweigerung von PatientInnen mit hohen Kostenrisiken.

Eine nie dagewesene Flut von Ratschlägen zum Monitoring des Gesundheitswesens wurde in den letzten Jahren von GesundheitsökonomInnen, medizinischen

Fachgesellschaften, Epidemiologen, Ethikern und der FMH publiziert (Referenzen 80-82, 88-93). Dabei kann ohne weiteres festgestellt werden, dass die Instrumente zur Erfüllung des Monitoring-Anspruchs mehrheitlich nicht validiert sind.

Das erklärte Ziel, Misstände als Folge von Globalbudgets und SwissDRG künftig und notabene ohne Verankerung dieser Aktivitäten im Krankenversicherungsgesetz aufzudecken, könnte jedoch weit verfehlt werden. Es ist sogar wahrscheinlich, dass die Monitoring-Instrumente – **insbesondere unter der simplen Verwendung von Mittelwertbeobachtungen und meist nicht validierten und arbiträren Korrekturfaktoren** – kontraproduktive Wirkung haben und zu einer Risikoselektion zum Schaden der PatientInnen führen, z.B. mit dem Mechanismus des „Risk Avoidance Creep“: Möglichkeit der Risikoselektion in Abhängigkeit von vermuteten Sterberisiken (Level of Evidence: IIb).



Im ökonomischen Diskurs über Ressourcen im Gesundheitswesen ist jedoch der **Einbezug von Kosteneffizienz-, Effektivitäts-, Qualitäts- und Gerechtigkeitsanalysen** notwendig (88).

Damit schafft die Gesundheitsökonomie professionelle und vor allem auch sachliche Diskussionsgrundlagen für politische Entscheide.

Davon abzugrenzen ist eine qualitativ ungenügende Gesundheitsökonomie, insbesondere wenn diese den publizierten Richtlinien (Hannoveraner Konsens Gesundheitsökonomie, 2007, Good Epidemiological Practice Konsens 2004) nicht genügen.

**Begleitforschung SwissDRG und das Problem
der pluralistischen Ignoranz in der Statistik, 18.10.2009**

Das grundsätzliche Problem betrifft den korrekten Gebrauch der verwendeten statistischen Methoden, die daraus ableitbaren Aussagen und die in den heutigen teils hysterisch geführten, ethischen Prinzipien entgegen gesetzten Aussagen in der Presse. Beispiel: Artikel „Last der kleinen Spitäler“ (NZZ am Sonntag, 23.08.2009): in diesem

Artikel ging es um eine bisher nicht publizierte (sic!) Studie, welche angeblich kantonale Unterschiede in der Behandlung von Brustkrebs im Frühstadium aufzeigte (Tabelle 1).

Tabelle 1: persönliche Mitteilung aus Predictors of geographic variation in adherence with minimal recommendations on diagnosis and treatment of early breast cancer: Results from a Swiss observational study. Silvia Ess, Markus Jörgler, Anita Savidan, Christine Bouchardy, Andrea Bordoni, Isabelle Konzelmann, Gernot Jundt, Nicole Probst, Harald Frick and Beat Thürlimann:

Table 1. Comparison of total scores, surgical score and non-surgical score according to the area of residence of the patient. The total score is the arithmetic sum of the surgical and not surgical component of the score

Region	N of cases	Total Score (MRS) (0-10 points)		Surgical Component (0-5 points)		Non-Surgical Component (0-5 points)	
		Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)
GE	900	9.13	(9.05-9.20)	4.36	(4.30-4.42)	4.76	(4.73-4.80)
VS	423	8.47	(8.35-8.58)	3.66	(3.56-3.76)	4.81	(4.76-4.86)
TI	617	8.24	(8.13-8.35)	3.72	(3.63-3.81)	4.52	(4.46-4.57)
BA	396	8.33	(8.18-8.48)	3.85	(3.73-3.96)	4.48	(4.41-4.56)
ZH	414	8.58	(8.45-8.72)	4.07	(3.97-4.16)	4.52	(4.43-4.61)
SGA	675	8.63	(8.54-8.73)	3.88	(3.80-3.96)	4.75	(4.71-4.79)
GG	319	8.30	(8.15-8.45)	3.79	(3.68-3.90)	4.51	(4.43-4.60)
Total	3744	8.60	(8.56-8.64)	3.95	(3.95-3.95)	4.65	(4.65-4.65)

GE: Geneva, VS: Valais, TI: Ticino, BA: Basel; ZH: Zurich, SGA: St. Gallen-Appenzell GG: Grisons-Glarus; CI: Confidence Intervall

Einerseits: Der Gesamt-Score von maximal 10.0 variiert im Mittel zwischen dem besten Kanton (9.1) und dem schlechtesten Kanton (8.2) mit 0.9 von 10 möglichen Punkten gering. Daraus prognostische Implikationen abzuleiten ist gewagt und eine - auch unkollegiale - Unterstellung. Zudem ist aufgrund der Abstracts – es sind deren zwei – nicht ersichtlich, wie die Krankengeschichten ausgewählt wurden, ob hier ein Selektionsproblem vorliegt und es ist nicht gesichert, dass die Unterschiede statistisch signifikant unterschiedlich sind. Ob die Autoren mit ihrer Arbeit den strengen Richtlinien genügend, müsste durch eine Peer-Review Prozess geklärt werden.

Andererseits: Das Hauptproblem – wie so häufig in Studien, die kantonale Unterschiede feststellen – ist auch hier das weitgehende **Fehlen der Ursachenforschung (Suche nach Confounders)**. Der Fragenkatalog an die behandelnden Ärzte, warum gewisse

Massnahmen nicht durchgeführt wurden, ist inexistent – auch in dieser Studie. Etwa so, wie wenn man nach den Ursachen von Lungenkrebs sucht und die Patienten nicht fragt, ob sie Zigaretten rauchen (**Zigarettenrauchen ist hier somit das Confounding**).

Statt also nach Ursachen zu fahnden, wurden im besagter Arbeit soweit erkennbar, Aussagen in der Tagespresse publiziert, dahingehend, dass kleinere Spitäler eine schlechtere Versorgungsqualität hätten, weswegen diese nun geschlossen werden sollen („Last der kleinen Spitäler“ vom 23.08.2009, NZZ am Sonntag).

Immerhin schreiben die Forscher in den Schlussfolgerungen (zweites Abstract): die Ursachen für geographische Unterschiede sind ungenügend erklärt, weiteren Studien müssen folgen.

**Begleitforschung SwissDRG und das Problem
der pluralistischen Ignoranz in der Statistik, 18.10.2009**

Zum Beispiel diese: das Problem für Unterlassungen in der Behandlung liegt mit hoher Wahrscheinlichkeit darin begründet, dass sich 70% der Aerzte (publiziert von Samia Hurst, Bioethisches Institut Universität Genf) in ihrer Freiheit, die beste Wahl für ihre Patienten zu treffen, unfrei fühlen (freedom under pressure, internationale Vergleichsstudie: auf die Frage „My job or financial security is threatened by current pressures to contain costs in health care“ antworteten in der Schweiz 73% der Aerzte mit ja, in Italien 69%, in Norwegen und England je 18%).

Forschung zu versteckter Rationierung ist weitgehend inexistent.

Stillschweigende Unterlassungen von wichtigen primär und sekundär präventiven Massnahmen aus Angst vor Regressforderungen der Krankenkassen bei „zu teurer“ Behandlung - gerade bei der teuren Krebsbehandlung - sind ein zentraler Confounder in der Ursachenforschung regionaler Behandlungsunterschiede und müssten als weitere Variable berücksichtigt werden.

Das zentrale Problem ist somit, dass die Monitoring-Methoden die Grundsatzfrage der Ursachen für beobachtbare Unterschiede ungenügend beleuchten und damit Raum lassen für unzulässige Schlussfolgerungen.

Tabelle 2 (aus 92a): Vergleich Outcome nach Herzinfarkt 2008: Angegeben sind die vom Universitätsspital mit der vom Bundesamt für Gesundheit vorgegeben Kodierung erhobenen Daten (BAG) sowie die von wissenschaftlichen Assistenten der Klinik für Kardiologie nach Durchsicht aller Berichte des Jahres 2008 sorgfältig erhobenen und durch Verlaufbeobachtungen ergänzten Daten zur 30-Tage-Mortalität (KAR). Beachtenswert sind die unterschiedlichen Fallzahlen in der kodierten und wahren Sterblichkeit.

Variable	Daten	Fallzahl	Todesfälle	Sterblichkeit (%)
Herzinfarkte total	BAG	125	14	11,2%
	KAR	415	11	2,7%
Altersgruppe 20–44	BAG	7	0	0,0%
	KAR	30	0	0,0%
Altersgruppe 45–64	BAG	56	3	5,4%
	KAR	211	5	2,3%
Altersgruppe 65–84	BAG	49	5	10,2%
	KAR	150	6	3,8%
Altersgruppe ≥85	BAG	13	6	46,2%
	KAR	13	0	0%
Fälle mit NSTEMI	BAG	14	0	0%
	KAR	135	2	1,5%

Internationale Konsensuspapiere in Gesundheitsökonomie und Epidemiologie fordern deshalb u.a. (93): „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.“

Das Problem des Confounding wird in den bisher zugänglichen Schriften zum Monitoring negativer Auswirkungen der SwissDRG aus den Kreisen der Ökonomie, Epidemiologie und Ethik nicht erwähnt (90,92b).

Ein weiteres Problem ist die Qualität der Daten. Wie kürzlich in einer Arbeit von Prof. Lüscher vom Universitätsspital Zürich gezeigt werden konnte, stimmen die Zahlen der Bundesamtes für Gesundheit betreffend Sterbehäufigkeit nach Herzinfarkt nicht mit den in einer Chartreview erhobenen Daten überein (Tabelle 2).

Monitoring Probleme: Beispiele zu Confounding und dem Positionspapier der FMH (90, Aktualisierte Versionen unter: http://www.fmh.ch/tarife/swissdrg/position_fmh.html)

Beispiel 1 (Aspirin als Qualitätsbenchmark für Spitäler bei der Behandlung eines akuten Herzinfarktes): Aspirin bei Herzinfarkt in den Empfehlungen der Schweizerischen Akademie der Medizinischen Wissenschaften zur Erhebung, Analyse und Veröffentlichung von Daten über die medizinische Behandlungsqualität (92b). In diesem Bericht wird davon ausgegangen, dass bei einem akuten Herzinfarkt die Gabe von Aspirin zur Vermeidung eines weiteren Herzinfarktes als Benchmark tauglich ist, da die Richtlinien diese Behandlung empfehlen. Im Bericht wird gefolgert, dass Spitäler, welche diese Richtlinie nicht befolgen, qualitativ ungenügende Herzinfarkt-Behandlungen anbieten und dank diesem Benchmark erfasst werden können. Hier wird das grundsätzliche Problem des Confounding nicht diskutiert: bevor nämlich der Aspirin-Benchmark Verwendung finden darf, müssen Confounders wie akute gastrointestinale Blutung, Allergie auf Aspirin oder nachgewiesene Unwirksamkeit von Aspirin im Thrombozyten-Funktionstest ausgeschlossen werden.

Die Frage, ob Aspiringabe beim Herzinfarkt als Benchmark für Qualitätskontrollen und als Monitoring für negative Auswirkungen von SwissDRG auf die Versorgungssicherheit tauglich ist, muss erst noch in weiteren wissenschaftlichen Studien geklärt werden.

Es kann nicht sein, dass reine Betrachtungen von Raten ohne breite Ursachenforschung (Confounderforschung) für Unterschiede verwendet werden, welche dann für politisch motivierte Aussagen herangezogen werden und letztlich der Versorgungsqualität mehr schaden als nützen.

Beispiel 2 (Falsche Benennung von ausländischen Beobachtungszahlen zur kardiovaskulären Hochrisikosituation in der Schweiz durch namhaften Gesundheitsökonom): Kardiovaskuläre Prävention ist eine der zentralen Aufgaben

unserer medizinischen Versorgung: sie senkt das Risiko für Herzschlag, Hirnschlag, Krebs und Demenz. Ungenügende Kontrolle der drei Risikofaktoren Rauchen, hoher Blutdruck und hohes Cholesterin führen zu einer Verminderung der Volksgesundheit, zu vielen vermeidbaren Fällen von Herz- und Hirnschlag, aber auch Krebs und zu einer Kostenerhöhung (besonders Pflegekosten) im letzten Lebensjahr. Behandlung von hohem Blutdruck und hohem Cholesterin ist kosteneffizient, senkt signifikant die vaskuläre Morbidität und Mortalität und ist in bestimmten Studien sogar in der Lage, die Gesamtsterblichkeit zu senken. Ein besonderes Beispiel für die Negierung von Evidenz wurde kürzlich in der Schweizerischen Aerztezeitung publiziert³. In diesem Artikel moniert der Autor, dass gemäss einer Studie aus Norwegen praktisch alle Personen ab dem 40. Lebensjahr Personen ein hohes vaskuläres Risiko hätten, was ein Schlaglicht auf die ungenügende Qualität der Risikovorhersage werfen würde⁴. Da diese Aussage auch noch in den Printmedien übernommen wurde (Irene Dietschi im Beobachter-Beitrag „Haben Sie Ihre Diagnose schon“.), ist dem Autor Aufmerksamkeit offensichtlich gewiss. Aus diesem Grund ist es hier angebracht, diese Aussagen zu kommentieren.

Die europäischen und internationalen Richtlinien für die Kategorisierung des Risikos in Risikoklassen für Herzinfarkt sind das Resultat einer grossen Leistung der Konsensfindung. Zur Risikovorhersage werden die wichtigsten unabhängigen Risikofaktoren verwendet, insbesondere Rauchen, Blutdruck, Blutfette, Alter, Geschlecht, Diabetes. Die Berechnung des Risikos erfolgt dann aufgrund der Ausprägung der Risikofaktoren anhand eines statistischen Modells, welches anhand von mathematischen Funktionen für jeden Risikofaktor das Risiko berechnet. Die Schweizerischen Richtlinien empfehlen dabei

³ http://www.saez.ch/pdf_f/2009/2009-24/2009-24-1267.PDF

⁴ <http://www.bmj.com/cgi/reprint/331/7516/551>

Begleitforschung SwissDRG und das Problem der pluralistischen Ignoranz in der Statistik, 18.10.2009

ausdrücklich, in einem ersten Schritt den elektronischen Rechner und nicht die publizierten Printtabellen zu verwenden, da diese je nachdem, wo bestimmte Risiken zu liegen kommen, in der Berechnung des tatsächlichen Risikos daneben liegen können. In der besagten Arbeit in Norwegen wurde aber genau so vorgegangen und das Resultat auch noch auf das hypothetische Risiko bis zum 60. Lebensjahr hochgerechnet, was den Systemfehler nochmals verstärkt. Zudem steht diese Studie mit ihren Aussagen weit und breit isoliert da. Für die Schweiz existieren zudem ganz andere Zahlen, welche aber von Domenighetti nicht erwähnt wurden. Gemäss einer eigenen Erhebung bei Personen ab 45 Jahren befinden sich nicht 90% sondern insgesamt 3% im Hochrisikosegment und die Indikation für ein Statin zur Senkung des Cholesterins kann nur in 6% der Fälle gestellt werden⁵. Da die Frage, ob eine Hochrisikosituation besteht, häufig in der Arztpraxis erfolgt, ist es zudem interessant, wie häufig diese bei Praxispatienten gefunden wird. Gemäss eigenen Zahlen aus einer kardiologischen Praxis fanden wir eine Hochrisikosituation bei asymptomatischen Patienten bei 0 % (anhand des AGLA-Rechners, welcher auf der PROCAM Studie basiert und für die Schweiz mit einem Faktor 0.7 adaptiert wurde). Zudem kann die Indikation für eine Statinbehandlung bei 12.2% der Patienten gestellt werden, selbst wenn die internationalen Richtlinien verwendet würden. Die gesundheitsökonomischen Bedenken von Prof. Domenighetti lassen sich für die Schweiz nicht objektivieren, die Prüfung der Aussagen durch einen unabhängigen Ethikrat für Gesundheitsökonomie wäre hier vor der Publikation sicher sinnvoll gewesen.

Beispiel 3 (untaugliche Statistik der santésuisse in der Erfassung von Überarztung, 97): Das Paradebeispiel für das Problem der pluralistischen Ignoranz in der Statistik ist die santésuisse Methode zur Beurteilung der Wirtschaftlichkeit einer Arztpraxis. Zur Beurteilung der Wirtschaftlichkeit eines Arztes gemäss KVG verwendet die santésuisse als Raster einen cutoff der mittleren Kosten pro Patient von +30%. Als Pseudostatistik muss die ANOVA-Methode in der Verwendungsart durch santésuisse bezeichnet werden, da diese vorgibt, Unterschiede unter den Mittelwerten auszugleichen und damit angeblich

Erklärungsgehalt zu liefern. Da für die ANOVA Methode der santésuisse jedoch lediglich die Variablen Alter und Geschlecht (sowie Wohnkanton) verwendet werden, beträgt der Erklärungsgehalt dieser Methode für die Kosten ca 11% (siehe: Glossar). Ärzte, welche über dem cutoff von 30% liegen, gelten trotzdem als („unökonomische“) Überarztter und können am Versicherungsgericht eingeklagt werden, wobei dieser cutoff als Beweismittel Verwendung findet. Es ist sehr schwierig, sich hier zu verteidigen, werden doch die Vergleichszahlen der Kollegen von santésuisse bewusst nicht verfügbar gemacht. Da beim Entscheid, ob die Verteidigungsstrategie des angeschuldigten Arztes aufgeht oder nicht, ist wiederum primär die santésuisse zuständig, so dass auf dieser Ebene der Beweisfindung die santésuisse als Klägerin und Richter in Personalunion auftritt. Die likelihood Analyse zeigt nun, dass selbst wenn dieser cutoff von +30% eine Sensitivität und Spezifität von 90% besässe (was für jeden Test ein hervorragender Leistungsausweis ist), die Nachttest-Wahrscheinlichkeit für Überarztung 16% beträgt, da generell davon ausgegangen wird, dass echte Versicherungs betrüger unter den Aerzten zu ca 2% (=Vortest-Wahrscheinlichkeit) zu finden wären. Das Gutachten des Seminars für Statistik der ETH Zürich, welches von der santésuisse in Auftrag gegeben wurde, beurteilt jedoch diese Fragen nicht, wie auch die Frage der Sensitivität und der Spezifität der santésuisse Statistik ausser acht gelassen wird. In diesem Gutachten werden weit herum bekannte statistische Grundsätze zitiert als Validierung der santésuisse Statistik, Aussagen welche auf Gymnasium-Niveau gelehrt werden. Damit stellt dieses Gutachten ein Gutachten zu allgemein statistischen Prinzipien auf, ohne selber das Vorgehen der santésuisse in der Erfassung von Überarztung wirklich zu prüfen. Es ist zudem bekannt, dass die santésuisse systematisch keine Kontrolle der Quelldaten gestattet, dies auf Ebene einer Studie von Schweizer Trustcenters als auch auf Ebene der Schweizerischen Aerztegesellschaft als auch auf Ebene des Schweizerischen Ethikrates für Statistik. Die Verweigerung, Quelldaten zugänglich zu machen, beweist bis zum Beweis des Gegenteils, dass santésuisse mit sensiblen Zahlen über das Aerzterating unfair umgeht.

Beispiel 4 (aus dem Positionspapier Begleitforschung der FMH 09/2009): Qualitätsförderung und therapeutische Freiheit:

⁵ <http://scopri.ch/CordicareI RomanensCDT012009.pdf>

Begleitforschung SwissDRG und das Problem der pluralistischen Ignoranz in der Statistik, 18.10.2009

Die Rechtsgrundlage der wohlgemeinten Äusserungen ist nicht definiert. Bisherige Erfahrungen belegen, dass gerade die therapeutische Freiheit je nach Sanktionsdruck erheblich eingeschränkt wird (S. Hurst, Bioeth. Institut, Genf: „freedom under pressure“).

Beispiel 5 (aus dem Positionspapier Begleitforschung der FMH 09/2009):

Deckung der Spalkosten. Ziel der DRG ist die Senkung der Fallkosten pro DRG. Die kostenmässigen Behandlungsrisiken bei Polimorbidität oder unvorhersehbaren Komplikationen ist gerade nicht Gegenstand des DRG Systems, da die Kostenrisiken der Behandlung auf die Institutionen verschoben werden soll. Alle Bemühungen, hier Kostenrisiken zurückzuverschieben weichen den brutalen Auftrag der DRG auf. Forderungen in Richtung solcher Aufweichungen müssen rechtlich geregelt werden, sollen sie überhaupt Gehör finden. Reine Absichtserklärungen werden im künftigen brutalen Überlebenskampf der Spitäler keinerlei Gehör finden. Willkürliche Senkungen der Baserate werden zu Spitalschliessungen führen.

Beispiel 6 (aus dem Positionspapier Begleitforschung der FMH 09/2009):

Transparente Leistungserfassung und Vereinfachung des Verfahrens: Die Leistungserfassung basiert auf Diagnosen. Die DRG sind grundsätzlich manipulierbar. Hier Transparenz zu fordern oder eine Vereinfachung des Verfahrens ist den Aussagen zu den Spalkosten diametral entgegengesetzt. Es sind gerade die Simplifizierungen, welche Transparenz und gerechte Bezahlung von Behandlungsrisiken verhindern. Es ist deshalb zu fordern, dass parallel zu den DRG ein PCG-basiertes System validiert wird, welches dann für die DRG Überprüfung eingesetzt werden kann. Es ist nachgerade unverständlich, dass der Trumpf der PCG als Schutz vor ungenauen DRG nicht verwendet wird. Selbst der Bund bearbeitet mit wissenschaftlichen Instrumenten die Validierung von PCG im Aramis-Projekt

„Eine der Hauptlücken bei den Statistiken zum Gesundheitswesen besteht darin, dass bei der ambulanten Versorgung keine Angaben über die Diagnosen erhoben werden. Zudem werden die heutigen Statistikangaben aus der Perspektive einzelner Leistungserbringer oder der Kostenträger erhoben, ohne dass die jeweiligen Einzeldaten der Betroffenen Patienten zusammengeführt werden können. Dementsprechend sind weder die den Behandlungen zugrunde liegenden Diagnosen noch deren Kosten bekannt. Damit fehlen die grundlegenden Voraussetzungen zur Beurteilung der Behandlungserfolge und der Behandlungskosten. Im Rahmen der „Machbarkeitsstudie Datenintegration zur Analyse von Erkrankungsverläufen und zur Darstellung der Diagnose bezogenen Leistungsangaben der ambulanten Versorgung“ (im Folgenden: Machbarkeitsstudie Datenintegration) wurden die methodischen und informatikmässigen Voraussetzungen bereitgestellt, um anhand vorhandener Daten der Krankenversicherer, des BFS und der abgegebenen Arzneimittel (PCG!) eine Klassifizierung der Behandlungen vornehmen zu können. Auf dieser Basis ist im Rahmen der vorliegenden Vertrags anhand der Beispiele der Krebspatienten, der Diabetiker und der Psychatriepatienten der Nachweis zu erbringen, dass die Methoden in der Praxis erfolgreich anwendbar sind.“

<http://www.aramis.admin.ch/Default.aspx?page=Texte&projectid=20046>

Beispiel 7 (aus dem Positionspapier Begleitforschung der FMH 09/2009):

Massnahmen bei Verlagerungen stationär ↔ ambulante. Die hier gemachten Äusserungen sind keine Schilderung von Massnahmen sondern Gummisätze ohne jeglichen praktischen Umsetzungswert. Mit reinen Absichtserklärungen – die sicher generell mit heftigem Kopfnicken beantwortet werden – kann die ungerechte und teils gefährliche Entwicklung des DRG Abgeltungssystems nicht verhindert werden.

**Begleitforschung SwissDRG und das Problem
der pluralistischen Ignoranz in der Statistik, 18.10.2009**

Tabelle 3: Schätzung der Sicherheit (Präzision) einer Aussage (modifiziert nach Circulation. 2009;119:1977-2016)

	Klasse I: Nutzen >>> Risiken.	Klasse IIa: Nutzen >> Risiken.	Klasse IIb: Nutzen ≥Risiken. Weitere Studien notwendig.	Klasse III: Risiko ≥ Nutzen.
Level A: mehrere Populationen untersucht; Daten stammen aus mehreren randomisierten klinischen Studien oder Metaanalysen	Empfehlung: Massnahme ist nützlich / effektiv Evidenz: ausreichend	Empfehlung: Massnahme spricht für nützlich / effektiv. Evidenz: teilweise widersprüchlich	Empfehlung: Massnahme ist wenig etabliert Evidenz: mehrere Studien mit erheblicher Widersprüchlichkeit	Empfehlung: Massnahme ist nicht nützlich / effektiv, ev. gefährlich Evidenz: ausreichend
Level B: Wenige Populationen untersucht. Daten stammen nur aus einer Studie.	Empfehlung: Massnahme ist nützlich / effektiv Evidenz: nur wenige Studien	Empfehlung: Massnahme spricht für nützlich / effektiv. Evidenz: widersprüchliche Studienlage	Empfehlung: Massnahme ist wenig etabliert Evidenz: widersprüchliche Studienlage	Empfehlung: Massnahme ist nicht nützlich / effektiv, ev. gefährlich Evidenz: wenig ausreichend
Level C: Sehr wenige Studien. Nur Daten aus einzelnen Berichten und Konsensus von Experten.	Empfehlung: Massnahme ist nützlich / effektiv Evidenz: nur Expertenmeinungen.	Empfehlung: Massnahme spricht für nützlich / effektiv. Evidenz: teils divergierende Expertenmeinung	Empfehlung: Massnahme ist wenig etabliert Evidenz: nur divergierende Expertenmeinungen	Empfehlung: Massnahme ist nicht nützlich / effektiv, ev. gefährlich Evidenz: nur Expertenmeinung

**Begleitforschung SwissDRG und das Problem
der pluralistischen Ignoranz in der Statistik, 18.10.2009**

Ruf nach nationalem Monitoring-Konsens im Gesundheitswesen

Eine Publikation inkorrektur Daten oder eine irreführende Interpretation – insbesondere von Mittelwert-Statistiken – sind ethisch nicht vertretbar. Ein Monitoring-Konsens ist zurzeit in der Schweiz inexistent, insbesondere scheint die Gefahr, welche von Mittelwertvergleichen ohne gleichzeitige Confounding-Forschung ausgeht, ungenügend erkannt.

Das Positionspapier für die Begleitforschung aus Anlass der Einführung von SwissDRG der FMH belegt dies deutlich.

Grundsätzlich sollte die Validität einer Evidenz gemäss internationalen Kriterien geprüft werden (Tabelle 3).

Dies gestattet, Aussagen zur Validität einer Monitoringvariable genau zu definieren. Damit kann die Gefahr eingegrenzt werden, dass Klasse IIc Aussagen zu Klasse Ia Aussagen umgewandelt werden. In einer Publikation zu Monitoringresultaten müsste somit zu jeder Aussage der Evidenzgrad erwähnt werden, bevor politische Massnahmen eingeleitet werden.

Aus dem gesagten ergeben sich zusammenfassend folgende Forderungen an eine qualitativ hochstehende Durchführung des SwissDRG Monitorings:

**Begleitforschung SwissDRG und das Problem
der pluralistischen Ignoranz in der Statistik, 18.10.2009**

Forderungen an eine qualitativ hochstehende Durchführung des SwissDRG Monitorings und des Benchmarkings unter den Spitälern:

1. Aus der Sicht der Patientensicherheit ist das Positionspapier Begleitforschung DRG der FMH zu wenig ausgereift. Die zur Qualitätskontrolle vorgesehenen Instrumente sind zu einseitig auf harte Endpunkte anhand von Raten ausgerichtet, entsprechen nicht den heutigen Qualitätsstandards gesundheitsökonomischer und epidemiologischer Methodik und schaffen Anreize, welche dazu führen werden, dass hohe Risiken nicht mehr behandelt werden aus Angst vor der Generierung einer institutions-bezogenen erhöhten Mortalitätsrate pro Eingriff.
2. Für die Definition von Qualitätsindikatoren im stationären Bereich sind die vorab gesundheitsökonomischen Standards zu definieren, zu erforschen und zu begründen. Dabei kann es nicht genügen, im Benchmarking zwischen den Spitälern eine gewichtete „Manhattendistanz“ anhand von (zu?) stark gewichteten Casemix-Indexes zu verwenden, da damit die Leistungsorientierung der Spitäler zu wenig die Struktur bedingten Kosten einer Institution berücksichtigen könnte (95). Vor dem Abbau von 100 Spitälern in der Schweiz wäre zudem eine Bedarfsanalyse sinnvoll, welche Fehlbelegungen von Patienten in Akutspitälern untersuchen würde (ähnlich einer Vollerhebung von 1983 im Kanton Basel-Stadt, 96). Mangel an intellektuellem Leistungswillen ist für die Bewältigung der anstehenden Probleme ungeeignet (siehe dazu aus Ref. 95: „Zur Beschreibung der Leistungen würden eigentlich auch Indikatoren zu deren Qualität gehören. Leider ist es aufgrund der verfügbaren Bundesstatistiken noch nicht möglich, derartige Indikatoren abzuleiten, weshalb wir auf diese Dimension verzichten und ersatzweise davon ausgehen, dass das Qualitätsniveau konstant ist.“
3. Die Hauptträgerin der Begleitforschung kann nicht die SwissDRG AG sein (Interessenskonflikt).
4. Die Hauptstossrichtung des Qualitätsmonitorings muss unter Einschluss der pharmaceutical cost groups auch im stationären Bereich erfolgen. Es ist eine Kontrolle (Validierung) der DRG mit Hilfe der PCG notwendig.
5. Es muss eine nationale Taskforce „Patientensicherheit und DRG“ zur Qualitätskontrolle der DRG und der Spitalleistungen ernannt werden. Diese sollte weitere Gremien berücksichtigen wie z.B. die SAMW, die Arbeitsgruppe um Yves Egli, Lausanne, das Zürcher Hortonzentrum unter der Leitung von Prof. J. Steurer und ev. dem Verein Physician Profiling, dem Verein Outcome, Verein Spitalbenchmark, der FMH, der Swiss DRG, der H+, dem Bund / BAG, der Kantone u.a.m. Die Aufgaben der Taskforce sind:
 - a. Definition eines nationalen Standards für die Monitoring-Methoden unter Einbezug einer „level of evidence“ berücksichtigenden Klassifizierung
 - b. Daraus abzuleiten und zu definieren: wissenschaftlich validierte Messinstrumente unter höchstmöglichen Qualitätsstandards der Messmethode selbst, speziell unter Berücksichtigung des Confounding Problems
 - c. Einbezug von PCG in die Qualitätssicherung auf stationärer und ambulanter Ebene zur Kontrolle der manipulierbaren DRG; damit Erfassung von „risk avoidance creep“ Effekten und (erzwungenen?) Fehl-Codierungen von DRG in den Spitälern
 - d. Ausarbeitung und Validierung nationaler Standards zur Qualitätskontrolle bis 2017
 - e. Erlass gesetzgeberischer Massnahmen auf Ebene KVG zur einstweiligen Verhinderung von Fehlentwicklungsmöglichkeiten, welche sich aus den DRG ab 2012 ergeben, insbesondere unter Berücksichtigung, dass die Finanzierung von Leistungen in „dubio pro reo“ bis auf weiteres auch gesetzlich festgelegt sind und dass bis 2017 politisch motivierte (künstliche) Senkungen der Baserate gesetzlich verboten sind.
 - f. Überprüfung der Auswirkungen von DRG bis 2017. Anschliessend weitere gesetzgeberische Anpassungen.
 - g. Allgemeine Zugänglichkeit der Forschung zur Generierung von relevanten Daten zu Spitalvergleichen, Baserate Erzeugnissen und Monitoring Resultaten.
 - h. Validierung der Benchmark-Methode zwecks Vergleich von Spitälern untereinander, z.B. anhand eines Tageszensus.

Schlussfolgerungen

Das heutige, qualitativ hochstehende Schweizerische Gesundheitswesen erfährt zurzeit einen Umbauprozess, dessen Auswirkungen auf die Versorgungssicherheit unbekannt sind. Der Ruf nach einem Monitoring zur Erfassung negativer Auswirkungen dieses Prozesses ist mehr als berechtigt. Die bisher präsentierte Methodik des Monitorings und des Benchmarkings zwischen den Spitälern ist zu wenig breit abgestützt, teilweise geheimniskrämerisch und berücksichtigt zudem in ungenügender Weise das methodische Problem des Confoundings,

also der eigentlichen Ursachenforschung für ermittelte Performance-Unterschiede. Ohne diese Grundlagenforschung können die Resultate des Monitorings für einen Abbau der medizinischen Versorgungsqualität und der Patientensicherheit missbraucht werden. Sämtliche Studien sollten eine Schätzung des Erklärungsgehalts der Daten (R^2) berechnen und daraus ableiten, ob die Realität in den Aussagen mit genügend Anstand berücksichtigt bzw. gemessen wurde.

Begleitforschung SwissDRG und das Problem
der pluralistischen Ignoranz in der Statistik, 18.10.2009

Literaturangaben

(auf www.physicianprofiling.ch/references.htm)

- 8a Implicit Rationing. Vortrag Samia Hurst. Bioethisches Institut der Universität Genf, 09/2008
<http://physicianprofiling.ch/rationierung.pdf>
- 80 Sterberaten BAG 2009
<http://physicianprofiling.ch/SterberatenBAGOriginal042009.pdf>
- 81 Sterberaten BAG 2009, Beilage
http://physicianprofiling.ch/SterberatenBAGQ_IndikBAG_Spez_V11_d.pdf
- 82 Sterberaten Kardiologie
<http://physicianprofiling.ch/PublicHealthHazardsMittelwerteSterberatenJACC032009.pdf>
- 88 Hannoveraner Konsens zur Gesundheitsökonomie 2007
<http://physicianprofiling.ch/PHYHannoveranerKonsensGesundheitsoekonomie2007.pdf>
- 89 Gesundheit kostet, in soziale Medizin 09/2009, Dr Alex Schwank
<http://physicianprofiling.ch/GesundheitHatIhrenPreisDrSchwankSozialeMedizin092009.pdf>
- 90 SwissDRG: Monitoring für eine nachhaltige Gesundheitspolitik erforderlich aus der Sicht der FMH
<http://physicianprofiling.ch/SwissDRGFMH092009.pdf>
<http://physicianprofiling.ch/SwissDRGKonzeptBegleitforschungFMH02009.pdf>
aktualisierte Version im Internet:
http://www.fmh.ch/tarife/swissdrg/position_fmh.html
- 91 SwissDRG: Monitoring aus ethischer Sicht
<http://physicianprofiling.ch/SwissDRGKriterienEthik092009.pdf>
- 92a SwissDRG: Vergleichszahlen können falsch sein, USZ hat nachgerechnet
<http://physicianprofiling.ch/SwissDRGStatistikProblemeLueschwer092009.pdf>
- 92b Qualität Datenerhebung der SAMW, Bericht
<http://physicianprofiling.ch/QualitaetDatenerhebungSAMW2009.pdf>
- 93 Richtlinien für die "Good Epidemiological Practice", 2004
<http://physicianprofiling.ch/GEPVollstaendig2004.pdf>
- 95 Benchmark Making zwischen den Spitälern, ein Konzept des BAG, BFS, Preisüberwachung
<http://physicianprofiling.ch/SpitalBenchmarkMethodeUndBeispieleBFS2006.pdf>
- 96 Pflegefälle und Akutbetten: Ergebnisse eines Tageszensus
<http://physicianprofiling.ch/RomanensTageszensusApril1983.pdf>
- 97 Moratoriumsforderung Wirtschaftlichkeitsverfahren der santésuisse
<http://physicianprofiling.ch/PHYRomanensSAeZ012009.pdf>