Inter-Institutional Profiling and Patient Safety: Which Method to Appropriately Assess the Future Risk of Flat Rates in Hospital Care?

Discussions and Recommendations of the ad hoc Taskforce "audits in medicine"

Michel Romanens, MD, Bernhard Hofmeier, MD, Franz Ackermann, MD

Corresponding author:

Michel Romanens, MD

Ziegelfeldstr. 1

CH-4600 Olten

info@kardiolab.ch

Abstract

Background

The Swiss Medical Society endorses a working group that has issued propositions to control for possible side effects of the introduction of flat rates as a hospital financing modality. These operate by law up from the year 2012.

Method

Review of the literature based upon the question, to what extent current evidence shows, that the coefficient of determination (r2) is sufficiently high to support the use of mortality rates to adequately profile institutions. For that purpose, we set the threshold for sufficient predictive value and sensitivity at > 0.90 and an r^2 (coefficient of determination) at > 0.90.

Results

None of the reviewed studies met the benchmark criteria for sufficient accuracy to detect poor hospital care performance. At its best, the profiling tools showed predictive values < 0.50. The potential of mislabelling was cited several times in the literature.

Conclusions

The Taskforce "audits in medicine" discusses the value of two principle ways to adequately profile institutions in Switzerland; (adjusted) mortality rates and audits of deaths and indications for medical interventions. The result of our mortality-focused review clearly indicates the non applicability of mortality rates to profile institutions; on the contrary, they may even decrease health care quality, increase costs and lead to mislabelling of institutions.

Word count: Abstract: 206; paper 4373

Key words: inter-institutional profiling; adjusted in-hospital death rates; benchmarking

Introduction

In view of the introduction of flat rates (diagnosis related cost groups, DRG) in Switzerland, questions about fairness and equity of health care performance and inter-institutional outcome-rates are overdue. Swiss specific flat rates will be used to cover hospital expenditures. Some patients will however be more costly than expected, e.g. with outliers (defined as longer than expected hospital stay) and inliers (defined as short time hospital stay with extreme costs) [1]. Therefore, treating such patients creates a cost risk to health care providers that ultimately may lead to institutions bank rot. Clearly enough, avoidance of treating potentially high cost risk patients, e.g. the sicker ones, might become a question of survival for a hospital. Based upon preliminary analysis of SwissDRG, DRG's explain hospital expenditures up to 48% [1]. Therefore, 52% of costs, if no other measures are taken, will be left as a cost risk to the treating hospitals. As reports from press media say, about one third ore about 100 hospitals with acute beds have to be closed. Since Switzerland disposes of a rather low number of acute hospital beds when compared to OECD countries (for the year 2007, 3.4/1000 beds are listed [2]), tear down 100 hospitals will create a scarcity of acute beds. Patient security within this sphere is likely to be threatened, since effects causing "good patient selection" and "waiting times" are likely to add up in a way decreasing patient safety. Within this potentially harmful setting of financing patient care in hospitals, the question, how to measure negative effects with highest accuracy at a lowest cost is of central importance. Basically, there are two ways to detect harmful effects: a) comparison of rates of side effects based upon institutional morbidity and mortality, eventually adjusted for case mix, age and gender, and b) by chart reviews (audits, either subjective or structured, e.g. using international guidelines) that identify preventable side effects of medical interventions and reviews of indications for interventions.

The following work stems from the ad hoc Taskforce "audits in medicine" and delivers its work and conclusions to the Swiss Medical Association (FMH) as an additional discussion and issues preliminary recommendations to better control for undesired, or even harmful, effects of flat rates on patient care and safety. Further, we aim at expanding the discussion furnished by the FMH in September 2009, expressed in a comprehensive position paper [3].

Methods

We give a brief overview on available evidence both for "rate comparisons" and "chart reviews", discuss advantages and disadvantages of these procedures.

We reviewed published and peer reviewed articles. Our focus was directed towards the explanatory content of statistical methods using "r^2 statistics" and/or accuracy analysis and tried to draw a conclusion about whether risk estimations were appropriately made. Since conclusions derived from benchmarking deal with the future of patients with respect to survival and resource allocation, the comparative quality of such benchmarks has to be known and extremely good in order to point to the inherent problem under investigation, e.g. poor quality of health care and in order to avoid misclassifications of institutions. Therefore, we define the explanatory content of a benchmark using the coefficient of determination (benchmark: $r^2 > 0.90$) and diagnostic accuracies like sensitivities or positive predictive values > 0.90).

Statistical Basics, methodological considerations, biases, and confounders

Basically, a benchmark is a test, for which accuracy should be known. As reviewed by Romanens et al [4], "a test result can be positive; indicating the presence of disease, or it can be negative, indicating the absence of disease". As a test may give erroneous results, it may be true positive (TP), false positive (FP), true negative (TN), and false negative (FN). This is the basic concept of the two-by-two table in its application to diagnostic tools." From this information, a number of key informations of test performance can be calculated (Table 1).

Table 1:

SE =TP/(TP + FN)

SP= TN/(TN + FP)

PPV = TP/(TP + FP)

NPV =TN/(TN + FN)

pLR =SE/(1 –SP)

nLR = (1 –SE)/SP

ACC= (TP +TN)/(TP +TN + FP+ FN)

For the calculation of probabilities, a range from zero to 1.00 is used. Multiply by 100 to obtain percentages and accordingly, sensitivities and specificities are not expressed in percent but in percent divided by 100. ACC, accuracy; FN, false negative; FP, false positive; nLR, negative likelihood ratio; NPV, negative predictive value; pLR, positive likelihood ratio; PPV, positive predictive value; SE, sensitivity; SP, specificity; TN, true negative; TP, true positive.

A frequently used second way to generate a perspective of a tests value is the coefficient of determination (r^2). "R^2 can be seen to be related to the unexplained variance (variance of the model's errors) with the total variance. The fraction of variance unexplained (or FVU) in the context of a regression task as performed with r^2 statistics is the amount of variance of the regressand Y which cannot be explained, i.e., which is not correctly predicted, by the explanatory variable X, e.g. due to an important but not measured variable (confounder). In statistics, a confounding variable (also confounding factor, lurking variable, a confounder, or confounder) is an extraneous variable in a statistical model that correlates (positively or negatively) with both the dependent variable and the independent variable. The methodologies of scientific studies therefore need to control for these factors to avoid a type 1 error; an erroneous 'false positive' conclusion that the dependent variables are in a causal relationship with the independent variable. Such a relation between two observed variables is termed a spurious relationship" (from Wikipedia.org, accessed 30.Dec. 2009). Thus, confounding is a major threat to the validity of inferences made about cause and effect, i.e. internal validity, as the observed effects should be attributed to the confounder rather than the independent variable.

A good benchmark has to have a very high explanatory content for the cause of a result or a difference that is searched for. E.g. reporting higher in-hospital 30 day mortality rates for cardiac surgery in institution A when compared to institution B can lead to the conclusion, that the quality of institution A is superior to institution B. The question here is: to what extent does this benchmark reflect poor delivery of medical treatment? Is it 5%, 50%, or 90%? Unfortunately, the benchmark "in-hospital mortality after cardiac surgery" is influenced by a number of factors that affect outcome and its predictive value for poor performance institutions. These factors can be categorized as outlined in Table 2:

Table 2: biases and confounding variables in benchmark statistics not due to poor quality of institutional care delivery

Type 1. Data source and coding inaccuracies

     a) error in reporting deaths as caused by a medical condition (classification bias)

     b) error in finding enumerator cases, leading to increased death rates (non-response bias)

     c) change of diagnostic criteria for a medical condition over time

     d) upcoding of diagnostic severities in order to obtain higher reimbursement in institution financed by flat rates

     e) inclusion of second diagnosis codes into the primary code

Type 2. Pre admission inaccuracies

     a) unknown medical interventions and unknown quality of care prior to hospital admission, e.g. housedoctors with work overload in rural areas or inappropriate medications applied to treat the life threatening condition

     b) regional differences influencing severity of established disease, e.g. amount of regional air pollution

     c) referral bias, e.g. higher risk patients sent to institutions with perceived better delivery of medical care

     d) local preferences of a population with respect to the indication for a hospitalization

Type 3. Admission inaccuracies

     a) blurred diagnosis, e.g. antibiotic therapy without obtaining previous blood cultures in a patient with acute endocarditis

     b) patient denies accurate diagnostic measures or therapeutic interventions

Type 4. Post admission inaccuracies

     a) medical complications not measured by admission diagnosis, e.g. unrecognized pulmonary embolism leading to death in a patient with myocardial infarction

     b) patient denial to intervene medically in a condition that leaves the patient with a high risk for death.

     c) patient referral to hospices reduces in-hospital mortality rates

Review of benchmark studies and benchmark quality assessment studies

The Swiss Federal Office of Public Health (www.bag.admin.ch) has published death rates from several hospitals and for several causes of death. For the year 2008, BAG published a post myocardial infarction death rate for the University hospital Zurich of 11.2%, where in fact the real death rate was 2.7%. This apparent discrepancy was due to an enumerator error, in that the BAG found 125 cases with myocardial infarctions (with 14 deaths), whereas in reality there were 415 cases with myocardial infarctions (with 11 deaths) [5].

A further problem stems from the cause of death (COD). By coding a COD, errors may occur on the physicians level, on the coders level – and where available – on the report of the pathologist. It was found, that 41% of COD were not appropriately coded when compared to other clinical and autopsy derived findings [6]. This issue was further discussed in a paper counting in-hospital deaths in Switzerland due to stroke. Apart from the coding bias, stroke was not only counted for by the first clinical diagnosis, but also for the clinical diagnosis two to five which is known to decrease specificity and accuracy of such codings [7]. Further, there were no data on coding errors, although audits on coding are regularly performed [8]. Therefore, the accuracy of death due to stroke as measurable from nation data bases (HOST) is unknown. Eventually breaking down death rates due to stroke to single hospitals with the attempt to create a benchmark for stroke care is therefore hampered by an unknown amount of inaccuracy and may mislabel institutions as poor performers [8].

Attempts to profile hospital quality based upon death rates have been performed for decades [9-20]. An elegant way to look at random variance of death rates is a measure of the correlation coefficients and kappa values (for agreement) looking at several linked causes of death. The "link" implies, that the same doctors care for different diagnoses (e.g. the pair "pneumonia and obstructive lung disease" or the pair "acute myocardial infarction and congestive heart failure"). This work was elegantly performed in a large study involving 5505 US hospitals in the year 1991 [21]. For each diagnosis, the number of patients, the observed (i.e., actual) mortality rate within 30 days of admission, and the predicted mortality rate were obtained. Predicted mortality rates were determined from multivariable models developed by HCFA and were based on age, gender, prior hospitalizations, reason for admission (based on the primary International Classification of Diseases [9th edition] Clinical Modification [ICD-9-CM] diagnosis code), and the presence of specific comorbid illnesses (e.g., cancer, diabetes) identified by ICD-9-CM codes [9]. COD under investigation were acute myocardial

infarction (=946), congestive heart failure (N=1914), Pneumonia (N=1559), obstructive lung disease (N=199), Stroke (N=1775), coronary bypass surgery (N=516) and hip fractures (N=628). The Pearson's correlation coefficients between death rates for the pair "acute myocardial infarction and congestive heart failure" were 0.18 and for the pair "pneumonia and obstructive lung disease" were 0.22, with similar results for the standardized mortality ratios quintiles assessed by Spearmans coefficients (range for all possible 21 pairs between 0.00 and 0.22). Further, weighted kappa statistics for agreement of pairs ranged between -0.01 and + 0.22 and was 0.12 for the pair "acute myocardial infarction and congestive heart failure". The authors concluded: "the findings suggest that mortality rates for different diagnoses may be poorly related and that it may not be valid to judge hospital quality on the basis of a single diagnosis or even a few diagnoses".

The influence of referral rates to hospices on the magnitudes of in-hospital mortality rates was analysed on a state-level in the US. Authors found that higher number of hospices per state explained in-hospital reduction of standardized mortality ratios by 22 to 24% (p=0.01) [22].

A nationwide attempt to profile hospitals using mortality rates was performed between 2003 and 2005 for all Dutch hospitals [23]. Major predictors of death reduction rates were the year of examination with a steady reduction of 8% per year, accounting for an r^2 of 0.45 for the rate of unexpected death rate reductions. Authors also observed that higher number of GP's per 10'000 habitant showed an r^2 of 0.17 and that academic hospitals had a higher unexpected death rate than second type hospitals with an r^2 of 0.26. In this sample, therefore, academic hospitals would provide a poorer care, which is counterintuitive. Consequently, authors invalidate their findings as a useful inter-institutional profiling tool by the following statement: "We may not have captured all the case-mix differences; rendering an HSMR comparison with other hospitals invalid. Model misspecification could be due to measurement errors, misspecified functional forms and omitted variable bias. One example of such an omitted variable is the readmission rate per hospital. Hospitals with high readmission rates may have more severe patients. However, the variable "readmissions" was not included due to underreporting". This gives also a hint to another bias: if readmissions may reflect a higher level of morbidity, how can this variable ever be used to identify "bloody discharges" from institutions sending their patients home to early for saving hospitalization costs.

An important confounder for mortality is the "emergency admission" [24], as assessed in 166 deaths having occurred in 8501 inpatients in 1993. Authors concluded: "In agreement with national results, the great majority of deaths (N=145) on this unit occurred among elderly patients admitted as emergencies. Many were unfit, with significant cardiac, respiratory or

renal comorbidity and high ASA grades. Standard physiological and operative scoring systems confirmed that those in whom death was inevitable were more ill as a group than those in whom death rates might reflect the quality of surgical care. However, there was considerable overlap between the groups. Such fitness assessments do not take account of the nature of the presenting surgical pathology, which in many emergency cases only becomes apparent at operation. Several patients presenting with carcinomatosis, extensive intestinal infarction or gross faecal peritonitis had reasonable fitness assessments although death was inevitable. In other instances death was inevitable because patients declined emergency operations or surgeons decided against them on account of great age or poor quality of life. This disparity between measurable physiological fitness and true prognosis in surgical patients complements previous observations in patients with non-surgical conditions."

Routine data collection may be used to screen for possible quality problems [25]. However, authors explain: "A note of caution is required when examining comorbidities. Jencks [26] et al showed how some comorbidity, such as diabetes and hypertension, may fail to be coded for in cases where there is very severe illness as they are displaced by complications of the principal diagnosis from the limited number of spaces on the record for diagnostic codes." Further: "Currently available systems to adjust for risk of adverse outcome, whether they use routine data or information from case notes, can explain no more than about 25% of observed mortality, and many of those working in this field have concluded that it is not yet possible to make valid inferences from risk adjusted outcomes" [25,26].

Weir et al aimed at monitoring the quality of hospital-based stroke services in 2001 [27]. After adjusting for risk, case fatality still remained significantly different (p=0.045) in 2724 patients. Authors discussed the following 3 major points: "First, we have shown that the large differences in crude case fatality between the study hospitals could be largely explained by important differences in case mix and that much of the remaining differences in case fatality could be attributed to chance. In other words, most of the variation in outcome after stroke between the study hospitals was due to factors outside their control rather than differences in the quality of the care they provided. (…) Second, after adjustment for important differences in case mix, case fatality at Hospital A was significantly higher than that at Hospitals B through E and was associated with the failure of Hospital A to provide any organized and specialized rehabilitation and, compared with the other hospitals, a much lower use of CT head scanning and a lower standard of documentation (according to the RCPSAP criteria). (…) Our third main finding is that adjusted case fatality data are clearly unable to differentiate between hospitals with moderate differences in stroke care."

The issue of Hospital Standardized Mortality Ratio, a global measure of hospital death rates for diagnoses accounting for 80% of in-hospital mortality has been discussed extensively [28]. Authors concluded: "Developing satisfactory performance measures will require substantial investment. Mortality measures will need to focus on specific conditions and procedures with known connections to elements of care that providers are able to control. Adjusting mortality measures for case mix can sometimes be done with the use of administrative data alone, but efforts to obtain clinical information from charts will be required in many situations."

Jarman looked at the explanatory content for death rates using routinely collected hospital data [29]. Counting 7.6 Mio admissions, the emergency admissions (4.6 Mio) accounted for 93% of observed deaths in the hospitals in England between 1991-95. Significant predictors of death were smaller numbers of medical doctors, comorbidities, and numbers of hospitals per 100'000 persons. In this very large study it was acknowledged, that a lack to correctly assess the severity of illness may account for the apparent poor performance of case-mix variables used (regression coefficient of -0.18 only). Authors concluded: "Most of the significant predictors in our two models are outside the direct influence of hospital policy (except doctor numbers per bed), and adjustment for these external factors narrows the range of mortality ratios". Therefore, a smaller number of hospitals, as planned for Switzerland, has been shown to increase patient in-hospital mortality.

William et al performed an analysis of the accuracy of risk-adjusted mortality rates as a measure of hospital quality of care [30]. Based on an analytical model of random measurement error, sensitivity and predictive error of mortality rate indicators of hospital performance were estimated. Authors found "under virtually all realistic assumptions for model parameter values, sensitivity was less than 20% and predictive error was greater than 50%. Reports that measure quality using risk-adjusted mortality rates misinform the public about hospital performance."

Guru et al [31] performed extensive audits to detect preventable deaths in randomly selected patients having undergone cardiovascular surgery and correlated their audits with risk-adjusted mortality statistics in 347 deaths at 9 Canadian institutions. Risk-adjusted in-hospital all cause mortality rates showed a Spearman correlation of -0.42 (p=0.26). Although 1/3 of deaths were judged as preventable, risk adjusted mortality rates were not correlated with preventable deaths. Importantly, preventable deaths occurred in 122 patients with an operative risk of <5% and <in 89 patients with an operative risk of >5%. Further, audits revealed 52 not indicated operations in 694 audits (=7.5%). Since incentives to perform operations with a poor indication might increase with flat rates in order to lower case-mix risk and increase the

number of cases – also termed risk avoidance creep [32], this particular variable has a great monitoring potential to detect undesired expansions of indications – without being measurable when mortality rates or readmission rates are used to monitor quality.

Discussion

The main finding of this review lies in the fact, that all reported attempts to correctly identify a poor hospital performance based upon risk-adjusted death rates is a clinical failure. In no study reviewed, neither predictive accuracy (sensitivity) nor predictive power (r^2 statistics) revealed high enough numbers in order to promote such tools for an inter-institutional quality of care process.

There has been a development of public health institutions to attempt to monitor and qualify the effect of medical interventions based upon inter-institutional differences of rates, sometimes corrected for co-variables such as patient morbidity at hospital entry, or simply corrected for age and sex. The performance of such benchmarks in terms of accuracy or confounding variables has rarely been a matter of research itself and such benchmarks have a statistical problem with small numbers [5]. The poor performance of such benchmarks – as planned to benchmark institutions in Switzerland by the FMH – is most likely due to a lack of sufficient access to patient charts and financing. Where such informations are available and directly compared to such benchmarks, accuracy of differences in rates (e.g. adjusted hospital mortality rates) was generally very low with respect to sensitivity and predictive values. Adding expected to observed mean values of mortality is yet another non validated approach with greatest mislabelling potential.

Especially intriguing is the fact, that costly but precise chart reviews [29-31] unmasked afore mentioned profiling tools and benchmarking attempts as inappropriate.

Implications for patient safety and monitoring: how can risk of flate rates best be assessed?

The effect of such profiling benchmarks on patient safety has not been studied extensively and is discussed elsewhere [32]. It is however likely, that by avoiding high risk interventions, quality and safety problems, especially in the perioperative situation [31] may disappear by a reduction of the prevalence of in-hospital deaths. This in turn may increase the number of high-risk patients, for whom risky interventions will be avoided. This clearly puts these patients at a higher risk of death, if the intervention avoided were appropriate and clinically indicated. Further, it is worrisome that smaller numbers of hospitals have been significantly correlated with an increased in-hospital death rate [29].

Risk for patients has to be detected with adequate tools; otherwise good hospitals will suffer mislabelling. This can not be a legislative intention. Since audits can discover preventable deaths – albeit "post festum" only – and can discover inappropriate use of medical

interventions, national databases on hospital deaths and indications for interventions should become available to institutionalized audits that regularly perform quality controls.

Within such a framework, we want to discuss the following key issues:

1) By the use of structured and subjective audits, medical doctors would not be profiled by inappropriate tools that may put patients at increased risk of a non-intervention, but would act within their framework of "professional ethics". Mortality statistics and similar profiling tools are in fact born of distrust in medical doctors, that they increasingly treat their patients for economical reasons only. Paradoxically enough, flat rates and mortality-rate profiling exactly promote such unethical attitudes.

2) As within the framework of scientific studies, any time possible audits performed by independent specialists will "per se" increase the quality of health care, if they are not followed by inappropriate sanctioning or attack the "professional ethics" of caring doctors. On a practical level it is very difficult for those independent of the profession to monitor practice, leaving the possibility that a code of practice may be self serving. This is because the nature of professions is that they have almost a complete monopoly on a particular area of knowledge. Therefore, quality control cannot be performed without the knowledge of specialists that come in for audits.

3) Increases in preventable deaths and inappropriate interventions can be quantified by regular audits and therefore serve as a tool to monitor undesired or harmful effects of flat rates.

4) Cost-efficiency studies are not existing to compare costs of audits versus public health derived observation tools, but it is likely that a sufficient number and network of audits will both increase quality of care and avoid unnecessary medical interventions, therefore be cost-saving, whereas profiling tools that may even promote costs and reduce patient safety, are very unlikely to be cost-efficient.

Limitations

This overview on physician and institution profiling in order to detect negative effects on medical performance when flat rates are eventually to be introduced has several limitations. First, it was not possible to review all available literature in that field and we had to focus on a rather limited field looking at mortality as a benchmark. However, mortality as a benchmark is likely to create the largest bias on patient safety, and therefore, the focus appears appropriate. Other benchmarks, such as patient satisfaction or level of luxury are quite less likely to affect patient safety.

Second, we can not provide scientific validations of our propositions, since we lack corresponding studies completely. Therefore, we create a new benchmark, e.g. "audits in medicine" without being able to clearly define a golden standard to which such audits could be compared to. However, since "audits in medicine" are likely to reduce the burden of unnecessary complications and costs, the call for such a golden standard may be misguided. Third, we have to leave it to the authorities, to eventually adopt the idea of "audits in medicine" as a new standard of health care controlling in Switzerland, but at that point of time, we cannot provide more details on how to elaborate such a network.

Conclusion

For benchmarking clinicians and institutions, based upon what is known from the literature, comparing mortality or disease rates gives inappropriate results and may lead to avoidance of high risk interventions, which is a serious threat against patient safety and patient survival during severe medical episodes.

At the moment, only reviewing clinical charts in patients with deadly course of intervention appears appropriate and cost-efficient. This approach helps to identify appropriate measures to install quality improvements that will avoid preventable deaths in the future, and this not only for high risk patients.

Further, appropriateness of interventions should also be checked in order to monitor the effect of flat pricing on increased and eventually unnecessary therapies in front of low risk conditions.

Independent chart reviews (audits) are necessary to detect and avoid inappropriate interventions and deaths. Audits should be performed by independent, eventually foreign experienced medical doctors without creating a financial incentive, thus creating a true independent observatory, ideally with two observers who agree, and in case of disagreement, a third observer, who discusses the problem and leads to a final conclusion.

For quality assurance, every deadly course of intervention and every indication for intervention have to be reported to a central database and we call for starting such a database as a pilot study as soon as possible, e.g. looking at cardiovascular interventions. Audits do not have necessarily to cover everything, but should perform random sample surveys. The risk, that audits may come to place, are very likely to reduce harmful or unjustified incentives due to flat pricing and serve as a preventive tool to excesses of over-medicalization and poor management of medical conditions, that might eventually lead to a patients death.

References

[1] http://physicianprofiling.ch/SwissDRGRisikoausgleich2007.pdf

[2] Berner Zeitung, 08.10.2009

[3] http://physicianprofiling.ch/SwissDRGKonzeptBegleitforschungFMH02009.pdf

[4] Romanens M, Ackermann F, Spence JD et al. Improvement of cardiovascular risk prediction: time to review current knowledge, debates, and fundamentals on how to assess test characteristics. European Journal of Cardiovascular Prevention and Rehabilitation 2010, 00:000–000 (ahead of print)

[5] Lüscher T. Lies, Damn Lies, and Statistics. Bemerkungen zu den Outcome-Zahlen des Bundesamtes für Gesundheit. Kardiovaskuläre Medizin 2009;12(9):229–233 20.

[6] Smith A, Hutchins G. Problems with Proper Completion and Accuracy of the Cause-of-Death Statement. Arch Intern Med. 2001;161:277-284

[7] Meyer K, Simmeta A, Arnold C. et al. Stroke events and case fatalities in Switzerland based on hospital statistics and cause of death statistics. SWISS MED WKLY 2009;139(5–6):65–69

[8] Kirkman M, Albert A. Letter to the Editor. SWISS MED WKLY 2009;139;463-4.

[9] Dubois R, Rogers W, Ill M et al. Hospital inpatient mortality: is it a predictor of quality? N Engl J Med. 1987;317:1674-1680.

[10] Green J, Passman LU, Wmtfeld N. Analyzing hospital mortality: the consequences of diversity in patient mix. JAMA. 1991;265: 1849-1853.

[11] Green J, Wintfeld N, Sharkey P, Passman U. The importance of severity of illness in assessing hospital mortality. JAMA. 1990; 263:241-246.

[12] Greenfield R, Aronow H, Elashoff R et al. Flaws in mortality data: the hazards of ignoring comorbid disease. JAMA. 1988;260:2253-2255.

[13] Hartz A, Gottlieb M, Kuhn E et al. The relationship between adjusted hospital mortality and the results of peer review. Health Serv Res. 1993;27:765-777.

[14] Kahn K, Brook R, Draper D. Interpreting hospital mortality data: how can we proceed? JAMA. 1988;260:3625-3628.

[15] Rosenthal G. Potential for bias in severity adjusted hospital outcomes data: an analysis of patients with rheumatic disease. J Rheumatol. 1994;21 :721-727.

[16] Luft H, Romano P. Chance, continuity, and change in hospital mortality rates: coronary artery bypass graft patients in California hospitals, 1983 to 1989. JAMA 1993;270:331-337.

[17] Blumberg M. Biased estimates of acute myocardial infarction mortality using Medis

Groups admission severity groups. JAMA. 1991;265:2965-2970.

[18] Thomas J, Holloway J, Guire K. Validating risk-adjusted mortality as an indicator for quality of care. Inquiry. 1993;30:6-22.

[19] Kahn K, Rogers W, Rubenstein L, et al. Measuring quality of care with explicit process criteria before and after implementation of the DRG-based prospective payment system. JAMA 1990;264:1969-1973.

[20] Rubenstein L, Kahn K, Reinisch E, et al. Changes in quality of care for five diseases measured by implicit review, 1981 to 1986. JAMA1990;264:1974-1979.

[21] Rosenthal G. Weak Associations between Hospital Mortality Rates for Individual Diagnoses: Implications for Profiling Hospital Quality. AmJ of Public Health 1997; 87:429.

[22] Chang C, Steinberg S. Influence of hospice use on hospital inpatient mortality. A state-level analysis. Accessed at http://healthecon.memphis.edu/Documents/Other%20Studies/Hospital_Mortality_and_Hospice_Chang_and_Steinberg_1_22_2005.pdf.

[23] Heijink R, Koolman X, Pieter D, et al. Measuring and explaining mortality in Dutch hospitals; The Hospital Standardized Mortality Rate between 2003 and 2005. BMC Health Services Research 2008, 8:73

[24] O'Leary D, Hardwick R, Cosford E, Knox A. Does hospital mortality rate reflect quality of care on a surgical unit? Ann R Coll Surg Engl 1997; 79: 46-48

[25] McKee M. Routine data: a resource for clinical audit? Quality in Health Care 1993;2:104-1 11

[26] Brook R, Iezzoni L, Jencks S et al. Symposium: case-mix measurement and assessing quality of hospital care. Health Care Financ Rev. 1987 Dec;Spec No:39-48.

[27] Weir N, Dennis M. Towards a National System for Monitoring the Quality of Hospital-Based Stroke Services. Stroke 2001;32:1415-1421

[28] Shojania D, Forster A. Hospital mortality: when failure is not a good measure of success. CMAJ 2008;179:153-157.

[29] Jarman B, Gault S, Alves B, et al. Explaining differences in English hospital death rates using routinely collected data BMJ 1999;318:1515-1520

[30] Thomas W. Accuracy of Risk-Adjusted Mortality Rate As a Measure of Hospital Quality of Care. Med Care. 1999;37:83-92.

[31] Guru V, Tu J, Etchells E, et al. Relationship Between Preventability of Death After Coronary Artery Bypass Graft Surgery and All-Cause Risk-Adjusted Mortality Rates

Circulation. 2008;117:2969-2976

[32] Frederic S. Resnic, MD, MSC, Frederick G. P. Welt. The Public Health Hazards of Risk Avoidance Associated With Public Reporting of Risk-Adjusted Outcomes in Coronary Intervention. J Am Coll Cardiol 2009;53:825–30.