Prüfung der Wirtschaftlichkeitprüfung von santésuisse

Übersichts-Gutachten im Auftrag der Stiftung für Fairness im Gesundheitswesen (www.fairfond.ch)
unter
Einbezug des mathematischen Gutachtens (Modul 1 des Gutachtens 2), erstellt von Hochschuldozent Dr. rer. nat. habil. Walter Warmuth (Gesundheitsforen Leipzig) unter Beteiligung von Helmut Dahl und Dr. David Matusiewicz (ForBiG Forschungsnahe Beratungsgesellschaft im Gesundheitswesen)

Impressum

VEMS Autoren: Dr. med. Michel Romanens, Dr. med. Franz Ackermann

Version: 6.6

Nutzungsrecht: nur mit Quellenangabe:
www.physicianprofiling.ch/WZWStatistik2014.pdf
http://physicianprofiling.ch/WZWStatistics2014engl.pdf
www.physicianprofiling.ch/wzwwarmuth2014.pdf

Inhalt

Santésuisse-Prüfmethode führt zu Überarztung 5
Disclosure 6
Zweck 6
Finanzierung dieses Gutachtens 6
Öffentlichkeitsprinzip 7
Websites 7
Historisches 7
Weitere Schriften des VEMS 8
English Version 9
Appropriate Benchmarks for Swiss Physician Profiling – A Brief Review 9
Benchmarks between appropriateness and utility 9
Legislative Intentions 10
"Cost Comparison" at the Physician Level 11
The Anova Index of Health Care Insurers 12
The Appropriateness of the ANOVA "Cost Index" 13
International Profiling Standards 15
Conclusions 21
Zusammenfassung 23
Einleitung 27
Juristische Problemzonen 30
Verstoss gegen das Legalitätsprinzip 30
Nicht vom Index betroffene Kassen als Klägerinnen 31
Die Wurzel des Problems 35
Weitere „Fallacies" 48
Statistische Täuschungsproblematik („ecological fallacy and aggregation bias") in der ANOVA-Methode von santésuisse 48
ZSR Problematik 51
Grundsätzliche Problematik des RSS-Index 57
Implikationen der Versicherer Statistik auf die RSS Statistik 59

Hintergrund 60
Fragestellung 61
Ausführungen 61
Diskussion 62
Weiteres Vorgehen 62
Potentielle Implikationen 63
Wissenschaftliche Lösungsansätze und Gutachten 66
WZW santésuisse – LEVEL I/II 68
Regelkreise künftiger WZW-Modelle Level I/II 69
Basismodell mit Morbiditätsvariablen 70
Grundsätzliche Anforderungen an die Prüfmethode 72
Weitere Direktiven 73
Die Chance WZW 78
Weiteres Gutachten bestätigt VEMS 80
Einleitung 81
Frage 1. 82
Frage 2. 87
Frage 3. 89
Frage 4. 90
Frage 5. 94
Frage 6. 94
Frage 7. 98
Frage 8. 101
Frage 9. 101
Frage 10. 102
Frage 11. 103
Frage 12. 105

Frage 13. 106
Frage 14. 107
Frage 15. 108
Frage 16. 109
Schlussfolgerungen..... 110
Forderungen des VEMS 114
Weitere Informationen 115
Konstantin Beck zum Problem der aggregierten Daten und weiteren Aussagen vom 02.09.2013..... 115
Weiterführende Literatur: 117
Risikostrukturausgleich Für Versicherer und AerztInnen mit diversen Modellen und mit umfassenden RSA Beschreibungen und Empfehlungen 117
Grundlegende Arbeiten von C.Beck 117
Hohe Wichtigkeit..... 117
Literatur zur ZSR Frage (pdf) 117
Literatur Internet:..... 118
Weiterführende Literatur zu Physician Profiling 118
Anhang zu Schwenkglenks Mittelwert..... 119
Email Korrespondenz mit ETH Zürich (Seminar für Statistik) 124
Santésuisse verweigert die Herausgabe von zugesagtem Zahlenmaterial:..... 127

Santésuisse-Prüfmethode führt zu Überarztung

Im Gesundheitswesen gilt der empirische ermittelte Grundsatz, dass wenige Personen hohe medizinische Versorgungskosten benötigen um adäquat behandelt werden zu können („heavy-tail“ Verteilung der Kosten). Die Krankenversicherer erhalten die finanziellen Mittel über Prämien zwecks Begleichung der Rechnungen, welche in der solidarischen Krankenversicherung anfallen. Gleichzeitig sind die Krankenversicherer gesetzlich beauftragt, die Wirtschaftlichkeit zu prüfen. Können die das?

Die Krankenversicherer müssen darauf achten, dass die Ausgabenseite unter Kontrolle bleibt, um nicht defizitär zu werden. Wenn die Kosten ausufern, kann man entweder die Einnahmen erhöhen oder versuchen, die Kosten zu senken. Können Wirtschaftlichkeitsverfahren Kosten senken?

Aus Sicht der Versicherer wird das bejaht. Durch Bestrafung von Ärzten, die pro Patient zu hohe Kosten verursachen (Durchschnittskosten-Methode) meint man die Überarztung erfasst und gemassregelt zu haben. Doch das Gegenteil ist der Fall, denn hohe Kosten sind praktisch immer indiziert (z.B. Behandlung von Nierenschwäche, Depression und Krebs) und nicht Ausdruck von Überarztung.

Wo aber findet Überarztung statt? Wenn zu viele Aerzte zu viele Untersuchungen und Operationen bei Gesunden durchführen, dann sind die Kosten pro Patient immer noch viel niedriger als bei den schwer kranken Personen. Das bedeutet: die Durchschnittskosten sinken, die Gesamtkosten steigen. Die Wirtschaftlichkeitsprüfungen von santésuisse erzeugen damit Überarztung bei Gesunden und verhindern adäquate Behandlung bei den Kranken. Damit erfüllt die Methode zur Prüfung der Wirtschaftlichkeit, welche santésuisse zusammen mit den Vorgänger-Organisationen seit über 30 Jahren anwendet, nicht die gesetzlichen Vorgaben. Sie erzeugt das, was sie zu verhindern vorgibt: die Überarztung. Denn so kann sich der Arzt vor Regressforderungen als Folge der Durchschnittskosten-Methode retten. Funktionale Aerzte, welche ihre Patienten möglichst korrekt in Eigenversorgung und unter hohen Kosten behandeln, die allerdings weit höhere Kosten als Folge von unnötigen Weiterweisungen verhindern, werden zu dysfunktionalen Ärzten, welche ihr Können den gesünderen Personen vorbehalten und die teuren Patienten abstossen.

Der VEMS kritisiert die Wirtschaftlichkeitsverfahren von santésuisse seit 11 Jahren und hat die mathematische Obsoleszenz der santésuisse Methode von Anfang an thematisiert. Das nun vorliegende mathematische Gutachten (Dr. Warmuth u.a.), welches sich in seiner Unbestechlichkeit wohltuend abhebt von vorgefassten statistischen und ökonomischen Zahlenspielereien der Krankenversicherer, zeigt eines: die Prüfung der Wirtschaftlichkeit muss komplett neu erfunden werden.

Ungeeignete statistische Verfahren, welche allenfalls zur Kostenkontrolle bei Gesunden oder wenig kostenintensiven Kranken funktionieren könnten, können nicht über jene wenigen gestülpt werden, welche die meisten Kosten im Gesundheitswesen verursachen. Diese müssen durch andere Versicherungsformen, welche diese hohen Kostenrisiken abdecken, finanziert werden. Diese Grundregel muss neu im Schweizerischen Gesundheitswesen bestand haben. Nicht nur im ambulanten Bereich, sondern besonders auch im stationären. Erst dann können wir davon ausgehen, dass unmoralische Rationierung, häufig in verdeckter Form, sich nicht noch weiter ausbreitet. Jeder Durchschnittsvergleich von Kosten hat letztlich zum Ziel, dass sich diese einmitten. Dies bedeutet: Vermeiden von hohen, aber notwendigen Kosten. Diese Methode enthält damit einen diskriminierenden Aspekt gegenüber jenen, welche das Gesundheitswesen am meisten benötigen: den Kranken.

Disclosure

Keine Interessenskonflikte

Dr. Romanens hatte nie auffällige Indizes in der RSS Statistik, hatte nie Kostenauffälligkeitsprobleme mit der RSS Statistik und wurde auch nie von santésuisse wegen Kostenüberschreitungen angeschrieben.

Zweck

Die vorliegende Zusammenstellung dient dem VEMS und interessierten Personen als Grundlagenpapier für die künftige Ausgestaltung eines wissenschaftlich besser abgesicherten Physician Profiling. Zudem enthält dieses Prüfpapier die relevanten Informationen für eine Moratoriumsforderung für die Durchschnittskosten-Methode. Ferner dient der vorliegende Text der Erstellung weiterer juristischer und ökonomischer Gutachten im Auftrag der Stiftung für Fairness im Gesundheitswesen.

Finanzierung dieses Gutachtens

Stiftung für Fairness im Gesundheitswesen (www.fairfond.ch)

Die Stiftung Fairfond bittet Gönner, sich im folgenden Link weitere Details zu beachten:

http://www.fairfond.ch/goenner

Bisher nicht gedeckte Gesamtkosten des Projekts: CHF 25‘000.

Öffentlichkeitsprinzip

Die Kontrolle der WZW-Regeln im Rahmen der obligatorischen Krankenpflege erfolgt gemäss dem Bundesgesetz über das Öffentlichkeitsprinzip der Verwaltung vom 17.12.2004, worin die Krankenversicherer und deren Dachverbände subsidiär öffentliche Funktionen wahrnehmen und entsprechend zur Offenlegung und Transparenz der Daten und Prozesse verpflichtet sind. Aus diesem Grunde will der VEMS ein absolutes Öffentlichkeitsprinzip, um zu verhindern, dass Eigeninteressen der Versicherer die Versorgungssicherheit gefährden. Es ist für den VEMS nicht akzeptabel, wenn zentrale Fragen der Versorgungssicherheit hinter verschlossenen Türen diskutiert werden, dies umso mehr als die Verbände und ihre Vertreter sich in erheblichen Interessenskonflikten befinden, welche unsachgemässe Absprachen fördern oder ermöglichen.

Websites

www.physicianprofiling.ch
www.vems.ch
https://sites.google.com/site/physicianprofiling2012/

Historisches

Der VEMS engagiert sich zusammen mit den Vorgängerorganisationen seit 2003 für die Verbesserung der Wirtschaftlichkeitsverfahren. Aus dieser Arbeit entstand die Notwendigkeit einer breiteren Optik, tangieren doch die Wirtschaftlichkeitsverfahren zentrale Aspekte der ärztlichen Tätigkeit, die WZW-Regel und das Berufsethos. Daraus hat

der VEMS gerade in jüngster Zeit mehrere Grundlagen-Papiere erarbeitet, welche das vorliegende Grundlagenpapier zum Thema Wirtschaftlichkeitsprüfungen ergänzen und erweitern.

Weitere Schriften des VEMS

AG WZW FMH Official Educational Slides zu Aggregation Bias (M. Romanens et al.)
www.physicianprofiling.ch/FMHAnova102013V5.pdf
VEMS Trilogie zu ethischen Fragen im Gesundheitswesen (Zusammenfassung):
http://www.physicianprofiling.ch/VEMSTrilogie2014.pdf
Rationierung, QALY und mathematische Maschinerie:
http://www.physicianprofiling.ch/VEMSRationierung2014.pdf
Ethik-Carta des VEMS:
http://www.physicianprofiling.ch/VEMSEthikCarta2014.pdf
Zweckmässigkeitsforschung:
http://www.physicianprofiling.ch/VEMSForschung2014.pdf

English Version

Appropriate Benchmarks for Swiss Physician Profiling – A Brief Review

Romanens Michel, MD

Ziegelfeldstr. 1

CH-4600 Olten

Reprints only with the following citation:
www.physicianprofiling.ch/WZWStatistics2014engl.pdf

Version 0.3 (Draft, open for further co-authors)

Count: 18’576 paragraphs

Benchmarks between appropriateness and utility

The Swiss Health Care System is a public service that is obliged by federal law to use appropriate decisions to maintain and treat medically defined human conditions. Medical services have to choose effective assessments and treatments that apply appropriately to an individual’s condition. Since prevention and treatment of disease may not achieve the desired effects in a single patient, *ineffective treatments may nevertheless be appropriate*. When applying measures of utility to appropriate medical decisions, correct resource allocation is easily mislabeled as useless and wasteful in the case

where appropriate resource utilization does not yield the expected effect. Therefore, utility measures per se are not an acceptable benchmark for physician profiling from the view of the medical profession. Inherently, utility in physician profiling, where cost issues are mixed up with effectiveness and appropriateness[1] has a harmful potential for both needed appropriate medical services and the development of evidence based medicine that is studied based on individual single medical conditions and observations.

Legislative Intentions

According to Swiss federal law, medical activities have to be effective, appropriate and economic. Violations of this rule may be penalized. The reasons for the financial burden of a health care system can be viewed at the level of *treating physicians* or at the level of *treated patients*. From a view that incorporates the societal value of the health care system annual costs per capita should be minimized for a maximum of effectiveness. This goal is certainly not in conflict with medical activities at the individual level and is therefore universally acceptable. During treatment episodes, however, higher short term costs may prevent health derailments, thus preventing exploding costs over time. Therefore, effective and appropriate utilization of higher costs at the physician level may prove to be both effective and economical at the societal level. On the other hand, if *physician profiling acts as cost containment force* on the physician level, creating a sort of *global budget*, higher costs may be avoided in the short term, but this may lead to higher costs at the societal level

[1] http://physicianprofiling.ch/WZWBAG2011.pdf

(higher per capita annual costs). Thus, observation of costs without observation of effectiveness and appropriateness has important drawbacks and may pervert the legislative intention. Correct models of physician profiling should therefore ideally include all annual treatment episodes and costs, in order to remain functional benchmarks.

"Cost Comparison" at the Physician Level

Physician profiling comparing costs among physicians, e.g. by using average costs per annum and capita caused by a single physician, completely eliminates the societal interest with respect to effectiveness and appropriateness of his medical interventions at the overall cost perspective (*omitted variable bias*). Using costs at the physician level, the number of costly patients treated at the same costs by the comparator group of physicians, increase the "cost index" of this single physician. Therefore, according to the Simpsons Paradoxon, a high cost physician may be mislabeled as violating economic specifications, where in fact *this physician treats high cost patients to the same price but more frequently*. Therefore, cost observations at the physician level are not appropriate and frequently hampered by a *non-normal cost distribution*, which precludes the use of comparative statistics. These fundamental problems in current Swiss physician profiling practice actually pervert both the mandate of insurers to profile physicians as well as the mandate between patient and physician: ultimately, there is no guarantee – as defined by legislation – that high cost patients can be treated according to the central medical criteria of effectiveness and appropriateness. This is therefore an *illegal approach to physician profiling* and causes

perverted incentives for patient treatment episodes, mainly in ambulatory medicine, where this "cost index" is also used with vigor in order to prevent higher costs. The Swiss parliament has recognized this fundamental drawback of current physician profiling and decided, that better methods of physician profiling are to be developed between insurers and physicians, expecting to avoid the potential for malfunctioning of ambulatory medical services in the future.

The Anova Index of Health Care Insurers

In order to overcome the acknowledged mislabeling method "cost comparison", insurers introduced a few patient characterizing variables to correct for a physician's "cost index". They decided to use the ANOVA method, age/sex cells and the working canton of the physician, and the comparator group was extended from cantonal to national. In the insurer financed report from the ETH Zurich (Prof. W. Stahel) it was stated, that this method functions technically correct[2], however, when this method was tested in reality in the Canton Ticino, a second report from the ETH Zurich casted serious doubts on the validity based upon observed detrimental effects[3]. During the course of the examination of the insurers benchmark "ANOVA Index" by the "technical working group" of the Swiss Medical Association, insurers presented explanatory contents of their ANOVA model (expressed as adjusted r^2)[4]: at the national level, costs of anesthesiologists were explained by 93%, costs of surgeons by 71%, of psychiatrists by 42%, of cardiologists by 73% and of internal medicine physicians by

[2] http://physicianprofiling.ch/HealthEconomicsGutachten%20Anova_MDA_ETH_d.pdf
[3] http://physicianprofiling.ch/PHYGutachtenNr2DrRothETH052008.pdf
[4] http://physicianprofiling.ch/ZSRKanton2.pdf

82%. Based on their use of the ANOVA method, which is normally used to compare a maximum of 3 explaining non-continuous variables among to different groups, insurers first used a fixed effect variable based on each physicians "cost index"[5]. The variability of this "cost index" variable termed ZSR accounts for the observed high explanatory content of this variable, found to range between 33% and 67%. The addition of the age/sex cells increased the explanatory content of this insurer's model by another 16% to 62%. In this model, the canton effect was only marginal. Therefore, insurers correct to the major extent the "cost index" by the "cost index" (the ZSR variable) and to a lesser extent the "cost index" using age/sex cells. The rationale behind this is twofold: first, physicians are profiled as primary responsible for the costs, second by a morbidity variable age/sex, where insurers assume that increases in age for both sexes increase costs.

The Appropriateness of the ANOVA "Cost Index"

From a scientific point of view, one should ask what might be the rationale to explain medical costs (the "cost index") by the physician. One could argue that differences in skills and intelligence may have cost effects, or differences in the numbers of treated patients per physician or some other unknown but important explaining variables. But such explaining variables lie outside the scope of the insurers ANOVA model: the physician effect on costs remains the prevailing cost effect on the "cost index" and therefore, the legislative intentions

[5] http://physicianprofiling.ch/AGWZWZSRAgeSex.pdf

remain violated to a major and statistically important extent. Further of interest is the high explanatory content of the age/sex cells. Based on all available literature worldwide, the explanatory content of age and sex on prevailing or expected costs is about 4-5%. How can the insurers present explanatory effects of up to 62%? The solution of this problem lies in the aggregation bias. When individuals ages are aggregated into cells (e.g. age cells of five years intervals with averaged numbers for each individual), the variability (expressible as standard deviation of the cell) of the dataset is reduced to the average value of the age/sex cell. By eliminating the variability within each cell, the aggregation bias is mere inflation of the explanatory content and creates high correlations without a real meaning. Therefore, cost explaining by the age/sex cells is artificially inflated and should not be used for cost corrections, since the magnitude of such cost-index corrections are not due to real age effects on costs. In summary, the insurers ANOVA model is not appropriate (a more detailed discussion with additional references is given here[6]). Based upon the legislative intent, the operating insurer's physician profiling method is malfunctioning and should be abandoned.

[6] http://physicianprofiling.ch/WZWStatistik2014.pdf

International Profiling Standards

The future of appropriate physician profiling depends on accepted and validated models that are further improved by revalidation techniques that incorporate measures of reliability and of the mislabeling potential. The base for this procedure is a multivariate stepwise regression model that by its intelligent design avoids all the biases actually encountered in Swiss physician profiling activities (Table 1).

Table 1: Swiss Physician profiling list of biases (selected statistical fallacy problems)

Simpsons Paradoxon	Cost Index	More high cost patients treated with the same amount of costs penalize doctors who care for more high cost patients than the comparator group.
Omitted Variable Bias	Cost Index, ANOVA Index	"Fixed effects regressions are very important because data often fall into categories such as industries, states, families, etc. When you have data that fall into such categories, you will normally want to control for characteristics of those categories that might affect the LHS variable. Unfortunately, you can never be certain that you have all the relevant control variables, so if you estimate a plain vanilla OLS model, you will have to worry

		about unobservable factors that are correlated with the variables that you included in the regression. Omitted variable bias would result. If you believe that these unobservable factors are time-invariant, then fixed effects regression will eliminate omitted variable bias."[7]
Tautology Bias	Cost Index	costs are explained at the physician level without further differentiation of the physicians reasons to cause differences in average costs
Multicollinearity Bias	Regression, ANOVA Index	costs are explained by several independent variables that are correlated among them.
Aggregation Bias	ANOVA Index	costs are corrected using aggregated cells, e.g. of cost explaining variables such as age, where the effect is loss of the variance around the mean value, which creates a meaningless inflation of the r^2 statistics
Insurers Bias	Cost Index	Because of risk selection among insurers, some have patients with higher costs. Physicians caring for more patients insured by insurers with higher costs may generate higher costs
Cohort Size Bias	Cost Index	Physicians who care for less patients than the comparator group may have an

[7] http://www.physicianprofiling.ch/FixedEffectModelBias.pdf

		artificially high cost index
Insurers Cell Bias	Cost Index	A physician may have observed increased costs at the Insurer A, but not B, C and D. Therefore, a non-constant notable increase of average costs is a strong argument against a consequent problem of fraudulence
Comparator Group Bias	Cost Index. ANOVA Index	Differences of medical practice among doctors may lead to cost differences without violation of the economical rules. Such biases may be introduces by the physicians age, subspecialty, working place (part time in-hospital), amount of emergency activities, gender, vendor of pharmaceutics.
Endogeneity from omitted variable bias	Cost Index. ANOVA Index	"In a statistical model, a parameter or variable is said to be endogenous when there is a correlation between the parameter or variable and the error term.[1] Endogeneity can arise as a result of measurement error, autoregression with autocorrelated errors, simultaneity and omitted variables. Broadly, a loop of causality between the independent and dependent variables of a model leads to endogeneity. For example, in a simple supply and demand model, when predicting the quantity demanded in equilibrium, the price

		is endogenous because producers change their price in response to demand and consumers change their demand in response to price. In this case, the price variable is said to have total endogeneity once the demand and supply curves are known. In contrast, a change in consumer tastes or preferences would be an exogenous change on the demand curve."[8]

The basic concept of acceptable physician profiling incorporates *first of all* individual data *from patients*. This is due to defect that a physician does not treat patients but a patient's disease(s). These patients can be characterized by several cost explaining variables, e.g. treatment episodes for a certain disease, disease severity and socio-economic status, since all of them are correlated with a certain amount of treating time and medication costs. Medication may be used as a proxy for a disease (e.g. insulin is a proxy for insulin dependent diabetes mellitus) in so-called pharmaceutical cost groups (PCG). The list of selected possible cost explaining variables is given in Table 2.

[8] http://en.wikipedia.org/wiki/Endogeneity_(applied_statistics)

Table 2: independent cost explaining variables at the individual patient level (Modified from[9]). The r^2 statistic is a frequently used but problematic measure of the explaining content of the models

Variable	Description	r^2 (%)
Age / sex	Demographic variable of not aggregated information	4
ACG	Ambulatory care groups	16
CRG	Clinical risk groups	15
DCG	Diagnostic cost groups	17
CDPS	Chronic illness and disability payment system	12
Medicaid Rx	Prescription based model including PCG's	13
PRG	Ingenix Pharmacy Risk Groups	17

Using such models, each individual patient is characterized by a set of morbidity variables, eventually extended by a set of disease, socio-economic and disability severity variables. From this, both prospective and retrospective (concurrent) models may be created that reflect the *expected costs per patient*. Such results may further be refined using costs from previous years, e.g. for hospitalizations, for utilization of physiotherapy or total costs of medications. By allocating the sum of costs from the models to a specific physician, *costs may now be compared to the effective costs encountered in the patients of this specific physician*. The simple division of these two

[9] www.physicianprofiling.ch/ProfilingStateAndImprove2013Overview.pdf

cost amount creates the predictive ratio. Ideally, the predictive ratio is 1.0 or 100%. The great advantage of such an approach is the now possible abandonment of the comparator group needed to define the cost-index.

A further *essential and inevitable* refinement of such procedures is described by Dudley et al in 2003[10]: "There remains considerable uncertainty about whether prospective or concurrent risk adjustment (RA) is preferable. Although concurrent models have better predictive power than prospective models, the large payments associated with concurrent RA create incentives for fraudulent coding. A hybrid strategy - in which prospective payments were used for patients with low expected costs and concurrent payments were available upon the diagnosis of a small number of common, expensive conditions - might improve predictive performance while requiring less auditing than fully concurrent RA. In addition, within condition RA (using clinical data) for the selected conditions could further improve predictive power." Further, Dudley proposes targeted auditing of high cost patients in order to add a strong reality component the profiling process.

Finally, models have been developed in order to assess the reliability and the potential and amount of mislabeling using physician profiling models[11]. In fact, the golden standard in physician profiling is *as yet* not easy to define, but nevertheless urgently needed. According to data from Swiss insurers, the positive predictive value of the "cost index" for the correct detection of fraudulent medical behavior may

[10] www.physicianprofiling.ch/ProfilingDudleyMedCare2003.pdf
[11] www.physicianprofiling.ch/RAMisclassificationRAND.pdf

lie between 11% and 17%, assuming that in 100 tested doctors true positives are 3, true negatives are 70 to 80, false positives are 15 to 25 and false negatives are 2.

Using the pretest probability or prevalence of about 5% fraudulent doctors and the Bayes Theorem to calculate posterior probabilities of a positive test defined by a cutoff of 130% in the "cost index", the posterior probability is still just 11%[12]. Such observations should have juridical implications in Switzerland[13].

These numbers clearly show that the detection of fraudulent doctors has to be accompanied by an adequate tool to avoid mislabeling. Such a tool has been developed by the RAND Corporation[14]: "The error variance is specific to a physician and is a function of the number of episodes assigned to the physician, the mix of episodes, and risk adjustment. A physician who had a high proportion of episode types characterized by large variations in cost would have a large physician specific variation".

Conclusions

In this brief overview about the possible ways to profile Swiss physicians in the future, we found, that current profiling used by insurers has an unacceptably high potential for mislabeling and an irrational cost control mechanism which contains elements that hinder the normal functioning of the medical profession, especially in ambulatory medicine. Both the "cost index" as well as the "ANOVA

[12] www.ncbi.nlm.nih.gov/pubmed/19952758
[13] www.physicianprofiling.ch/FallacyJuryBayes2000.pdf
[14] www.physicianprofiling.ch/RAMisclassificationRANDNEJM2010.pdf

Index" is clearly flawed by test inaccuracies, inflated correlations and a too long list of avoidable biases. The goal of this communication is to *increase the sensibility in those responsible for the ultimate correct configuration of physician profiling in Switzerland*. As is hoped to be shown, cost explosion may be partly due to dysfunctional physician profiling. We propose that insurers and physicians start to talk seriously about physician profiling on a scientific level and based upon individual patient data sets for annually observed total costs.

But the real innovative aspect goes beyond the boundaries of physician profiling at the physician level. Future research should try to define, *how medical activities that are at the same time effective, appropriate and economic are to be defined and observed over the whole chain of Health Care System activities.* This work may show that higher than expected cost at the physician level in ambulatory medicine may save even higher costs at the stationary hospital level or even reduce mortality in primary care[15]. And therefore, any future "cost index" or predictive ratio should be corrected for the beneficial effect of costly but effective and appropriate medicine at the societal level.

[15] http://physicianprofiling.ch/VersorgungOutcomeStillfried022012.pdf

Zusammenfassung

Die Qualität des „physician profiling" hängt ab von der gewählten Methode, ein ideales Abbild des Patienten, also des „patient profiling" zu sein.

Statistisch einwandfrei Profiling-Instrumente erreichen jedoch nie einen höheren Erklärungsgehalt als 45% für die notwendigen Behandlungskosten.

In der heutigen von santésuisse praktizierten Methode des RSS-Index und ANOVA-Index besteht eine Kostenerklärung von bis zu 90%, dies Folge von verschiedenen Biases wie „confounding, aggregation, simpson's paradox, early treatment bias, collinearity, tautology, endogeneity from omitted variable bias[16]". Santésuisse perpetuiert eine bisher wissenschaftlich nicht begründete Methode unter unzulässiger Fokussierung auf die Arztkosten selbst, welche ja erklärt werden sollten, während die Behandlungsnotwendigkeit und die damit verbundenen statistischen Messgrössen wie Morbidität und Morbiditätsgrad nicht berücksichtigt werden.

Diese Unterlassung entspricht bei weitem nicht dem heutigen Standard und hat ungünstige, weil ausschliesslich ökonomische Anreize auf die Zweckmässigkeit ärztlicher Entscheide. Dies entspricht nicht der Intention des Gesetzgebers und blendet die weiteren Kriterien der Beurteilung, nämlich die Wirksamkeit und die Zweckmässigkeit aus.

Das Problem der „misclassification" ist wegen der schlechten „reliability" (RAND Corporation) der Methode von santésuisse

[16] http://www.physicianprofiling.ch/FixedEffectModelBias.pdf

erwiesenermassen sehr hoch (ca. 25% falsch positive im Sinne einer falschen Anschuldigung wegen Überarztung) und damit schädlich.

Insbesondere unter Berücksichtigung der ganzen OKP-Behandlungskette wird die Aerzteschaft, welche gewohnt ist, unnötige Kosten als Folge von Krankheiten und Hospitalisationen zu vermeiden, statistisch auffällig.

Damit werden die Benchmarks von santésuisse unzweckmässig und widersprechen den gesetzlichen Vorgaben, indem garantiert bleiben muss, dass der Arzt auch teure Patienten behandeln kann.

Dort, wo echte Überarztung hingegen häufiger wahrscheinlich ist, nämlich bei der medizinischen Überversorgung im niedrigen Risikobereich, wo die gesamten Behandlungskosten ansteigen, der Mittelwert pro Patient aber wegen niedriger Behandlungskosten sinkt, ist die santésuisse Methode überhaupt nicht sensitiv, im Gegenteil: sie fördert geradezu solche Mechanismen und ist damit selbst Ursache von Überarztung. Dies widerspricht klar dem gesetzlichen Auftrag.

Dies hat zudem das Potential für vermeidbare Kostenfolgen und zeigt das hohe Rationalisierungspotential, welches ein korrektes „physician profiling" erschaffen wird.

Zudem wird eine rein statistische Überprüfung immer problematisch bleiben, zeigen doch alle Studien, dass der Erklärungsgehalt, wie eingangs erwähnt, für Arztkosten nie den 40% Benchmark erreicht. Es wird deshalb unumgänglich sein, auffällige Kosten zu auditieren, um damit anstelle statistischer Vorverurteilungen einen *Realitätsbezug zur Zweckmässigkeit der eingesetzten Mittel* zu schaffen.

Zudem ist die statistische Methode (ANOVA, Regression) bei extrem schiefer Verteilung der Arztkosten per se inakzeptabel.

Ohne die Bereitschaft der Aerzteschaft, eine effiziente Selbstkontrolle durchzuführen, ist es unmöglich, adäquate WZW Verfahren durchzuführen. Wohin fachfremde Kontrolle führt, hat santésuisse zur Genüge vorgeführt.

In der Zwischenzeit müssen *rasch Lösungen erarbeitet werden*, welchen die aktuell verwendete statistische Methode und deren implizite unverantwortbare Kostenfolgen ablösen.

Dazu gehört auch Anpassung der Rechtssprechung an *realitätsbezogene Beurteilungen* und die Bereitschaft der Versicherer, *die primitive Datenbank der santésuisse durch Individual-Daten zu ersetzen*, um die gravierendsten Statistik-Missbräuche von santésuisse zu elimieren (insbesondere der „aggregation bias", die zahlreichen impliziten „Simpsons' Paradoxes" und die tautologische Erklärung der Arztkosten durch die Arztkosten anhand der „ZSR Variable".)

Auf der methodischen Ebene wird ferner die ANOVA Methode von santésuisse zur Beurteilung der Wirtschaftlichkeit, welche für zwei bis drei erklärende Variablen für den Vergleich des Mittelwertes von Gruppen entwickelt wurde, durch adäquate statistische Methoden ersetzt werden müssen, unter Ausschluss der Hochkosten-Patienten, welche praktisch nie wegen Überarztung teuer sind.

Hauptverantwortlicher für das heutige WZW-Verfahren ist nicht der Dachverband santésuisse sondern jeder einzelne Krankenversicherer in der Schweiz. Zudem bezahlen in der Schweiz traditionell häufig ihre Rechnungen „out-of-the-pocket", gemäss

eigener Studie erfolgen rund 15% der Kosten über Selbstzahlungen[17], welche die Kassen nicht erfassen und den Mittelwert verzerren.

[17] http://physicianprofiling.ch/rsspssstatistikbericht052010.pdf

Einleitung

Mindestens seit 1998 konnten sich der Durchschnittskosten-Index der Krankenversicherer und deren Dachverbände in Form des RSS-Index als Beweismittel für Überarztung in der Rechtsprechung durchsetzen. Die offensichtlichen statistischen und juristischen Mängel dieser einfachen Durchschnittskosten-Methode zur Beurteilung von Arztkosten betreffend Wirksamkeit, Zweckmässigkeit und Wirtschaftlichkeit hat zu einer breiten Ablehnung dieses Benchmarks geführt.

Santésuisse hat deswegen im Jahr 2005 die ANOVA-Methode entwickelt und von der ETH Zürich begutachten lassen. Im Jahr 2008 hat die santésuisse den ANOVA-Index als neues Beurteilungsinstrument im Rahmen des WZW Screening-Verfahrens eingeführt. Im Wesentlichen geht es darum, eine Korrektur des RSS-Index (ZSR Variable) betreffend erklärenden Variablen wie Alter und Geschlecht durchzuführen, in der Annahme, das ältere PatientInnen auch mehr Kosten verursachen würden. Zudem wurde die Vergleichsgruppe statt wie im RSS-Index kantonal auf die ganze Schweiz ausgedehnt.

Die Methode wurde damals durch ein Gutachten der ETH Zürich überprüft und als ausreichend beurteilt[18]. Allerdings: in einem zweiten Gutachten im Auftrag der Aerztegesellschaft des Kantons Tessin zeigte sich, dass die Neu-Beurteilung (ANOVA-Index anstelle des Rechnungssteller-Index) zu einer deutlichen Verschiebung der Zahl auffälliger Aerzte bei Allgemeinmedizinern, Internisten und Pädiatern führte[19]. Die Gutachter der ETH Zürich befanden daraufhin,

[18] http://physicianprofiling.ch/HealthEconomicsGutachten%20Anova_MDA_ETH_d.pdf
[19] http://physicianprofiling.ch/PHYGutachtenNr2DrRothETH052008.pdf

dass trotz der ANOVA-Methode, welche ja für kantonale Unterschiede korrigieren soll, weitere Besonderheiten vorliegen könnten, weswegen die Methode ungenügend sei. Dieses zweite Gutachten wurde durch den VEMS 2008 publiziert, worauf eine juristische Intervention durch den Tessiner Kantonalpräsidenten Dr. Denti erfolgte, das Gutachten sei aus dem Internet zu entfernen. Die FMH hat sich nie dafür eingesetzt, das ETH Gutachten durch unabhängige Experten beurteilen zu lassen.

Materiell hat der ANOVA-Index bis heute zudem überhaupt keinen Einfluss auf die Rechtsprechung. Er wurde bisher gemäss Aussage santésuisse in dem Sinne verwendet, dass zur Entlastung der AerztInnen ein auffälliger RSS-Index nur dann berücksichtigt wurde, wenn der ANOVA-Index über 130% läge. Dies führte gemäss Ausführungen der santésuisse im Jahr 2011 zu einer Reduktion von auffälligen AerztInnen von ca. 2500 auf ca. 2300.

In Anbetracht der anhaltend unbefriedigenden Situation hat das Parlament nach der Motion Heim / Meyer / Cassis vom 11. Mai 2010 beschlossen, den Art. 56 KVG dahingehend zu ändern, dass die WZW Verfahren zusammen mit der Aerzteschaft verbessert werden müssen.

Seither erarbeiten die FMH, santésuisse, diverse Versicherer, der VEMS und andere Vertreter die Grundlagen für die neue Ausgestaltung des WZW Verfahrens. Ziel dabei ist es, ein Verfahren zu entwickeln, welches die Zahl falsch positiver und falsch negativer Aerzte im Screening-Verfahren betreffend Überarztung senkt. Seitens der Versicherer besteht dabei, die Verantwortung des Arztes für die Kosten im Analysemodell zu bevorzugen, während für die Aerzteschaft ein Analysemodell im Vordergrund steht, welches primär

eine auf den einzelnen Patienten zentrierte Aussage gestattet: denn der Arzt behandelt nicht die Patienten sondern die Krankheiten der Patienten. Ziel ist es, die aktuell unbefriedigende Situation möglichst schnell auch materiell zu ändern.

Juristische Problemzonen

Verstoss gegen das Legalitätsprinzip

Die santésuisse muss im Auftrag der Krankenversicherer die Wirtschaftlichkeit überprüfen. Gemäss FMH war die damit verbundene Rechtsprechung jedoch problematisch[20]: „Diese Rechtsprechung stiess in der Lehre immer wieder auf Kritik. Die Lehre sah in Art. 56 Abs. 2 KVG keine genügende gesetzliche Grundlage für die «Rückforderung der veranlassten Kosten und erachtete diese Rechtsprechung damit als Verstoss gegen das in der Bundesverfassung verankerte Legalitätsprinzip[21] [22]."

Es war deshalb im Sinne der Aerzteschaft, dass die santésuisse versuchte, statistische Korrekturverfahren für die verursachten Arztkosten zu benutzen, in Analogie zum Risikostrukturausgleich (RSA) unter den Krankenkassen[23]. Ziel der Prüfung muss dabei jedoch letztlich nicht nur die Wirtschaftlichkeit sein, sondern eben auch die Prüfung der Wirksamkeit und der Zweckmässigkeit. Da dies mit reinen Arztkostenvergleichen nicht zu leisten ist, forderte die Rechtsprechung für Index-Unterschiede eine Sicherheitsmarge von 20-30%. Der ANOVA-Index sollte nun die Arztkosten adjustieren, beispielsweise für Alter und Geschlecht, in der Annahme, dass diese Variablen als unabhängige Variable einen statistisch signifikanten Einfluss auf die Arztkosten haben. Wegen angeblicher Datenschutzprobleme lieferten dann die Krankenversicherer die Daten in aggregierter Form. Dies bedeutet, dass pro Arzt Mittelwerte

[20] http://www.saez.ch/docs/saez/archiv/de/2011/2011-37/2011-37-809.PDF
[21] Das Legalitätsprinzip besagt, dass jegliches staatliche Handeln einer gesetzlichen Grundlage bedarf.
[22] BGE 137 V 43 E. 2.4 m.w.H.
[23] http://physicianprofiling.ch/PHYBeck2.pdf

der Arztkosten pro Alters- und Geschlechtskategorie übermittelt wurden.

Nicht vom Index betroffene Kassen als Klägerinnen

Der RSS-Index wird aus Versichererdaten zusammengesetzt, jeder Versicherer kann den Index jeden Arztes berechnen (RSSV). Dabei stellen sich zwei Probleme:

1) Ein Arzt kann zwar im RSS-Index von santésuisse auffällig sein (> 130%), dies muss aber nicht bedeuten, dass dies bei allen Kassen, an welche er Rechnung stellt, der Fall ist. Kommt es zur Anklage, kann es sein, dass santésuisse Versicherer vertritt, welche selber vom Index>130% gar nicht betroffen sind. Dies ist verfahrensrechtlich problematisch;

2) der RSA (gemeinsame Einrichtung KVG) zeigt, dass bestimmte Kassen teurere Patienten versichern. Dies muss für die Bildung des RSS-Index berücksichtigt werden (RSA auch für den Arzt). Santésuisse soll aufgefordert werden, zu diesen beiden Problemen Stellung zu beziehen (erfolgte am 18.02.2014 in der technischen Arbeitsgruppe durch den Autor) und ihre künftige Strategie dazu offen legen. Da es sich um ein Verfahrensproblem zur WZW Regelung handelt, müssen diese beiden Probleme quantifiziert und anschliessend zusammen mit der FMH neu überdacht werden.

Es kann sein, dass beispielsweise ein RSS-Index von 150% zusammengesetzt ist aus einem RSSV bei Kasse A von 90%, bei Kasse B von 120% und bei Kasse C von 250%. Der resultierende RSS-Index von 150% wäre dann ausschliesslich verursacht durch Kasse C.

Falls Kasse A und B als Klägerin auftritt, kann dies problematisch sein, insbesondere Kasse A hätte allenfalls einen Gewinn denn einen Schaden zu beklagen. Falls Kasse A und B als Klägerin auftreten würden, ergeben sich folgende Überlegungen: „In der Praxis hängt die Klageberechtigung nicht davon ab, ob eine Kasse die Marke von 130 Indexpunkten überschreitet. Nach einem Urteil des Eidgenössischen Versicherungsgerichts (K 39/95 11.07.1996 E. 5e) muss der geprüfte Arzt gegenüber einem Krankenversicherer nicht einen Index von >100% aufweisen oder gegenüber allen verfahrensbeteiligten Krankenversicherern den Durchschnitt überschreiten; es bleibe der internen Ausmarchung der Krankenversicherer überlassen, unterschiedlichen Fallwerten von Versicherer zu Versicherer im Rahmen der internen Aufteilung der Rückerstattung Rechnung zu tragen. Das ist natürlich aus Sicht der juristischen Dogmatik des Prozessverfahrens alles andere als korrekt, wird aber von der Rechtsprechung im Interesse einer einfachen und wirksamen Wirtschaftlichkeitsprüfung in Kauf genommen. Diese Nachsicht liess sich in den Jahren, als die Prüfpraxis des Durchschnittskostenvergleichs entwickelt wurde, rechtfertigen. Bei den heutigen Mitteln der Informatik dürfte das nicht mehr der Fall sein. Diese Rechtspraxis kann somit materielle Auswirkungen auf die Prüfergebnisse haben.

Gefordert ist, dass eine klagewillige Kasse in der Prüfperiode Leistungen zu Lasten einer Versicherung abgerechnet hat (Urteil des Eidgenössischen Versicherungsgericht EVG K 61/99 E. 3). In K 6/06 E. 3.3 hat sich das EVG nun aber zur Aussage verstiegen, es schade auch nicht, wenn einzelne Kläger in der fraglichen Zeit keine Leistung erbracht hätten; diese werde gegebenenfalls bei der internen

Aufteilung eben leer ausgehen. Das ist aber eine völlig unnötige Missachtung von Prozessrechtsgrundsätzen. Das gegenteilig lautende Urteil EVG K 61/99 E. 3 hatte das EVG offensichtlich übersehen. Das KVG sagt ferner nicht aus, unter welchen Voraussetzungen ein Krankenversicherer oder ein Verband zu einem Bussenantrag legitimiert ist. Es kann begründet werden, dass ein einzelner Krankenversicherer als Antragsteller durch das fehlbare, zum Bussenantrag Anlass gebende Verhalten unmittelbar betroffen sein muss oder mit grosser Wahrscheinlichkeit früher oder später in seinen tatsächlichen Interessen als KVG-Versicherer beeinträchtigt wird. Santésuisse kann aus eigenem Recht Antrag stellen, wenn es durch das fehlbare Verhalten in einem eigenen, das KVG betreffenden Sachanliegen in qualifizierter Weise betroffen ist. Völlig offen ist dagegen, ob Bussen in einer Art Pauschalbeanstandungsverfahren überhaupt zulässig sind und wie in einem solchen Falle eine Busse aufzuteilen wäre. Schwierig zu begründen wäre die Zahlung einer Busse an einen Verband. Die Antragsberechtigung der Verbände nach Art. 59 Abs. 2 KVG schliesst nicht auch ein Recht auf Bussenvereinnahmung mit ein. Krankenversicherer-Verbände sind keine Organe der OKP, aber dem KVG dennoch bekannte und teilweise für die Durchführung der OKP unentbehrliche Einrichtungen, die mitunter auch öffentlich-rechtliche Aufgaben wahrnehmen können. Gleichwohl kann die Auffassung vertreten werden, dass Bussen grundsätzlich den vom Verband vertretenen Versicherern zugesprochen werden müssen, es sei denn, der Verband sei in einem eigenen, das KVG betreffende Sachanliegen in qualifizierter Weise berührt."[24]

[24] Gebhard Eugster, Mail vom 18.01.2014 an den VEMS, leicht gekürzter Text.

Somit ergeben sich grundsätzlich zwei Diskussionspunkte. Zum ersten stellt sich die Frage, ob der „Überarztungsverdacht" höher ausfällt, wenn ein Arzt bei allen Kassen etwa gleich „überarztet" als wenn ein Arzt z.B. bei Kasse C einen Index von z.B. 250% aufweist, bei den übrigen Kassen jedoch einen Index von < 130%? Zweitens stellt sich die Frage, ob der RSS-Index nicht aufgrund der Daten der einzelnen Versicherer adjustiert werden sollte, zumal auch dies auf die Rechtsprechung materiellen Einfluss nehmen könnte.

Die Wurzel des Problems

Die zweckmässige Verwendung der OKP Gelder ist die Wurzel des Problems. Ohne dieses Problem grundsätzlich verstanden zu haben, fehlt auch der Pilot durch den Dschungel und die Irrungen in der Handhabung der WZW-Verfahren durch santésuisse.

Im Grundsatz will der Gesetzgeber den rationellen und zweckmässigen Einsatz des Prämienfrankens für eine zweckmässige Medizin. Das Gesundheitswesen muss diesen Auftrag so gut es geht erfüllen und handelt nicht zuletzt in seiner zweckmässigen Erfüllung rechtlich korrekt und verfassungskonform. Im Obligationenrecht, im Medizinalberufegesetz und in der Standesordnung ist zudem klar definiert, dass die Zweckmässigkeit eines Indikationsentscheides nie durch zweckfremde – ökonomische - Interessen korrumpiert werden darf.

Die ideale Versorgungsforschung umfasst die lebenslange Begleitung der Gesundheitskosten eines Individuums und überprüft die Zweckmässigkeit der eingesetzten finanziellen Mitteln anhand qualitativ hochstehenden, facheigenen Wissens und leitet daraus Korrekturen zur Verbesserung des rationellen Einsatzes der verfügbaren Mittel ab. In diesem idealen System wird es ohne echten Mangel (z.B. zu wenig Organspender) keine Rationierung geben. Die statistischen Mittel für die ideale Versorgungsforschung basieren auf der Beobachtung realer „treatment episodes" (Krankengeschichten). Die Hochrechnung daraus gewonnener Erkenntnisse auf die Bevölkerung gestattet ein Abbild der Realität von Zweckmässigkeit und Missbrauch in der Medizin und mithin im Gesundheitswesen. Diese facheigene Kontrolle ist die Grundvoraussetzung für eine

realitätsbezogene Beurteilung des gesetzlich geforderten Zweckmässigkeits-Imperativs (siehe auch Eckwertepapier der FMH)[25].

Die Realität weicht hingegen von diesen Idealen ab. Die Betrachtung der Kosten erfolgt in der Regel über ein Jahr. In den HMO Praxen wurden Kapitationmodelle erzeugt, welche die erwarteten Kosten pro Patienten-Behandlungsjahr berechneten und eine gewissen Toleranzmarge für Kostenüberschreitungen gewährleisteten. Da gleichzeitig kein effektiver Anreiz gesetzt wurde, die Kosten auf der ganzen OKP Ebene zu reduzieren, erwies sich das Vergütungsmodell „HMO-Globalbudget" als Modell der Unterfinanzierung und wurde weitgehend verlassen[26]. Parallel dazu entwickelten sich die Managed Care Modelle, welche ihre Leistungen generell über den Tarmed abwickelten, gleichzeitig aber Boni erhielten, wenn die OKP Kosten pro Jahr mehr als 10% unterschritten werden konnten[27]. Mit diesem Modell konnten die OKP Jahreskosten in der Regel um ca. 15% gesenkt werden, jedoch ohne begleitende Outcome-Forschung und der Gefahr des Sparens am falschen Ort. Tatsache bleibt, dass sowohl HMO- wie Managed Care-Modelle die Kostenproblematik auf der Patientenebene analysierten, und zwar unter Berücksichtigung der Einzeldaten.

Somit ist festzuhalten: die Versicherer analysieren den Kostenbedarf und damit auch die Kostennotwendigkeit der OKP-versicherten Personen basierend auf den jährlichen individuellen Patientendaten. Die mathematische Methode für die Analyse der Kostennotwendigkeit ist die Regressionsanalyse in einem

[25] http://www.physicianprofiling.ch/Eckwertepapier_WZW_2013_11_26.pdf
[26] http://www.physicianprofiling.ch/MCCapitationBeck2012.pdf
[27] http://www.physicianprofiling.ch/ManagedCareOKP012014.pdf

multivariaten Modell, stellvertretend die Arbeiten von Prof. Beck[28] und eine Übersicht zur internationalen Literatur[29]. Idealerweise müssen aber die heutigen Modelle zwecks Vermeidung der Multikollinearität mittels „stepwise regression" durchgeführt, werden, um Variablen, welche in der „simple regression" statistisch signifikant sind wie etwas das Alter, allenfalls eliminieren zu müssen (die Variable würde dann auf Kosten besserer Erklärungsvariablen aus dem Modell eliminiert).

Das Statistikverfahren von santésuisse stellt nun gewissermassen die erwähnten Methoden auf den Kopf. Anstelle von Analysen, welche auf individuellen Patientendaten beruhen, wird nun eine Vergleichsgruppe von Ärzten gebildet, welche einen Kostenindex definieren der als Normalwert gilt (=Index 100%) und Kostenüberschreitungen auf der Arztebene von mehr als 20 bis 30 Prozent gelten als WZW-Verletzung.

Man will nicht mehr wissen, wer behandelt wird, sondern ob der behandelnde Arzt teurer ist als die Kollegen in der Vergleichsgruppe. Auch will man nicht mehr wissen, ob der Arzt dank teureren aber zweckmässigen Kosten noch teurere zweckmässige Nachfolgekosten verhindert.

Damit geht santésuisse nicht mit einer unvoreingenommenen statistischen Haltung an eine wissenschaftliche Fragestellung. Sie benutzt die Statistik in ihrem Sinn: die teuren Aerzte sollen als im gesetzlichen Sinn unzweckmässig handelnde gebrandmarkt werden, damit die Kosten sinken. Hier missverstehen santésuisse und die

[28] http://www.physicianprofiling.ch/Beck2.pdf
[29] www.physicianprofiling.ch/ProfilingStateAndImprove2013Overview.pdf

durch sie vertretenen Krankenversicherer in unverantwortlicher Weise komplett die Intentionen des Gesetzgebers.

Der Durchschnittskosten-Benchmark kann in seiner mathematischen Simplizität nicht überboten werden. Seine Anwendung kann für die Beurteilung eines zweckmässigen Mitteleinsatz schon wegen der nicht normalen Verteilung der Kosten (schiefe Verteilung) nicht verwendet werden. Schlimmer noch: wenn die Aerzteschaft dem 130% Benchmark der Kosten entfliehen will, müssen die teuren billiger werden, damit sinkt der absolute Kostenwert des 130% Benchmarks in einer logischen Abwärtsspirale.

Konsequent zu Ende gedacht bedeutet dies: der Arzt arbeitet am Schluss gratis. Dieser Mechanismus, wo der Mittelwert das erklärende Instrument einer Beobachtung wird – anstelle der Analyse von erklärenden Variablen – erzeugt den „medicus santésuissensis": dieser wird immer den Benchmark von 130% unterbieten müssen, er ist ein gewollte Züchtung dieses Benchmarks und damit gewissermassen Opfer und Vollstrecker der Kosten-Eugenik von santésuisse.

Diese Analogie zu Francis Galton[30] ist in seiner Qualität nicht nur erschreckend sondern zeigt den diskriminierenden Aspekt der santésuisse Statistik auf; nur der billige Arzt ist der gute Arzt. Logisch aus Sicht der Versicherer, inkompatibel mit der Realität und dem Versorgungsauftrag der Aerzteschaft.

Zudem fördert der Kostenbenchmark 130% eine Überarztung im niedrigen Risikobereich, wo auch gleichzeitig die Verarztungskosten geringer sind. Damit steigen die Gesamtkosten im

[30] http://www.physicianprofiling.ch/MathematischeMaschinerieEidenbenz1993.pdf

Gesundheitswesen, während gleichzeitig der Kostendurchschnitt pro Patient sinkt.

Leider ist die von santésuisse ohne öffentliche Kontrolle eingeführte Methode keineswegs der erhoffte Fortschritt. Ganz im Gegenteil: der ANOVA-Index korrigiert die Arztkosten durch die Arztkosten mit der ZSR Variable als Merkmal für die Durchschnittskosten und korrigiert für das Patientenalter mit aggregierten Daten.

Die Folge davon ist eine betreffend die Fragestellung unbrauchbare, realitätsfremde Methode. Der VEMS fordert deshalb schon seit 2004 ein Moratorium[31]. Die Verantwortung der Versicherer, welche santésuisse mit primitiven (aggregierten) Datenbanken beliefern, ist ebenfalls nicht von der Hand zu weisen.

Erstaunlich ist, dass die Partner der Krankenversicherer, insbesondere die FMH, sich nie auf einen echten wissenschaftlichen Diskurs zur Verbesserung der WZW Verfahren eingelassen hat; sie verblieb damit immer im Rahmen der Vorgaben und Denkweise der Krankenversicherer, ohne selber mit eigenen Ideen und Methoden aktiv zu werden.

Erstaunlich ist auch, dass Fragen der Versorungssicherheit (teure Patienten werden tendenziell unterversorgt, billige Patienten tendenziell übermedikalisiert) die verantwortlichen Gesundheitsdirektoren nicht auf den Plan gerufen hat. Diese haben sich bisher in keiner Weise zur Problematik geäussert und scheinen sich des Problems noch nicht einmal bewusst zu sein.

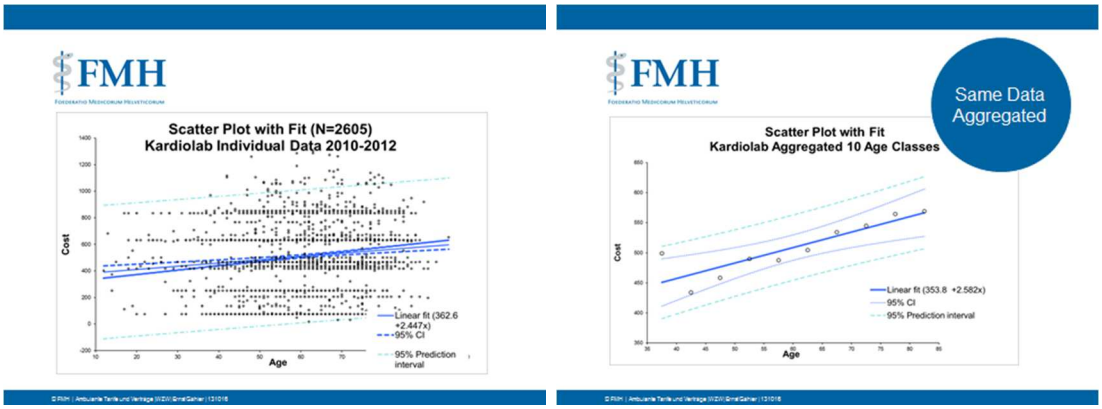
Das gleiche gilt auch für die Schweizerische Akademie der Medizinischen Wissenschaften, welche trotz entsprechenden Anfragen

[31] http://physicianprofiling.ch/VEMSAufsichtsbeschwerde2004.pdf

des VEMS bisher in keiner Weise zur Sinnlosigkeit der WZW Prüfungen der Krankenversicherer in der Schweiz konkret Stellung bezogen hat.

Betreffend die ANOVA Methode besteht das grundlegende statistische Problem unter anderem bei der Datenaggregation („aggregation bias"). Was geschieht dabei?

Im Wesentlichen werden die Informationen einer Wolke von Daten auf einen Mittelwert projiziert, mit dem Effekt, dass die Information von der Standardabweichung wegfällt. Ohne diese Detailinformation (der Individualdaten) erscheint der Einzelne als der Mittelwert der Gruppe, was in der Regressionsanalyse zu sehr hohen Korrelationen führt[32].



Santésuisse hat nun versucht, die Kostenrisiken für den Arzt mit dem Anova-Index transparenter zu gestalten. Sie geht dabei davon aus, dass die Arztpraxis mit älteren Patienten in teureren Kantonen mehr Kosten generiert und korrigiert den RSS Index für die Arztpraxis dann aufgrund des Anova-Index nach unten.

[32] www.physicianprofiling.ch/FMHAnova102013V5.pdf

Vorab sind dabei zwei Fragen zu beantworten: welche Variablen sind für die zu untersuchende Arztpraxis betreffend Kosten prädiktiv und welche Daten stehen überhaupt zur Verfügung. Oder mit anderen Worten: stimmt das tatsächlich, dass in einer Arztpraxis X die älteren Patienten teurer sind und wie bestimmt man das?

Die Korrelation zwischen Alter und Kosten kann mit der Regressions-Methode abgeschätzt werden. Wir tragen also auf der X-Achse das Alter und auf der Y-Achse die Kosten auf und erhalten eine Beziehung y=x (formel). Der Determinationskoeffizient „R2" zeigt an, wie viel Prozent der Kosten durch das Alter erklärt werden können. In der Regel sind das 4-5 Prozent.

Doch santésuisse verfügt nicht über diese Einzeldaten der Patienten sondern über Mittelwerte vorbestimmter Altersklassen, also aggregierte Daten anstelle von Einzeldaten. Dies bedeutet, dass jedem Patienten der Altersgruppe anstelle des Originalwertes der Kosten ein Mittelwert der Kosten zugeordnet wird. Dies führt zu einer künstlichen Erhöhung des R2, welcher nun bis zu 80% Erklärungsgehalt für die Kosten erreichen kann. Dieser ökologische Trugschluss ist in der Literatur unter dem Begriff „ecological fallacy" bekannt und betrifft in diesem speziellen Fall das Problem des „aggregation bias".

Dies bedeutet, dass santésuisse seit der Einführung des ANOVA-Indexes ein Statistik betreibt, die einem erheblichen Trugschluss ausgesetzt ist, denn die Korrelationsstärke der Beziehung zwischen Alter und Kosten hat direkten Einfluss auf den Anova-Index.

Wir beobachten also, dass der Anova-Index von santésuisse in unbekanntem Ausmass einen statistischen Trugschluss oder Aggregation-Bias aufweist, dass somit die Kostenrisiken aus unterschiedlichen Altersstrukturen per defintionem nicht adäquat berechnet werden können, mithin, dass der Risiko(struktur)ausgleich (RSA oder RA) zwischen der Arztpraxis X und seinem Vergleichskollektiv nicht funktionieren kann.

Selbstverständlich wird ein korrekter RSA für die altersbedingten Kostenrisiken generiert, beispielsweise im RSA unter den Krankenkassen gemäss dem Vorgehen „gemeinsame Einrichtung KVG", wo mit einem Budget von 500 Mio. Franken jährlich der Risikoausgleich berechnet wird. Dabei werden die klassischen Zellen der Altersgruppen verwendet, wie wir in folgender vereinfachender Tabelle sehen können:

Tabelle 1: Zweizellen Modell zur Korrektur von zwei Altersgruppen (alt / jung)

Zelle 1 Vergleichsgruppe:	Zelle 2 Vergleichsgruppe:
1000 jungen Patienten zu 50 Fr. pro Jahr	1000 älteren Patienten zu 100 Fr. pro Jahr
Santésuisse-Index: 75 Franken = 100%	
Zelle 1, Arzt X	Zelle 2 Arzt X:
100 junge Pat. zu 100 Fr., odds ratio 2.0 x 100	100 älteren Pat. zu 100 Fr., odds ratio 1.0 x 100
Santésuisse-Zellenindex: 100 Franken=133.333%	Santésuisse-Zellenindex: 100 Franken=100.000%
Zellen-bereinigter Index: [100 OR 2.0 x 100 OR 1.0] = Index 150.000%	

Dieses Beispiel belegt, dass der Arzt jüngere Patienten teurer behandelt (behandeln muss?), die älteren Patienten jedoch gleich

teuer wie die Vergleichsgruppe behandelt hat. Unter Berücksichtigung der Anzahl Patienten pro Zelle steigt sein zellenbereinigter Index von 133.333% auf 150.000% an, also um 16.667%.

Beim RSA unter den Krankenkassen erhält also der Versicherer der Patienten von Arzt X Geld aus dem Risikoausgleichstopf, und zwar 100x50=5000 für die Zelle 1 und 100x0=0 Franken für die Zelle zwei. Für den Arzt X hingegen wird der RSA umgekehrt betrachtet, die höheren Kosten werden ihm als Überarztung ausgelegt, er kann gebüsst werden.

Zählt man nun alle santésuisse verfügbaren 48 Zellen zusammen, wird der Zellen-Index von Arzt X entsprechend den RSS Index von Arzt X beeinflussen. Dieses Ausmass muss mit dem Anova-Index verglichen werden. Dies wurde bisher immer nicht offengelegt (26.02.2014).

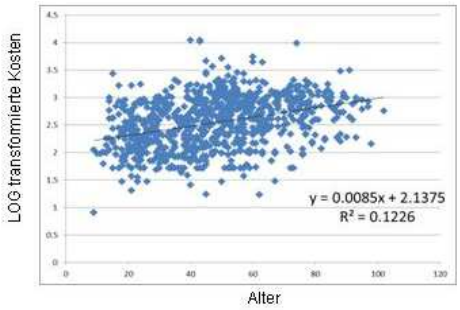
Somit sehen wir zweierlei: die Zellentechnik erlaubt ohne grosse Statistik eine einfache und verständliche Korrektur von Kostenunterschieden in Alters- und Geschlechtsklassen. Sind die Kosten gleich, gibt es keinen Risikoausgleich. Beim Anova-Index hingegen muss eine Beziehung zwischen Alter und Kosten a) überhaupt existieren und b) aufgrund von nicht aggregierten Einzeldaten für die Vergleichspraxen und für die Arztpraxis X berechnet werden.

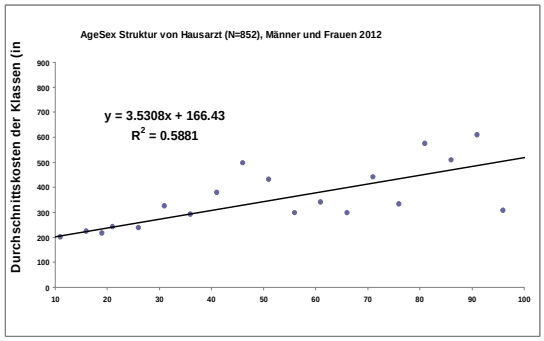
Dass dies im Einzelfall eines Allgemeinpraktikers problematisch ist, sehen wir am folgenden Beispiel:

Graphik 1A: log transformierte Kosten aus Graphik 1B: Aggregierte Alter-Kostenregression

Einzeldaten: R2=12% Aggregierte Daten: R2=59%

Mikrolevel-Analyse Makrolevel-Analyse





Wir sehen also hier, dass der „aggregation bias" aus einer schlechten Korrelation eine gute Korrelation erzeugt, bei identischem Datensatz.

Das zweite statistische Problem unter dem Kapital „ecological fallacy" ist das Simpson's Paradoxon: „Analysefehler treten auf, wenn das zu untersuchende Problem aus seinem Kontext herausgerissen und in der statistischen Terminologie analysiert wird" [33].

Mit der Durchschnittskostenmethode ist santésuisse erfolgreich bei den Bundesrichtern durchgedrungen. Hilfeleistung erfolgte durch gefühlte Vorverurteilung: die Statistik besagte, der Arzt ist schuldig, damit waren seine Chancen minimiert, Vergleiche erfolgten zu Hauff,

[33] www.physicianprofiling.ch/SimpsonParadoxUnraveled.pdf

um drohende Blamagen vor Gericht zu umgehen. Ein perfektes System.

Die Problematik des RSS-Index wurde von den Richtern nicht begriffen. Sie erklärten Kostenüberschreitungen zu häufig als Versicherungsbetrug.

Die Gerichtsbarkeit war möglicherweise Opfer ihres Übereifers, welcher gestützt auf reine Mittelwertbetrachtungen die klassische Situation des Simpson's-Paradox erzeugt. Das Simpson Paradox besagt, dass Beobachtungen sich in das Gegenteil verkehren, *wenn anhand von erklärenden Variablen der Mittelwert neu gerechnet wird*. Konkret: ein Arzt erscheint in der RSS Statistik als auffällig, in der Realität ist er unterdurchschnittlich teuer oder umgekehrt.

Dies sei an einem Beispiel mathematisch hergeleitet[34]: die Vergleichsgruppe behandelt zwei Kategorien, teure und billige Patienten. Die billigen kosten 50 Franken pro Jahr, die teuren 100 Franken pro Jahr. Die Vergleichsgruppe behandelt im Schnitt in jeder Kategorie 1'000 Patienten, also insgesamt 2'000 Patienten, der RSS Index beträgt also 75 Franken, gemäss santésuisse 100%.

Nun zwei Beispiele. Der Arzt 1 hat wenig billige Patienten, z.B. 50, er behandelt diese ebenfalls zu 50 Franken pro Jahr, dann hat er 100 teure Patienten zu je 100 Franken pro Jahr. Der Index für billige und teure Patienten beträgt je 100%, da er billige und teure Patienten zu den gleichen Preisen verarztet wie die Vergleichsgruppe. Da er aber weniger billige Patienten hat, beträgt der RSS Index 111%. Der Arzt 1 wird also einen höheren Mittelwert haben in der

[34] http://www.physicianprofiling.ch/WZWoddsratio.xlsx

RSS Statistik, obwohl er billige und teure Patienten im Schnitt genau gleich teuer behandelt wie die Patienten aus den Vergleichspraxen. Das Paradox entsteht in diesem Fall daraus, dass der Arzt 1 weniger billige Patienten behandelt. Allein dies genügt schon, seinen Index um 11% in die Höhe zu treiben. Nehmen wir nun Arzt 2, welcher im folgenden Beispiel gemäss RSS Index unterdurchnittlich viel kostet Index 99%, effektiv aber einen Index von 135% hat: dieser Arzt behandelt im Schnitt 1000 billige Patienten zu 70 statt 50 Franken, er hat aber wenige teure Patienten, nämlich nur 150, und behandelt diese gleich teuer wie die Vergleichsgruppe, nämlich mit durchschnittlich 100 Franken. Er behandelt also 1000 billige Patienten im schnitt 40% teurer (odds ratio 1.4).

Arzt 2 behandelt zwei Klassen von Patienten (billige und teure):

	Billige	Teure	Alle
Anzahl behandelte	1000	150	1150
Durchschnittskosten	70	100	73.91
Kosten pro Jahr	70'000	15'000	85’000
Odds ratio	1.4000	1.0000	**0.9855**
Odds ratio bereinigt	**1.3478**		

Arzt 2 gilt also als wirtschaftlich, obwohl er gemäss santésuisse Kriterien das Gegenteil davon wäre, er hat nämlich einen Kostenmittelwert von 135%. Das erklärt sich nun damit, dass er teure Patienten gleich teuer behandelt wie die Vergleichspraxen, hingegen erzeugt er bei den billigen Patienten, welche in der Überzahl sind – und das ist der springende Punkt - im Schnitt 40% höhere

Kosten. Das Problem kann man selber nachrechnen mit der von uns publizierten Tabellenkalkulation im Internet[35].

[35] http://www.physicianprofiling.ch/WZWoddsratio.xlsx

Weitere „Fallacies"

Statistische Täuschungsproblematik („ecological fallacy and aggregation bias") in der ANOVA-Methode von santésuisse

Bei der Durchführung des RSA hatten die Krankenversicherer ursprünglich aggregierte Daten benutzt[36], dies aber wegen des inhärenten Aggregationsbias wieder verlassen, wie dies in einem Bericht zu Handen des Schweizerischen Nationalfond festgehalten wurde[37].

In diesem Bericht wurde auch empfohlen, die Analysen auf Individualdatenebene durchzuführen, also ohne aggregierte Mittelwerte oder ähnliches.

Obwohl man versuchte, Aggregationsregeln einzuführen, blieb die Diskussion auf der Ebene der AP-DRG (Diagnosen) und Hospitalisation im Vorjahr stecken[38].

Die Bedeutung des Aggregationsbias liegt bei der Bildung des ANOVA-Indexes in der Inflation des Bestimmtheitsmasses (auch Determinationskoeffizient genannt) R^2 in der verwendeten Regressionsanalyse. Dies bedeutet in Bezug auf die Bildung von Altersklassen aus Durchschnittswerten (z.B. Alter 40-44, 45-49 ecc), dass die Bedeutung der Variable Alter auf die Erklärung der

[36] Die Modelle basierten auf folgenden Variablen: Soziale Klasse, Alter, Geschlecht, Arbeitslosigkeit, Tätigkeit im öffentlichen Dienst, Sterblichkeit, Behinderung, Urbanisierungsgrad, Anzahl Angehöriger, Einkommen, medizinisches Angebot.
[37] http://www.snf.ch/SiteCollectionDocuments/medienmitteilungen/mm_04sep15_2_d.pdf
[38] http://www.snf.ch/sitecollectiondocuments/nfp/nfp45/nfp45_hollye_sb.pdf

Arztkosten weit höher geschätzt wird als wenn die nicht aggregierten Individualdaten verwendet würden.

Es existieren Formeln um den Aggregationsbias abzuschätzen[39]. Beträgt beispielsweise das R2 aus 22 aggregierten Altersklassen 0.12[40] und die Menge der Individualdaten 500'000, dann würde das R2 aus den Originaldaten 0.0000054545 betragen[41]. Je höher das R2 in der RSA Formel ausfällt, desto stärker ist dann der Effekt auf den ANOVA-Index. In der Praxis würde dies bedeuten, dass ein Arzt mit deutlich älteren PatientInnen eine Senkung seines ANOVA-Indexes erfahren würde und umgekehrt.

Die höhere Präzision nicht aggregierter Datensätze wird ebenfalls in den Capitationmodellen der amerikanischen Versicherer erwähnt[42].

Selbst in Metaanalysen sind aggregierte Daten nur speziellen statistischen Verfahren mehr oder weniger sinnvoll einzusetzen.[43]

Zudem kann mit aggregierten Daten allenfalls eine Aussage über generelle Tendenzen gemacht werden, *kausale Zusammenhänge können jedoch nicht erstellt werden*[44].

Rückrechnungen von aggregierten Datensätzen auf die individuelle Ebene sind zudem äusserst kompliziert[45] und für unsere

---

[39] http://www.physicianprofiling.ch/AggregationBias1992.pdf
[40] Dies würde bedeuten, das 12% der Arztkosten durch Alterseffekte erklärt werden können.
[41] In diesem Fall wäre die Bedeutung des Alters als Prädiktor für Arztkosten statistisch inexistent. Die Berechnungen sind als Excel Datei erhältlich: http://www.physicianprofiling.ch/AggregatedR2.xls
[42] http://www.actuary.org/pdf/health/Risk_Adjustment_Issue_Brief_Final_5-26-10.pdf
[43] http://www.ncbi.nlm.nih.gov/pubmed/19254930
[44] http://www.uclouvain.be/cps/ucl/doc/demo/documents/Shpenev.pdf
[45] http://people.bu.edu/pekoz/approximate.pdf

Fragestellung nicht validiert. Aggregation als Ersatz für individuelle Datensätze ist problematisch.[46]

Dieses für Big Data typische Problem wurde auch schon in der NZZ unter dem Titel „Die Erkenntnis von nichts" diskutiert[47]/[48], ein umso bedeutsameres Problem, als geschätzt wurde, dass rund 20% der Big Data Analysen das Gesundheitswesen betreffen.

Zudem ist die Faszination, welche von Big Data ausgeht nicht neu und führte bereits vor über 100 Jahren letztlich zur realitätsfremden Eugenik, welche den Mittelwert als Beweismittel für Zusammenhänge betrachtete[49].

Aus Sicht der Bedeutung von epidemiologischen Variablen als unabhängige Variablen zur Erklärung der Arztkosten kommt der statistischen Analyse enorm hohe Bedeutung zu, umso mehr, als nun geplant ist, die ANOVA-Methode mit weiteren unabhängigen Variablen wie „Hospitalisation im Vorjahr" und „pharmaceutical cost groups" (PCG) zu füttern.

Allein schon aufgrund der teils massiven Überschätzung des R^2 als Folge des Aggregationsbias ist die Beweiskraft der ANOVA-Methode zur Erfassung von Unwirtschaftlichkeit nicht gegeben[50].

Bereits im Jahr 2009 wurde die Problematik der ANOVA-Methode aus juristischer Sicht kritisiert[51] und im Jahr 2012 weiter substanziiert[52]. Diese Bedenken erhalten aufgrund des

---

[46] http://ajph.aphapublications.org/doi/abs/10.2105/AJPH.91.4.632
[47] http://www.nzz.ch/meinung/debatte/die-erkenntnis-von-nichts-1.18212517
[48] http://www.physicianprofiling.ch/BigDataNZZ122013.pdf Gastbeitrag
[49] http://www.physicianprofiling.ch/MathematischeMaschinerieEidenbenz1993.pdf
[50] http://www.physicianprofiling.ch/AGWZWZSRAgeSex.pdf
[51] http://www.physicianprofiling.ch/PhyWZW2Eugster2009.pdf
[52] http://www.physicianprofiling.ch/WZWEugster2012.pdf

Aggregationbias den statistischen Beweis für die falsche Anwendung des ANOVA-Indexes durch santésuisse.

Mathematisch sieht die ANOVA Methode der santésuisse vereinfacht gesagt folgendermassen aus: Zu erklärende Arztkosten Y = Arztkosten (ZSR) + Korrektur für Alter und Geschlecht = Erklärungsgehalt von Y =40+40%=80% (der Effekt pro Kanton dürfte in diesem Kontext marginal sein, wurde aber bisher auch nicht im Detail offengelegt). Dies bedeutet: rund 40% des ANOVA Modells von santésuisse wird erklärt durch die Arztkosten, was einer Tautologie entspricht und die weiteren 40% durch die Alters- und Geschlechtsstruktur.

Gemäss Literatur erklären Alter und Geschlecht (demographische Variable) normalerweise 4-5 Prozent der Arztkosten.

Der Aggregation-Bias bauscht also den Erklärungsgehalt von Alter- und Geschlecht im ANOVA Modell um den Faktor 10 auf. Im Ergebnis ist die ANOVA-Statistik der santésuisse eine Ansammlung von „statistical fallacy" Effekten. Dazu noch näheres betreffend die Arztkosten, welche die Arztkosten erklären, anhand dem folgenden Kapitel ZSR Problematik.

ZSR Problematik

Das ANOVA Modell von santésuisse beinhaltet aggregierte Alters/Geschlechtszellen (N=42), Korrektur von Wohnkanton und die ZSR Nummer. Die Resultate werden facharzt-spezifisch berechnet[53], hier die Formel zum ersten Schritt (Arzt + Alter/Geschlecht): [Ai = μ

[53] www.physicianprofiling.ch/WZWAnovaGutachtenStahelRoth.pdf

+ αi + (Σ βj) / 42], wobei μ allgemeines Niveau der Durchschnittskosten, αi Effekt des i-ten Arztes (i = 1, ..., n) auf die Durchschnittskosten, βj Effekt von j-ter Patientenkategorie (j = 1, ..., 42).

Im ANOVA Modell sind dann die Effekte auf den Erklärungsgehalt der Kosten zu beobachten[54]. Das R2 für Alter/Geschlecht variiert in den Beispielen zwischen 11% bis 62% (je nach Facharztgruppe), für die ZSR beträgt die Variation 33% bis 67%, für die Kombination 42% bis 89%. Dies bedeutet, dass die abhängige Variable Arztkosten durch diese beiden Variablen im ANOVA Modell der santésuisse erklärt sind.

Gemäss Literatur erklären Alter und Geschlecht aus nicht-aggregierten Daten, wie bereits im vorigen Kapitel erwähnt, in der Regel 4-5 Prozent der Kosten. Wie bereits angemerkt, besteht hier wegen dem Aggregation Bias eine Überschätzung dieser Erklärungsvariablen.

Aus der Diskussion mit santésuisse ist nun die Handhabung der Erklärungsvariable ZSR im ANOVA Modell von *zentraler Bedeutung*.

Zunächst lautet die Begründung für die gemäss Literatur nie verwendete Erklärungsvariable ZSR=Arzt auf die Arztkosten[55]:

> *„Es ist nicht so, dass die ZSR eine eigentlich Erklärende ist (wie Alter, Geschlecht, etc.) und sie wird schon gar nicht „noch dazu genommen" – sie ist das Kernstück des ANOVA-*

[54] www.physicianprofiling.ch/AGWZWZSRAgeSex.pdf
[55] www.physicianprofiling.ch/ZSRDangelo122013.pdf

Ansatzes. Ziel ist, ein Leistungsmittel pro Arzt zu ermitteln, man startet also mit „Bruttoleistung ~ Arzt", d.h. jeder Arzt ist direkt und ausschliesslich für die Leistungen seiner Patienten verantwortlich. Da sich zeigt, dass die Zusammensetzung des Patientengutes einen entscheidenden Einfluss auf diese Leistungen hat, fügt man dem Modell weitere Variablen (echte Erklärende) hinzu um diese externen Einflüsse herausrechnen zu können. Welche Variablen das im Detail sind, wird sich im Laufe der Weiterentwicklung herausstellen. Alter und Geschlecht sind bei vielen Facharztgruppen sicher dabei. Lässt man die Variable ZSR weg, werden lediglich Patientenkosten innerhalb einer Facharztgruppe betrachtet. Der Bezug zum entsprechenden Arzt fehlt. Bei der Wirtschaftlichkeitsprüfung wird der Fokus aber auf die Arztkosten unter Berücksichtigung der Zusammensetzung seiner Patienten gelegt. Wenn man hingegen nur die ZSR als Variable berücksichtigt erhält man Ergebnisse ähnlich der Rechnungsstellerstatistik auf gesamtschweizerischer Ebene, ohne Berücksichtigung der Patientenzusammensetzung."

Aus diesen Mitteilungen ergibt sich somit, dass das ANOVA Modell der santésuisse die ZSR Nummer in Form der „Durchschnittskosten des Indexarztes" berücksichtigt (=RSS-Index).

Dies bedeutet, dass eine intraindividuelle Variable (der Arzt selber) in die Varianzanalyse einbezogen wird. Dass dieser Effekt tatsächlich relevant ist, wird durch das R2 bewiesen. Damit werden nicht randomisierte Variablen des Arztes (z.B. seine Intelligenz) auf

die Kosten miteinbezogen. Vom Konzept her wird gedacht, dass der Arzt dessen Arztkosten erklärt, also eine Art der Tautologie.

Ferner wirkt sich der Aggregation bias auch auf die Gesamtkosten des Arztes aus und *beinhaltet darüber hinaus auch die veranlassten Kosten*, was eigentlich gemäss gängiger Gerichtspraxis aus dem Index eliminiert werden sollte.

Der weltweite wissenschaftliche Standard des Physician Profiling kennt dieses Betrachtungskonzept nicht, sondern verwendet eine statistische Methode, welche unter dem Namen multiple Regressionsanalyse bekannt ist[56]. Es existiert auch keine Literatur darüber, wie weit Arzt-spezifische Variablen wie etwa seine Intelligenz sich auf die Kosten auswirken.

Das Grundkonzept des Physician Profiling wurde damit von santésuisse offensichtlich ungenügend berücksichtigt. Dieses beinhaltet nämlich, dass patientenrelevante Variablen wie Krankheiten, sozioökonomischer Status, Behandlungsdauer (treatment episodes) usw. für die Prüfmethode unabdingbar sind, *will man nicht solche unabhängige Effekte fälschlicherweise dem Arzt zuordnen („omitted variable bias")*.

Durch das Weglassen zentral bedeutsamer „random effect variables" wird die Arztbedeutung auf die Kosten unter Verwendung der „Durchschnittskosten des Indexarztes" künstlich überhöht. Das ausschliesslich mit der ZSR eine genaue Zuordnung der Kosten pro Arzt möglich sei, wie santésuisse behauptet, ist als Argument natürlich falsifizierbar. Zudem funktionieren multiple lineare

[56] www.physicianprofiling.ch/ProfilingStateAndImprove2013Overview.pdf

Regressionsmodelle im Physician Profiling zur Arzt-Identifikation ja auch ohne das von santésuisse gewählte „Durchschnittskosten des Indexarztes".

Es geht santésuisse also darum, ein Modell durchzusetzen, welches im ersten Schritt und a priori eben gerade jenen Mittelwertvergleich beinhaltet, der per se einer wissenschaftlichen Prüfung in keiner Weise genügen kann. Korrigiert man ein Modell, welches 40% auf falschen Annahmen gründet, kann das Resultat nie überzeugen.

Der Aerzteindex der CSS verwendet ebenfalls einen Teil der auf den Individualdaten der Patienten basierenden Arztkosten (nämlich die nicht-veranlassten Arztkosten) als Random-Variable in einem multivariaten Regressionsmodell[57]. In diesem Modell werden aber auch die gesamten Medikamentenkosten als weitere erklärende Variable sowie PCG verwendet, womit sich leider die Gefahr der Multikollinearität ergibt[58].

Erwähnenswert ist ferner, dass die gesamten Medikamentenkosten den höchsten Korrelationskoeffizienten aufweisen (0.764 bei Allgemeinmedizinern) und damit als Merkmal für die Behandlungsnotwendigkeit durchaus sinnvoll eingesetzt werden können.

Ungünstig am CSS Modell ist jedoch, dass auch hier die Arztkosten selber eine zu starke Gewichtung erhalten (Tautologie-Bias) und dass Patienten mit geringerer Anbindung (z.B. nach

[57] http://www.physicianprofiling.ch/CSSAerzteIndex.pdf
[58] http://de.wikipedia.org/wiki/Multikollinearität

einfachen Abklärungen bei Spezialisten) für das Prädiktionsmodell dieses Aerzteindexes nicht verwendet werden können.

Auch in diesem Modell ist somit die ZSR Nummer und die damit verbundenen Arztkosten keine conditio sine qua non für die Durchführung des Profiling. *Somit kann der CSS-Aerzteindex nur als Vorläufer-Prüfmethode betrachtet werden, welche aber ohne grossen Aufwand den internationalen Standards angepasst werden könnte.*

Ein weiteres zentrales Problem betrifft die Überprüfbarkeit des ANOVA Indexes von santésuisse betreffend die „reliability and misclassification"[59]. Dies kann nicht ausreichend abgeschätzt werden, da entsprechende Rechnungsmodelle fehlen.

Der ANOVA-Index ist zudem nicht publiziert und auf seine diagnostische Performance hin nicht validiert. Die mit Gewissheit geäusserte Erklärung der ETH-Gutachter, dass dieses Modell Rechtsfähigkeit erlangen wird, hat sich, zu Recht, nicht bewahrheitet.

Es ist auch zu fragen, warum die ETH-Gutachter nicht selber die Limitationen der ANOVA Methode von santésuisse, insbesondere den sofort ins Auge springenden Aggregation bias, diskutiert haben oder zumindest eine Einordnung der Methode in andere verfügbare Methoden des Physician Profiling vorgenommen haben (Limitationen eines durch santésuisse bezahltes Gutachten, welches auf der OKP Ebene durchaus geeignet ist, Schaden anzurichten).

[59] www.physicianprofiling.ch/RAMisclassificationRANDNEJM2010.pdf

Grundsätzliche Problematik des RSS-Index

Santésuisse verwendet eine Mittelwertüberschreitung der Kosten von > 130% als Hinweis für Überarztung, welcher im Fall einer Verhandlung vor dem Versicherungsgericht als Beweisgrenze gewertet wird. Da es sich dabei um einen Screening-Test für Überarztung handelt, stellen sich eine Reihe von Fragen. Wie gut ist beispielsweise dieser santésuisse-Test für die Erfassung echter Überarztung, oder anders formuliert, was ist der Goldstandard für den Test?

Leider ist santésuisse der Goldstandard für ihren santésuisse-Test nicht bekannt. Bekannt ist hingegen, dass selbst santésuisse von ca 3% echt überarztenden ausgeht, produziert jedoch rund 25% auffällige. Aufgrund dieser Informationen kann die Performance des santésuisse-Tests abgeschätzt werden.

Der Test kann nicht mehr als diese 3% echt überarztende feststellen (=true positive), die 25% wären die falsch auffälligen (=false positive). Gehen wir davon aus, dass nochmals 2% echte Überarzter sind, welche jedoch unter dem Mittelwert „verschwinden", dann haben wir falsch negative von 2% (=false negative). Am Schluss bleiben somit die richtigerweise nicht überarztenden übrig (=true negative), rund 75%.

Aus dieser Information ergeben sich für den Santésuisse-Test eine Sensitivität von 60%, eine Spezifität von 74%, eine Accuracy von 73%, sowie ein positiv und negativ prädiktiver Wert von 11% und 97%. Somit bestehen zwei Probleme: der niedrige positiv prädiktive Wert bedeutet, dass das Verhältnis der falsch positiven zu den richtig positiven rund 1:10 beträgt, 90% der auffälligen wären dann in Wirklichkeit gar nicht auffällig. Santésuisse behauptet auch mal, dass nur 15% der Aerzteschaft auffällig sei (!). In diesem Fall hätten wir einen positiv prädiktiven Wert von 17%, es ändert sich also nicht viel.

Tabelle zur diagnostischen Genauigkeit des RSS-Index

	TP	TN	FP	FN	SENS	SPEZ	PPV	NPV	ACC	ALL	pLR	nLR	PRE%
RSS-Index	3	70	25	2	60	74	11	97	73	100	2.3	0.54	5
RSS-Index	3	80	15	2	60	84	17	98	83	100	3.8	0.48	5

TP= true positive, TN=true negative, FP=false positive, FN=false negative, SENS=Sensititivät, SPEZ=Spezifität, PPV=positive predictive value, NPV: negative predictive value, ACC=Accuracy, pLR=positive likelihood-ratio, nLR=negative likelihood-ratio

Das zweite Problem betrifft die Anwendung dieses Tests bei niedriger Prävalenz von Überarztung (in diesem Beispiel gehen wir realistischerweise von höchstens 5% aus). Unter Verwendung des Bayes Theorem kann nun die Nachtest-Wahrscheinlichkeit bei einem positiven Test (="Cutoff von 130% überschritten") unter Verwendung der folgenden Formel berechnet werden:

Bayes Theorem[60]:

PTP pos: (PV x SE) / [PV x SE + (1 – PV) x (1 – SP)]

PTP pos = posttest Probabilität bei positivem Test; PV = Vortest-Wahrscheinlichkeit oder Prävalenz. SE = Sensitivität; SP = Spezifität.

Unter Anwendung des Bayes Theorem steigt bei V.a. Überarztung wegen der Indexüberschreitung die Wahrscheinlichkeit (Probabilität), dass echte Überarztung vorliegt, von 5% auf 11%. Das Problem ist nun, dass im Gerichtsverfahren diese Wahrscheinlichkeit deutlich höher geschätzt wird, ein bekanntes Täuschungsproblem in der Jurisprudenz[61]. Möglicherweise unbeabsichtigt, kann nun santésuisse ihren santésuisse-Test präventiv einsetzen, indem nun jeder mit RSS-

[60] www.ncbi.nlm.nih.gov/pubmed/19952758
[61] www.physicianprofiling.ch/FallacyJuryBayes2000.pdf

Index Überschreitung von 130%, und sei dies nun zu 90% unwahrscheinlich, verdächtigt werden kann und wird.

Das Problem liegt hier in der Betrachtung der Gesamtkosten pro Arzt, ohne unabhängige erklärende Variablen zu berücksichtigen („omitted variable bias"). Damit entsteht eine Situation, wo die Wahrscheinlichkeit für statistische Täuschungen, wie im Kapitel „Wurzel des Problems" dargelegt, in Form des Simpson's Paradoxon erheblich ansteigt.

Somit ist der RSS-Index für die Fragestellung, ob im Sinne des Gesetzgebers die WZW-Regeln verletzt wurden, aus der Sicht des VEMS als missbräuchlich zu beurteilen.

Auch weitere Voraussetzungen wie Normalverteilung der Kosten, ca. 80% Abdeckung (Selbstzahlerproblem, lückenhafter Datensatz bei der Rechnungsübermittlung) sind an dieser Stelle zu nennen.

Implikationen der Versicherer Statistik auf die RSS Statistik

Der RSS-Index wird aus Versichererdaten zusammengesetzt, jeder Versicherer kann den Index jeden Arztes berechnen (RSSV).

Dabei stellen sich zwei Probleme: 1) Ein Arzt kann zwar im RSS-Index von santésuisse auffällig sein (> 130%), dies muss aber nicht bedeuten, dass er bei allen Kassen, an welche er Rechnung stellt, der Fall ist. Kommt es zur Anklage, kann es sein, dass santésuisse Versicherer vertritt, welche selber vom Index>130% gar nicht betroffen sind. Dies ist verfahrensrechtlich problematisch; 2) der RSA (gemeinsame Einrichtung KVG) zeigt, dass bestimmte Kassen teurere Patienten versichern. Dies muss für die Bildung des RSS-Index berücksichtigt werden (RSA auch für den Arzt). Santésuisse soll aufgefordert werden, zu diesen beiden Problemen

Stellung zu beziehen und ihre künftige Strategie dazu offen legen. Da es sich um ein Verfahrensproblem zur WZW Regelung handelt, müssen diese beiden Probleme quantifiziert und anschliessend zusammen mit der FMH neu überdacht werden.

Hintergrund

Die Schweizer Krankenversicherer haben seit 1998 eine Zunahme des Versichertenbestandes von 5.1 Millionen auf 8.0 Millionen Personen erreicht. Parallel dazu hat der Gesetzgeber beschlossen, dass Kostenunterschiede, welche durch Alter und Geschlecht erklärbar sind (neuerdings unter Einbezug der Variable „Hospitalisation im Vorjahr") mittels der „Gemeinsamen Einrichtung KVG" jenen Kassen rückvergütet werden, welche höhere Kostenrisiken pro Alters- und Geschlechtszelle (aggregierte Klassen) vorzuweisen haben.

Dieser sogenannte Risikostrukturausgleich generiert pro Jahr Umverteilungen unter den Kassen in der Höhe von 505 Millionen Franken für das Jahr 1996, Tendenz stetig steigend auf 1496 Millionen Franken im Jahr 2012. Wegen schlechter Risiken haben für die Jahre 1996-2012 folgende Kassen mehr als eine Milliarde Franken erhalten: Helsana Versicherungen AG,

Visana AG, CSS Kranken-Vers. AG, Wincare Versicherungen AG., CONCORDIA AG. In den Ausgleichsfond haben in den Jahren 1996-2012 folgende Kassen mehr als eine Milliarde Franken einbezahlt: Progrès Versicherungen AG, Intras Krankenversicherung

AG, Sanitas Grundversicherungen AG, ASSURA[62]. Die genaue Aufteilung ist im Anhang, Tabelle 2 aufgezeigt.

In der Rechnungsstellerstatistik RSS werden jedem Arzt die entsprechend den aggregierten Alters- und Geschlechtsklassen verursachten Kosten zugeordnet. Dabei werden die Kosten pro Patient und Kasse aufsummiert. Durch Bildung von Vergleichsgruppen führt santésuisse eine Wirtschaftlichkeitsprüfung pro Arzt durch. Sind seine Gesamtkosten höher als 130%, gilt er als auffällig.

Fragestellung

Im Rahmen der an SASIS, dem Datenpool von santésuisse, gelieferten Daten kann auch eine RSS Statistik auf Versicherer-Ebene (RSSV) erstellt werden. Diese Variable hat das Potential, auf verschiedenen Ebenen den Überarztungsverdacht weiter zu adjustieren, wirft aber auch automatisch Fragen auf.

1. Ist der Arzt bei allen Kassen im RSSV auffällig?
2. Kann der RSS-Index anhand des Risikostrukturausgleichs (RSA) adjustiert werden?

Ausführungen

RSSV Auffälligkeit: dieser Abschnitt wurde im Kapitel juristische Problemzonen abgehandelt.

RSSV Adjustierung: Unter Berücksichtigung des Simpsons' Paradox ist es denkbar, dass Kassen, welche Risikoausgleichszahlungen erhalten auch auf der Arztebene zu höheren Kosten führen (Selektion erhöhter Risiken). Es ist deshalb zu

[62] www.physicianprofiling.ch/RSSV012014.xls

überlegen, wieweit ein solcher Effekt sich auf den RSS-Index auswirkt. Anhand der in Tabelle 2 aufgeführten Berechnungen könnte durchaus die Variable „Angehörigkeit des Patienten zur Kasse X" relevante Effekte auf den RSS-Index haben. Dieser Effekt ist dann grösser, wenn die Vergleichsgruppe eher Patienten von billigeren Kassen behandelt hat. Es wäre mathematisch einfach, für solche Effekte eine Korrektur durchzuführen.

Diskussion

In der bisherigen Rechtsprechung wurde der Überarztungsverdacht aufgrund eines aggregierten Mittelwertes der Kosten definiert (sogenannter RSS-Index). Die Aggregation findet dabei auf verschiedenen Ebenen statt, unter anderem werden die Informationen des Indexes auf der Kassenebene (RSSV-Index) unberücksichtigt gelassen, obwohl es hier jährlich um Milliardenbeträge geht. Da die Kostenstruktur der Patienten pro Kasse X erhebliche Unterschiede aufweist, sind detailliertere Informationen notwendig.

Diese Informationen gestatten es, den RSS-Index zu bereinigen, analog zum Vorgehen beim RSA unter den Kassen. Dabei wird der Kasse X eine Zelle zugeordnet. Dies kann weiter durch Alters-/Geschlechtszellen differenziert werden.

Weiteres Vorgehen

Um den Effekt eines RSSV-Index bereinigten RSS-Index besser abschätzen zu können und um über die Häufigkeit der Kostenüberschreitung auf Ebene RSSV-Index mehr Informationen zu erhalten, wird um eine klärende Analyse gebeten. Diese sollte

allenfalls durch eine unabhängige wissenschaftliche Institution erfolgen, einerseits systematisch für alle Versicherer und Aerzte und andererseits anhand bereits verurteilter Fälle.

Potentielle Implikationen

Die RSSV-Adjustierung und die Feststellung der Homogenität von Kostenüberschreitungen durch den Arzt auf der Versichererebene haben ein grosses materielles Potential auf die Rechtssprechung. So wäre es wenig plausibel, wenn ein Arzt der Überarztung angeklagt würde, wenn er nur bei einer oder zwei von z.B. 10 Kassen eine Indexüberschreitung des RSSV aufweisen würde. Demnach ist auch zu überlegen, wieweit die bisherige Rechtssprechung hier eine Fehlbeurteilung in einzelnen Fällen vollzogen hat. Zudem ist es näher zu begründen, warum bei den Kosten auf der Versichererebene ein gesetzlich verankerter RSA legitim ist, diese Variable jedoch bei den Kosten, die der Arzt für seine Patienten verursachen muss oder sollte, überhaupt keine Berücksichtigung fand. Es ist klar, dass diese Problematik auch in den laufenden Verfahren thematisiert werden sollte, um juristisch problematischen Fehlbeurteilungen zuvorzukommen. Zudem kann der ANOVA-Index der santésuisse diese Adjustierungen nicht leisten, da dieser ebenfalls nicht auf der RSSV-Ebene operiert.

Tabelle 3: RSS-Index Veränderungen als Folge der Anzahl behandelten Patienten pro Kasse (bei gleich bleibenden Kosten pro Patient). Der Effekt des Simpsons' Paradox auf den RSS-Index. Kasse A ist die Billigkasse, Kasse C ist die Kasse mit eher teuren Patienten. Kasse B liegt dazwischen.

	Kasse A	Kasse B	Kasse C	Summe
Vergleichsgruppe Anzahl Patienten	1000	1000	1000	3000
Vergleichsgruppe Durchschnittskosten/ Patient	50	75	100	
Summe pro Zelle	50000	75000	100000	225000
Kosten pro Patient (Arzt-Index = 100%)				75
Beispiel 1: gleiche Häufigkeit				
Arzt1 Anzahl Patienten	100	100	100	300
Arzt1 Durchschnittskosten /Patient	50	75	100	75
Summe pro Zelle	5000	7500	10000	
Odds ratio pro Zelle	1.0000	1.0000	1.0000	
Kosten pro Patient				75
Index unbereinigt	100	100	100	**100**
Beispiel 2: Mehr Billigkassen-Patienten				
Arzt2 Anzahl Patienten	**500**	100	100	700
Arzt1 Durchschnittskosten /Patient	50	75	100	75
Summe pro Zelle	25000	7500	10000	
Odds ratio pro Zelle	1.0000	1.0000	1.0000	
Kosten pro Patient				61
Index unbereinigt	500	100	100	**81**
Arzt3 Anzahl Patienten	100	100	**500**	700
Arzt1 Durchschnittskosten /Patient	50	75	100	75
Summe pro Zelle	5000	7500	50000	
Odds ratio pro Zelle	1.0000	1.0000	1.0000	
Kosten pro Patient				89
Index unbereinigt	100	100	500	**119**

Im Ergebnis kann somit festgestellt werden: die Mittelwert-Statistik von santésuisse ist eine Ansammlung von Simpson's Paradox Effekten.

Die Folge ist eine Täuschungsstatistik. Die von santésuisse angebotene Remedur mittels ANOVA Index führt nun anstelle der

oben erwähnte statistischen Täuschungen eine neue Täuschung ein: den „aggregation bias" gepaart mit dem „tautology bias"[63,64].

Der Preis dieser Statistiken bezahlen die PatientInnen in Form von unzweckmässigen Behandlungsentscheiden, verzögerten Abklärungen, vermeidbaren Kosten und schliesslich eine Zunahme der Kosten über die gesamte OKP.

Dieser ökonomische, auch präventiv auf die Kosten wirksame Eingriff in die medizinische Profession ist a priori inakzeptabel.

Welche gesetzgeberischen Massnahmen noch notwendig sind, um die statistischen Methoden zeitgemäss und adäquat zu gestalten, wird sich noch zeigen müssen. Die bisherigen Verlautbarungen lassen daran zweifeln, dass die Krankenversicherer und mit ihnen santésuisse ihre für sie lukrativ erscheinende Statistikmethoden freiwillig aufgeben werden.

[63] http://hyper.ahajournals.org/content/50/2/424.full.pdf
[64] http://www.physicianprofiling.ch/TautologyBias BMJ.pdf

Wissenschaftliche Lösungsansätze und Gutachten

Für die Qualitätsbeurteilung jeglicher ärztlicher Tätigkeit sind professionelle Methoden notwendig. Qualitätsmessung (QM) hat in der Medizin zunehmend hohe Bedeutung. Dies impliziert, dass das Potential für Mislabeling, also Fehlbeurteilung durch QM, steigt. QM hat somit auch das Potential, korrekte ärztliche Tätigkeit als unkorrekt zu beurteilen.

Beispiele für QM mit hohem Fehlerpotential sind Durchschnittsbetrachtungen mit Rückprojektion auf Einzelfälle. Typische Beispiele für solche QM stammen aus dem Umfeld des Public Health, welche Mittelwertunterschiede beobachten, deren Ursachen jedoch nicht analysieren und trotzdem nicht validierte Schlussfolgerungen daraus ziehen, welche dann häufig falsch sind, wenn sie auf Individuen zurückprojiziert werden. Typischer Vertreter solcher QM ist z.B. G. Domenighetti, u.a.m.

Auch das BAG führt QM Statistiken mit hohem Mislabeling-Potential und minimalem Erklärungsgehalt (Sterberaten der Spitäler für verschiedene Krankheiten). Weitere ähnliche Statistiken sind die aggregierten Datenbanken der santésuisse sowie die aggregierten Studiendaten, welche für die HTA Beurteilung der Zweckmässigkeit verwendet werden.

Der Durchdringungsgrad durch QM mit ungenügendem Realitätsbezug als Folge nicht weiter validierter

Mittelwertbetrachtungen hat in den letzten Jahren deutlich zugenommen und wird zunehmend unkritisch auch für die Versorgungsforschung eingesetzt.

Der QM mit Mittelwertbeobachtungen (QMM) müssen künftig QMA, als QM basierend auf "random audits" als neues QM-Instrument entgegengestellt werden.

Dabei wird ein Paradigmawechsel auf mehreren Ebenen möglich. Anstelle der Rückprojektion des Mittelwerts auf das Individuum erfolgt die Prüfung anhand zufälliger Stichproben auf der Ebene des Individuums, welche dank der „random selection" verallgemeinert werden dürfen. Zudem erfolgt die Verallgemeinerung auf der Basis einer fachlich hochstehenden Prüfung und nicht auf häufig zu wenig fachlich nachvollziehbaren Schlussfolgerungen.

Somit sind zwei neue Paradigmen operativ: 1. Vom Individuum zur Verallgemeinerung und nicht umgekehrt, und 2. Fachlich hochstehende statt fachfremde Beurteilung.

Zufällige professionelle Beobachtung und Validierung auf der klinischen Ebene (z.B. interinstitutional audits), besonders bei teuren oder umstrittenen Fällen, muss als neue Methode mit Realitätsbezug der häufig schlechte Qualität unterstellenden QMM Methode entgegengesetzt werden. QM ohne Messung der Morbidität müssen zudem als obsolet bezeichnet werden und sollte im Armamentarium zur Steuerung des Gesundheitswesens eliminiert werden. Ihre Verwendung dient reinen Beobachtungszwecken, nicht jedoch Erklärungs- oder Steuerungszwecken.

So kann die schleichende Erosion ärztlicher Glaubwürdigkeit gestoppt und der zunehmenden Frustration bei Aerzten wegen nicht sachgerechter Aburteilung begegnet werden. Damit ist auch ein QM entsprechend internationalen Standards eine Option.

Gerade die WZW Verfahren der santésuisse zeigen exemplarisch, wie dringend notwendig der doppelte Paradigmawechsel bei der QM ist.

WZW santésuisse – LEVEL I/II

Die santésuisse Methode zur Prüfung der Wirtschaftlichkeit ist nicht zielführend und erfüllt nicht den gesetzlichen Auftrag; sie ist eine einseitige Kostenbetrachtung, welche nur zu Kostenverschiebungen führt. Das Resultat mit 20-30% auffälligen Aerztinnen und Aerzten pro Jahr ist nicht realistisch. In der Deklaration des Erklärungsgehalts ihres Modells (R2) ist santésuisse dem ecological fallacy Problem erlegen. Dies das vorläufige Fazit aus der Überprüfung der WZW Statistik von santésuisse.

Die Eckwerte einer sachgerechten WZW Prüfung sollten künftig internationalen Standards genügen. *Reine Mittelwertbetrachtungen auf der Arztebene sind dabei ohnehin schon lange obsolet*. Für die Modellentwicklung künftiger Verfahren gilt der folgende Regelkreis:

Regelkreise künftiger WZW-Modelle Level I/II

Grundlage bildet die Wahl der erklärenden Variablen aus den Datenbanken der Versicherer. Diese erklären die Kosten anhand von Messinstrumenten (R2, mean absolute prediction error MAPE, Predictive Ratio). Die Wahl der erklärenden Variablen wird nach dieser Validierung angepasst (Morbi-RSA), bessere Modelle erneut validiert. Die schliesslich trotz verbesserter Modelle nicht erklärten Kosten können Anlass für einen Überarztungsverdacht bilden.

Besonders teure Patientenbehandlungen wären dabei zumindest in Form von Stichproben zu auditieren[65] und dann aus dem Prozess zu eliminieren, *um zu garantieren, dass ein Arzt auch teure Patienten behandeln kann* (dieser Auftrag gilt im Rahmen der gesetzlichen Vorschriften uneingeschränkt).

Damit kann das Modell anhand der professionellen Realitätsprüfung weiter angepasst und validiert werden. Das Modell wird zudem weiter validiert durch aerzteeigene Datenbanken, welche insbesondere bei einem Überarztungsverdacht operativ werden, nachdem diese Prozesse selber validiert wurden, z.B. bei den „out-of-the-pocket" Zahlungen.

Demnach kann ein auffälliger Arzt im Screeningverfahren durch das aerzteeigene Modell revalidiert werden.

---

[65] www.physicianprofiling.ch/ProfilingDudleyMedCare2003.pdf

Hybride Datenbanken bilden eine weitere unabdingbare Qualitätskontrolle, indem zum Beispiel Medikamentendaten der Versicherer in die Datenbanken der Aerzte integriert werden.

Basismodell mit Morbiditätsvariablen

Tabelle 4 zeigt eine Auswahl möglicher Variablen, welche aktuell bei der RSS Statistik und im Modell der CSS (K. Beck, U. Kunze) Verwendung finden. Dieses Modell kann nun durch weitere Variablen ergänzt werden und bildet die Basis für einen internen und externen Validierungsprozess gemäss Regelkreis.

Die Problematik des RSS Indexes wird besonders verdeutlicht anhand des Vergleichs mit Methoden der Risiko-Adjustierung (risk adjustment) unter Berücksichtigung international gültiger Standards. Am meisten fortgeschritten ist in der Schweiz das vom CSS-Institut von Prof. Konstantin Beck und Frau Ute Kunze entwickelte CSS-eigene Modell[66]. Dieses wird explizit auch verwendet für Wirtschaftlichkeitsprüfungen.

[66] www.physicianprofiling.ch/WZWCSS072013.pdf

Tabelle 4: Vergleich der Variablen für die Bildung des RSS- und des CSS Indexes [predictive ratio = index /100]

Variable	RSS	CSS	Bemerkung
Direkte Arztkosten t	☑	☒	t=aktuelles Jahr
Direkte Arztkosten t-1/-2	☒	☑	t-1 und t-2=Vorjahr(e)
Medikamentenkosten t-1/-2	☒	☑	
Physiotherapie t1	☑	☒	
Physiotherapie t-1/-2	☒	☑	
Indirekte Laborkosten t1	☒	☑	
Indirekte Laborkosten t-1/-2	☒	☑	
Behandlungsanteil Arzt t	☑	☑	Übereinstimmung
Übrige Behandlungskosten t-1	☒	☑	
PCG / Tarmedpositionen	☒	☑	
Spitalaufenthalt T-1/-2	☒	☑	
Altersgruppe (nicht aggregiert)	☒	☑	
Geschlecht	☑	☑	Übereinstimmung
Geboren t/t-1	☒	☑	

Während der RSS-Index lediglich die Durchschnittskosten vergleicht, wird der CSS-Index mittels einer Regressionsanalyse errechnet. Anhand des Regressionsmodells kann die Bedeutung einzelner Variablen statistisch beschrieben und in ein Erklärungsmodell der Kosten eingebaut werden.

Fehlt eine Variable im CSS-Index, taucht jedoch mit RSS-Index auf, bedeutet dies, dass das Modell der CSS diese Variable für die

Erklärung der Kosten verworfen hat. Wie Tabelle 4 zeigt, dass der RSS-Index ausser beim Geschlecht auf anderen Variablen als der CSS-Index basiert. Insbesondere schliesst der CSS-Index Morbiditätsvariablen ein und erhöht die Konfidenz anhand der Berücksichtigung von Variablen der Vorjahre. Der resultierende Index ist das Resultat des Vergleichs von erwarteten und tatsächlichen Kosten (predictive ratio), ein internationales Mass für das Physician Profiling.

Der CSS Aerzteindex hat jedoch seine eigenen Bias: Multikollinearität, Tautologie, Aggregation.

Grundsätzliche Anforderungen an die Prüfmethode

Die internationale Literatur hat für das Physician Profiling nur die multivariable Regressionsmethode ausreichend validiert und akzeptiert[67].

Analysefehler (Bias) sind:

- Aggregation Bias: aggregierte Variablen (z.B. Altersklassen) bewirken eine Inflationierung des R2. Analysen müssen ausschliesslich unter Verwendung von Einzeldaten der Individuen erfolgen.
- Multikollinearität-Bias: Variablen, welche untereinander korreliert sind, führen zu Fehlbeurteilungen bei der Aussagekraft des Modells.

[67] www.physicianprofiling.ch/ProfilingStateAndImprove2013Overview.pdf

- Tautologie-Bias: Erklärung der Arztkosten durch die Arztkosten (z.B. sogenannten ZSR-Variable).
- Kohortengrössen-Bias: Predictive Ratios sind in Arztpraxen mit vergleichsweise weniger Patienten (kleine Kohortengrössen) deutlich geringer[68].

Weitere Direktiven

- Definition eines Goldstandards mit dem Zweck der Berechnung der Testgenauigkeit (Sensitivität, Spezifität, Accuracy) der Prüfmethode anhand der 4-Felder-Tafel[69].
- Auditing Outliers (sehr teure Patienten z.B.), welche klinisch beurteilt werden und konsekutiv aus der Prüfung ausgeschlossen werden[70].

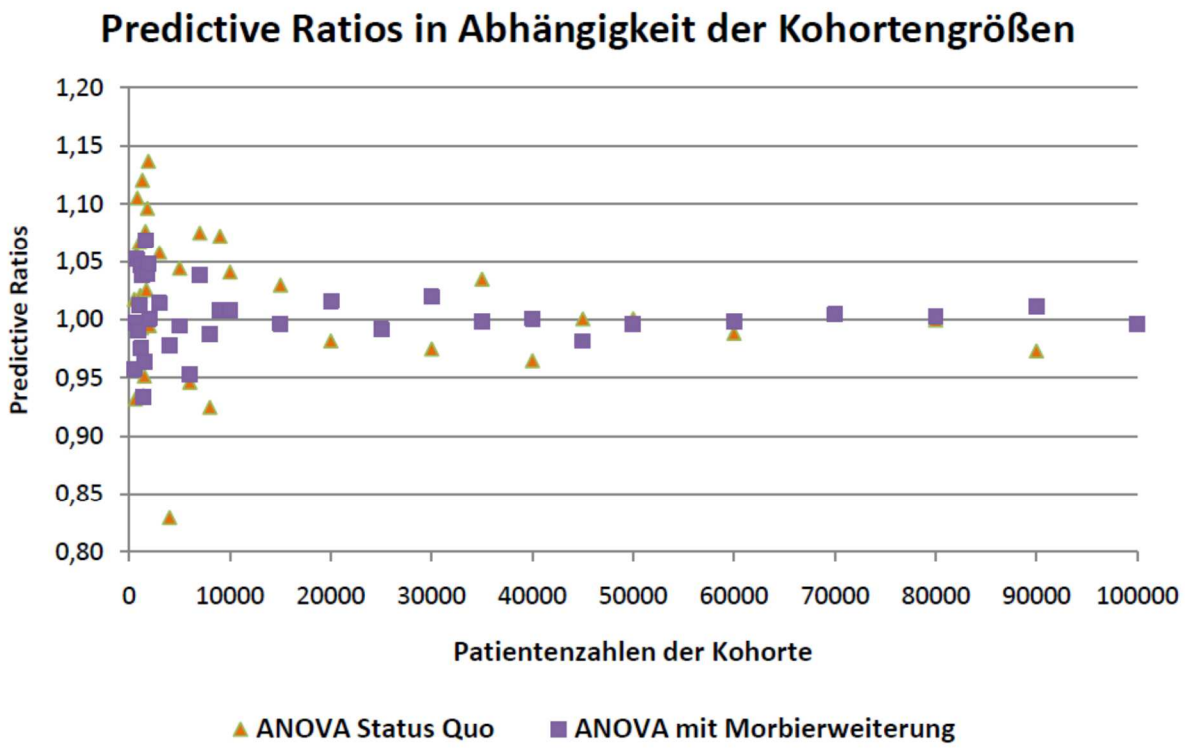
Die Praxisgrösse wurde von ***Prof. Jürgen Wasem im Gutachten für den VEMS aus dem Jahr 2010 als relevant definiert*** (S. 23[71]). Die predictive ratios waren in Arztpraxen mit vergleichsweise wenigen Patienten (kleine Kohortengrösse) deutlich schlechter (Abbildung 1).

Entsprechend muss dann auch die als auffällig geltende predictive ratio an die Kohortengrösse angepasst werden, wie dies der CSS-Index tut (Folie 17). Für grosse Praxiskohorten gilt ein Index ≥ 120% als auffällig, während bei kleineren Praxiskohorten ein auffälliger Index bei ≥150% angesetzt wird. Dies bedeutet auch, dass trotz Einbezug zahlreicher Morbiditätsvariablen in der Regel eine Toleranzgrenze von ≥130% abgeleitet wird.

---

[68] www.physicianprofiling.ch/gutachtenwasem2010.pdf
[69] www.physicianprofiling.ch/RAMisclassificationRAND.pdf
[70] www.physicianprofiling.ch/ProfilingDudleyMedCare2003.pdf
[71] www.physicianprofiling.ch/gutachtenwasem2010.pdf

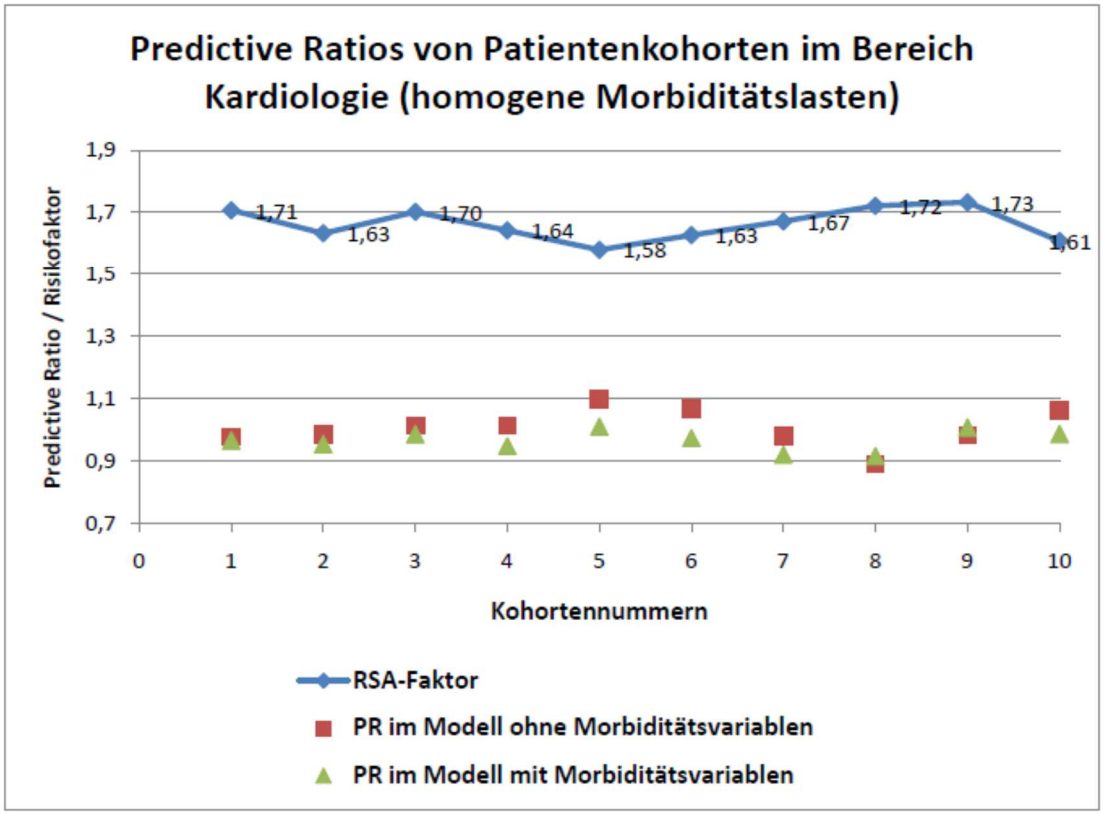
Abbildung 1:



Unter Verwendung solcher internationaler Standards unter Einschluss von Morbiditätsvariablen resultiert dann gemäss CSS-Index in den meisten Fällen (97%) betreffend Überarztungsverdacht ein neutraler oder negativer Befund. Dies bedeutet, dass jährlich im Aerztekollektiv der CSS Krankenversicherung lediglich 3% der Aerzte als auffällig gelten. Dies kontrastiert scharf mit den 25% auffälligen Aerzten, welche der RSS-Index jährlich produziert. Hochgerechnet auf die ca 15‘000 ambulant tätigen Aerztinnen und Aerzte in der Schweiz sind dann statt 3750 nur noch 450 Aerzte weiter prüfenswert.

Die Bedeutung des Morbi-RSA wurde auch im Gutachten von Prof. Wasem 2010 klar herausgearbeitet[72]: sind die Krankheitsschweregrade zufällig ausgewählt, genügt die RSS Statistik zwar:

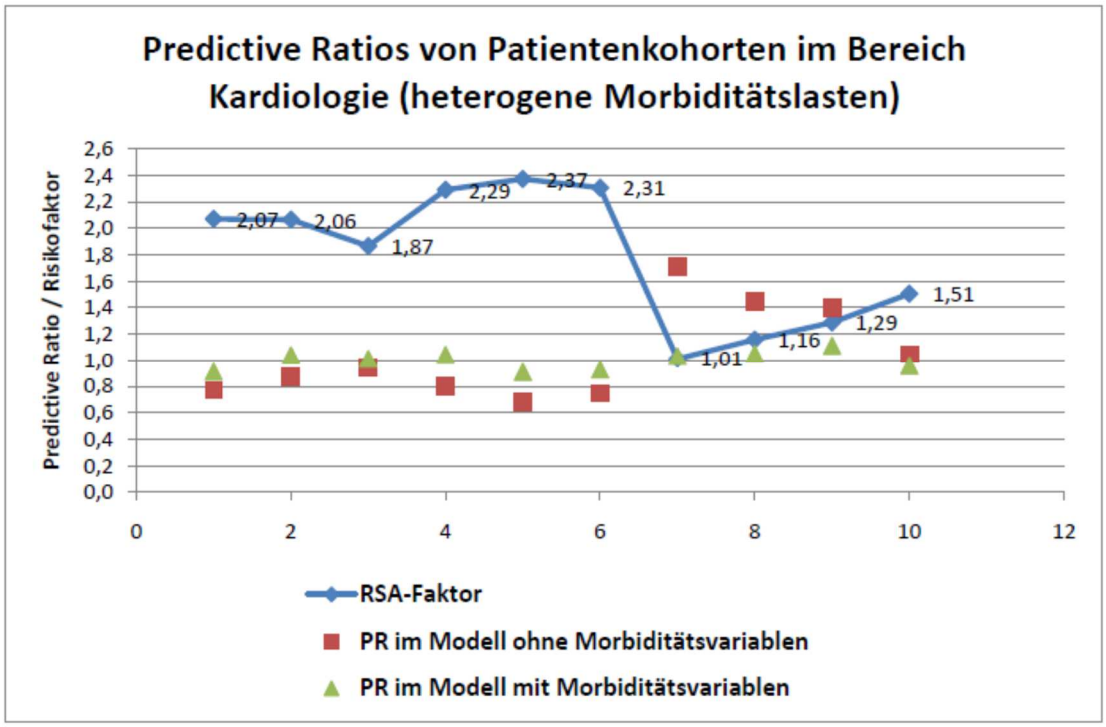
Abbilung 2: Zufällige Morbidität, RSS Statistik ausreichend



[72] http://physicianprofiling.ch/gutachtenwasem2010.pdf

Wird hingegen die reale Morbidität im Modell angeschaut, ist die RSS Statistik irreführend:

Abbildung 3: Irrführung durch RSS Statistik ohne Abbildung der Morbidität:



Gemäss internationaler Literatur ist der Risikoausgleich RA anhand von Morbiditätsvariablen, ev. unter Zuzug von pharmaceutical cost groups und Krankheitsschweregraden betreffend R2, Prädiktionsfehler und prädiktive Ratio durchwegs relativ niedrig (Tabelle 5).

Tabelle 5: Vergleich von R2, MAPE und Predictive Ratios

(Quelle: www.physicianprofiling.ch/ProfilingStateAndImprove2013Overview.pdf)

Tool	As offered	100K truncation	250K truncation	Recalibrated + 250K truncation	Prior Costs + 250K truncation	Concurrent +250K truncation
R-square (%)						
ACGs	16.2	20.8	19.2	19.6	23.0	31.5
CDPS	12.4	17.6	14.9	17.7	24.6	36.8
CRGs	14.9	19.3	17.5		20.5	
DCGs[1]	17.4	22.3	20.6	21.3	26.5	54.5
RxGroups	16.8	23.8	20.4	20.5	27.1	36.9
PRGs	17.2	25.0	20.5	21.2	27.4	
Medicaid Rx	12.9	19.3	15.8	17.7	26.3	31.0
Impact PRO	21.3	26.3	24.4	25.6	27.2	
ERGs	16.2	23.7	19.7	20.0	26.5	43.3
Mean absolute prediction error (MAPE, %)						
ACGs	90.4	87.7	89.9	88.8	86.2	76.6
CDPS	95.8	93.4	95.3	91.9	85.6	79.6
CRGs	91.4	88.7	90.9		86.6	
DCGs	88.0	85.3	87.5	87.0	82.5	63.4
RxGroups	85.9	82.9	85.3	85.3	80.7	72.4
PRGs	86.4	83.4	85.8	85.6	80.9	
Medicaid Rx	90.2	87.3	89.6	88.4	81.9	78.0
Impact PRO	82.4	79.3	81.8	81.6	80.6	
ERGs	87.0	84.1	86.4	86.1	81.2	68.0
Predictive ratio for patients in top 1 percent of current year cost[2]						
ACGs			22.1	21.8	27.1	55.9
CDPS			14.6	18.2	24.2	51.6
CRGs			22.0		28.4	
DCGs[1]			23.4	20.5	25.2	65.3
RxGroups			19.9	18.2	24.9	51.9
PRGs			21.1	19.2	25.0	
Medicaid Rx			16.0	15.9	24.2	44.1
Impact PRO			30.0	26.9	29.7	
ERGs			17.7	18.0	24.3	54.7
Predictive ratio for patients with mental health diagnoses in prior year						
ACGs			92.3	98.0	99.2	100.6
CDPS			92.5	91.1	94.6	80.2
CRGs			89.0		101.8	
DCGs			95.9	98.5	100.3	94.7
RxGroups			88.6	89.2	96.8	79.6
PRGs			87.1	87.4	94.9	
Medicaid Rx			94.0	88.1	96.8	79.9
Impact PRO			98.0	97.8	100.0	
ERGs			91.9	92.3	97.2	85.2

Die Chance WZW

Einer gemeinsamen Einrichtung WZW könnten neben der Übertragung von Wirtschaftlichkeitsprüfungen auf Level I/II weitere wichtige Aufgaben übertragen werden.

Voraussetzung bildet der Einschluss von Diagnosecodes unter Berücksichtigung von Schweregraden von Krankheiten und Polymorbidität.

Aus Beobachtungen von Veränderungen können wichtige epidemiologische Fragestellungen, Aspekte der Patientensicherheit und validierbare Steuerungsentscheide in der medizinischen Versorgung erforscht und verfügbar gemacht werden.

Diese Aufgabe wäre idealerweise ebenfalls gemeinsam mit den Ärzten, den Versicherern und den Behörden zu lösen. Dabei sind auch durchaus Modelle zu prüfen, welche auf eine Aerzte-Vergleichsgruppe komplett verzichten, da die Homogenisierung der Vergleichsgruppen immer problematisch bleiben wird.

In diesen künftigen Modellen würde für die Schweiz der Kostenbedarf anhand der Patientenvariablen ermittelt und mit den tatsächlichen Arztkosten verglichen. Dies entspricht internationalen Standards.

Solche Modelle eliminieren somit die RSS Statistik komplett aus dem Fokus und ermöglichen auch die korrekte Erfassung von Kostenüberschreitung bei geringerer Arzt-Patientenbindung (siehe Modell Beck / Kunze).

Diese Eckwerte zeigen, dass die Beurteilung von dem, was wirksam, zweckmässig und wirtschaftlich ist, zunehmend aus dem klinischen Alltag verschwindet und ersetzt wird durch Verallgemeinerungen mit Mittelwertstatistik, welche als beobachtende Forschung kein Wissen schafft, somit unwissenschaftlich ist.

Die fachfremde Beurteilung der WZW Kriterien nimmt groteske Züge an, ordentliche Behandlungen teurer PatientInnen werden absichtlich verunmöglicht (und an andere Institutionen weitergewiesen oder inadäquat behandelt – versteckte Rationierung).

Vor diesem Hintergrund muss neu gedacht werden, wer künftig bestimmt, was WZW ist. Dies kann unmöglich eine Sache der Schweizerischen Bundesverwaltung bleiben[73].

Mit der Diskussion um verbesserte WZW Verfahren mit santésuisse eröffnet sich eine grosse Chance für die Aerzte und ihre PatientInnen, mithin für die FMH, hier entscheidende Verbesserungen für die zweckgerechte medizinische Versorgung zu erarbeiten.

[73] http://physicianprofiling.ch/WZWBAG2011.pdf

Weiteres Gutachten bestätigt VEMS

Der VEMS stellte den Gesundheitsforen in Leipzig / ForBiG ***16 Fragen*** im Sinne eines Kurzgutachtens, welches in Ergänzung zum Wasem / ForBiG Gutachten 2010 gedacht ist. Die Antworten erstellt von H Doz. Dr. rer. nat. habil. Walter Warmuth (Gesundheitsforen Leipzig) unter Beteiligung von Helmut Dahl und Dr. David Matusiewicz (ForBiG Forschungsnahe Beratungsgesellschaft im Gesundheitswesen)

Die Fragen ergeben sich aus der bisherigen Beobachtungen und wurden von Michel Romanens formuliert. Das Gutachten ist im Internet im Original abrufbar

www.physicianprofiling.ch/wzwwarmuth2014.pdf

Einleitung

Es geht im Gutachten wesentlich um die Überprüfung der Anwendbarkeit zumindest der Nützlichkeit[74] mathematischer Modelle in Gesundheitswesen. Behinderungen ergeben sich daraus, dass

1. die empirische Sozialforschung (auch Sozialstatistik) gewöhnlich von empirischen Mittelwerten von Anzahlen, Dauern und Kosten ausgeht und annimmt, dass sich Beobachtungswerte, die um diese schwanken, „gleichberechtigt“ ausgleichen,
2. der Fundus der mathematischen Methodenstatistik fast ausschließlich ein mathematisch normalverteiltes Schwanken um Mittelwerte bedingt und es so zu einem überirdischen Konsens angewandter mathematischer Statistik und empirischer Sozialforschung kommt,
3. risikomathematische Herangehensweisen (versicherungsmathematische, finanzmathematische, extremwerttheoretische, …) Sozialforschern und Statistikern[75] weitgehend unbekannt sind.

Problematisch bleibt, dass ein Gesundheitswesen konstituierendes System als Organisationssystem nicht von der Finanzierung eines Gesundheitswesens getrennt wird.
Versorgung und Versorgungsforschung könnten nach der Theorie sozialer Systeme ihre Heimat im Organisationssystem haben. Gesundheitsökonomie

---

[74] „Alle mathematischen Modelle sind falsch, manche sind nützlich.“ (Sachs, L., Statistische Methoden, Springer-Verlag, Berlin 1993, S. 146.)

[75] Die ETH verfügt auf diesem Gebiet über besondere Traditionen; ich möchte auf den Schweizer Paul Embrechts und den Deutschen Hans Föllmer verweisen. Meine liebe Frau Elke hat mit dem Schweizer ETH-Absolventen Moritz Adelmeyer das für die Lehre in der Schweiz und Deutschland „gültige“ Buch „Finanzmathematik für Einsteiger“ geschrieben.

meint sich mit der Finanzierung sozialer Systeme abgeben zu müssen und zu können, ohne die unterschiedliche bis fehlende Ausdifferenzierung zu beachten.

Die Fragen wurden von Dr. Michel Romanens, VEMS, den deutschen Gutachtern zur Beantwortung vorgelegt:

Frage 1.

Wie beurteilen Sie einen fixen RSS-Index > 130% betreffend dem Beweisgrad für Überarztung generell (also für alle Aerzte gleichbleibend)? (Anm.: hier geht es darum, ob eine fixe Grenze über alle Facharztrichtungen hinweg aus Ihrer Sicht je zielführend war, mithin, ob ganz grundsätzlich die Methode überhaupt je als juristische Beweisgrenze validiert und damit verwendbar war). Siehe auch link A (ZSRKanton.pdf) und link B (ZSRKanton2.pdf).

Die seit den 1970er Jahren in der Schweiz praktizierte „statistische Durchschnittsmethode" wird als zweckrational ausgegeben, sie ist es nicht. Sie orientiert sich am Mittelwert und unterstellt, dass es Abweichungen gibt, nach oben und nach unten, die sich letztendlich ausgleichen. Einer Asymmetrie des Ausgleiches meint man mit einem Rechnungsstellerstatistik-Index zu begegnen, der in einer Vorstellung von einem Konfidenzintervall hoher Sicherheit bis zu 130 % des Indexes reicht.

Kostenverteilungen in Gesundheitswesen haben eine spezifische Gestalt. Zunächst gebe ich ein Beispiel nach der Arztgruppe an.

Abbildung 1. Heavy-tail bei der Gestation (Mittelwert der Kosten 4.455,28 €, Standardabweichung 4.355,72 rund 100% des Mittelwertes), 62 % der Fälle liegen unter dem Mittelwert, 71 % der Fälle liegen unter dem Mittelwert + 10% des Mittelwertes), veröffentlicht in Degenkolbe, B.; Gette, E.; Seiffert, A.; Warmuth, W.: Justierung durch Vergleich, Predictive Modeling News, (2011)1. S. 12-14. Die Anwendung des Modells der Normalverteilung schließt den

Kostenausgleich der Kosten bis ca. 20.000 € durch negative Kosten ab ca. -10.000 € ein. Die hohen Kosten sind medizinischen Notwendigkeiten geschuldet, die Konzentration vor dem Mittelwert resultiert aus der extremen Überarztung durch medizinisch nicht notwendige Kaiserschnitte (mehr als das 10-fache des medizinisch Notwendigen). Bezüglich dieser Kostengruppe gleichen sich das deutsche und das schweizerische Gesundheitswesen schon seit etwa fünf Jahren.

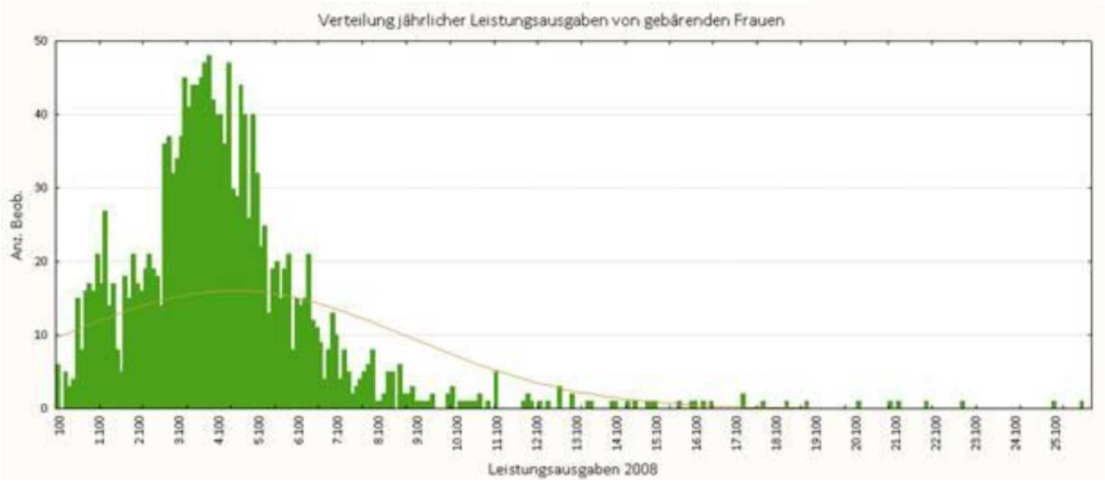


Abbildung 2. Behandlungskostenverteilungen in Gesundheitswesen (aus Warmuth, W.: Sie wollen doch nur spielen: Gesundheitsökonomie in der GKV, Versicherungswirtschaft, 8(2013)13. S. 20-24.)

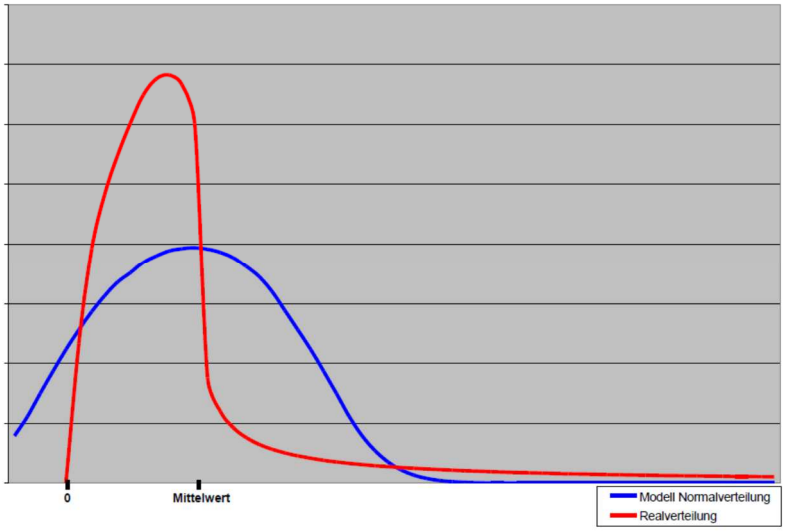
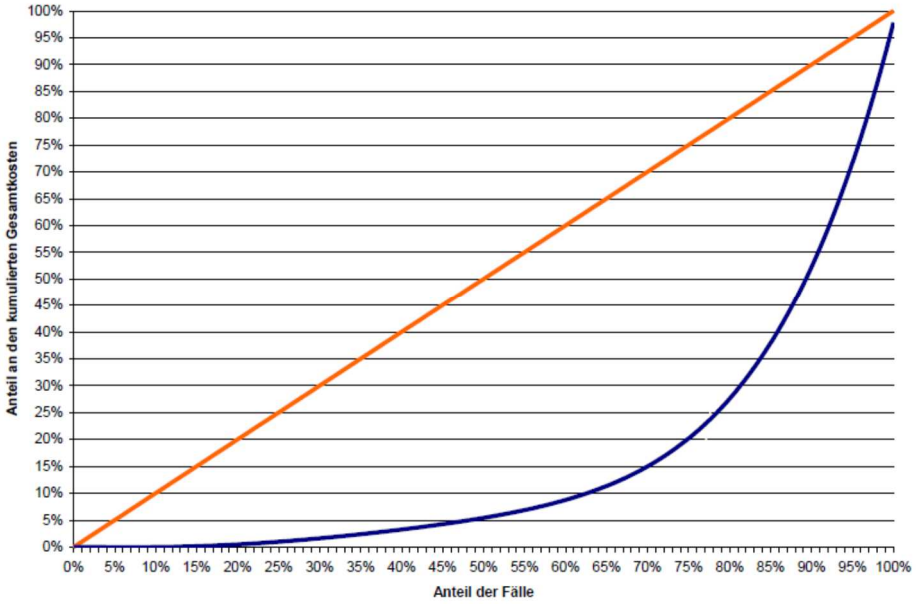


Abbildung 3. Konzentration der Behandlungskosten in Gesundheitswesen (aus Warmuth, W.:Sie wollen doch nur spielen: Gesundheitsökonomie in der GKV, Versicherungswirtschaft, 68(2013)13. S. 20-24.)



Da die Kosten in der Abbildung 3 der Größe nach sortiert sind, entspricht der RSS3-Index von 130 % gerade 65 % der kumulierten Kosten, das sind etwa 94 % der gesamten Fälle. Es ist also „normal", dass 6 % der Fälle noch 35 % der Kosten ausmachen. Kein einziger der 6 % teuersten Fälle steht a priori für eine Überarztung.

Überarztung im niedrigen Indexbereich? Von einer Überarztung in erster Linie kann ausgegangen werden, wenn 94 % der kleinsten Fallkosten kumuliert merklich die prognostizierten 65 % der kumulierten Kosten übersteigen. Es gibt Methoden, die eine auffällige Konzentration im niedrigen Indexbereich belegen. Diese Methoden habe ich in großen Projekten im Zusammenhang mit der Betrugserkennung bei Versicherern (Allianz Versicherungs-AG) und Gewerbesteuererklärungen (Landesamt für Steuern Bayern) benutzt.

Es wird dabei jedoch nur die Vermutung des Faktes gestärkt, der einzelne Fall oder die einzelnen Fälle sind nicht belegbar, da sie in der Masse maskiert sind. Veröffentlicht sind viele Methoden in Warmuth, W.: Und sie betrügt uns doch, Versicherungswirtschaft, 62(2007)3. S. 200-203 und Warmuth, E.;Warmuth, W.: Was haben Aktienkurse und Fibonacci-Zahlen gemeinsam? Der Mathematikunterricht, 58(2012)1. S. 34-38 (mit weiteren ETH-Zitaten zu Methoden bei der Betrugserkennung).

Überarztung bei höheren Kosten? Eine Überarztung in zweiter Linie bei den 6 % höheren Kosten ablesen zu können, halte ich für schwierig bis unmöglich. Diese Kosten sind nicht unbedingt allen Arztgruppen, allen Regionen, … zuordenbar. Diese höheren Kosten sind teilweise managebar, teilweise poolbar, ganz sicher aber rückversicherbar u. U. sind diese Risiken verbriefbar.

Ein Realkostenausgleich für die 94 % der Fälle ist äquivalent zu einem Ausgleich der Kosten für die 94 % nur zu je 65 % von den mittleren Kosten. Das gilt aber nur in Bezug auf große Bereiche des Anfallens der Kosten, wo auch die höheren Kosten bereichsanteilig vorkommen. Vermutlich ist das unterhalb der Schweiz nicht praktizierbar, für den einzelnen Arzt gilt das keinesfalls.

Die Abbildung 3 variiert nur leicht bei Abrechnungen der Leistungserbringer gegenüber deutschen Krankenkassen, wenn auf Versichertengruppen fokussiert wird: Bestandsversicherte, Kündiger, Verstorbene, Neugeborene, Neuversicherte (ein Jahr oder älter) wenn auf Leistungsbereiche der Krankenversicherer fokussiert wird: stationäre Leistungen, ambulante Leistungen, Arzneimittel über Apothekenabrechnungszentren, sonstige Leistungen (z. B. Heil- und Hilfsmittel, aber auch Arzneimittel bei z. B. Enzymersatztherapien, außerbudgetäre Leistungen, Rehabilitations-Leistungen, Pflege-Leistungen nach dem SGB V, …), zahnärztliche Leistungen, Arbeitsunfähigkeits-, Entgeltfortzahlungs- und Krankengeldleistungen (in Deutschland werden diese versicherungsfremden Leistungen aus den Krankenversicherungsbeiträgen quersubventioniert) kassenbezogen spielt dabei die Alters- und Geschlechtsstruktur, die Beschäftigtenstruktur (dabei nur wenig die

Erwerbminderungsrentenstruktur), die Dauermedikations- und Nachbehandlungsstruktur eine Rolle.

Eine fixe Grenze über alle Facharztrichtungen hinweg ohne Berücksichtigung der spezifischen Behandlungskostenstruktur in Gesundheitswesen war von Anfang an nicht zielführend, sie kann so auch nicht evaluiert worden sein. Wenn die spezifische Behandlungskostenstruktur auch weiterhin nicht einbezogen werden soll, dann ist es unerheblich, ob eine facharztspezifische Verfeinerung gewählt wird oder nicht. Die hohen Kosten (6 % der Fälle mit 35 % der Kosten) lassen sich kaum auf die Arztrichtungen oder auf Ärztegruppen mit spezifischer Behandlungskostenstruktur aufteilen, auf einzelne Ärzte überhaupt nicht. Es ist deshalb auch nicht mit arztgruppenspezifischen Absenkungen für die abrechenbaren Fälle zur Stützung der Fälle oberhalb der „Toleranzschwelle" zu rechnen. Die Ungleichbehandlung der Fachrichtungen, Arztgruppen und Ärzte ist beträchtlich. Wenn die Ärzte zur Selbstregulierung aufgefordert werden, dann können sie das nur, indem sich z. B. Ärzte verschiedener Arztrichtungen zusammenschließen, indem z. B. Leistungsangebote ein- oder ausgeschlossen werden, indem z. B. nur gut abrechenbare Versichertenrisiken zur Behandlung selektiert werden usw.

Unter Berücksichtigung der spezifischen Behandlungskostenstruktur sind sogar einige Spezifikationen unterhalb der Arztrichtungen sinnvoll. Sie können den Ausrichtungscharakter der Arztgruppe bzw. die Bestandszusammensetzung der Patienten bzw. Versicherten berücksichtigen, die hohen Kosten bleiben aber auch hier außen vor.

Frage 2.

Halten Sie, und falls ja, warum, eine Anpassung der Beweisgrenze (nach oben oder nach unten) für eine unabdingbare Vorbedingung (Flexibilisierung der Beweisgrenze) für ein sachgerechtes Beurteilungsverfahren?
(Anm.: hier geht es um die Frage, anschliessend an Frage 1, welche Argumente ganz grundsätzlich dafür sprechen, dass eine Flexibilisierung eine Grundvoraussetzung für ein korrektes Verfahren darstellt, und zwar unabhängig davon, ob mit oder ohne Vergleichsgruppe gerechnet wird).

Eine fixe obere Grenze über alle Facharztrichtungen hinweg ohne Berücksichtigung der spezifischen Behandlungskostenstruktur in Gesundheitswesen war von Anfang an nicht zielführend.

Wird an der nicht sinnvollen Interpretation „Beweisgrenze" ohne Berücksichtigung der spezifischen Behandlungskostenstruktur in Gesundheitswesen festgehalten, dann erscheint das Finden von oberen Grenzen zumindest nach Arztrichtungen als Verbesserung, es ist für diesen Zweck notwendig. Die „Toleranzschwellen" in den einzelnen Arztrichtungen stellen u. U. ein Konstrukt dar, das mit einer großen Menge von Abrechnungen problemlos umgehen kann. Von den normalen hohen Kosten sind die Arztrichtungen aber nur sehr vage vorhersehbar betroffen, die Volatilität nimmt mit dem Abnehmen der Anzahl der Ärzte sehr stark zu. Der einzelne Arzt ist schicksalhaft „Toleranzgrenzen" ausgeliefert.

Auch die Ergänzung durch untere Grenzen ist methodisch sinnvoll, aus der Sicht der „Steuerungswirkung" sollte das aber wohlüberlegt werden. Aus den Erfahrungen bei der Weiterentwicklung des DRG-Systems gibt es sehr wohl Anlässe, wo auf eine untere Grenze zu verzichten ist. Mit einer Wahl einer unteren Grenze werden bei den dominanten symptomatischen Anlässen für Arztkontakte m. M. n. die Aufwendungen bei der Diagnostik zunehmen.

Ohne Berücksichtigung der spezifischen Behandlungskostenstruktur in Gesundheitswesen ist ein „korrektes" Verfahren unmöglich, anders ausgedrückt: ein Verteilmechanismus, der ausschließlich auf das Kompensieren von Überschreitungen durch das Unterschreiten vom Mittelwert setzt, ist für einen Kostenausgleich in Gesundheitswesen nicht nützlich.

Ein kleiner Index deutet auf eine „selektive" Versorgung hin. Das kann dann ein Indiz für eine Unterversorgung sein, mehr noch wird es sich aber um eine „nicht zu Ende erfolgte" Versorgung bei diesem Arzt handeln, die Weiterversorgung erfolgt bei einem anderen Arzt. Eine vorsichtige Annäherung an die Bewertung solcher Situationen setzt das Liefern von patientenbezogenen Abrechnungen an die santésuisse voraus. Aber auch dann würde versteckt die Vorstellung von der Möglichkeit einer generellen „korrekten" Behandlung eine Rolle spielen.

Frage 3.

Ist die Prüfung sämtlicher Aerzte zielführend oder wäre auch eine zufällige Stichprobe (z.B. 5% pro Facharztgruppe) ausreichend? (Anm.: hier geht es um die Frage, ob ein geringerer Screening-Aufwand Gelder frei machen würde, welche für eine verbesserte Screening Methode verwendet werden könnten bzw. was das Rationale in Deutschland ist, pro Jahr nur 5% pro Facharztgruppe zu screenen).

Zufällige Referenzstichproben sind ausreichend. Dabei bezieht sich die Referenz aber nicht allein auf Arztgruppen, sondern auch auf Konstellationen von Arztgemeinschaften, die mit einer gruppenübergreifenden Ausstattung (z. B. zur Diagnostik) versehen sind (Nutzenvorteile vs. Verwertungsdruck), auf Patientengruppen, auf Regionen – ganz wichtig ist aber dabei auch die Abbildung des Ausmaßes, das durch die Patienten bedingt ist, die zum Überschreiten der „Toleranzschwellen" bei den Ärzten führen. Solche Referenzkonstellationen sind sicherlich vereinbar. Hierbei ist eine Zufallsstichprobe in dem Sinne zu stratifizieren, dass über der „Toleranzschwelle" keine „realen" Ärzte in der Referenz verbleiben, Referenzkonstellationen übergreifend werden z. B. virtuelle Ausgleichpools gebildet.

Wie bei den DRG-Weiterentwicklungen könnte eine Referenzkonstellation für ein Kalenderjahr gelten. Die Zufallsauswahl in den Schichten ist Bedingung.

Frage 4.

Ist die ANOVA Korrektur anhand der ZSR Variable eine ausreichend erklärende Variable im ANOVA Modell betreffend der Problematik des „tautology bias“ und wie ist diese Variable in der Literatur begründet, bzw. was beschreibt sie eigentlich?
(Anm.: anhand der ZSR Nummer ist jeder Arzt mit seinen Kosten identifizierbar. In der ANOVA Korrektur beginnt das Modell mit der ZSR Nummer als fixed effect variable zu rechnen und ist nichts anderes als der RSS_6-Index des Arztes. In der Gleichung RSS-Index=ANOVA[ZSR] ist der RSS-Index enthalten, man erklärt also die Arztkosten durch die Arztkosten, was wir als Tautology Bias bezeichnen. Wichtig zu wissen: die ZSR Variable erklärt als fixed effect variable je nach Facharztrichtung ebenfalls ca. 40% der Arztkosten).

Mittelwertunterschiede in den ärztlichen Leistungen sollen mit der ANOVA Korrektur ausgeglichen werden. Kosten werden zu korrigierten Kosten indem die Einflussfaktoren Facharztgruppe, Patientengut und der Standort-Kanton „herausgerechnet“ werden. Die Korrektur erfolgt so, dass die korrigierten Mittelwerte a posteriori hypothetisch als gleich anzusehen sind. Diese Methode bedingt bei der Schätzung von Parametern zumindest annähernd symmetrisch um die Mittelwerte schwankende ärztliche Leistungen, und das bei gleicher Varianz, letztendlich muss es sich um Normalverteilungen handeln. Auch wenn beide Annahmen nicht erfüllt sind, ist die Herangehensweise in Teilen bei Sozialwissenschaftlern üblich, bei Gesundheitsdaten eher selten hilfreich, bei Gesundheitskostendaten nur bedingt interpretierbar. Die ANOVA Korrektur sagt nur etwas über die Kosten aus, die um einen von höheren Kosten bereinigten Mittelwert schwanken und sich dabei ausgleichen. Im Rahmen der in der Schweiz letztendlich nicht zur Diskussion stehenden „statistischen Durchschnittsmethode“ stellt die ANOVA Korrektur eine Veränderung dar, die, dort wo sie anwendbar ist, weder gut noch schlecht ist. Die bedingte Anwendbarkeit bezieht sich nur auf den sich ausgleichenden Kostenbereich. Es wird an keiner Stelle untersucht, ob eine Varianzanalyse überhaupt anwendbar ist.

Die Methode gründet sich auf eine Zerlegung der Varianz in als identisch postulierte Schwankungen um die zunächst noch unterschiedlichen Mittelwerte und die durch die Unterschiede der Mittelwerte bedingten Schwankungen. Die unterschiedlichen Mittelwerte werden zu gleichen Mittelwerten „korrigiert". Methodisch ist das im gesetzten Rahmen soweit in Ordnung. Nun ist die Kostenverteilung aber weit entfernt davon, normalverteilt zu sein. Wenige hohe Kosten in Arztreferenzkonstellationen beeinflussen sehr stark den Mittelwert, die Varianzen sind ebenfalls hochsensibel betroffen. Von identischen Varianzen auszugehen ist nicht lauter.

Die Berücksichtigung der Facharztgruppe, des Patientengutes und des Standort-Kantons erscheint zunächst als Verbesserung. Es ist nicht ermittelbar, zu welchen Lasten die Standardisierung führt:

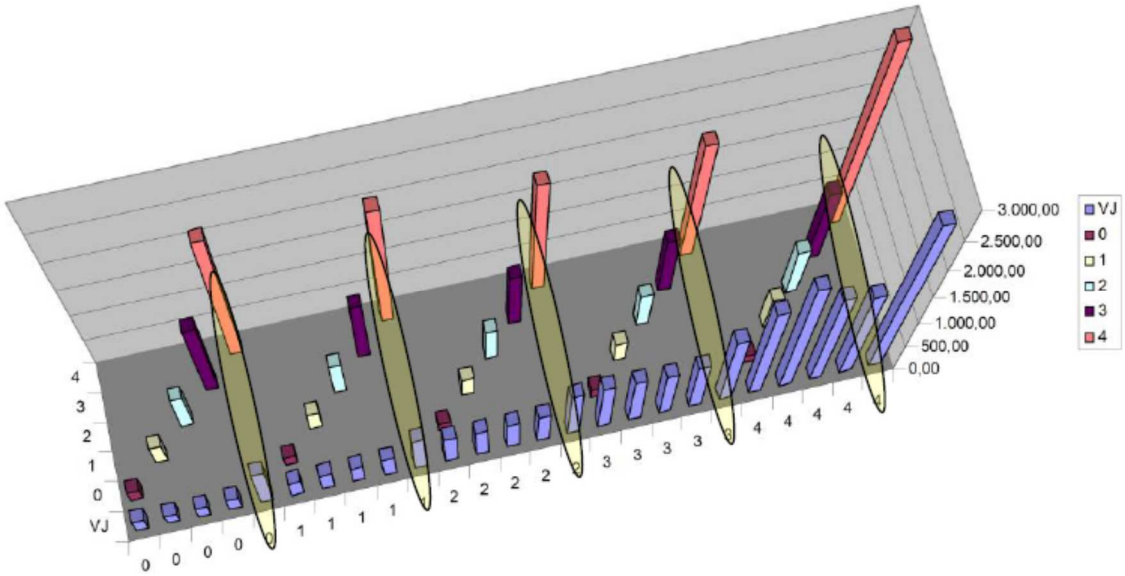
Beispiel: Eine Überarztung aus einem Überangebot von Leistungserbringern führt zu geringeren Kostenmittelwerten und größeren Varianzen. Die geringeren Kostenmittelwerte resultieren aus der Mengenausweitung bei den eher gesunden Versicherten, bei den intensiv zu behandelnden Patienten kommt es in der Regel nicht zu einer Mengenausweitung. Die größeren Varianzen entstehen vor allem durch die Verschiebung des Mittelwertes weg von den höheren Kosten. Die Anzahl der Ärzte über der „Toleranzschwelle" nimmt bei geringeren mittleren Kosten zu.

Beispiel: Die Tätigkeiten der verschiedenen Facharztgruppen variieren nicht nur sehr stark, es sind die gruppenspezifischen Behandlungskosten, die stark unterschiedlich streuen. Die präziseren Mittelwerte sind als Vorteile bei der Standardisierung anzusehen, die großen sehr sensibel auf wenige Fälle reagierenden Streuungen lassen Auswirkungen schlecht vorhersagen.

Einen Tautologie-Bias vermuten Sie aber m. M. n. zu Unrecht. Es ist sehr wohl so, dass bestimmte hohe Kosten spätere hohe Kosten erklären, dass Kostenentwicklungen sich in Vorzeiten andeuten.

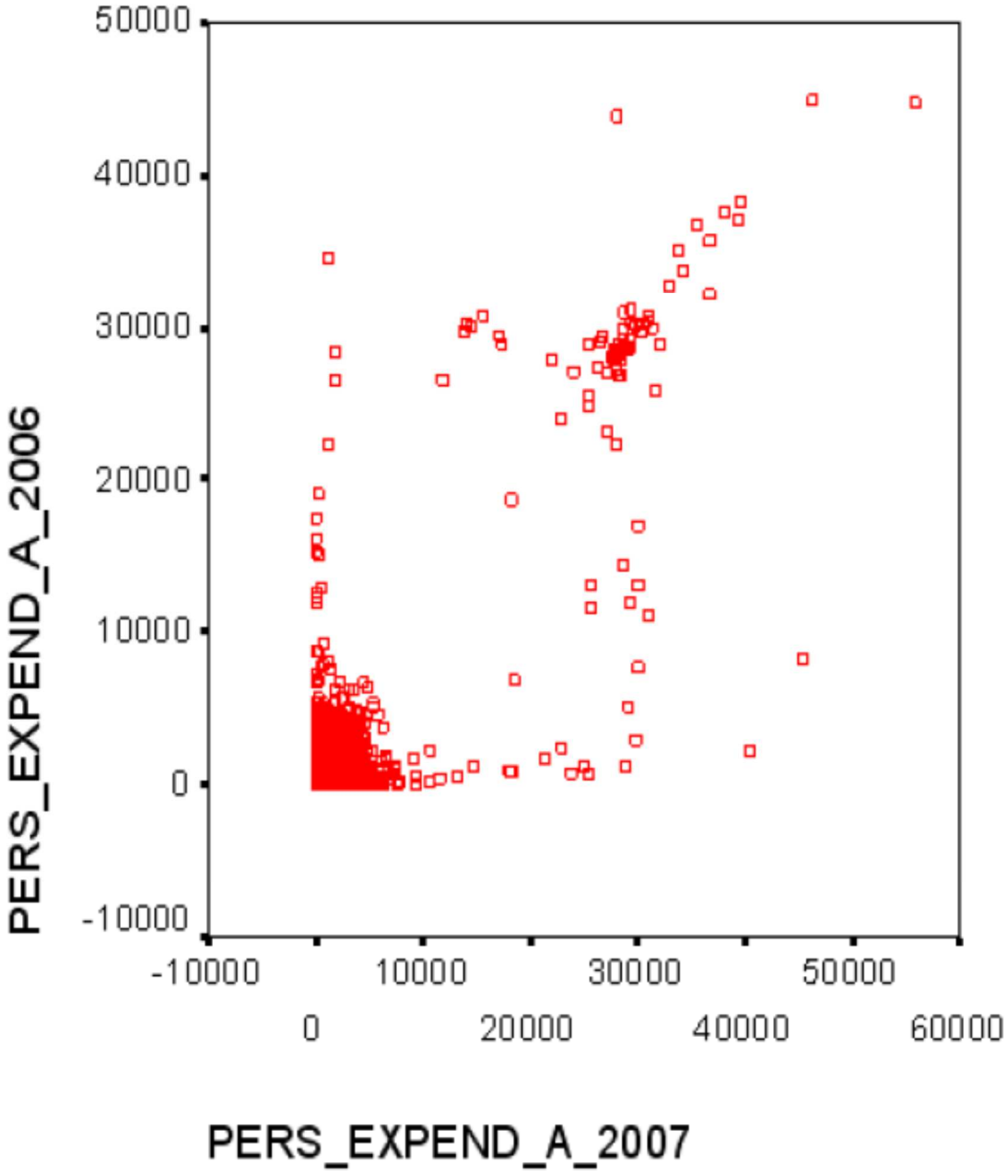
Ich führe zwei Beispiele aus der Untersuchung deutscher Gesundheitskosten an. Zunächst betrachte ich mittlere Jahreskosten gekoppelt an die Anzahl der Arztkontakt-Quartale aufeinanderfolgender Jahre.

Abbildung 4. Hohe Arztkosten kündigen sich im Vorjahr an. Die Achsen bezeichnen die Anzahl der Arztkontakt-Quartale. Die unterste Säulenzeile gibt die mittleren Jahresgesamtkosten im Vorjahr an. Die von den Ellipsen noch erreichten Säulen sind die Jahreskosten der „Hochnutzer" (4 Quartale) im Folgejahr.



Im Weiteren betrachte ich nur die ambulanten Jahreskosten in aufeinanderfolgenden Jahren.

Abbildung 5. Hohe Kosten sind im ambulanten Sektor relativ selten. Das Aufeinanderfolgen von hohen Kosten ist nicht selten. Auffällig ist das „Nest" der Kosten im Bereich von 30.000€ pro Jahr (Dialyse, Strahlentherapie, künstliche Befruchtung, …).


50000
40000
30000
20000
10000
0
-10000
PERS_EXPEND_A_2006
-10000
10000
30000
50000
0
20000
40000
60000
PERS_EXPEND_A_2007

Frage 5.

Ist die ANOVA Korrektur mittels aggregierten Alters-/ und Geschlechtsklassen statistisch korrekt? (Anm.: hier geht es um das Ausmass des Aggregation Bias, wonach rund 40% der Kosten durch Alter/Geschlechts Klassen erklärt werden, obwohl nur 4% anhand von Individualdaten zu erwarten wären, und was das für Effekte auf das Endergebnis haben wird, das heisst, wie weit die ANOVA Korrektur die fixed effect variable ZSR korrekt oder eben verzerrt korrigiert).

Wenn alle Patientenkosten eines Arztes berücksichtigt werden sollen, dann ist die ANOVA Korrektur keinesfalls statistisch korrekt. Es können aber für bestimmte Therapien, denen ein ähnlicher Diagnostikaufwand gegenübersteht, u. U. Kostenbereiche gefunden werden, wo sich Kosten um einen spezifischen Wert alters- und geschlechtsabhängig ausgleichen. Hier wäre die Anwendung der ANOVA Korrektur nützlich. Nur hier könnte gutgläubig eine Beziehung zum Aggregation Bias hergestellt werden.

Frage 6.

Sollte eine Korrektur für die Arztkosten auf Individualdaten der Patienten erfolgen, um einen Aggregation Bias zu vermeiden? (Anm.: dies ist eine wichtige grundsätzliche Frage, da santésuisse nur über aggregierte Alters/Geschlechts-Klassen verfügt, im Gegensatz zu den Versicherern, welche sich weigern, die Einzeldaten aus „Datenschutzgründen" [Zitat santésuisse] zu liefern).

Eine Berücksichtigung von Individualdaten der Patienten trägt zum besseren Verständnis bei den Arztkosten bei. Da der RSS-Index und die ANOVA Korrektur jedoch nicht auf die Entstehung von Kosten durch Behandlung von Patienten ausgerichtet ist, kann diese Berücksichtigung von Individualdaten unterbleiben. Ein Entgeltsystem, das nur zur Kenntnis nimmt, dass ein „Mittelwert" von Kosten steigt (was nicht einmal etwas mit einer Überarztung zu tun hat), bestenfalls noch das Steigen der Kosten bei einem „mittleren

Patienten" konstatiert, ist durch die Berücksichtigung von Individualdaten nicht zu verbessern.

Der mittlere Patient wird älter, ältere Patienten benötigen in der Regel höhere Behandlungsaufwendungen, für mehr ältere Patienten sind auch mehr Ärzte vonnöten. Ärzte behandeln Patientenkollektive konkreter Patienten, die sich weder im Mittel im Alter, in der Geschlechtszusammensetzung, in der Morbidität, … unmittelbar bei ihm ausgleichen, noch ist der Arzt mit einem mittleren Behandlungsbedarfsgebiet versehen, noch ist die Facharztmischung, noch der Umfang der Ärzteschaft bedarfsvergleichbar.

Beispiel: Die Fallpauschalen bezogene Krankenhausstatistik in Deutschland (DRG-Statistik – Diagnosen und Prozeduren der vollstationären Patienten und Patientinnen in Krankenhäusern) als Vollerhebung mit „weitgehend … vollständiger … Erfassung" (Selbsteinschätzung des Statistischen Bundesamtes) weicht erheblich von Statistiken ab, in denen Merkmale der Krankenhäuser und der Krankenhauspatienten in Bezug auf z. B. Diagnosen kumuliert erfasst werden.

Die letztendliche Ursache hierfür ist, dass nach ICD-10-Diagnosen keine Abrechnung von Leistungen durch die Leistungserbringer erfolgt, auch überhaupt nicht erfolgen kann („…noch etwas ausführen" MR). Die Behandlung (und dann auch Abrechnung) orientiert sich zunächst an Krankheitsgruppen MDG/DRG (erste drei Stellen) und die gruppeneigenen CCL, im zweiten Schritt kommen die (eine weitere) Diagnostik betreffenden OPS, therapeutische OPS und die PCCL hinzu. Es gibt nur wenige Kodierzusammenhänge (zu ICD-10), es gibt keinen Kodierzusammenhang, der nicht durch die vom Controlling benutzte Software bedient wird. Schon bei der Aufnahme eines Patienten erstellt das nach DRG abrechnende Krankenhaus eine eigene „Einlieferungsdiagnose" (die assoziiert ist mit der für möglich gehaltenen DRG-Abrechnung nach MDG/DRG [erste drei Stellen]), diese ist in der Regel nicht die Einlieferungsdiagnose eines einweisenden Arztes.

Beispiel: Wird ein aus dem Bett gestürzter älterer Patient mit einer (möglichen) Kopfverletzung in ein Krankenhaus zu einer Tageszeit eingeliefert, in der kein CT verfügbar ist, dann wird entweder an ein anderes Krankenhaus verwiesen oder auf die Abrechnung nach B80* (Kopfverletzung) fokussiert (B80Z, Risikofaktor 0,305, mittlere Verweildauer 2,3 Tage, Erlös weniger als 1.000 €).

Die gleiche Einweisung in ein Krankenhaus mit 24-Stunden-CT-Verfügbarkeit und Stroke Unit (OPS 8-981 Neurologische Komplexbehandlung) orientiert auf B39* (Schlaganfall) (B39A, Risikofaktor 5,593, mittlere Verweildauer 16,0 Tage, Erlös fast 20.000 €, im Bundesland Rheinland-Pfalz fast 25.000 €).

Auch bei einer nicht risikoaversen Arztvergütung sind Individualdaten der Patienten nicht leicht und in beliebiger Tiefe zur Korrektur hinzuziehbar.

Beispiel: Als eine Stichprobe der DRG-Statistik kann die Datenbank beim InEK (Institut für das Entgeltsystem im Krankenhaus) gelten. Freiwillige Kalkulationskrankenhäuser ergänzen fallbezogene Kostendaten zu den Grunddaten der DRG-Statistik. Die Struktur der in der „Stichprobe" abgebildeten Kostenträger, Leistungserbringer, Regionen, Patienten, Behandlungen, Kodierungen, …) weicht ebenfalls erheblich von DRG-Abrechnungen in Deutschland nach der DRG-Statistik ab. Diese nicht zufällige Stichprobe des InEK wird zwar zur Weiterentwicklung des DRG-Systems in Deutschland benutzt, für einen nicht so durchdacht geregelten Bereich wie die ambulante Versorgung sind aus einer solchen Datenbank abgeleitete Kennwerte angreifbar.

Arztkosten und Individualdaten sollten aus der gleichen Datenquelle gezogen werden, am geeignetsten ist hier die Datenquelle des Kostenträgers (die Kostenträger stellen der santésuisse nur nach Ärzten verdichtete Daten bereit, ein Patient kann in mehreren Abrechnungen vorkommen).

Durch ein Mehr an Individualdaten, die von den Patienten „preisgegeben" werden, steigt die Regressgefahr und damit die

Ausuferung von Behandlungsleistungen zur Verringerung dieser Gefahr. Der Arzt unterbricht den diagnostischen Prozess in der Regel mit seiner Entscheidung, die Therapie zu beginnen. „Dabei besteht die Entscheidung nicht in der Feststellung, ob die Krankheit nun tatsächlich vorliegt oder nicht. In der Praxis ist vielmehr von Bedeutung, ob das Maß der Wahrscheinlichkeiten für therapeutische Entscheidungen ausreichend ist.“[76]

[76] Köbberling, J.: Diagnoseirrtum, Diagnosefehler, Befunderhebungsfehler, Bewertungen und Vermeidungsstrategien, Verlag Versicherungswirtschaft, Karlsruhe 2013, S. 5.

Frage 7.

Die ANOVA Korrektur anhand der ZSR Variable und den Alters- und Geschlechtsklassen erklärt rund 80% der Arztkosten. Lässt sich diese Beobachtung mathematisch und gemäss Literatur begründen? (Anm.: dies ist eine Anschlussfrage an Frage 5, mit Fokus auf die Gesamtperformance der aktuellen ANOVA-Indexes, wie weit dies überhaupt in der Literatur beschrieben wurde und ob die ETH Gutachter hier mit der Erklärung, der ANOVA Index könne so gerechnet werden, aus gesundheitsökonomischer Sicht nicht nachvollziehbare Empfehlungen abgegeben haben, mithin, ob das Seminar für Statistik der ETH Zürich in seinem Gutachten 2005 überhaupt die Voraussetzungen mitbringt, die damalige Fragestellung adäquat zu beantworten, wenn man das Ergebnis betrachtet [hohes R2 des ANOVA-Indexes]).

Es gibt keine Literatur, es wird eine solche Literatur auch nie geben, die das Vorgehen des Seminars für Statistik der ETH Zürich im Zusammenhang mit Kosten bei der Behandlung empfiehlt[77].

Dreist sind die in dem Gutachten von 2005 auf Seite 12 abgegebenen Bewertungen. Diese Bewertungen sind alle falsch bzw. unnütz.

Kriterium Methode: Wenn nur auf einen Mittelwert aller Kosten in den Gruppen abgestellt wird, dann ist die Berücksichtigung von Alter, Geschlecht und Kanton sogar ein massiver Rückschritt. Die hohen Kosten verteilen sich nicht gleichmäßig auf die Arztpraxen, nicht entsprechend Alter und Geschlecht der Patienten. Würden die höheren Kosten a priori eliminiert, dann wären die neuen Restmittelwerte u. U. eine Verbesserung.

Kriterium Ausreisserproblematik: Hier zeigt sich das absolute Unverständnis beim Seminar für Statistik. Erstens handelt es sich bei

---

[77] „Kosten bei der Behandlung" steht hier für den „Leistungsumfang der [Krankenversicherungen], der Prävention, Diagnostik, Symptomatik, Kuration (Therapie von Krankheiten), Rehabilitation und Palliative Therapien" betrifft, nach Warmuth, W.: Sie wollen doch nur spielen: Gesundheitsökonomie in der GKV, Versicherungswirtschaft, 68(2013)13. S. 20-24.

einer Heavy-tail Verteilung[78] (eine solche haben die Gesundheitskosten) nicht um Ausreißer, sondern um einen Bereich höherer Kosten, der zweitens aber nur wenige Patienten betrifft, und so nicht bei jedem Arzt gemittelt oder gewichtet zu Buche schlägt. Bei als normalverteilt angenommenen Daten könnte das Seminar für Statistik so schließen, wie sie geschlossen haben. Es wird an keiner Stelle darauf hingewiesen, dass trotz des Nichterfülltseins von Voraussetzungen abgewogen wurde, dennoch die Methode für nützlich zu erachten.

Der Hinweis des Seminars für Statistik auf eine Logtransformation der Daten ist gut, die andere Kostenverteilungsart wird zur Kenntnis genommen, und zwar nicht nur die Schiefe sondern auch die Heavy-tail-Eigenschaft. Hiermit wird eine andere Betrachtungsrichtung eröffnet. Die ANOVA-Korrektur wird damit aber final beerdigt, weil höhere Kosten durch die Transformation heruntergewichtet, die Varianzkomponenten damit auch geschwächt und das Vorhandensein von höheren Behandlungskosten bei allen Ärzten Bedingung wäre.

Kriterium Vergleichsgruppengrösse: „Grössere Vergleichsgruppen bilden stabilere Vergleichswerte" (Zitat aus dem Gutachten), dieses gilt nur unter der Annahme, dass es sich um Patienten bei den Ärzten handelt, deren homogene Kosten sich jeweils im Mittel ausgleichen. Wenn die höheren Kosten eliminiert würden, dann geschähe dieser Ausgleich um einen kleineren Mittelwert auch nur hypothetisch, die Unterscheidung nach Arztgruppen wäre hier jedoch nützlich. Nach dem propagierten Inhalt des Gutachtens 2005 (Vergleichbarkeit

---

[78] Bei allen Verteilungen von Behandlungskosten in Deutschland handelt es sich um Heavy-tail Verteilungen, hohe Kosten sind hier „normal". Ein gutes Modell stellt die Pareto-Verteilung dar. Real sind mit 80 % der Behandlungen rund 20 % der Kosten gedeckt, im 95 % Konfidenzintervall liegen 65 % der Kosten, auf wenige nicht korrekt als Hochkostenfälle [aus gesundheitsökonomischer Sicht sind sie es] bezeichnete Fälle entfallen noch 35 % der Kosten. Siehe z. B. Brammer, K., Schmidt, A., Warmuth, W., „Morbiditätsorientierte Simulation des Ertrags bei der gesetzlichen Krankenversicherung auf der Basis einer kassenspezifischen zufälligen und marktabhängigen Zu- und Abgangs-Modellierung", 9. Data Mining Cup-Anwendertage, 04./05.06.2008; auch Klemd, F. „Modellierung des Heavy-tail Kostenbereiches für Krankenversicherungen – unter besonderer Berücksichtigung der Ausbreitung, der Abgrenzung zu ‚Ausreißern' und der Nähe zu Extremverteilungen –", BA-Arbeit, 2012.

sichern) müssten die Vergleichsgruppen so groß gebildet werden, dass alle Patienten, Arztgruppen, Kantone mit ihren niedrigeren und höheren Kosten repräsentiert wären, an ein gleichgroßes Gebilde könnte die Kostenerstattung als Block erfolgen. Ich vermute, dass ein solches Konstrukt die Größe der Schweiz hätte (nur die Schweiz ist sich vermutlich ähnlich, Selbstähnlichkeit), damit würde es zu einem Realkostenausgleich kommen.

Kriterium Berücksichtigung von Besonderheiten: Die benannte Berücksichtigung von Alter und Geschlecht betrifft einen „mittleren Patienten", hier verweise ich auf die Antwort zur Frage 6, wo das schon in seinen risikoaversen Auswirkungen dargestellt wurde.

Kriterium Spezifität: Gesundheitsökonomisch wird auf wenige „falsch-positive" Fälle gesetzt, wodurch hohe Gesamtkosten vermieden werden. Die Spezifität („richtig-negative" Fälle zu erkennen) wird allein hoch gewertet. Die Sensitivität („richtig positiv" Fälle zu erkennen) ist zumindest für Leid, das Patienten unnötig zugefügt wird, verantwortlich. Die Ethik bleibt hier aus gesundheitsökonomischer Sicht auf der Strecke, auch Ärzte klären ihre Patienten hier nicht immer genügend auf. Im Falle des ELISA-Tests auf HIV kommt es regelmäßig aus diesem Grunde zum Suizid[79].

Kriterium Weiterentwicklung: Die Methode steckt in ihrer praktizierten Anwendung in einer sich selbst einengenden Sackgasse.

Kriterium Abdeckungsgrad: Aus den gleichen Daten wird mit vorgeblichen Innovationen ein schlechteres Ergebnis erzielt, es sei denn es wird auf einen Realkostenausgleich abgestellt, wie er sich aus dem „Kriterium Vergleichsgruppengrösse" vermutlich ergibt.

---

[79] Gerd Gigerenzer, J. A. Muir Gray, Bessere Ärzte, bessere Patienten, bessere Medizin, Medizinisch Wissenschaftliche Verlagsgesellschaft

Frage 8.

Wie beurteilen Sie die Herangehensweise im ETH Gutachten Stahel/Roth 2005 und die Erkenntnisse aus dem ETH Gutachten Nr. 2 für den Kanton Tessin? (Anm.: hier geht es darum, in welchem Ausmass Sie feststellen können, dass die Gutachter sich nun aufgrund des zweiten Gutachtens widersprechen).

Insbesondere nach der Selbsteinschätzung im Gutachten 2005 vom Kriterium Methode und vom Kriterium Ausreisserproblematik her hätte das Ergebnis im Kanton Tessin überhaupt nicht auftreten können. Es ist einfach das eingetreten, was zu erwarten war, dass einfach etwas anderes herauskommt.

Ich vermute, dass durch die Herauslösung der höheren Kosten eine größere Ähnlichkeit der Kantone verbleiben würde.

Der Ausflug der Autoren in dem neueren Gutachten zu „mässig bis stark korellierten" Indexwerten für die beiden Methoden zeigt, dass sie sich selbst von ihrem Gutachten 2005 distanzieren.

Frage 9.

Wie beurteilen Sie ganz prinzipiell ein Modell, das primär die Arztkosten als erklärende Variable (fixed effect model) mit einbezieht? (Anm.: hier geht es darum, wieweit ein solches Modell für die Erklärung der Arztkosten in der Gesundheitsökonomie überhaupt diskutiert wurde, und falls ja, ob es validiert wurde).

Zur Prognose von Gesamtaufwendungen ist die Einbeziehung von Arztkosten als erklärende Variable hilfreich. Das habe ich in einer Reihe von Projekten demonstriert (z. B. Vorhersage von Krankenversicherungskosten eines Distrikts in Indien, z. B. Vorhersage von Programmkosten in speziellen Versorgungsinitiativen einer deutschen Krankenkasse). Für einen einzelnen Arzt kann mit hoher Wahrscheinlichkeit (>80 %) geschlossen werden, dass seine „Hochkostenpatienten" auch im Folgejahr diese hohen Kosten verursachen (die Abbildung 4 gehört zu diesen Untersuchungen). Für ein Indexverfahren auf Vergleichsgruppenbasis wäre hierfür die

Repräsentativität dieser Patienten mit höheren Kosten vonnöten, diese Vergleichsgruppe hätte vermutlich die Größe der Schweiz.

Frage 10.

Ist die Weiterentwicklung der Methode anhand weiterer erklärender Variablen mit der ANOVA Methode möglich (Literatur dazu?) oder wäre ein Regressionsmodell hilfreicher, falls ja, warum? (Anm.: hier geht es darum, ob künftige Variablen eben auch nicht kategorischer Art sein werden und damit die ANOVA Methode nicht mehr adäquat eingesetzt werden kann; zweitens um die Frage, ob überhaupt ein ANOVA Modell als primäre statistische Methode nicht primär mit einem REGRESSION-Modell hätte konzipiert werden müssen, mithin die Frage, warum die Erfinder des ANOVA-Indexes auf die ANOVA Methode gesetzt haben und was für Vorteile sie damit allenfalls davon erwartet haben könnten).

Wenn die höheren „normalen" Kosten den Mittelwert bei einzelnen Ärzten unvergleichbar verzerren, dann sind Weiterentwicklungen im Rahmen einer Varianzanalyse nicht möglich, dann ist eher ein Rückbau oder Nichteinsatz angebracht. Es ist nicht erkennbar, dass die ANOVA-Methode bei der Arztvergütung nützlich ist.

Ein Regressions-Modell zur Arztvergütung ist überhaupt ungeeignet.

Während für die Anwendung des Regressions-Models die Normalverteiltheit (im mathematischen strengen Sinne) Bedingung ist, kann die ANOVA-Methode auch teilweise auf Schätzungen zurückgreifen, die nur eine Nähe zu einer symmetrischen Verteilung aufweisen. In diesem Sinne ist der Versuch über die ANOVA-Methode vermutlich gestartet worden. Hier haben die „Erfinder" des ANOVA-Indexes vermutlich ihre Hoffnungen gesetzt. Zur Schätzung der Größe, des Gewichtes und der Kosten risikolos geborener Einlinge, oder zur Schätzung der Schwangerschaftsdauer bei Ausschluss der Frühgeburten und auch der Kaiserschnittgeburten wäre die ANOVA-Methode ein guter Griff.

Der Nützlichkeitsnachweis der ANOVA-Methode bei der Arztvergütung fehlt, er kann auch nicht erbracht werden, wenn auf den Mittelwert aller Kosten abgestellt wird.

Frage 11.

Ist eine verlässliche Beurteilung der Überarztung ohne Morbiditätsvariable möglich (RSS-Index) und ist ein Surrogatmarker Alter/Geschlecht für die Morbidität ausreichend? (Anm.: hier geht es um die Frage, ob der ANOVA-Index vom Konzept her und aufgrund bekannter Tatsachen nie hätte als statistische Verbesserung des RSS-Index bezeichnet werden dürfen, wie die Gutachter des Seminars für Statistik dies ja getan hatten).

Überarztung ist aus Mittelwerten wegen der Spezifik von Gesundheitskosten nicht ablesbar. Die Tatsachenvermutungsgrenze kann an einer Obergrenze von Kosten, relativ von einem Mittelwert abgeleitet, nicht festgemacht werden. Eine Überarztung wird durch gesundheitsökonomische Handlungen gefördert, das Vorgehen nach dem RSS-Index verschlechtert die Grundversorgung nicht hochmobiler Versichertengruppen, fördert selektive Versorgungsangebote für junge, gesundheitsbewusste Versicherte, verschiebt Behandlungen regional und zeitlich zu Lasten von Versicherten. Der ANOVA-Index, unter anderen Prämissen erdacht, war nie eine Verbesserung, sicher eine Veränderung, teilweise eine Verschlechterung.

Beispiel: Geringere Mittelwerte entstehen bei einer Mengenausweitung auf weniger morbide Versicherte, z. B. sind in zwei Regionen in Deutschland (Westfalen, Nordbaden) wegen eines Überangebotes an Leistungserbringern mehr Patienten anzutreffen, die dafür im Mittel nicht so hohe Kosten verursachen. Die Gesamtkosten fallen auf alle Versicherten bezogen höher aus, der Durchschnitt bei den Patienten ist geringer als in anderen Regionen.

Beispiel: Der Vorlaufzeit-Bias (z. B. bei Prävention) führt zu einer Vorverlegung des Zeitpunktes der Diagnose, damit zu einer Fallzahlerhöhung.

Der Überdiagnose-Bias erhöht durch das Aufgreifen nichtprogressiver Erkrankungen die Fallzahl.

Der Intervall-Bias überschätzt langandauernde Phänomene bei Zeitpunktbetrachtungen.

Alle diese Verzerrungen sind Ausdruck von gesundheitsökonomisch verursachter Überarztung.

Beispiel: Der Selektion-Bias drückt aus, dass bekannte und unbekannte Störgrößen/Einflussgrößen nicht gleichmäßig in z. B. der Interventions- und z. B. der Kontrollgruppe verteilt sind (in Bezug auf die höheren Kosten hatte ich das schon benannt). Hier kann ganz sicher nicht eine Überarztung belegt werden, auch wenn sie vorhanden sein sollte.

Beispiel: Die von mir in vielen Projekten bei Krankenkassen evaluierten Disease Management Programme in Deutschland führen zu einer Überarztung und Übermedikalisierung eher „gesunder" Versicherter. Ihr durchschnittliches Kostenniveau ist niedriger als bei Versicherten mit höherer Morbiditätslast und höherer notwendiger Behandlungsintensität, soviel auch hier zum Mittelwertvergleich. Die in Deutschland seltene Integrierte Versorgung in den gleichen Krankheitsgruppen ist erfolgreich.

Beispiel: Die Einführung ambulanter Operationen führte zu einer beträchtlichen Mengenausweitung bei dieser Art Operationen. Im Durchschnitt sind diese Operationen günstiger als vergleichbare stationäre Eingriffe, da auch weniger komplizierte und unnötige Fälle operiert werden. Summarisch sind die Kosten merklich höher. Eine Sonderidee des Spitzenverbandes Bund der Krankenkassen führt dann noch zu weiteren Kosten: „Blutige Entlassungen" aus ambulanten Operationen werden außerbudgetär (also nicht im Rahmen des Budgets der Kassenärztlichen Vereinigungen) direkt ungedeckelt bei den Krankenkassen abgerechnet.

Beispiel: Ein geringerer Einzelkostensatz erfordert aus Ertragsgesichtspunkten eine Erhöhung der Patientenanzahl, dieses führt zu einem Anwachsen der mittleren Wartezeit der Patienten[80].

Frage 12.

Basierend auf dem Überarztungsmodell Beck/Kunze ist trotz Einbezug von Morbiditätsvariablen entweder eine Flexibilisierung der Beweisgrenze notwendig oder es ist sogar je nach Facharztgruppe nicht möglich, verlässlich Kostenüberschreitungen zu erfassen. Was hat dies für Implikationen für den RSSIndex und den ANOVA-Index, welche beide ohne Morbiditätsvariable(n) auskommen? (Anm.: hier geht es um die Frage, dass die Beobachtungen Beck /Kunze eigentlich dazu führen müssten, die Validität des RSS-Indexes erheblich zu bezweifeln, dass also mit dieser Publikation durchaus Konsequenzen für ein Moratorium des RSS-Indexes und des ANOVA-Indexes hätten plausibel hergeleitet werden können).

In dem Artikel von S. von Rotz/U.Kunze/K.Beck wird ein „Instrument zur Beurteilung der Wirtschaftlichkeit von Grundversorgern" vorgestellt. Die zur Anwendung des Modells notwendigen Voraussetzungen werden nicht geprüft, sie sind auch nicht erfüllt. Dennoch handelt es sich um ein bedingt nützliches Herangehen, da nur Variablen einbezogen werden, die in einem Zusammenhang zu den Arztkosten stehen können/könnten (nicht etwa wie in der MorbiRSA-Regression in Deutschland, wo Kodierungen betrachtet werden, die überhaupt keinen Kostenbezug bei Ärzten haben). Das in dem Artikel verwandte R_2 ist natürlich nur eine Krücke, die nur unter normalverteilten Kosten betrachtet werden könnte. Es ist jedoch nach diesem Artikel überhaupt nicht zu erwarten, dass einzelne Ärzte mit

---

80 Dietrich, C. E. M., Walleitner, P.: Warteschlangen-Theorie und Gesundheitswesen, Einführende Betrachtungen für den Praktiker, Medizinische Informatik und Statistik 34, Springer Verlag Berlin Heidelberg New York 1982. Patienten, die im Mittel alle 10 Minuten eintreffen, von einem Arzt im Mittel in 9 Minuten behandelt werden, warten im Mittel 81 Minuten in der Praxis, siehe dort S. 43.

einem RSS-Index in wirtschaftliche und unwirtschaftliche Ärzte eingeteilt werden können. Vielleicht gelänge das für große Gruppen ähnlicher Ärzte, die dann als Gruppe sicherlich aber geschlossen merklich unter der 130 %-Schranke liegen würde. Der RSS-Index bleibt ein nicht zu rechtfertigender Rahmen. Eine vernünftige Grenze ist auch nicht verhandelbar. Da auch bei einer Berücksichtigung einer Facharzt-, Kanton-, Versicherten-, Leistungserbringerdichte-, … Spezifik die höheren Kosten über diese Gruppen nicht gleichmäßig gewichtet verteilt sind, ist auch eine anteilige Bemessung bis zu einer spezifischen Schranke bei großen Gruppen schwer möglich. Bei einzelnen Ärzten scheidet sie aus.

Frage 13.

Basierend auf der Tatsache, dass selbst Morbi-korrigierte Methoden (Beck/Kunze) nicht für alle Fachrichtungen anwendbar sind, müsste man da nicht davon ausgehen können, dass der RSS- und ANOVA-Index zuerst fachspezifisch auf die Validität geprüft werden müssen, bevor man überhaupt daran denken kann, dies als Beweismittel für Überarztung den Gerichten zu präsentieren?

Die fachspezifische Prüfung sollte bei der Bestimmung der Verteilung der individuellen Fallkosten beginnen. Zu berücksichtigen wären das Alter, das Geschlecht und die Art (Beschäftigter, Rentner, Betreuungsbedürftiger, …) der Versicherten zum Leistungsbeginnzeitpunkt, die aufgesuchte Erstarztgruppe, die Verfügbarkeit (zeitlich und räumlich) von bedingt für den Fall erforderlichen weiteren Leistungserbringern (u. U. im Netzwerk) unter Berücksichtigung weiterer Merkmale (z. B. muslimische Frauen gehen eher ungern zu einem Frauenarzt), die versorgungsrelevante Leistungserbringerdichte und die patientenbezogene Verfügbarkeit von Leistungserbringern und die Art der Leistung (Vorsorge, Diagnostik, Therapie, …).

Frage 14.

Basierend auf dem deutschen Modell, erfolgt anhand des Morbi-RSA (Erklärungsgehalt im R2?) eine patientenseitige Kostenzuteilung, ferner eine Auditierung von Kostenausreissern, welche im Morbi-RSA Modell nicht erklärt werden können. Ist es vor dem Hintergrund dieser Vorgehensweise überhaupt noch sachgerecht, mit einem RSS-Index oder ANOVA-Index zu arbeiten, oder sollte diese Methode vollständig verlassen werden? (Anm.: hier geht es vor allem um die Frage, ob überhaupt noch eine Vergleichsgruppe für ein adäquates „physician profiling" notwendig ist, mithin, ob auf die ZSR Variable in den statistischen Modellen komplett verzichtet werden könnte.

Aus methodischer und datenstatistischer Sicht und vermutlich erst recht aus fachlicher Sicht ist es nicht sachgerecht nach einem RSS-Index oder ANOVA-Index zu arbeiten, das dabei benannte „physician profiling" könnte jedoch im Rahmen eines „gemanagten" Vollkostenausgleichs der Krankenversicherungen mit den Ärzten Regelbestandteil sein.

In die Vorgehensweise im Rahmen des Morbi-RSA in Deutschland können keine Hoffnungen gesetzt werden. Die für den Morbi-Teil des RSA verwendete Modellierung ist sinnwidrig[81], über 60 Milliarden Euro werden Jahr für Jahr auf der Individualebene betrachtet falsch verteilt. Die Regression ist nicht anwendbar, in Bezug auf Alter, Geschlecht und Erwerbsminderung in Teilen jedoch nützlich. Davon betroffen sind aber nur die Krankenkassen. Bei den Ärzten erfolgt eine Vollkostenerstattung, die kassenärztlichen Vereinigungen verhandeln, mit Daten besser ausgestattet und nicht sozialwissenschaftlich methodisch beschränkt, unter Berücksichtigung von Alter und Geschlecht der Patienten, von weiteren soziodemographischen Merkmalen der Patienten, von regionalen

[81] Warmuth, W.: Sie wollen doch nur spielen: Gesundheitsökonomie in der GKV, Versicherungswirtschaft, 68(2013)13. S. 20-24.

Arztgruppen- und Versorgungsunterschieden, mit „Partnern" eine vollkostendeckende Patientenpauschale.

Mit der Nichtanwendbarkeit der Regression bei Gesundheitskosten hat auch das R_2 keinen Erklärungsgehalt.

Frage 15.

Schätzung erwarteter „predictive ratios" in Abhängigkeit der Praxisgrösse: A) 500 Patienten, B) 1.000 Patienten, C) 15.000 für folgende Modelle: 1) RSS Index, ANOVA-Index, ANOVA-Index + PCG; 2) Morbi-Index, Morbi-Index + Ausschluss teure Patienten. Hier geht es darum abzuschätzen, ob ein Systemwechsel von Modell 1 zu Modell 2 für die meisten Praxen (Grösse um 1000 Pat./Jahr) erwartungsgemäss mit einer signifikanten Verbesserung der „predictive ratios" verbunden sein wird.

Solche Schätzungen sind nur möglich, wenn die normalen höheren Kosten ausgespart bleiben. Hierbei sind die Arztgruppen, die Versichertenstruktur, die Kantone mit einer Unterteilung in Gebiete mit einer soliden Grundversorgung, mit einer Überversorgung und mit einer Unterversorgung zu berücksichtigen. Letztendlich könnten solche Schätzungen aber an realen Gesamtkosten validiert werden. Auf den einzelnen Arzt kann das aber sicherlich nicht heruntergebrochen werden.

Frage 16.

Ist die Homogenisierung von Vergleichsgruppen aus gesundheitsökonomischer Sicht zwecks Verbesserung des RSS- und ANOVA-Index je genügend trennscharf möglich, ohne dass erhebliche Confounders verbleiben? (Anm.: hier geht es um die Frage, ob auch gut erarbeitete Vergleichsgruppen innerhalb von Fachärzten nicht doch noch einen begründbar grossen (> 60%) Kostenspielraum belassen könnten, etwa in dem Sinne, dass beispielsweise Kardiologen mit Zuweisungspraxis sich sub-spezialisieren auf Herzinsuffizienz und andere auf Hypertonie, mit entsprechend differierender Kostenstruktur aufgrund der unterschiedlichen Morbiditäten / Co-Morbiditäten).

Geht man von Mittelwerten aus, die alle Patienten-Kosten eines Arztes berücksichtigen, dann ist die Willkür des RSS- und ANOVA-Index nicht beseitigbar.

Wie bei der Frage 15 halte ich eine Homogenisierung von Vergleichsgruppen (nach Arztgruppen, nach der Versichertenstruktur, nach den Kantonen mit einer Unterteilung in Gebiete mit einer soliden Grundversorgung, mit einer Überversorgung und mit einer Unterversorgung, für normale Kosten) für möglich. An diesen Vergleichsgruppen könnte ein Sub-Spezialisieren simuliert werden (ohne dass es erst zur Praxis wird), es könnte die Ausweitung genauso erkundet werden, wie die Bewirtschaftung des verbleibenden „Restes". Ebenso sind Confounder „fiktiv" ermittelbar, die unter Beachtung einer Bilanztreue belegen, dass der Kostenspielraum zwischen den Ärzten dennoch enorm ist, womit sich die Beweiskraft von RSS- und ANOVA-Index pulverisiert.

Schlussfolgerungen

Unbestritten ist: der von santésuisse verwendete RSS- und ANOVA-Index arbeitet mit aggregierten Daten und einem „Durchschnittskosten des Indexarztes" unter Einschluss des Arztes, führt zu nicht nachvollziehbaren, im Ergebnis betreffend diagnostischer Genauigkeit nicht überprüfbaren Effekten und letztlich zu aus ärztlicher Sicht nicht akzeptierbaren Statistik, weil teure Patienten nicht mehr behandelt werden können und Anreize gesetzt werden, unter Senkung der Durchschnittskosten pro Patient die Gesamtkosten im ambulanten Bereich des Gesundheitswesen ansteigen zu lassen.

Tatsache ist, dass die Durchschnittskosten-Methode von santésuisse deshalb nie hätte angewendet werden dürfen, weil sie den gesetzlichen Auftrag, nämlich die Überprüfung der WZW-Regeln, komplett im Sinne der Durchschnittskosten interpretiert, eine volkswirtschaftlich desaströse Optik, wie gezeigt wurde.

Im Ergebnis produzierte santésuisse in einem unkontrollierten und *durch externe Prüfer unkontrollierbarem* Rahmen eine ihrem Wunsch nach Kostenreduktion *willkürliches Benchmark-System*, welches schon auf der Datenbankebene Funktionsprobleme aufweist (Schwenkglenks Mittelwert Problematik) und alle Elemente enthält, das zu bewirken, was es vorgibt vermeiden zu wollen: unnötiger Kostenanstieg im Gesundheitswesen.

Dass es so weit kommen konnte liegt einerseits daran, dass in der Optik der Krankenversicherer nur die hohen Kosten pro Patient besonders gewichtet und beachtet werden, dass ferner einen öffentliche Kontrolle der santésuisse Statistiken vollständig fehlt und

dass sämtliche bezahlten Regressforderungen, insbesondere jene unsagbaren „Vergleiche" unter der Hand, aus diesen genannten Gründen rückforderbar sind.

Das blinde Vertrauen der Öffentlichkeit und insbesondere der beaufsichtigenden Behörden in die Statistiken der santésuisse ist aufgrund der hier liegenden Beobachtungen nicht mehr zu verantworten.

Da die Wirtschaftlichkeitsverfahren Implikationen für die Behandlungsqualität haben und in einer differenziert und korrekt angewendeten Ausprägung durchaus auch Aspekte der Versorgungsforschung berühren, sind die verantwortlichen Stellen, allen voran, die Gesundheitsdirektoren der Kantone direkt in der Pflicht, die festgestellten Missstände abzustellen. Ferner muss zunächst zwischen AerztInnen und Krankenversicherern ein offengelegter und überprüfbarer Konsens hergestellt werden, wie ein statistisch und wissenschaftlich korrekter RSA überhaupt theoretisch ausgestaltet werden müsste.

Eine substantielle Verbesserung des Prüfverfahrens kommt deshalb nicht umhin, auf allen Ebenen, auch unter umfassender Erfassung veranlasster Kosten im ambulanten und stationären Setting, aktiv zu werden. Dies ist mit den vergleichsweise primitiven Datenbanken der santésuisse nicht zu leisten.

Daraus folgt zwingend die Notwendigkeit für eine Nationale Datenbank aus Individualdaten der Versicherer und Trustcenters unter der Kontrolle durch eine Behörde entsprechend internationalen Standards. Zudem ist durchaus denkbar, dass künftige Modelle die Krankheit der Patienten ausreichend abbilden, so dass auf

Vergleichsgruppen vollständig werden kann und soll. Dies wäre eine das Prüfproblem erheblich vereinfachende Herangehensweise.

Ein weiterer Punkt betrifft die hohen Kosten einzelner Patientinnen und Patienten, welche zur heavy-tail Verteilung der Kosten führen. Die besonders teuren Krankheiten dieser Personen müssen aus einer separaten Versicherung finanziert werden, wenn Kostenberechnungen auf der Screening-Ebene weiterhin eine Option bleiben sollen. In diesem Fall sind die teuren Krankheiten aus der Statistik der WZW-Beobachtungen zu eliminieren.

Der heutige ANOVA-Index der santésuisse ist ferner als nicht rechtsfähig und als statistisch nicht sachgerecht zu beurteilen. Eine rechtswissenschaftliche Literatur, welche sich mit diesen nun aufgedeckten Umständen publizistisch auseinandersetzt, ist deswegen besonders dringend zu wünschen.

Die Literatur zeigt, dass das Problem der Überarztung nicht einfach zu lösen ist. Das Modell von Dudley 2003 scheint besonders interessant zu sein, aber auch hier beträgt das R2 < 40%, immerhin wären Audits vorgesehen, also eine Art PVK.

Die Hauptaufgabe wird also die Definition des Vertrauensintervalls für Überarztung bilden: Mittelwert-Betrachtungen vom Typ >130% können nicht a priori als Beweismittel akzeptiert werden, 2000 (oder 15% der ambulant tätigen AerztInnen) falsch positive pro Jahr sind denn auch eindeutig zu viele. Die Berechnung dieser Vertrauensintervalle pro Facharztgruppe könnte so ausgelegt werden, dass die Prozentzahl der auffälligen AerztInnen modelliert wird, z.B. < 2%, < 5% ecc. unter oder mit Verwendung von High UND Low Outlyers (High=Ueberversorgung, Low=Unterversorgung). Für die Beurteilung

des Mislabeling Potentials werden die grundlegenden Arbeiten der RAND hilfreich und wohl auch wegweisend sein. Doch auch hier gilt die Elimination von Kosten-Outlyers, weil bei diesen Überarztung die grosse Ausnahme bildet. Überarztung findet allenfalls im niedrigen Kostensegment statt. Zur Erfassung echter Überarztung, eben jener im niedrigen Kostensegment, ist die santésuisse Methode überhaupt vollkommen ungeeignet.

Forderungen des VEMS

1. Moratorium für die Durchschnittskosten-Methode per 01.06.2014.
2. Highcost Outlyers müssen aus der Statistik eliminiert werden.
3. Einsetzung einer paritätischen Steuerungsgruppe mit Weisungsbefugnis, welche die notwendigen Anpassungen rasch umsetzt.
4. Komplette Neuausrichtung der WZW Verfahren bis Ende 2014 im Sinne der Professionalisierung der Methode unter primärer Verwendung der Patientenmorbidität (siehe Gutachten VEMS ForBig / Gesundheitsforen 2014, Modul 1 und 2)
    a. Keine Mittelwert-Vergleiche
    b. Keine Anova-Statistik
    c. Keine Regressions-Statistik
5. Einstweiliges Prüfmoratorium
6. Prüfung weiterer gesetzlicher Anpassungen im KVG
7. Prüfung von Regressforderungen gegen santésuisse betreffend Vergleichszahlungen im Rahmen von Wirtschaftlichkeitsprüfungen.
8. Separate Krankenversicherung von Hochkosten-Patienten im ambulanten und stationären Bereich.
9. Auditierung ärztlicher Entscheide durch Fachpersonen.

Weitere Informationen

Konstantin Beck zum Problem der aggregierten Daten und weiteren Aussagen vom 02.09.2013

Erste Frage:

Basiert Tabelle 4 auf aggregierten Mittelwerten?

www.physicianprofiling.ch/HealthEconomicsPCGArticelBeckWorkingPaper2006.pdf

Die Antwort lautet: Nein, die R-Quadrat-Berechnung stützt sich auf individuelle Versichertendaten.

Zweite Frage:

Ist es korrekt, dass Sie das Regressionsmodell pro Facharztgruppe gebildet haben?

Ja, das ist korrekt

Dritte Frage:

Bei der Berechnung des r2 bzw. der errechneten Faktoren in der Regressionsformel: haben Sie da auf Patientenebene die Rohdaten oder aggregierte Daten verwendet (z.B. Mittelwerte von Altersgruppen)?

Gleiche Antwort wie Frage 1, wir berechnen praktisch ausschliesslich R2 mit Patientendaten und nicht anhand von Mittelwerten

Vierte Frage:

Haben Sie den Effekt von nicht aggregierten und aggregierten Daten auf das Erklärungsmodell (adj. R2) geprüft und was war das Resultat?

Da wir aus der Literatur wissen, dass die R2 bei aggregierten Daten strikt grösser sind, verwenden wir die aggregierte Version praktisch nie. Je nach Form der Aggregation bewegt sich das R2 anschliessend zwischen 80 und 99%. Da diese Zunahme aber völlig nichtssagend ist

(solange es um die Erklärung individueller Daten geht) haben wir ihr keine Beachtung geschenkt.

Fünfte Frage:

Wenn Sie die Patientengruppe eines Arztes dem Gesamtkollektiv der Kosten aller Patienten gegenüberstellen: benutzen Sie da eine Signifikanzberechnung für die erfasste Differenz und falls ja, welche?

Es handelt sich um eine grobe Signifikanzberechnung: Wir legen Interventionsschwellen fest und diese Schwellenwerte sind umso höher je kleiner das Versichertenkollektiv ist. D.h. ein Arzt mit kleinem Kollektiv wird erst nach Überschreitung einer hohen Schwelle als unwirtschaftlich eingestuft, bei einer Ärztin mit grossem Kollektiv genügt bereits die Überschreitung einer kleineren Schwelle. Die Logik hinter diesen Schwellen entspricht dem Prinzip der Signifikanzberechnungen. Die Anwendung ist allerdings einfacher.

Weiterführende Literatur:

Risikostrukturausgleich Für Versicherer und AerztInnen mit diversen Modellen und mit umfassenden RSA Beschreibungen und Empfehlungen

www.physicianprofiling.ch/RSAVerisk2010.pdf
www.physicianprofiling.ch/RSABAG2009.pdf

Grundlegende Arbeiten von C.Beck

www.physicianprofiling.ch/Beck2.pdf
www.physicianprofiling.ch/HealthEconomicsPCGArticelBeckWorkingPaper2006.pdf

Hohe Wichtigkeit

www.physicianprofiling.ch/ProfilingDudleyMedCare2003.pdf
www.physicianprofiling.ch/ProfilingStateAndImprove2013Overview.pdf
www.physicianprofiling.ch/ProfilingJAMA2009Overview.pdf
www.physicianprofiling.ch/Profiling2004.pdf
www.physicianprofiling.ch/RAMisclassificationRAND.pdf
www.physicianprofiling.ch/RAMisclassificationRANDNEJM2010.pdf
www.physicianprofiling.ch/RAMisclassificationRANDNEJM2010Appendix.pdf

Literatur zur ZSR Frage (pdf)

www.physicianprofiling.ch/AGWZWZSRAgeSex.pdf
www.physicianprofiling.ch/ ZSRDangelo122013.pdf
www.physicianprofiling.ch/ WZWAnovaGutachtenStahelRoth.pdf

Statistikliteratur:

www.physicianprofiling.ch/STATExcelEconometricsFixedEffectsModel.pdf
www.physicianprofiling.ch/RAMisclassificationRAND.pdf
www.physicianprofiling.ch/RAMisclassificationRANDNEJM2010.pdf
www.physicianprofiling.ch/RAMisclassificationRANDNEJM2010Appendix.pdf
www.physicianprofiling.ch/RAExpertPanel.pdf (risk adjustment and socio-economic status)

www.physicianprofiling.ch/RAAsthmaFixedVsRandom.pdf
www.physicianprofiling.ch/STATDatenanalyse.pdf
www.physicianprofiling.ch/STATMultivariateVerfahren.pdf
www.physicianprofiling.ch/STATMehrebenenanalyse.pdf

Literatur Internet:

Dummy Variablen und ANOVA: A regression model in which the dependent variable is quantitative in nature but all the explanatory variables are dummies (qualitative in nature) is called an Analysis of Variance (ANOVA) model.
http://en.wikipedia.org/wiki/Dummy_variable_(statistics) with the following reference:
Basic econometrics. McGraw Hill. p. 1002. ISBN 0-07-233542-4.
Paneldatenmodelle, Erklärung von fixed and random effect models:
http://de.wikipedia.org/wiki/Lineare_Paneldatenmodelle

Weiterführende Literatur zu Physician Profiling

www.physicianprofiling.ch/ProfilingMultilevel.pdf
www.physicianprofiling.ch/ProfilingPoland2011.pdf
www.physicianprofiling.ch/ProfilingAndMisclassification.pdf
www.physicianprofiling.ch/ProfilingReview.pdf
www.physicianprofiling.ch/Profiling1994.pdf
www.physicianprofiling.ch/Profiling2006.pdf
www.physicianprofiling.ch/Profiling2008.pdf
www.physicianprofiling.ch/ProfilingCMS2011.pdf
www.physicianprofiling.ch/ProfilingIssues2012.pdf
www.physicianprofiling.ch/ProfilingUncertainties2012.pdf
www.physicianprofiling.ch/ProfilingTeaching2008.pdf
www.physicianprofiling.ch/ProfilingTaiwan2010.pdf
www.physicianprofiling.ch/ProfilingMentalDisorders2012.pdf
www.physicianprofiling.ch/ProfilingMedicationDatabase2005.pdf
www.physicianprofiling.ch/ProfilingGroupLevel2011.pdf
www.physicianprofiling.ch/ProfilingDutch2009.pdf

Anhang zu Schwenkglenks Mittelwert

Wertebereiche der rückgerechneten Durchschnittskosten der Referenzkollektive

Datenbasis siehe: *Vergleich verschiedener Instrumente (Rechnungsstellerstatistik der santésuisse und Praxisspiegel der Trustcenter) zur Beurteilung der von Schweizer Ärzten in der Grundversorgung verursachten Behandlungskosten. Statistischer Studienbericht. Version 1.0 vom 8. Juli 2010*

Die Durchschnittskosten des Referenzkollektivs wurden berechnet als Durchschnittskosten pro Patient des jeweiligen Arztes geteilt durch seinen RSS Index. Für die Ärzte, die zum selben Referenzkollektiv gehören, müsste sich immer der gleiche Wert ergeben, d.h. der Wertebereich (Differenz von kleinstem und grösstem Wert) müsste null sein. Die Daten zeigen erhebliche Abweichungen von dieser Erwartung.

Wertebereiche der gesamten RSS-Kosten nach Kanton und Facharztbezeichnung (0 = Allgemeinmedizin, 1 = Innere Medizin ohne Zusatzbezeichnung)

	r_ref1~t	KantID	ArztFMH
1.	0	NE	1
2.	0	SH	1
3.	0	AR	1
4.	0	FR	1
5.	0	VS	0
6.	0	VS	1
7.	0	GE	1
8.	0	JU	0
9.	.0176392	NW	0
10.	1.686523	GL	1
11.	1.758972	UR	0
12.	2.070435	ZG	1
13.	2.113953	LU	0
14.	2.758911	LU	1
15.	2.7901	ZG	0
16.	3.989014	OW	0
17.	5.327148	SZ	0
18.	5.92218	AR	0
19.	7.242737	ZH	0
20.	9.951233	ZH	1
21.	11.97986	BL	1
22.	17.86804	BS	1
23.	22.81543	TI	1
24.	40.15503	NE	0
25.	58.80835	BL	0
26.	71.1546	FR	0
27.	75.31073	TG	1
28.	77.18066	TG	0
29.	94.94678	TI	0
30.	95.97986	GE	0
31.	148.9603	SO	1
32.	154.016	SG	1
33.	178.2499	SO	0
34.	191.0374	GL	0
35.	204.5066	BE	0
36.	218.0773	SG	0
37.	220.0635	BS	0
38.	269.4888	AI	0
39.	354.2109	AG	1
40.	429.3038	AG	0
41.	471.4733	BE	1
42.	544.387	GR	1
43.	557.7886	SH	0
44.	1238.857	GR	0

Wertebereiche der gesamten Arztkosten nach Kanton und Facharztbezeichnung (0 = Allgemeinmedizin, 1 = Innere Medizin ohne Zusatzbezeichnung)

	r_refl~t	KantID	ArztFMH
1.	0	JU	0
2.	0	SH	1
3.	0	VS	0
4.	0	GE	1
5.	0	FR	1
6.	0	AR	1
7.	0	NE	1
8.	0	VS	1
9.	.0101318	NW	0
10.	1.021606	SZ	0
11.	1.025024	OW	0
12.	1.736084	UR	0
13.	2.11972	NE	0
14.	2.62262	LU	1
15.	2.838379	GL	1
16.	3.219604	LU	0
17.	4.298157	ZG	1
18.	4.403015	ZG	0
19.	4.43927	AR	0
20.	4.573059	ZH	0
21.	4.946014	TI	1
22.	5.411224	BS	1
23.	7.877197	ZH	1
24.	9.309143	BL	1
25.	9.761627	FR	0
26.	26.20755	TI	0
27.	33.91864	BL	0
28.	42.24292	GR	1
29.	45.23306	GE	0
30.	48.95245	AI	0
31.	50.49304	BS	0
32.	65.12024	BE	0
33.	91.04993	TG	0
34.	105.3617	SO	1
35.	109.397	GR	0
36.	122.5464	SG	1
37.	175.7716	BE	1
38.	179.8399	TG	1
39.	212.3817	GL	0
40.	229.6093	SG	0
41.	322.3238	AG	0
42.	340.7361	SH	0
43.	358.6542	SO	0
44.	456.2288	AG	1

[1] Mail von Kristof Nagy vom 29.10.2013:

Subject:
Differenz gemäss Schwenkglenks
Date:
Tue, 29 Oct 2013 09:31:03 +0000
From:
Nagy Kristof - SWP ZH <Kristof.Nagy@santesuisse.ch>
To:
'Ernst Gähler' (Ernst.Gaehler@hin.ch) <Ernst.Gaehler@hin.ch>, '"michel.romanens@hin.ch"' <michel.romanens@hin.ch>
Cc:
Caminada Markus - GL_TS SO <Markus.Caminada@tarifsuisse.ch>, "D'Angelo Mirjam - SWP SO" <Mirjam.D'Angelo@santesuisse.ch>, Grolimund Oliver - IS SO <Oliver.Grolimund@sasis.ch>

Sehr geehrter Herr Gähler
Sehr geehrter Herr Romanens

Im Auftrag von Frau D'Angelo sende ich Ihnen die in Aussicht gestellte Abklärung in Bezug auf Ihre gestellten Fragen zum RSS-Index.

PD Dr. Matthias Schwenkglenks stellt in seiner Studie „Vergleich verschiedener Instrumente (Rechnungsstellerstatistik der santésuisse und Praxisspiegel der Trustcenter) zur Beurteilung der von Schweizer Ärzten in der Grundversorgung verursachten Behandlungskosten" vom 08. Juli 2010 fest, dass in der RSS die Division der Durchschnittskosten eines Arztes mit seinem Index nicht die Gruppenmittel im entsprechenden Kanton ergeben, respektive dass die so errechneten Gruppendurchschnitte eine z.T. breite Streuung aufweisen (vgl. Studie Schwenkglenks, S. 28).
Diese Unterschiede sind zurückzuführen auf die Rundung der Zahlen rsp. Indices, welche in der RSS-Statistik ausgewiesen werden. Um die korrekten Gruppendurchschnitte evaluieren zu können, müsste man die ungerundeten Zahlen verwenden, welche in der RSS nicht ausgewiesen sind. Gerne untermauern wir diesen Sachverhalt an folgendem Beispiel:

Index gemäss RSS des Arztes:

Index = 51

Total direkte Kosten: Kosten pro Erkrankten = 320.67

=> Gruppenschnitt gemäss Schwenkglenks: CHF 628.7647059

Ungerundete Werte des Arztes gemäss SASIS AG:

Index = 51.44264222

Total direkte Kosten: Kosten pro Erkrankten = 320.6668573

=> Gruppenschnitt CHF 623.3483419

Diesem Mail beigelegt finden Sie die offizielle Bestätigung dieses Sachverhaltes vor.

Email Korrespondenz mit ETH Zürich (Seminar für Statistik)

Werner Stahel **stahel@stat.math.ethz.ch**
27.08.2013
Sehr geehrter Herr Dr. Romanens
> Doch was wäre zum statistischen Vorgehen der santésuisse in der
> Regressionsanalyse ohne eine solche Aggregation zu sagen?
Ein Einfluss wäre vorhanden, da die Mittelung über die Patienten
in einer Klasse zuerst geschieht und nachher logarithmiert wird.
Wäre es umgekehrt, so wäre kein Effekt zu erwarten.
In welcher Richtung der Effekt für den einzelnen Arzt wirkt,
kann man nicht generell sagen. Die Wirkung hängt davon ab, ob
extrem hohe Kosten in "gut besetzten" Klassen anfallen oder
in solchen mit wenigen Patienten.
Da für alle Aerzte das gleiche Verfahren angewandt wird und
nachher nur Unterschiede zwischen Aerzten (resp. zwischen
einzelnen Aerzten und Mittelwerten über Aerzte) eine Rolle
spielen, ist davon keine systematische Verfälschung zu
erwarten.
Ich hoffe, dass Ihnen diese Antwort hilft.
Freundliche Grüsse
Werner Stahel

Werner Stahel, Seminar fuer Statistik, HG G 14.2
ETH, Raemistr. 101, CH-8092 Zurich, Switzerland
tel. +41 44 632 34 30, http://stat.ethz.ch/~stahel

Werner Stahel **stahel@stat.math.ethz.ch**
29.08.2013
Sehr geehrter Herr Romanens
Das Bestimmtheitsmass R2 ist ein bekanntes Mass für die "Güte"
eines Modells. Wie Sie feststellen, ist es sehr
interpretationsbedürftig. Allenfalls brauchbar ist es für den
Vergleich verschiedener Modelle für den gleichen Datensatz.
Theoretisch wäre es denkbar, via Anzahl der gemittelten
Patienten die Streuung in den ursprünglichen Daten
zurückzurechnen und damit ein R2 zu bestimmen, welches bei
Einzeldaten zu erwarten wäre. Das ist aber wieder mit Annahmen
über Unabhängigkeiten verbunden, ist nicht trivial und für sich
selbst ja auch nicht aussagekräftig.
Es wird sicher sinnvoll sein, ein Modell mit und eines ohne
Morbiditätsvariable zu vergleichen -- mit dem Datensatz, der
dereinst zur Verfügung stehen wird.
Freundliche Grüsse
Werner Stahel

Werner Stahel, Seminar fuer Statistik, HG G 14.2
ETH, Raemistr. 101, CH-8092 Zurich, Switzerland
tel. +41 44 632 34 30, http://stat.ethz.ch/~stahel

Werner Stahel **stahel@stat.math.ethz.ch**
13.01.2014

Sehr geehrter Herr Romanens
Das von Ihnen erwähnte Modell [1] bildet die Grundlage, auf der der Index beruht. Die Idee ist die, dass einem Arzt nicht deshalb "überhöhte" Kosten zugerechnet werden, weil er kostenintensive Patienten hat. Also werden, wie beim Risikoausgleich der Prämien für die Krankenkassen, die Kosten betreffend der Alters-/Geschlechtsklasse adjustiert.
Das geschieht mit Modell [1]. Es ermöglicht eine schärfere Bestimmung jener, die allenfalls "ohne Grund" höhere Kosten verrechnen als ihre Kolleg/innen.
Ein hohes R2 ist ja anzustreben, damit die Methode Sinn macht und mit erhöhter Wahrscheinlichkeit die "Richtigen" von der Ueberprüfung betroffen werden.
Natürlich ist das nur eine grobe Ajustierung. Es wäre ideal, die Morbidität irgendwie "objektiv" erfassen und dann ausgleichen zu können. Sonst gibt es Anreize für Taktiken, mit denen man das Hängenbleiben im Filter verhindern kann.
Formel [2] wirkt sich ja auf alle Aerzte (i) gleich aus:
Man addiert den Mittelwert der beta_j als Konstante.

> Wäre es möglich, die ANOVA Korrektur
> auf die Arztkosten ohne diese Indexvariable zu rechnen und was
> käme dann in Bezug auf den ANOVA-Index heraus?
Eine Arzt-Variable braucht es ja, um die Index-Variable zu rechnen. Man könnte höchstens die Kosten pro Patient für jeden Arzt mitteln, ohne die Alters-/Geschlechts-Verteilung zu berücksichtigen -- das ist wohl nicht in Ihrem Sinn.
Habe ich Ihre Frage beantwortet?
Freundliche Grüsse
Werner Stahel

Werner Stahel, Seminar fuer Statistik, HG G 14.2
ETH, Raemistr. 101, CH-8092 Zurich, Switzerland
tel. +41 44 632 34 30, http://stat.ethz.ch/~stahel

Tabelle 2: Risikostrukturausgleichszahlungen summiert für die Jahre 1996-2012. Angaben in Millionen Franken.

	Erhalten	Bezahlt
Helsana Versicherungen AG	-7811.63	0.00
Visana AG	-5384.92	0.00
CSS Kranken-Vers. AG	-2127.86	0.00
Wincare Versicherungen AG	-1753.69	0.00
CONCORDIA AG	-1107.84	0.00
Vivao Sympany AG	-189.13	0.00
Krankenkasse Institut Ingenbohl	-56.50	0.00
KRANKENKASSE SLKK	-55.11	0.00
kmu-Krankenversicherung	-18.54	0.00
Krankenkasse Turbenthal	-2.13	0.00
Krankenkasse Zeneggen	-0.92	0.00
Kranken- und Unfallkasse Einsiedeln	-1.81	0.55
Cassa da malsauns Lumneziana	-1.68	1.02
Krankenkasse Simplon	0.00	1.27
Krankenkasse Stoffel Mels KKS	-0.88	1.46
sanavals Gesundheitskasse	-0.41	1.79
KK Wädenswil	-8.22	2.71
maxi.ch Versicherungen AG	0.00	4.36
Krankenversicherung Flaachtal AG	-0.07	7.89
CM Vallée d'Entremont	-1.48	8.24
Krankenkasse Birchmeier	-1.47	9.12
vita surselva	-1.51	9.72
Krankenkasse Visperterminen	0.00	12.00
KLuG Krankenversicherung	-4.78	13.26
SUPRA	-555.54	14.68
Glarner Krankenversicherung	0.00	15.68
AMB	-3.39	16.08
GALENOS Kranken- und Unfallversicherung	-79.82	16.21
Aquilana Versicherungen	-21.07	16.87
AGILIA Krankenkasse AG	-2.37	17.16
Krankenkasse Steffisburg	-0.85	20.18
Publisana	0.00	22.59
rhenusana	0.00	26.92
innova Krankenversicherung AG	-24.01	28.55
Moove Sympany AG	-10.08	44.67
Krankenkasse Luzerner Hinterland	0.00	47.20
sodalis gesundheitsgruppe	0.00	59.19
Atupri	-111.64	66.80
PROVITA Gesundheitsversicherung AG	-3.75	67.73
Kolping Krankenkasse AG	-15.19	67.89
vivacare AG	0.00	72.34
Sanagate AG	0.00	91.08
Sumiswalder Krankenkasse	0.00	104.66
Avenir Assurance Maladie SA	-29.01	138.31
Compact Grundversicherungen AG	0.00	141.01
ÖKK	-10.20	195.64
Philos Assurance Maladie SA	-83.17	216.73
Easy Sana Assurance Maladie SA	-2.86	239.60
Arcosana AG	0.00	341.62
sana24 AG	0.00	350.89
EGK Grundversicherungen	-10.70	392.74
Agrisano Krankenkasse AG	0.00	402.00
KPT Krankenkasse AG	-80.83	410.26
SWICA Krankenversicherung AG	-167.17	414.06
Sansan Versicherungen AG	0.00	606.39
Avanex Versicherungen AG	0.00	716.10
Mutuel Assurance Maladie SA	-2.77	727.46
Progrès Versicherungen AG	0.00	1344.17
Intras Kranken-Versicherung AG	-30.60	1405.36
Sanitas Grundversicherungen AG	0.00	2141.71
ASSURA	0.00	3982.71

Santésuisse verweigert die Herausgabe von zugesagtem Zahlenmaterial:

Emails:

AW: Excel Tabellen Dr. R und Dr. H. / ZSR Statistik, Vergleichsgruppen

---

D'Angelo Mirjam - SWP SO <Mirjam.D'Angelo@santesuisse.ch>

3. März 2014
09:19

An: Michel Romanens <info@kardiolab.ch>

Cc: Caminada Markus - GL_TS SO <Markus.Caminada@tarifsuisse.ch>, "'Gähler Ernst ' (Gähler Ernst)" <egaehler@hin.ch>, Thomas Kessler <thomas.kessler@fmh.ch>, "Dr. med. Markus Trutmann" <markus.trutmann@fmch.ch>, Peter Frutig <peter.frutig@hin.ch>, Jürg Schlup <juerg.schlup@fmh.ch>, "Saxer Prof. iur." <saxer@steinlex.ch>, "Moline Philip (philip.moline@newindex.ch)" <philip.moline@newindex.ch>

Sehr geehrter Herr Romanens

Die technische Arbeitsgruppe ANOVA baut auf eine konstruktive Zusammenarbeit der Parteien im Hinblick auf die Weiterentwicklung der statistischen Methode. In diesem Zusammenhang setzen wir den sorgfältigen Umgang mit internen Informationen, Auswertungen und schützenswerten Daten voraus.

Da wir letzte Woche festgestellt haben, dass Sie auf Ihrer persönlichen Homepage http://www.physicianprofiling.ch/ interne Informationen wie E-Mailverkehr, Protokolle von Sitzungen und Auswertungen z. Hd. der Arbeitsgruppe publizieren, haben wir uns entschlossen Ihnen keine weiteren Informationen mehr zukommen zu lassen. Herr Dr. H. hat von uns wie versprochen bis Ende Februar eine Rückmeldung erhalten.

Freundliche Grüsse

Mirjam D'Angelo

---

santésuisse
Die Schweizer Krankenversicherer
Mirjam D'Angelo
Abteilung Wirtschaftlichkeitsprüfung
Projektleiterin
lic.rer.pol.
Römerstrasse 20
4502 Solothurn

Tel +41 32 625 4731
Fax +41 32 625 47 01
Mirjam.D'Angelo@santesuisse.ch
www.santesuisse.ch
www.monsieur-sante.ch

Von: michel.romanens@gmail.com [mailto:michel.romanens@gmail.com] **Im Auftrag von** Michel Romanens
Gesendet: Donnerstag, 27. Februar 2014 21:34
An: D'Angelo Mirjam - SWP SO; 'Gähler Ernst ' (Gähler Ernst); Thomas Kessler; Saxer Prof. iur.; Dr. med. Markus Trutmann; Moline Philip (philip.moline@newindex.ch); Peter Frutig; Jürg Schlup
Betreff: Fwd: Excel Tabellen Dr. R und Dr. H. / ZSR Statistik, Vergleichsgruppen

Sehr geehrte Frau D'Angelo

Ich stelle fest, dass die von Ihnen versprochenen Daten gemäss Auftrag der FMH noch nicht bei mir eingetroffen sind, damit ist die von mir gesetzte Frist seit bald 2 Tagen abgelaufen. Erstmals ermahne ich Sie deshalb, mir die Daten bis morgen, 28.02.2014, 12 Uhr, zugänglich zu machen (via Email als 2 excel Tabellen), damit das weitere ordnungsgemässe Funktionieren der technischen Arbeitsgruppe gewährleistet bleibt und die

Datenanalyse zeitgerecht erfolgen kann.

Zudem hat Dr. H. noch keine Stellungsnahme betr. des Verfahrens gegen ihn erhalten (santésuisse hat eine hohe Busse in Planung wegen erhöhten veranlassten Vergleichskosten, deren Ursache Dr. H. schon 6x einwandfrei belegen konnte, worauf die santésuisse von weitere Massnahmen absah, zudem haben Sie selber an der letzten Sitzung vom 18.02. festgestellt, dass die veranlassten Kosten durch die Praxisbesonderheiten erklärt sind). Diese Mitteilung haben Sie ebenfalls versprochen, bis spätestens morgen, aber bisher nicht eingehalten. Es würde darum gehen, das Verfahren wegen mangelnder Begründung unmittelbar einzustellen.

Michel Romanens