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 (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.

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

"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 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.

---

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

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 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

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

		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 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

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

pharmaceutical cost groups (PCG). The list of selected possible cost explaining variables is given in Table 2.

Table 2: accepted independent cost explaining variables at the individual patient level (Modified from[9]). The r^2 statistic is a 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 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

[9] www.physicianprofiling.ch/ProfilingStateAndImprove2013Overview.pdf
[10] www.physicianprofiling.ch/ProfilingDudleyMedCare2003.pdf

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 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".

---

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

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 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