Journal of Health Policy, Insurance and Management

No VII/I

Editorial Scientific Committee:
Helmut Brand (NL)
Maastricht University

Tatiana Gruzeva (Ukr)
National Medical University of Kiev

Paweł Górski (PL)
Medical University of Lodz

Claudia Habl (A)
Austrian National Institute of Healthcare (ÖBIG)

Romuald Holly (PL)
Medical University of Lodz, Insurance Institute in Poland

Yves Jorens (BE)
Ghent University

Thomas Massey (USA)
Clark University

Kai Michelsen (NL)
Maastricht University

Anna Staszewska (PL)
Insurance Institute in Poland

Yves Jorens (BE)
Ghent University

Thomas Massey (USA)
Clark University

Kai Michelsen (NL)
Maastricht University

Anna Staszewska (PL)
Insurance Institute in Poland

Yves Jorens (BE)
Ghent University

Richard Maquire (USA)
University of Massachusetts – Medical School

Kai Michelsen (NL)
Maastricht University

William Mosakowski (USA)
Public Consulting Group

Jacques Scheres (NL)
Maastricht University

Lukasz Sulkowski (PL)
Academy of Management

Journal of Health Policy, Insurance and Management

No VII/I

Editorial Board:
Romuald Holly (PL)
(editor in chief of the Journal)
Maastricht University

Thomas Massey (USA)
Clark University

Kai Michelsen (NL)
Maastricht University

Anna Staszewska (PL)
(assistant editor of the Journal)
Insurance Institute in Poland

Romuald Holly (PL)
Medical University of Lodz, Insurance Institute in Poland

Yves Jorens (BE)
Ghent University

Richard Maquire (USA)
University of Massachusetts – Medical School

Thomas Massey (USA)
Clark University

Kai Michelsen (NL)
Maastricht University

William Mosakowski (USA)
Public Consulting Group

Jacques Scheres (NL)
Maastricht University

Lukasz Sulkowski (PL)
Academy of Management

Authors of the issue

Journal of Health Policy, Insurance and Management No VII/I

Bertko John – FSA, MAAA,
Center for Consumer Information and Insurance Oversight, Centers for Medicare & Medicaid Services, USA

Bridges John E.P. – PhD,
Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

Drahik Anna – Assistant Professor at the Institute of Health Economics and Clinical Epidemiology, University Hospital of Cologne, Germany

Georgieva Lidia – MD, PhD, MPhil,
Associate Professor, Department of Social Medicine and Healthcare, Medical University, Sofia, Bulgaria

Głogowski Cezary – MD, MSE, MBA,
Director of Insurance Institute in Poland

Holly Romuald – Professor of Medical University of Łódź (Head of Health Care Policy Chair) and Warsaw School of Economics, Director of Insurance Institute in Poland

Hristov Nikolai – MD, PhD,
Medical University, Sofia, Bulgaria

Ilchev Petko – MD, PhD, MPH,
Assistant Professor, Head of Department of Social Medicine and Healthcare, Medical University, Sofia, Bulgaria

Jorens Yves – Professor of Social Security Law, Faculty of Law of the Ghent and Antwerp University in Belgium

Kardas Przemysław – MD, PhD,
Head of the First Department of Family Medicine, Medical University of Łódź, Poland

Marczak Michał – Professor of Medical University of Łódź (Health Care Policy Chair), Poland

Lewek Paweł – MD,
Assistant at the First Department of Family Medicine, Medical University of Łódź, Poland

Lüngen Markus – PD Dr. rer. pol., Institute of Health Economics and Clinical Epidemiology, University Hospital of Cologne, Germany

Michelsen Kai – PhD,
Assistant Professor, Faculty of Health, Medicine and Life Sciences, Department of International Health, Maastricht University, Netherlands

Moskalenko Vitaly Fedorovych – Rector of Bogomolets National Medical University, Doctor of Medical Sciences, Member of the National Academy of Medical Science of Ukraine, Professor of National O.O. Bogomolets Medical University, Kiev, Ukraine

Salchev Petko – MD, PhD, MPH,
Associate Professor, Head of Department of Social Medicine and Healthcare, Medical University, Sofia, Bulgaria

Scheres Jacques – Professor of Public Health Genomics, National Institute of Public Health, Warsaw, Poland, Coordinator Euregional Care, University Hospital Maastricht azM, the Netherlands, Deputy Chair European Centre for Disease Prevention and Control, ECDC, Stockholm, Sweden

Sierocka Aleksandra – PhD,
Medical University of Łódź (Health Care Policy Chair), Poland

Skibińska Edyta – PhD student, Assistant at the Medical University of Łódź (Health Care Policy Chair), Poland

Stock Stephanie – MD, PhD,
Institute for Health Economics and Clinical Epidemiology, University Hospital of Cologne, Germany

Sułkowska Joanna – MD, PhD,
Department of Public Health, Academy of Management, Poland

Sułkowski Łukasz – Professor of Economics, Prorector of Academy of Management, Poland

Szczepaniak Magdalena – PhD student at the Medical University of Łódź, Poland

Tsai Alexander C. – PhD,
Robert Wood Johnson Health and Society Scholars Program, Harvard University, Cambridge, USA

Więckowska Barbara – PhD,
Warsaw School of Economics, Social Insurance Chair, Poland
Risk Adjustment in Health Care Systems: Opportunities and Challenges
Table of contents

Foreword ................................................................. 5

I Risk Adjustment in Health Care Systems
– idea, theory and methods of application

Romuald Holly, Medical University of Łódź and Warsaw School of Economics, Insurance Institute in Poland
Kai Michelsen, Maastricht University (Netherlands)

Risk Adjustment as a principle and a tool of the social change
in Health Care System: from “passive” to “active” adjustment .......... 7

Cezary Glogowski, Insurance Institute in Poland

The concepts and methods of Risk Adjustment
in selected European countries ............................................. 23

Alexander C. Tsai, Harvard University (USA)
John F. P. Bridges, Johns Hopkins Bloomberg School of Public Health (USA)

Statistical and economic Risk Adjustment methods
for measuring the quality of hospitals ........................................ 45

Anna Drabik, Markus Lüngen, Stephanie Stock, University Hospital of Cologne (Germany)

Risk Adjustment in the German health insurance system
– Does the risk compensation system leave incentives
for risk selection? ........................................................... 63

Petre Iltchev, Medical University of Łódź (Poland)

Application of data warehouse for Risk Adjustment algorithms ......... 77

Barbara Więckowska, Warsaw School of Economics (Poland)

Effectiveness of health care systems on the competitive
health care market ............................................................ 93

Petko Salchev, Nikolai Hristov, Lidia Georgieva, Medical University of Sofia (Bulgaria)

Possible approaches to benchmarking
voluntary health insurance funds in Bulgaria ......................... 101
Romuald Holly, Magdalena Szczepaniak, Medical University of Łódź (Poland)
The development of the private health insurance and medical services market in Poland in the years 2005-2010 ............... 111

Vitaly F. Moskalenko, National Academy of Medical Science and Bogomolets National Medical University of Kiev (Ukraine)
Conceptual approaches to formation of new preventive strategy in public health ....................................................... 121

Research reports

John Bertko, Center for Consumer Information and Insurance Oversight, CMS (USA)
Current and future use of Risk Adjustment in the U.S. ......................... 135

Łukasz Sulkowski, Joanna Sulkowska, Academy of Management SWSiZ (Poland)
Cultural dualism of Polish hospitals ................................................. 139

Paweł Lewek, Przemysław Kardas, Medical University of Łódź (Poland)
Polish doctors’, pharmacists’ and patients’ opinions about generic drugs – results of a survey ........................................ 145

Aleksandra Sierocka, Michał Marczak, Medical University of Łódź (Poland)
Risk management of adverse events in hospitals (case study) ............... 153

Edyta Skibińska, Medical University of Łódź (Poland)
Design of health insurance products as a prerequisite for competitive advantage in the health care market ......................... 163

Instructions for manuscript preparation

Authors of the issue № VIII/I
Foreword

The *Journal of Health Policy, Insurance and Management* is the result of an initiative and venture undertaken by a semi-formal consortium of researchers from Poland, the Netherlands, Belgium, Austria, Ukraine, and the USA. It focuses on the social and economic consequences of civilizational diseases and related priority issues that present a challenge contemporary health care systems. This journal and its formula have been established on the basis of editorial experience gained from publishing the Polish periodical *Polityka Zdrowotna*, which has dealt with the most important and urgent problems of health policy, health insurance, and health management since 2004.

All over the world, health care systems come under great pressure. Confronted with an increasing demand of health care, they do not have resources that would even remotely suffice to meet the growing needs of our ageing populations. On the one hand, it is true that there is still substantial room to be gained by increasing the efficiency of the existing systems of health care provision, which has been proven by myriads of successful innovative projects in primary and hospital care. Especially the application of eHealth and telemedicine is promising, as are prevention and the promotion of a healthy lifestyle. Some positive and mitigating effects will also be brought by other systemic improvements such as the introduction of market mechanisms to create financial stimuli for all players in the field, including the patient’s own choices and responsibility. It may, however, be doubted whether these policies and strategies will yield enough savings to close the budget gap which is expected to occur within two decades, and which could lead to the bankruptcy of many health care systems. Even as we write this foreword, the Hungarian Presidency of the European Union summons European health ministers to Budapest for a conference on a health care system for the future, initiating a quest for a way to cope with the unprecedented and challenging growth in health care demand in the 21st century. Out of the same concern, the Polish Presidency of the EU in the second half of 2011 will give priority to “emphasiz[ing] the improvement of public health, with particular reference to the narrowing of differences in the health situation of EU countries and countries of the Eastern Partnership. A review of the health policies of the EU states will be a significant goal of the Presidency.” Needless to say, such an ambitious goal cannot be realized without a critical analysis of the economic and financial incentives of the various public health and health care systems.

The Europeanisation of health care should be looked at with great attention but is also confronted with a certain ambiguity. While health care is on the one hand an economic market that should be scrutinized under the internal market rules, on the other hand the European Union also recognises the fundamental right to health endorsing the fundamental goal of access to necessary health care for the whole population, irrespective of individual health and financial status.

The first international issue of the journal focuses on the idea and methods of risk adjustment, which is one of the most promising ways of developing and improving contemporary health care systems in both the “old” EU Member States (like Germany, the Netherlands, France, and Belgium) and the “new” ones (like Poland, the Czech Republic, Hungary, and Bulgaria). In all of these countries health care efficiency depends on the same health-based risk drivers and risk adjustments, although their role, importance, hierarchy, and interactions in each of the countries are different. That is why we must look for different algorithms for using RA methods in each individual EU country.

In parallel, we should be looking for an “algorithm of algorithms” as an overriding rule and procedure which will allow the implementation of the ideas of cooperation, harmonization and convergence of European health care systems. Taking into consideration this background, the Editors of the international *Journal of Health Policy, Insurance and Management* have decided to release a special issue on risk adjustment theory and methods of application.
An unequivocal and generally accepted definition of what risk adjustment exactly means is still lacking. Nevertheless, it is becoming increasingly apparent that risk adjustment represents an instrument which can substantially contribute to ensuring a more efficient use of the available budget, maintaining solidarity in the system, and guaranteeing universal access to high-quality health care. Experiences from pilot countries such as the Netherlands and Germany, which have opted for explicit (regulated) market principles in their systems, have indeed shown that this is the case. Risk adjustment is also a necessary element for the creation of a level playing field in the health insurance and provider market. The reduction of “the risk of risk selection” by insurers and providers is another aspect where risk adjustment is helpful. In more centrally financed health care systems with less pronounced market elements, risk adjustment can also be used to facilitate a rationalized and balanced (regional) budget distribution in order to meet the demands of the respective populations.

At the same time, the experiences of several pilot countries have made it clear that a vast amount of data from all players in the health care field must be collected and made timely available. This is a prerequisite for the development of validated algorithms and for the identification of useful parameters and effective “adjustors” which constitute the foundations for the correct application of risk adjustment instruments. While offering our contribution, we are encouraging new research efforts in this field (e.g., the first findings concerning competition as one of the possible crucial “adjustors” that should be included into RA algorithms brought different results from those suggested by some authors monitoring risk adjustment systems in Western European countries).

Health care is not static but, fortunately, full of progress; diagnostic, technological and pharmaceutical innovations often lead to significant, sudden impacts. Risk adjustment must therefore be constantly monitored and evaluated, and must permanently be adjusted itself. And, needless to say, the experiences of the pilot countries are especially interesting for the many countries which have not introduced risk adjustment to date, but are considering this move. Experts from such countries also share their vision in this issue.

There is still a long way to go before our health care systems may fully benefit from perfect risk adjustment instruments, if they are ever available. Even when they seem to be perfect on theoretical, mathematical, and ergonomic grounds, they will still have to be accepted by all players, and especially by the politicians and society. Will people accept being weighted on the basis of e.g. their body mass, age, family size, and healthy as well as unhealthy behaviors, especially if the latter make their lives happier? The idea and method of risk adjustment can be used in many fields and in many aspects. For a geneticist it can be very interesting to see whether, and how, risk adjustment systems will handle the increasing knowledge about a person’s genetic susceptibility to develop a certain disease, be it a rare one or a very common one, such as obesity, diabetes, a cardiovascular disorder, cancer, rheumatoid arthritis, or dementia. Lawyers practicing social law would like to assess the usefulness of risk adjustment e.g. for the implementation the new Directive on Patient Mobility, while to social politicians risk adjustment may be a practical tool for advancing e.g. the ideas of harmonization and convergence.

As said before, the right questions are still to be posed and answers found, and this issue of the journal is hoped to promote the debate.

Please feel invited to join in and contribute.

Romuald Holly, Yves Jorens, Jacques Scheres
ABSTRACT

The past two decades have witnessed an expansion in efforts to publicly disseminate data on hospital performance based on comparisons of risk-adjusted outcomes for the purpose of affecting reimbursement or patient choice. While much is known about which risk factors should be adjusted for, less is known about the appropriate statistical methods that should be used in deriving such quality measures.

We discuss the literature on profiling and risk adjustment, with an emphasis on recent econometric and statistical methods, highlighting key assumptions involved in the various analytical techniques. Particularly problematic for the traditional methods of analysis are inadequate sample sizes and unobserved severity of illness. We illustrate how these issues affected recent public profiling initiatives and highlight how recent contributions from the econometric and statistical literature may be helpful in ameliorating these problems.

Keywords: risk adjustment methods, health care quality

INTRODUCTION

Measuring health care quality efficiently and unbiasedly has been driven by different concerns in various health care settings, such as facilitating health services research, identifying exemplary providers as “best practice” models, and motivating internal quality improvement initiatives. The information obtained through these processes can be used to create “provider profiles” – also referred to as “report cards”, “consumer reports” or “quality measures” – that can in turn be used to identify low-quality outliers, assist policymakers with “pay for performance” schemes, or guide consumer decision-making in competitive markets. Examples of these profiling initiatives abound. Profiling initiatives face four major tasks: to gather data accurately, to analyze the data appropriately, to disseminate the reports effectively, and to highlight their meaning clinically. The practice of making comparisons of health care quality across hospitals requires meticulous adjustments to account for their varying patient populations, and this can be an extremely hazardous undertaking if not performed well. We therefore undertook this literature synthesis to focus on the appropriate analysis of data obtained from profiling initiatives. This paper will be of particular interest to health services researchers and medical professionals interested in methodological contributions from other fields.

In this literature review, we first present a conceptual model to guide our introduction of the material. Second, we discuss some of the traditional methods used to construct quality measures for report cards. We then identify two central limitations of these traditional methods: small sample size and biased estimation. These problems may be corrected with the use of hierarchical models and instrumental variables estimation. Fourth, we identify useful features of the newer methods that may be particularly helpful in policymaking contexts: monitoring quality over time and generating policy-relevant statistics. We conclude with some critical observations on the literature and practical suggestions for researchers.

NEW CONTRIBUTION

Previous reviews of the profiling literature have emphasized the dissemination of reports. Other reviews have emphasized data collection and model selection, which is a related field of inquiry, as the development of appropriate risk adjustment models is a necessary prerequisite for the development of comparative performance measures. Our aim is to discuss the literature on comparative performance...
measures—with a specific emphasis on hospitals—identifying the differences, strengths, and weaknesses of these methods. A distinguishing characteristic of our review is that we draw from a comprehensive range of literature of potential interest to health services researchers, including methodological refinements from the fields of medicine, econometrics, and biostatistics.

Second, by explicitly highlighting the assumptions involved, we identify methodological issues that must be addressed by health services researchers doing work in this area. Third, we emphasize gaps in the current literature and suggest where empirical and theoretical advances can be made.

LITERATURE SEARCH

Although our objective was to critically review the hospital profiling literature, the methods employed in this area of research are fairly general and can be applied broadly. We therefore adopted several strategies to identify peer-reviewed, English-language studies related to provider profiling methodology and provider profiling initiatives based in the United States. All searches were conducted by a single author (A.C.T.). Our strategy included searches of the PubMed electronic database with the following Medical Subject Heading (MeSH) terms: report cards, public performance reports, provider profiling, and consumer reports. From these search results, we branched out by using the PubMed “Related Articles” feature, which retrieves similar articles (based on a comparison of words from the title, abstract, and MeSH terms) using a word-weighted algorithm. We searched the EconLit electronic database using the same keywords. Finally, we identified additional articles—e.g., those published as book chapters or as conference proceedings—by hand-searching the reference lists of all identified articles.

Figure 1. A Simplified Conceptual Framework of the Relationship between Patient Outcomes and Three Explanatory Sources of Variance

CONCEPTUAL FRAMEWORK

The Institute of Medicine defined quality as “the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge” (Institute of Medicine 1990) (p. 21), suggesting a continuum of quality, clinical standards against which care can be compared, and multiple perspectives from which quality can be evaluated. Classically, the ways in which quality has been operationalized in the health services research literature generally have some relation to Donabedian’s focus on the structures, processes, and outcomes of care, with structures hypothesized to influence processes, and processes in turn hypothesized to influence outcomes. Although health care quality is multidimensional, due to the analytical complexity most empirical work has concentrated on single measures of structure, process, or outcome chosen on a priori grounds. For example, some studies have attempted to validate the identification of statistical mortality outliers by determining health-improving service inputs and then assessing the extent to which the statistical mortality outliers deliver these services.

Variation in outcomes of care is attributable to three sources20,21 as depicted in the Figure 1, and typically modeled using a regression framework with mortality as the dependent variable. Analysts must accurately characterize these three sources of variation so as to hold providers accountable only for those factors directly under their control.

1. Systematic variance consists of clinical and non-clinical factors that can be measured and included in the regression model. It is typically attributed to patient clinical factors, i.e., observed severity of illness. Hospitals that attract sicker patients would be expected to have greater mortality rates than hospitals that attract healthier patients, so mortality comparisons between hospitals require meticulous adjustments for severity of illness22 and, potentially, other patient characteristics, e.g., socioeconomic status.23 Risk adjustment permits estimates
of the mortality that would be expected of a hospital given its patient population and other characteristics that might be presumed to affect outcomes. Processes of care and non-clinical factors24 such as hospital characteristics25 may also explain some of the variation in outcomes.

2. Random variance may contribute to unfavorable (favorable) outcomes even if appropriate (inappropriate) care is administered. The model is therefore not deterministic. When the dependent variable is dichotomous – e.g., equal to unity if the patient died within 30 days of admission, equal to zero otherwise – the random error term is typically assumed to follow a binomial process.

3. Valid variance, the remaining variation in outcomes not explained by systematic variance or random error, is typically attributed to true differences in hospital quality. However, accounting for observed severity of illness may be inadequate for risk adjustment if substantial unobserved severity of illness contributes to variations in patient outcomes.

Implicitly, these potential sources of variance are assumed to be additive. Our review of the literature revealed no studies that assumed a non-linear relationship between the sources of variation.

TRADITIONAL APPROACHES TO RISK ADJUSTMENT

Early work26-28 accomplished risk adjustment by fitting a least squares regression model to hospital-level data, with the hospital’s mortality rate as the dependent variable. The use of patient-level discharge data is now the norm. A linear probability model can be fit to the patient-level data, with mortality as the dependent variable. Since this procedure frequently generates predicted probabilities that lie outside of the unit interval, the logit model is typically used. The mean predicted deaths for a particular hospital, then, is the estimated deaths that would be expected had that hospital’s mortality experience with its patients been comparable to the entire sample’s mortality experience with similar patients.29

Traditional quality assessments based on mortality determine the extent to which a hospital’s observed deaths differ from its expected deaths. According to this view, hospital quality is equal to patient outcomes (“observed”) minus systematic variance (“expected”) (see Figure 1). In other words, hospitals that have fewer deaths than expected are deemed to be of higher quality than hospitals that have more deaths than expected. The early literature focused on constructing either absolute (i.e., observed deaths minus expected deaths) or relative (i.e., observed deaths divided by expected deaths) measures of quality.

The absolute difference between observed and expected deaths can be standardized, that is, converted into a z-score that measures the difference in units of the standard deviation. The z-score then permits one to determine the probability of the observed number of deaths (or more) by chance alone.30 Health services researchers frequently assume that the quantity of predicted deaths is known without error – a reasonable assumption since it is usually generated by fitting a logit regression model to a data set far larger than the sample size at a particular hospital and its estimation error is likely to be small relative to the random error about the number of observed deaths. Given this assumption, the normal approximation to the binomial distribution can be used to estimate the variance in observed deaths, provided that the number of expected deaths is five or more.31 This particular z-score approach is commonly used in the literature on profiling.32-37 More appropriate variance formulas that account for the correlation between observed and expected deaths have been proposed but are rarely used.38 Other studies use this z-score but also employ a continuity correction factor, which adjusts for the fact that the continuous normal distribution is being used to approximate the discrete binomial distribution.39,40 When the number of expected deaths is less than five, the normal approximation can yield substantially biased results, and simulation is recommended.41

A second approach to computing z-scores, based on the Poisson approximation, implies a probability calculation that depends only on the expected number of deaths.52 It has been shown that the Poisson approximation yields a more conservative z-score than the normal approximation but that the two formulations are very similar when the mortality rate is low.43 Furthermore, as with the normal approximation, the Poisson approximation yields biased results when the number of expected deaths is less than five.44 A third approach has been to assume heterogeneous rather than homogeneous binomial processes45, but this method has been shown to yield less conservative z-scores as the variance in the estimated probabilities of death increases.46
The result of dividing the number of observed deaths by the number of expected deaths is known as the standardized mortality ratio (SMR). The SMR is frequently used in studies of hospital profiling. Some studies multiply the SMR by the observed mortality rate for the pooled sample to obtain a “risk-adjusted mortality rate,” interpretable as the mortality rate that would have been observed at that hospital if its patient population had been similar to that of the average hospital in the sample. A 95% confidence interval for the SMR can be computed by applying the normal approximation to the binomial distribution or by using the bootstrap method. Prior work suggests that the parametric method yields variance estimates with good coverage probabilities. Notably, although choosing between absolute and relative measures of quality involves invoking strong assumptions about whether the effect of quality is additive or multiplicative with respect to patient risk, this area of research has received little attention in the literature. In the absence of theoretical guidance, the choice between the two measures should depend on previous empirical findings about how quality interacts with patient risk. For example, one study noted that the greatest variation in quality between hospitals occurred for low-risk (as opposed to medium- and high-risk) patients. Ultimately, the choice between absolute and relative measures of quality may depend on the objective of the analysis. For example, if the objective is to identify low quality statistical outliers for some kind of punitive assessment, prudence would be warranted. A conservative strategy would therefore be to compute both absolute and relative quality measures and label a particular provider as a statistical outlier only if so identified by both measures.

**OVERCOMING PROBLEMS WITH SMALL SAMPLE SIZE**

These traditional quality measures, which are based on conventional regression modeling, have been shown to have a limited ability to capture much of the systematic component of variation in patient mortality. This problem is magnified especially when profiling smaller hospitals or even individual physicians in light of previously published work showing that random processes account for a substantial portion of variation. Thus, defining hospital quality as “outcomes minus systematic variance” (see Figure 1) becomes extremely problematic when dealing with small sample sizes. Combining data on outcomes other than mortality and combining data on outcomes across different conditions have been suggested as ways to improve the power of a study without increasing sample size. Most studies have not done so in a systematic fashion. Silber, Rosenbaum, and Ross analyzed rates of in-hospital deaths, serious complications, and in-hospital deaths following serious complications but considered these outcomes independently of one another.

The use of hierarchical models addresses many concerns with small sample sizes. Hierarchical modeling is so named because it deals with data that feature more than one level of random variation. Conventional regression models, such as those discussed previously, are a special case in which there is only one level of random variation, i.e., between individual patients. Hierarchical modeling can permit, for example, hospital-to-hospital variation in the overall mortality rate but constrain the effect of patient clinical characteristics on patient mortality to be the same in all hospitals, i.e., random effects. Alternatively, the effect of patient clinical characteristics can be permitted to vary across hospitals, i.e., random slopes. The key underlying assumption, known as the assumption of “exchangeability,” is that the patient- and hospital-specific intercepts and slopes can be regarded as drawn from a normally distributed population of patient- and hospital-specific intercepts and slopes. Provider characteristics can also be included in the model, although one study found that nearly all of the variation in outcomes was fitted using patient characteristics only. Furthermore, whether the effect of patient clinical characteristics on mortality is truly exchangeable across institutions deserves careful consideration.

Bayesian methods are used to estimate the hospital profiles. In particular, a “shrinkage factor” is applied to each hospital’s maximum likelihood-estimated profile. In essence, the shrinkage factor weights the hospital profile by its reliability and is larger when the data provide little information about a particular hospital. The lower the reliability, e.g., if the hospital has a small procedure volume or if there is a lot of random noise, the more the hospital profile is shrunk towards the overall mean, thereby removing regression-to-the-mean bias. Thus, small-volume hospitals need not
be dropped from the analysis, a practice that is common in profiling analyses based on conventional regression models.

In practice, hierarchical modeling has been found to be more conservative in identifying quality outliers, although in other studies hierarchical modeling has been found to yield estimates largely in agreement with analyses based on conventional regression models.

OVERCOMING PROBLEMS WITH OMITTED VARIABLES

A common assumption is that the risk adjustment model accounts for the entire systematic component of variation in mortality. This assumption is necessary because of the possibility that severely ill patients will be referred to higher-quality providers, an example of “selection bias”. If the assumption is true, and the component of variation attributable to random error can be satisfactorily characterized, then the remaining component of variation in mortality can be attributed to legitimate differences in hospital quality. However, if there is unobserved severity of illness, it will be folded into the error term, thus introducing correlation between the systematic variance and the random error term. In the language of the Figure, the “random variance” will then be not so random.

The response of the medical literature has largely consisted of efforts to collect as much detailed clinical data (“systematic variance,” see Figure) about the patient as possible, so as to adjust for observed differences in severity – a prerequisite for developing comparative measures of performance. Because reviewing patient charts to construct large, condition-specific databases containing clinical information is an extremely expensive undertaking, there still remains considerable interest in using administrative data – which are typically generated automatically for billing purposes – for risk adjustment. The weaknesses of administrative data have been extensively documented and catalogued. Many studies have tried to assess whether administrative data perform “well enough,” potentially for profiling purposes, compared to clinical data. These studies have generated conflicting conclusions. Some conclude that risk adjustment models derived from administrative data perform reasonably well compared to data obtained through detailed chart review, while others contend that claims data are inadequate.

Of note, neither administrative nor clinical data are capable of accounting for unobserved heterogeneity or unmeasurable differences in severity. Even the most rigorous clinical data collection efforts have been criticized for not accounting for all systematic differences in patient severity, thereby penalizing providers for accepting more severely ill patients.

In the absence of a randomization experiment – in which patients are randomly allocated to hospitals, thus ensuring that unobserved severity will, on average, be evenly balanced across hospitals – this problem must be dealt with by confidently asserting that the risk adjustment model is adequate, examining health conditions for which the problem of selection bias is likely to be minimized, demonstrating that the data do not suffer from the problem of selection bias, or using econometric methods to make the necessary adjustments.

A first approach to accounting for unobserved variables is the use of fixed effects regression modeling, i.e., including a dummy variable for each hospital when fitting a regression model to the data. These hospital-specific effects sweep out all unobserved variables that are invariant across patients, so the effect of hospital-specific characteristics (such as procedure volume or revascularization capability) on patient mortality cannot be estimated. Any unobserved severity will be folded into the error term. A significant weakness of the fixed effects approach is that the fixed effects themselves may be endogenously determined. For example, if sicker patients are referred to higher quality hospitals, and patient severity of illness is not completely accounted for with the administrative and/or clinical variables included in the risk adjustment model, then the hospital-specific effects will be inconsistently estimated.

Fortunately, the parameters of interest can be consistently estimated using the method of instrumental variables – an econometric method that has been used with some frequency in the econometrics literature but less frequently in health services research.

The instrumental variables strategy hinges upon identifying a so-called “instrumental variable” (IV) that is correlated with the endogenously determined variable but uncorrelated with the error term. For example, IV estimation has been used to assess the effect of cardiac catheterization on mortality after hospital admission for AMI. Randomizing patientsto receive catheterization would have been unethical, and a conventional regression
analysis would have yielded biased estimates because the decision to pursue catheterization is closely related to the patient’s severity of illness (not all of which could be captured with the administrative data available to the investigators). The instrument used by McClellan, McNeil, and Newhouse was the incremental distance to the nearest hospital with cardiac catheterization facilities beyond the distance to the nearest hospital without cardiac catheterization facilities. The intuition here is that differential distance is correlated with hospital “choice” but is uncorrelated with patient severity of illness (i.e., because patients with AMI tend to be admitted to the nearest hospital).

Recently, investigators have IV estimation for comparing hospital performance. Gowrisankaran and Town treated mortality as a continuous variable and directly applied linear IV methods, producing estimates of hospital quality that differed substantially from those obtained by fitting a fixed effects linear probability model to the data. Of the 51 hospitals identified as quality outliers using conventional regression modeling, only 19 were identified as quality outliers using IV estimation. Geweke, Gowrisankaran, and Town extended these methods to develop non-linear internally consistent models of hospital quality and choice. They described hospital choice using a multinomial probit model and mortality with a binary probit model, and their selection model permitted correlation between the two equations. Bayesian inference was used to estimate the selection model, but a supercomputer and several days of computing were required to obtain the results.

In this specific context, distance between the patient and a given hospital is considered to be a good instrument, because it is correlated with hospital “choice” but assumed to be uncorrelated with the unobserved component of illness severity. In empirical studies, linear distance has been found to be correlated with hospital choice, thus validating the first assumption. Little research has been undertaken to explore the robustness of the second critical assumption, although it cannot be examined directly.

**MONITORING QUALITY OVER TIME**

Less attention has been paid to characterizing provider performance over time. Several studies have documented substantial variation in outcomes for individual providers across years. Even if hospital quality is stable over time, observed mortality rates will still vary substantially due to year-to-year random processes (see Figure 1), thereby obscuring the true correlations in performance across years. These problems are amplified substantially when considering the year-to-year performance of smaller-sized hospitals or individual physicians. One critical review of the New York State CSRS noted that “Changes in rank were so extensive that in one year 46% of the surgeons had moved from one half of the ranked list to the other.” Increased stability can be had by simply aggregating data for hospitals across years. Doing so, however, can have the effect of suggesting a stability in outcomes that may not truly exist. Moreover, profiles based on two-year-old data may not reflect current performance and therefore may be irrelevant for policymaking.

To address these shortcomings, some researchers have incorporated both cross-sectional and longitudinal information to make inferences about provider quality. In one study, a conventional logit model was used to generate O/E measures for 43 hospitals based on 11 consecutive 6-month periods; then a Poisson regression model was fit to the data with the hospital as the unit of analysis. Bronskill used hierarchical modeling to identify “multivariate outliers,” i.e., providers that deviate from the sample mean along more than one dimension of performance, relying on a quality measure based on the squared Mahalanobis distance. In both of these studies, a provider could be classified as a quality outlier based on its mortality rate as well as its trend. Thus, a provider judged to have a non-statistically significant deviation from the mean quality profile might nevertheless be flagged as warranting closer scrutiny if the data also indicated a downward-sloping trend. McClellan and Staiger developed a method that resembles the hierarchical Bayes approach but is less computationally intensive. They used a vector auto-regression (VAR) framework to decompose data on hospital mortality into “signal variance” (variation due to legitimate differences in hospital quality), persistence in the signal over time, and random error. These
estimated variance components were then used to construct “filtered” estimates that exploited the longitudinal information to optimally estimate a hospital’s quality in a given year. The “filtered” estimates also incorporated the previously discussed “shrinkage” concept by weighting information differently depending on hospital size, the estimated signal variance, and the degree of persistence over time. Because their method relies on fixed effects, however, it is still subject to the criticism, discussed previously, that the fixed effects may be endogenously determined.

Newer time-series hierarchical models have combined many of these innovations. Using data from the U.S. Department of Veterans Affairs, West et al. studied patients returning for outpatient follow-up visits within 30 days of discharge for general psychiatric, substance abuse psychiatric, and basic medical/surgical care, where low rates of return were taken to indicate low quality. They imposed an auto-regressive structure on the hospital random effects to allow hospital quality to persist over time and also permitted correlation between the three outcomes. A high degree of persistence in quality over time was noted, but they estimated low correlations (i.e., on the order of 0.1) between the three outcomes. However, this method should be especially relevant to the assessment of small providers, where collecting data on multiple outcomes can potentially make up for paucity of data on any single outcome.

GENERATING POLICY-RELEVANT STATISTICS

Rankings based on traditional quality measures discussed previously have been shown to be a particularly unreliable source of information. More informative performance indices have been proposed that rely on the ability of Bayesian hierarchical models to directly estimate the probability that a particular hospital has performed acceptably. The advantage of such an approach is that, if all of the hospitals under scrutiny are performing at a high level of quality – as in the mythical town of Lake Wobegon, “where all the children are above average” – an arbitrary number of hospitals are not automatically labeled as quality outliers (as would be the case under conventional hypothesis testing at the 0.05 significance level). For example, a hospital might be classified as a quality outlier if the probability of mortality for an “average” patient treated at that hospital exceeded some specified threshold. Hierarchical models can be used to generate a flexible range of performance indices. Interest may focus on high-risk patients instead of “average” patients, or the threshold can be set at an absolute clinical standard determined by professional consensus.

Alternatively, hierarchical models can be used to calculate the SMR for each hospital, and a hospital would be classified as a quality outlier if its SMR exceeded some specified threshold with a specified probability.

Two other approaches use the probability of mortality or the SMR to generate rankings. Under these methods, a hospital would be classified as a quality outlier only if its 95% credible interval (the Bayesian analog of a 95% confidence interval) fell entirely within the bottom or top quartile of ranks. The indices based on mortality probabilities have been found to be more conservative than the indices based on SMRs, but the thresholds for the SMR-based indices can be modified to make their corresponding indices more conservative in identifying outliers.

CONCLUSIONS

The shortcomings of traditional quality measures based on conventional regression models can be illustrated with a brief discussion of two instances in which profiling initiatives did not accomplish their intended objectives. Analyzing the methods used in these public profiling initiatives, in light of our literature review, suggests a number of areas where the traditional models can be improved. Although significant gaps remain, the newer methods can aid in avoiding unintended adverse policy consequences.

Early analyses of data from the Cleveland Health Quality Choice (CHQC) program suggested that the profiling initiative was associated with declines in in-hospital mortality. However, subsequent research demonstrated a “sicker and quicker” phenomenon, in which the location of deaths had simply shifted from in-hospital to post-hospital settings and that few bona fide changes in quality had occurred. Among a particularly vulnerable group of patients – those for whom a do-not-resuscitate order was
Cutler, Huckman, and Landrum analyzed in CABG surgery in-hospital mortality. The profiling initiative stimulated a decline in post-discharge mortality actually increased. Moreover, the CHQC program did not accomplish its intended objectives, as hospitals publicly identified as high-quality (low-quality) outliers failed to gain (lose) market share. Ultimately, the CHQC program was discontinued after a prominent hospital system withdrew its participation. Despite the extensive adjustment for severity of illness in the CHQC reports, some asserted that their patients were sicker in ways that could not be captured even with detailed clinical data.

In contrast to the findings of Baker, Cutler, Huckman, and Landrum, analyzed data from the New York State Cardiac Surgery Reporting System (CSRS) and found that hospitals identified as low quality outliers did lose relatively healthier patients to competing facilities. Other investigators concluded that the profiling initiative stimulated a decline in CABG surgery in-hospital mortality. Subsequent research challenged this conclusion, charging that the public nature of the reports motivated cardiac surgeons to avoid high-risk cases or engage in “gaming.” A significant limitation of these studies is their focus on the recipients of CABG surgery rather than on the entire population at risk for CABG surgery. A recent econometric study addressed this limitation by studying Medicare beneficiaries hospitalized for AMI, drawing on the plausible assumption that the composition of the AMI population was not affected by the introduction of report cards. This study concluded that the report cards prompted providers to engage in risk selection, resulting in higher costs for all patients, unchanged outcomes for healthy patients, and worse outcomes for sicker patients.

Overcoming problems with small sample size will continue to remain a significant problem for researchers and policymakers. Notably, one recent study based on a nationally representative sample of 994 U.S. hospitals demonstrated that few hospitals had high enough procedure-specific case volumes for increases in mortality to be detected statistically.

The problem of mistakenly identifying low quality outliers is frequently encountered when comparing smaller hospitals (or individual physicians), but this problem can be ameliorated using hierarchical statistical models. The newer VAR and hierarchical time-series models may permit aggregation of multiple performance measures per hospital (if clinically appropriate), or the aggregation of data across years. As long as the multiple performance measures are not perfectly correlated within hospitals, and/or as long as hospital performance is not perfectly correlated from year to year, these methodological refinements will help in ameliorating sample size limitations.

The issue of unobserved severity of illness may be more problematic. Although collected detailed clinical data may allay some objections to comparisons of performance, even the most meticulous adjustments for risk may be unable to satisfy some participants, as was observed in the CHQC program. IV estimation may overcome the limitations of administrative databases. However, many questions remain. Studies in the health services research literature have used distance as an instrument, but it remains an open question as to whether these studies’ results are robust to alternative instruments. Additionally, no studies have undertaken to compare IV estimation to conventional regression when more detailed clinical data are available. That is, if detailed clinical data are available for risk adjustment, is the IV correction necessary? And finally, the strong assumptions invoked by this method, e.g., that distance is uncorrelated with unobserved severity of illness, have yet to be assessed using simulation studies. Predicting performance would be of relevance in settings where the public release of quality data typically lags its collection by one to two years. Current longitudinal methods should be extended to permit prediction, so that providers expected to perform poorly in the future can be identified in the present.

In addition to focusing resources on making sure that quality data are released and published in a timely manner, “predicting” current performance on the basis of past data may be one way to amplify its relevance. The hierarchical Bayesian models discussed in this review would be particularly suitable for this purpose, given their ability to generate policy-relevant statistics of interest to policymakers concerned with prediction. While the evidence is mixed regarding whether publicly released performance measures are actually used by patients, physicians, policymakers, or other organizations, the statistics generated by hierarchical Bayesian models may be more intuitive (and therefore more useful) to potential users.

In light of our review, we conclude that the newer models should be used whenever possible. The traditional models may suffice.
for situations in which risk adjustment is adequate (i.e., selection bias is expected to be of little concern) and sample sizes are large, but these situations are likely to be rare. Even with the use of newer methodological refinements, however, we argue that the literature is not sufficiently developed to support the automated identification of quality outliers. Instead, as Luft and Hunt suggest, providers might be labeled as statistical outliers “not as a final classification, but as a yellow flag that warrants investigation”.

Although agreement on final classification of quality outliers requires consensus on how to conduct “follow-up diagnostic testing”, in some cases emphasizing hospital performance on measures of process may be more appropriate than emphasizing outcomes.

Many methodological contributions have been made to the statistical literature on measuring quality in recent years, particularly in the areas of improving the efficiency and unbiasedness of quality estimates. These methods will become increasingly helpful as health systems undergo improvements in information technology that will potentially enable the provision, in a timely fashion, of large amounts of accurately coded clinical data. In the meantime, caution is indicated. Measuring the quality of health care providers in an accurate and timely manner should be regarded as a first step towards improving the quality of health care service delivery. In this regard, the potential benefits from provider profiling are enormous, but much work remains to be done.

REFERENCES


78. Rice N, Jones A. Multilevel models and health economics, Health Econ 6, no. 6, 1997: 561-575.


This issue of the *Journal of Health Policy, Insurance and Management* has been financed by the research program on the causes and socioeconomic consequences of civilizational diseases supported by:

Krajowy Instytut Ubezpieczeń
Insurance Institute in Poland
POLAND
02-656 Warsaw
ul. Ksawerów 30/92

Faculty of Health Sciences
Health Care Policy Chair
POLAND
90-131 Łódź
ul. Lindley’a 6

GSK Commercial Sp. z o.o.
POLAND
02-697 Warsaw
ul. Rzymowskiego 53

Sanofi-Aventis Sp. z o.o.
POLAND
02-672 Warsaw
ul. Domaniewska 41
INSTRUCTIONS FOR MANUSCRIPT PREPARATION

In the *Journal of Health Policy, Insurance and Management* (*Journal*) are published:

a. Original research papers  
b. Dissertations  
c. Reports  
d. Reports from research  
e. Abstracts of postdoctoral, doctoral research programs  
f. Reports of international and national conferences  
g. Reviews  

Please prepare the proposal of the article in accordance with the following guidelines:

I. General Instructions:

   a. **Software and format** – Microsoft Word 6.0 is preferred  
   b. **Page setup and paper format** – not exceed 27,000 signs without space (max. 15 pages A4)  
   c. **Line space** – 1.5 times  
   d. **Font** – Times New Roman or Times-Roman (12 pt)

II. Data of authors, affiliation and title of the paper on the first page:

   a. Title of the paper  
   b. Name(s) and surname of the author(s) with initials  
   c. Academic/university degrees, function performed in the chair/department/institution  
   d. Affiliation/full name and address of the institution where author(s) is employed  
   e. E-mail address, telephone number of the first author

III. Abstract:

   a. Approximately 250 words  
   b. Figures, tables, and references should not be cited in the abstract  
   c. Keywords – not exceed 8 words

IV. Footnotes:

   Immediately below the text - if the number and volume occupies more than 1/3 of the page, footnotes placed at the end of the article (endnotes)

V. References:

   a. Applied uniformly, without division of the literature (e.g. books and articles)  
   b. Placed at the end of the article  
   c. All the references must be in the alphabetic order

VI. Important guidelines:

   a. The articles should be submitted through e-mail  
   b. The e-mail id for submission of articles as an attached file is: **kiu@kiu.com.pl**  
   c. **The Copyright Transfer Statement** should be sent by post to the Insurance Institute in Poland as Consortium Editor (according to the attached schema). Address for correspondence:

   **Insurance Institute in Poland**  
   **POLAND 02-656 Warsaw, ul. Ksawerów 30/92**

   d. The article may be published in the Journal provided that:
      * the paper was not published, nor has been submitted to the any other publisher  
      * the paper was approved by all coauthors  
      * author(s) transfer all copyright ownership in and relating to the article, in all forms and media, to the *Journal* in the event that the article is published. However, this agreement will be null and void if the article is not published in the *Journal*  
      * all funding sources of the research were disclosed  
      * author(s) waives fee

   Publisher is not responsible for the legal consequences arising of a breach of these principles

VII. Additional Information:

   For additional information, please contact the Editor-in-Chief: **Romuald Holly**  
   at Insurance Institute in Poland (as Consortium Editor)  
   telephone: +48 22 628 99 30, fax: +48 22 628 99 31  
   e-mail: kiu@kiu.com.pl
Journal of Health Policy, Insurance and Management

Editorial Scientific Committee:
Helmut Brand (NL)  
Maastricht University
Tatiana Gruzeva (Ukr)  
National Medical University of Kiev
Paweł Górski (PL)  
Medical University of Lodz
Claudia Habl (A)  
Austrian National Institute of Healthcare (ÖBIG)
Romuald Holly (PL)  
Medical University of Lodz, Insurance Institute in Poland
Yves Jorens (BE)  
Ghent University
Pawel Górski (PL)  
Medical University of Lodz
Claudia Habl (A)  
Austrian National Institute of Healthcare (ÖBIG)
Romuald Holly (PL)  
Medical University of Lodz, Insurance Institute in Poland
Yves Jorens (BE)  
Ghent University
Richard Maquire (USA)  
University of Massachusetts – Medical School
Thomas Massey (USA)  
Clark University
Kai Michelsen (NL)  
Maastricht University
William Mosakowski (USA)  
Public Consulting Group
Jacques Scheres (NL)  
Maastricht University
Łukasz Sułkowski (PL)  
Academy of Management

Editorial Board:
Romuald Holly (PL)  
(editor in chief of the Journal)  
Medical University of Lodz, Insurance Institute in Poland
Thomas Massey (USA)  
Clark University
Kai Michelsen (NL)  
Maastricht University
Łukasz Sułkowski (PL)  
Academy of Management
Anna Staszewska (PL)  
(secretary of the Journal)  
Insurance Institute in Poland

Authors of the issue

Bertko John – FSA, MAAA, Center for Consumer Information and Insurance Oversight, Centers for Medicare & Medicaid Services, USA

Bridges John EP – PhD, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

Drahik Anna – Assistant Professor at the Institute of Health Economics and Clinical Epidemiology, University Hospital of Cologne, Germany

Georgieva Lidia – MD, PhD, MPhil, Associate Professor, Department of Social Medicine and Healthcare, Medical University, Sofia, Bulgaria

Głogowski Cezary – MD, MSc, MBA, Insurance Institute in Poland

Holly Romuald – Professor of Medical University of Łódź (Head of Health Care Policy Chair) and Warsaw School of Economics, Director of Insurance Institute in Poland

Hristov Nikolaí – MD, PhD, Medical University, Sofia, Bulgaria

Illichev Petko – MD, PhD, Associate Professor, Head of Department of Social Medicine and Healthcare, Medical University, Sofia, Bulgaria

Kardas Przemysław – MD, PhD, Head of the First Department of Family Medicine, Medical University of Łódź, Poland

Marczak Michał – Professor of Medical University of Łódź (Health Care Policy Chair), Poland

Lewek Paweł – MD, Assistant at the First Department of Family Medicine, Medical University of Łódź, Poland

Lüngen Markus – PD Dr. rer. pol., Institute of Health Economics and Clinical Epidemiology, University Hospital of Cologne, Germany

Michelsen Kai – PhD, Assistant Professor, Faculty of Health, Medicine and Life Sciences, Department of International Health, Maastricht University, Netherlands

Moskalenko Vitaly Fedorovych – Rector of Bogomolets National Medical University, Doctor of Medical Sciences, Member of the National Academy of Medical Science of Ukraine, Professor of National O.O. Bogomolets Medical University, Kiev, Ukraine

Salchev Petko – MD, PhD, MPH, Associate Professor, Head of Department of Social Medicine and Healthcare, Medical University, Sofia, Bulgaria

Scheres Jacques – Professor of Public Health Genomics, National Institute of Public Health, Warsaw, Poland, Coordinator Euregional Care, University Hospital Maastricht azM, the Netherlands, Deputy Chair European Centre for Disease Prevention and Control, ECDC, Stockholm, Sweden

Sierocka Aleksandra – PhD, Medical University of Łódź (Health Care Policy Chair), Poland

Skibińska Edyta – PhD student, Assistant at the Medical University of Łódź (Health Care Policy Chair), Poland

Stock Stephanie – MD, PhD, Institute for Health Economics and Clinical Epidemiology, University Hospital of Cologne, Germany

Sułkowska Joanna – MD, PhD, Department of Public Health, Academy of Management, Poland

Sułkowski Łukasz – Professor of Economics, Prorector of Academy of Management, Poland

Szczepaniak Magdalena – PhD student at the Medical University of Łódź, Poland

Tsai Alexander C. – PhD, Robert Wood Johnson Health and Society Scholars Program, Harvard University, Cambridge, USA

Więckowska Barbara – PhD, Warsaw School of Economics, Social Insurance Chair, Poland
Risk Adjustment in Health Care Systems: Opportunities and Challenges

Insurance Institute in Poland

POLAND
02-656 Warsaw
ul. Ksawerów 30/92

www.kiu.pl