

## Validity of the risk adjustment approach to compare outcomes

Validade da abordagem do ajuste de risco para a comparação de resultados de serviços de saúde

Leticia Krauss Silva <sup>1</sup>

---

<sup>1</sup> Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz. Rua Leopoldo Bulhões 1480, Rio de Janeiro, RJ 21041-210, Brasil. leticiak@ensp.fiocruz.br

**Abstract** *This paper focuses on the issue of the extent to which the present mainstream risk adjustment (RA) methodology for measuring outcomes is a valid and useful tool for quality-improvement activities. The method's predictive and attributional validity are discussed, considering the confounding and effect modification produced by medical care over risk variables' effect. For this purpose, the sufficient-cause model and the counterfactual approach to effect and interaction are tentatively applied to the relationships between risk (prognostic) variables, medical technology, and quality of care. The main conclusions are that quality of care modifies the antagonistic interaction between medical technologies and risk variables, related to different types of responders, as well as the confounding of the effect of risk variables produced by related medical technologies. Thus, confounding of risk factors in the RA method, which limits the latter's predictive validity, is related to the efficacy and complexity of associated medical technologies and to the quality mix of services. Attributional validity depends on the validity of the probabilities estimated for each subgroup of risk (predictive validity) and the percentage of higher-risk patients at each service.*

**Key words** *Risk Adjustment, Outcomes, Validity*

**Resumo** *Este trabalho focaliza a questão da validade da metodologia usual de ajuste de risco para comparar resultados de serviços de saúde e de sua utilidade para implementar atividades de melhoria da qualidade. As validades preditiva e atributiva do método são discutidas, considerando o confundimento e a modificação de efeito produzidas pela atenção médica sobre o efeito das variáveis de risco. Nesse sentido, as abordagens de causa suficiente e contrafactual para efeito e interação são tentativamente aplicadas às relações entre variáveis de risco, tecnologias médicas e qualidade da assistência. As principais conclusões são de que a qualidade do cuidado modifica a interação antagonista entre tecnologias médicas e variáveis de risco, relacionada a diferentes tipos de respostas, assim como o confundimento do efeito das variáveis de risco produzido pelas tecnologias médicas associadas. Assim, o confundimento dos fatores de risco no método de ajuste de risco, que limita sua validade preditiva, estaria relacionado à eficácia e à complexidade das tecnologias médicas associadas (considerando indicação e performance e sua relação com qualidade) e ao espectro de qualidade dos serviços. A validade atributiva dependeria da validade das probabilidades estimadas para cada subgrupo de risco (validade preditiva) e da porcentagem de pacientes de alto risco de cada serviço.*

**Palavras-chave** *Validade; Ajuste de Risco; Resultados*

## Introduction

The argument that will be pursued in this paper relates to the question of the extent to which the present mainstream methodology of risk adjustment for measuring outcomes (the RA approach) is a valid and useful tool for quality-improvement activities. To help clarify the subject being examined, another related issue is considered first, related to the question of the extent to which that method can estimate effectiveness/effectiveness gap.

Estimation of effectiveness and of the effectiveness gap is fundamental for estimating the cost-effectiveness (as opposed to cost-efficacy) of alternative medical care interventions – information which is necessary to support adequate decisions regarding the dissemination and financing of medical technologies – as well as for carrying out quality-improvement activities (Banta & Luce, 1993; Brook & Lohr, 1985). This is particularly relevant for low-effectiveness (low-quality) environments and “low-effectiveness” technologies (technologies whose benefit tends to be considerably lower than expected according to efficacy findings when applied under ordinary conditions).

However, estimation of effectiveness is associated with a defined individual/set of technologies and estimation of the effectiveness gap depends on the development of an outcome standard, associated with good quality of care. Besides providing an estimate of efficacy, randomized clinical trials (RCTs) carried out at state-of-the-art services can provide an outcome standard to allow for the measurement of effectiveness (and effectiveness gap) for the conditions and groups of patients included in them, based on the rates observed in RCT-treated patients. Known major shortcomings of such an approach to estimate outcome standards are that RCTs: are limited in number and scope; restrict admission; frequently differ in relation to inclusion-exclusion criteria, intervention, and endpoints for the same technology and general condition; and are supposedly carried out by motivated staff. Meanwhile, for non-tested conditions – which correspond to the majority of patients – and for non-tested subgroups and related overall conditions, there is no consensus on how to derive valid outcome standards (Colditz et al., 1988; D’Agostino & Kwan, 1995; GAO/USA, 1992; Hlatky et al., 1988; Maklan et al., 1994; Moses, 1995; Selker et al., 1992).

On the other hand, risk assessment for comparison of outcomes could be viewed as a descriptive problem, where outcome occurrence

is related to risk (prognostic) factors, without aiming at causal interpretation of the relationship. However, the rational intervention underlying the RA approach for comparison of outcomes (quality improvement activities) demands information that can be interpreted in causal terms (lower-quality care causes or produces worse severity-adjusted outcomes than higher-quality care) (Miettinen, 1985).

To interpret differences in outcomes of medical services in causal terms, it is important that the services’ populations be comparable. Despite its limitations (Miettinen, 1985; Rothman, 1986), indirect standardization of risk factors, taking one or two risk factors into account, was commonly used until the mid-1980s to assess outcomes, as for example with neonatal mortality rates adjusted to birth weight and/or gestational age (Bowes Jr. et al., 1984; Helliher, 1977).

The present RA approach is based on multivariate models – equations whose terms represent patients’ risk categories/variables. The RA equation supposedly derives an “average” effect of each of the different levels of selected risk factors, producing a predicted probability of in-hospital death (or another outcome) for each patient. The expected deaths for each hospital are subtracted from the observed deaths, which are then transformed into a z score that is used to rank hospitals.

In a classical review of risk adjustment of health outcomes methodology, which can be seen as a form of indirect standardization (the number of expected deaths being calculated by multiple regression), Blumberg (1986) states that the interpretation of the standardized ratio for a given hospital or other sub-sample must take into consideration the *nature of the standard*: If, e.g., “*the standard is based upon a population for which care was presumed to be excellent and outcomes very good, then a value for a particular hospital of over 1, say 1.10, might still represent a favorable outcome. In contrast, if the standard is based upon rather old experience before modern improvements took place, then a value of 0.8 might reflect a poor outcome*” (Blumberg, 1986:372).

However, by the mid-1980s, the need to have a large set of observations for the standard population (development data set) in order for the prediction model to be capable of producing statistically reliable rates of expected adverse outcomes for each category of patient attributes was pressing towards the trend of using all cases in the study universe with the subject clinical care (all cases of acute myocardial infarct, for example) as the standard (Blumberg, 1986).

Current studies involving a risk-adjustment approach to compare outcomes, although generally focusing on specific clinical conditions, derive their equations from development databases consisting of patients with different admission risks and treated at services with different levels of quality and probably with different levels of technological complexity (Hartz et al., 1993; Iezzoni, 1997a; Knaus et al., 1993; Selker et al., 1991).

What would be the consequences of such a trend for both the validity of the RA method and the interpretation of its findings?

The present RA approach implies that if the expected and observed numbers of deaths in a health service are similar, the service displays average quality. Therefore, the present mainstream risk adjustment approach does not elaborate and use a standard of good quality care or an absolute standard of quality of care (Daley & Schwartz, 1997; Hannan et al., 1990; Selker et al., 1991). Thus, despite the intention of examining the effectiveness and comparing outcomes of medical care (Iezzoni, 1997b), the RA approach is not capable of assessing effectiveness.

Furthermore, the RA method does not deal with process of care (procedures/performance), which is rather treated as a black box. RA studies fail to explain either how they approach or conceive of the relationship between the selected risk (prognostic) variables at admission, which constitute the terms of the RA equation, and medical technologies, or the way they are performed. They also fail to inform how the model expresses such a relationship or how the approach captures it, in order for findings on expected/observed outcomes to be interpreted accordingly.

Since the RA method bears no explicit relationship to medical technologies, it can serve as only a blunt tool for estimating effectiveness, or effectiveness gap. On the other hand, how well does the RA approach compare outcomes?

Our argument focuses on two dimensions of validity in the RA method: (a) predictive validity, or the extent to which the method accurately predicts the probability of death (and which patients have died) and (b) attributional validity, or the extent to which the method allows one to attribute differences in mortality to quality of care (Daley, 1997; Donabedian, 1980).

## Predictive validity of the risk-adjustment approach

In his review, Blumberg (1986) considered, as did Selker (1993) and Knaus (1993), that the regression techniques used by most RA studies tend to underestimate high-risk patients' probability of death and to overestimate low-risk patients' probability of death, although they did not perform validation procedures. These observations could be associated with a lower standardized mortality ratio for low-risk admissions services than the ratio corresponding to services with a relatively high percentage of high-risk patients. Dubois et al (1987) found that the average severity score of the high outliers (services with more deaths than expected according to the RA equation) significantly exceeded the score of the low outliers. Some strategies have been devised to correct such a problem, but why should it occur? Our question thus refers to the accuracy of the (multiple) standard per risk stratum produced by the RA model and the validity of such a standard as a quality assessment tool.

Reported multiple correlation coefficients ( $R^2$ ) of RA equations, which frequently involve dozens of variables, are generally smaller than 0.30, whether cross-validated or not (Iezzoni et al., 1992, 1998; Knaus et al., 1993), higher rates being related to indices that include in-hospital risk information like APACHE III and APR-DRG. The remaining (residual) variation should then be attributed to poor adjustment (due to restriction on the number and inadequate selection and transformation of risk variables, non-inclusion of interaction terms, and data quality), random variation (which becomes small with large databases), and supposedly, variation in the process of care, including quality of care.

## Confounding

Risk-adjustment regression equations do not generally include process-of-care variables (technologies and their performance), so as supposedly not to adjust for the process of health care, thus preventing comparison of outcomes according to differences in process of care (including quality of health care) (Iezzoni, 1997; Knaus et al., 1993). However, it could be argued that if medical care is supposed to be a relevant predictor of hospital outcomes, then the RA equation should include it; otherwise, risk variables' effects could possibly be confounded by it (Miettinen, 1985).

Moreover, the patient's status both for each factor and the risk conditions as a group de-

marcates the component elements of the referents (Donabedian, 1982): the indications (and contraindications) of the respective medical procedures are therefore associated with them (i.e., with their utilization). Some such factors, either by themselves or through their proxy variables, are probable candidate terms for a related RA approach equation (Daley, 1997; Hannan et al., 1990; Selker et al., 1992).

Considering that both risk/severity variables and medical care variables have an effect on the study outcome, and that medical care variables are associated with risk factors and are not intermediate steps between risk factors and their effect, one can conclude that the effect of risk/level of risk factors in RA equations would be confounded by the effect of medical technologies if they are not terms in the RA equation (confounding by indication). By modifying medical technologies' effects, quality of care is also involved in such a confounding mechanism.

To further examine the issue of the relationships between risk (prognostic) variables, medical technology, and quality of care, it seems relevant to consider *the sufficient cause model and the counterfactual approach to effect and to interaction* (Greenland, 1993; Miettinen, 1982, 1985; Rothman, 1976). Such approaches have been applied to understand interaction types in risk and preventive mechanisms, but not to therapeutic actions. In most of the remaining paragraphs, we attempt to apply them to the problem of comparing the outcomes of medical care.

In addition to the risk variables at hospital admission, variables associated with a clinical outcome at hospital discharge (e.g., 7-day mortality rate) include the related medical technologies and the way they are utilized and performed (including their institutional determinants). A different outcome should be expected if hospital admission had not taken place for that same population, that is, if the counterfactual condition had occurred. The counterfactual condition is a reference condition, contrary to fact, against which the treatment (or exposure) will be evaluated (Rothman & Greenland, 1998).

If medical care variables are supposed to interfere with the effect of risk variables in such a way as to diminish the number of deaths corresponding to the counterfactual condition (no medical care), how does the RA approach grasp the relationship among such related groups of variables?

To develop a RA equation (and the related severity score), observed deaths at the end of

the episodes of care under study, *and not the supposed deaths had medical care not taken place (counterfactual condition)*, are considered. Accordingly, the equation estimates an expected number of deaths at the end of hospital stay (which does not correspond to the counterfactual condition). Thus the question is: how is medical care captured by the RA equation? Is such apprehension valid and useful to interpret risk adjustment findings when comparing medical care outcomes?

### Effect modification

The counterfactual approach to biological interaction deals with counterfactual response (interaction) types (that will be presented below), where the effect of one factor depends on the person's status for the other factor. The counterfactual approach is logically related to the sufficient-cause approach to interaction, where sufficient cause means a set of minimal (necessary) conditions and events that inevitably produce disease/death.

Medical care may also be analyzed as a modifier of the risk of death (at admission) through avoidance of the completion of sufficient causes of death (or morbidity) (Miettinen, 1974; Rothman & Greenland, 1998) related to the condition under study. Taking acute myocardial infarction (AMI) as an example, medical care can do this by removing certain fatal conditions such as primary ventricular fibrillation, through the use of defibrillators, or by sparing myocardial muscle through the use of drugs like beta-blockers, nitrates, and thrombolytics. Medical care therefore decreases the average risk of death associated with conditions at admission, like arrhythmia and heart failure – clinical conditions which themselves could constitute the selected risk factors or which could be represented by proxy risk factors. But how can such modification of effect be envisioned?

The relationship between medical care (or medical technologies) and the sufficient causes of death/disease (risk factors) could be conceived as antagonistic interaction (response), that is, medical care could be seen as a powerful effect modifier which renders the effect of a sufficient cause non-existent, i.e., which makes the exposure causally inoperative regarding the study outcome (Miettinen, 1974).

To the extent that the antagonistic interactions succeed in blocking the corresponding sufficient causes, they decrease the death rate, e.g., of the assisted population. Medical technologies may therefore modify risk factors' effect through successful antagonistic interac-

tions with certain sufficient causes of death/disease. In order to take such effect modifications into account, the RA equation should include the corresponding product terms, though it would not be a simple endeavor.

Since the equation does not incorporate the antagonistic interactions between medical care variables and risk-at-admission variables (organized in the main sufficient causes), rate differences associated with different levels of risk variables may thus be poorly estimated. Moreover, considering that those antagonistic interactions present, for each sufficient cause, a certain configuration and degree of specificity that may change with time and place, the use of a general severity score in a RA equation implies a blurring of such diversity of relationships and of the way the related effect modification and confounding are captured.

On the other hand, the appropriate use of medical technologies, including the corresponding indication, sequencing, dosage, and ability to perform, is very important in determining/changing the probability of benefit/success (efficacy) of a set of technologies and changing the natural history of a disease. Quality of care is thus a relevant modifier of the effect of medical technologies (efficacy vs. effectiveness). Therefore, the effect of risk-at-admission factors (without the benefit of medical care) is modified by medical technologies whose effect is in turn modified by quality of care. But how would the antagonistic interaction between medical technologies and risk factors be conceived regarding different degrees of severity?

Considering a certain outcome, the antagonism to different types of sufficient causes may be expressed through independent actions as well as through different interactions among medical technologies – in the cooperative sense, for synergistic action, as proposed by Greenland (1993) and Miettinen (1982) under the counterfactual model of interaction.

Medical care means thus either independent technological action or different kinds of more or less complex (cooperative) interactions among technologies (and the way they are performed) and antagonistic interactions with risk factors. For some patients, the available set of medical technologies hardly changes their risk at admission, i.e., they either die or survive regardless of good or inadequate medical care (i.e., they are “doomed” or they survive as their probability of death at admission is extremely high or very low). Interaction among technologies can result from different mechanisms, the effect of one factor depending on the patient’s status for the other factor:

- a procedure is effective for a patient only when another procedure or procedures are also (adequately) performed; this mechanism represents synergistic interaction;
- two (or more) procedures are effective alone or when both are present, but only one exerts its effect; this mechanism corresponds to competitive interaction;
- a procedure is effective depending on the non-utilization of another procedure or on the value of the patient for a certain risk variable (contra-indication); the related mechanism corresponds to an antagonistic response (or interaction).

In the case of medical care, synergistic interaction is usually an intended effect, since procedures are prescribed for patients with certain (referent) characteristics. Synergistic interactions correspond to synergistic responders. However, it is common for the precise demarcation of patients for whom an individual procedure is needed not to be known, since knowledge about the mechanisms of action of an individual/set of intervention(s) may be limited – even factorial designs are limited regarding the presence of synergistic responders – in addition to the difficulty, for example, in translating micro-mechanisms of action into easily diagnosed clinical characteristics (Rothman & Greenland, 1998). Besides, independent action may be sufficient regarding a certain outcome, e.g., 30-day mortality, as compared to a more ambitious result (such as 1-year mortality), where synergistic interaction may be required. In other words, the response involved in a successful antagonistic action between medical care and severity conditions varies according to the study outcome.

The competitive interaction response includes risk conditions over which knowledge about the need of each respective technology may be non-existent or inconclusive. Competitive interaction may also result from lack of information about the sufficiency of an independent action (low-quality care).

A form of antagonism between medical technologies is seen when one procedure blocks the effect of another. Another form of antagonistic interaction between medical technologies involves a treatment that is unsafe for patients with a certain value for an attribute (qualitative interaction), interfering with the effect of other potentially necessary procedures.

Antagonistic interaction between medical technologies may result in less benefit than the effect obtained from the utilization of only one of the respective technologies and, given medical knowledge, may correspond to low quality

of care. Antagonistic interaction among medical technologies may also comprehend unanticipated adverse effects. This kind of interaction among medical technologies may result in synergistic interaction between technologies and the sufficient cause.

Low-risk at admission (without considering the benefit of medical care) patients generally demand fewer interventions to reach the effect survival at discharge, e.g., than do high-risk patients. Medical care may also succeed through independent action in low-risk patients. Furthermore, low-risk patients seem to frequently reach a same given outcome through the effect of "alternative" treatments, in the sense that any among a set of suitable procedures would change their fate into hospital survival (outcome), with the other "alternative" treatments being unnecessary, considering the chosen outcome (regardless of whether the other procedures would improve other outcomes such as patient's life expectancy). In other words, low-risk patients seem to frequently present a competitive interaction response; scientific evidence in this respect is scarce.

It is interesting to note that competitive interaction here, although decreasing the effect of the respective technologies (considering the sum of individual effects), does not generally diminish the total number of salvaged patients, regarding the chosen outcome; however, it does imply a waste of resources.

On the other hand, patients under higher risk at admission frequently present more than one sufficient cause of death (morbidity), and commonly require several interventions, representing synergism, particularly those involving complex technologies (considering again both indication and performance). Also, high-risk patients are probably more subject to antagonistic interaction between medical technologies (and to the resulting interaction between medical technologies and the sufficient causes), due to the higher number of technologies involved in their care and to their side effects and contra-indications (Miettinen's law of nature).

Modification of the effect of medical technologies by the level of quality seems generally more relevant for relatively complex sets of technologies (considering both their indication and performance). Higher-quality services present much higher probability than lower-quality services of appropriately utilizing relatively complex sets of technologies, in such a way as to produce their maximum possible benefit. For less complex sets of technologies, quality seems to play a less striking role. An ex-

ception are life-saving emergency technologies, like resuscitation procedures, which may not qualify as complex, but may present frequent quality problems given their high promptness requirement.

For those reasons, synergistic responders regarding a certain outcome tend to be more affected by quality of care: good-quality services tend to produce nearly all the benefit expected from each of the respective technologies, while low-quality services tend to produce almost none of the benefit expected from the related technologies. Meanwhile, competitive (as well as independent and eventually less complex synergistic) responders, in general, probably benefit in a much less differential way from different levels of quality.

It is possible to infer from the previous paragraphs that low-quality care generally results in less confounding of risk factors' effect (counterfactual condition) than higher-quality care, especially in relation to higher-risk patients, generally due to synergistic and antagonistic interactions related to multiple/complex technologies.

Thus, for each service, the average probability of death (or morbidity) after medical care (predictive validity) for each subgroup of patients presenting specific sufficient cause of death depends on the probability of death at hospital admission without medical care (risk at admission in a counterfactual condition), on the efficacy (rate ratio) and complexity of associated medical technologies (considering their indication and performance, and their relation to quality). Now, considering a group of services, the corresponding average probability of death (morbidity) after medical care for each sufficient-cause stratum would additionally depend on the technological level mix and on the quality mix of services. The confounding of risk factors in the RA method (after medical care), which limits its predictive validity, is therefore related to the efficacy and complexity of associated medical technologies, and to the quality mix of services.

Because the RA approach for comparing outcomes does not include process-of-care variables (medical technologies, their interactions and antagonistic actions, and their complexity) and utilizes just one severity score, the shape and effect of such powerful confounders/effect-modifiers remain obscure, limiting the predictive validity of the method.

### Attributional validity of the risk-adjustment approach

Several studies have specifically addressed the attributional validity of RA methods by using different strategies, such as comparison of their findings with those of peer review and of structure and process (using explicit and implicit criteria) analyses (Dubois et al., 1987; Hannan et al., 1990; Hartz et al., 1993; Park et al., 1990; Thomas et al., 1993). Their findings generally indicate that risk-adjusted outcomes have low or limited attributional validity. In addition, studies comparing the predicted mortality derived from different methods of estimating severity found that, for all study conditions, agreement on hospital performance based on different severity measurements was low (Lezoni et al., 1995, 1996a, 1996b, 1998).

Using a different and interesting strategy, two similar simulation models were developed to measure the accuracy of mortality rates in detecting low-quality hospitals, finding low accuracy (Hofer & Hayward, 1996; Thomas & Hofer, 1999). However, the models failed to conceptualize the relationships among factors contributing to hospital mortality, not only in their assumptions and parameters, like no case mix or severity differences across hospitals or a fixed ratio between the probabilities of death of patients receiving poor and good quality care; also, models' parameters were varied independently.

According to the reasoning set forth in the previous section, attributional validity of the RA method is limited because it depends on the validity of the probabilities of death after medical care estimated for each subgroup of sufficient cause at a service mix (predictive validity), and also on the percentage of patients belonging to each subgroup, especially on the percentage of higher-risk patients (synergistic patients) at each service. Standardized mortality ratios related to services with high proportions of high-risk patients, as classified by the RA severity score, and low or average quality regarding such patients are higher than those corresponding to services with low proportions of high-risk patients and similar quality level. This shortcoming limits the attributional validity of any indirect standardization method.

Thus, the predictive validity of the above studies, including the ones that used the conventional RA approach, was limited by their assumptions and parameters, which consequently compromised their attributional validity. A different distribution of higher-risk (synergistic) patients among services may have further limited the studies' attributional validity.

To overcome the problem of differential distribution of higher-risk patients, a well-known alternative is to proceed with a stratified analysis, considering a standard for each risk stratum. Schwartz et al. (1997) assume the above limitation in the RA method and point also to stratified analysis as an alternative to the original method. However, stratified analysis resolves only partially the problem of differences in the structure of risk factors (sufficient causes), as the question related to the validity of the standard for each risk stratum (predictive validity) remains.

Although generally disregarded by RA studies, differences in hospitals' technological level and corresponding efficacy should be taken into account when evaluating the performance of hospitals per se, especially in developing countries: services with lower technological levels will present relatively higher z scores than those with higher technological levels if they assist higher risk patients.

### Conclusions

Considering that the standard per risk stratum estimated by the current RA methods represents an average, unspecified regarding sufficient cause, technology, or quality, then the analysis is limited to classifying the relative position of the services regarding such an average, whether working within a stratified analysis or with the entire risk spectrum (adjusted outcomes). Since that standard is estimated without taking into account process-of-care interaction and confounding mechanisms associated with prognostic variables, then its predictive and attributional validity are compromised.

On the other hand, for an equation to take into account process-of-care variables, considering the relations outlined above, one would have to know, among other things, about services' quality of care, which is being assessed. No easy, practical solution is foreseen in the field of comparing services' quality. One way out of such a paradox could be to base the equation and the corresponding product terms on a high quality care services' database. High-quality services could be selected from periodic, evidence-based, process-of-care reviews, which could be directed to a non-random sample of the supposedly best services. The problem in developing countries may be to find a high-quality subgroup of services, at least for most of the involved sufficient causes of death/morbidity.

## References

- BANTA, H. D. & LUCE, B. R., 1993. *Health Care Technology and its Assessment*. London: Oxford University Press.
- BLUMBERG, M. S., 1986. Risk adjusting health care outcomes: A methodological review. *Medical Care Review*, 43:351-393.
- BOWES Jr., A. W.; FRYER Jr., G. E. & ELLIS, B., 1984. The use of standardized neonatal ratios to assess the quality of perinatal care in Colorado. *American Journal of Obstetrics and Gynecology*, 148: 1067-1073.
- BROOK, R. H. & LOHR, K. N., 1985. Efficacy, effectiveness, variations, and quality: boundary-crossing research. *International Journal for Quality in Health Care*, 23:710-722.
- COLDITZ G. A.; MILLER, J. N. & MOSTELLER, F., 1988. The effect of study design on gain in evaluations of new treatments in medicine and surgery. *Drug Information Journal*, 22:343-352.
- D'AGOSTINO, R. B. & KWAN, H., 1995. Measuring effectiveness: What to expect without a randomized control group. *Medical Care*, 33:AS95-AS105.
- DALEY, J. & SHWARTZ, M., 1997. Developing risk-adjustment methods. In: *Risk Adjustment for Measuring Healthcare Outcomes* (L. I. Iezzoni, ed.), pp. 279-330, 2<sup>nd</sup> Ed. Chicago: Health Administration Press.
- DALEY, J., 1997. Validity of risk-adjustment methods. In: *Risk Adjustment for Measuring Healthcare Outcomes* (L. I. Iezzoni, ed.), pp. 331-364, 2<sup>nd</sup> Ed. Chicago: Health Administration Press.
- DONABEDIAN, A., 1980. *Explorations in Quality Assessment and Monitoring*. v. 1. Ann Arbor: Health Administration Press.
- DONABEDIAN, A., 1982. *Explorations in Quality Assessment and Monitoring*. v. 2. Ann Arbor: Health Administration Press.
- DUBOIS, R. W.; ROGERS, W. H.; MOXLEY III, J. H.; DRAPER, D. & BROOK, R. H., 1987. Hospital inpatient mortality: Is it a predictor of quality? *New England Journal of Medicine*, 26:1674-1680.
- GAO/USA (United States General Accounting Office), 1992. *Report to Congressional Requesters. Cross Design Synthesis: A New Strategy for Medical Effectiveness Research*. Document GAO/PEMD-92-18. Washington, DC: Program Evaluation and Methodology Division, GAO/USA.
- GREENLAND, S., 1993. Basic problems in interaction assessment. *Environmental Health Perspectives*, 101 (Sup. 4):59-66.
- HANNAN, E. L.; KILBURN Jr., H. & O'DONNELL, J. F., 1990. Adult open heart surgery in New York State. *JAMA*, 21:2768-2774.
- HARTZ, A. J.; GOTTLIEB, M. S.; KUHN, E. M. & RIMM, A. A., 1993. The relationship between adjusted hospital mortality and the results of peer review. *Health Services Research*, 27:765-777.
- HELLIER, J., 1977. Perinatal mortality: 1950 and 1973. UK Office Population Census & Surveys. *Population Trends*, 10:13-25.
- HLATKY, M. A.; CALIFE, R. M.; HARRELL Jr., F. E.; LEE, K. L.; MARK, D. B. & PRYOR, D. B., 1988. Comparison of predictions based on observational data with the results of randomized controlled clinical trials of coronary artery bypass surgery. *Journal of the American College of Cardiology*, 11:237-245.
- HOFER, T. P. & HAYWARD, R. A., 1996. Identifying poor-quality hospitals. Can hospital mortality rates detect quality problems for medical diagnoses? *Medical Care*, 34:737-753.
- IEZZONI, L. I., 1997a. Risk and outcomes, In: *Risk Adjustment for Measuring Healthcare Outcomes* (L. I. Iezzoni, ed.), pp. 1-42, 2<sup>nd</sup> Ed. Chicago: Health Administration Press.
- IEZZONI, L. I., 1997b. Data sources and implications. In: *Risk Adjustment for Measuring Healthcare Outcomes* (L. I. Iezzoni, ed.), pp. 169-242, 2<sup>nd</sup> Ed. Chicago: Health Administration Press.
- IEZZONI, L. I.; ASH, A. S.; COFFMAN, G. A. & MOSKOWITZ, M. A., 1992. Predicting In-Hospital Mortality: A comparison of severity measurement approaches. *Medical Care*, 30:347-359.
- IEZZONI, L. I.; ASH, A. S.; SHWARTZ, M.; DALEY, J.; HUGHES, J. S. & MACKIERNAN, Y. D., 1996a. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. *American Journal of Public Health*, 86: 1379-1387.
- IEZZONI, L. I.; ASH, A. S.; SHWARTZ, M.; LANDON, B. E. & MACKIERNAN, Y. D., 1998. Predicting in-hospital deaths from coronary artery bypass graft surgery: do different severity measures give different predictions? *Medical Care*, 36:28-39.
- IEZZONI, L. I.; SHWARTZ, M.; ASH, A. S.; HUGNES, J. S.; DALEY, J. & MACHIERNAN, Y. D., 1995. Using severity-adjusted stroke mortality rates to judge hospitals. *International Journal for Quality in Health Care*, 7:81-94.
- IEZZONI, L. I.; SHWARTZ, M.; ASH, A. S.; HUGNES, J. S.; DALEY, J. & MACHIERNAN, Y. D., 1996b. Severity measurement methods and judging hospital death rates for pneumonia. *Medical Care*, 34:11-28.
- KNAUS, W. A.; WAGNER, D. P.; ZIMMERMAN, J. E. & DRAPER, E. A., 1993. Variations in mortality and length of stay in intensive care units. *Annals of Internal Medicine*, 118:753-761.
- MAKLAN, C. W.; GREENE, R. & CUMMINGS, M. A., 1994. Methodological challenges and innovations in patient outcomes research. *Medical Care*, 32 (Sup. 7):JS13-JS21.
- MIETTINEN, O. S., 1985. *Theoretical Epidemiology*. New York: John Wiley & Sons.
- MIETTINEN, O. S., 1982. Causal and preventive interdependence. *Scandinavian Journal of Work, Environment & Health*, 8:159-168.
- MIETTINEN, O. S., 1974. Confounding and effect modification. *American Journal of Epidemiology*, 99:350-353.
- MOSES, L. E., 1995. Measuring effects without randomized trials? Options, problems, challenges. *Medical Care*, 33(Sup. 4):AS8-AS14.
- PARK, R. E.; BROOK, R. H.; KOSECOFF, J.; KEESEY, J.; RUBENSTEIN, L.; KEELER, E.; KAHN, K. L. & ROGERS, W. H., 1990. Explaining variations in hospital death rates, randomness, severity of illness, quality of care. *JAMA*, 264:484-490.
- ROTHMAN, K. J., 1976. Causes. *American Journal of Epidemiology*, 104:587-592.



- ROTHMAN, K. J., 1986. *Modern Epidemiology*. Boston: Little, Brown and Co.
- ROTHMAN, K. J. & GREENLAND, S., 1998. *Modern Epidemiology*. 2<sup>nd</sup> Ed. Philadelphia: Lippincott-Raven.
- SELKER, H. P., 1993. Systems for comparing actual and predicted mortality rates: Characteristics to promote cooperation in improving hospital care. *Annals of Internal Medicine*, 118:820-822.
- SELKER, H. P.; GRIFFITH, J. L.; BEHANSKY, J. R.; CALIFE, R. M.; D'AGOSTINO, R. B.; LAKS, M. M.; LEE, K. L.; MAYNARD, C.; WAGNER, G. S. & WEAVER, W. D., 1992. The thrombolytic predictive instrument (TPI) project: Combining clinical study data bases to take medical effectiveness research to the streets. In: *Medical Effectiveness Research Data Methods* (M. L. Grady, ed.), pp. 9-31, Rockville: Agency for Health Policy and Research.
- SELKER, H. P.; GRIFFITH, J. L. & D'AGOSTINO, R. B., 1991. A time-insensitive predictive instrument for acute myocardial infarction mortality: A multicenter study. *Medical Care*, 29:1196-1211.
- SCHWARTZ, M.; ASH, A. S. & IEZZONI, L. I., 1997. Comparing Outcomes across Providers. In: *Risk Adjustment for Measuring Healthcare Outcomes* (L. I. Iezzoni, ed.), pp. 471-516, 2<sup>nd</sup> Ed. Chicago: Health Administration Press.
- THOMAS, J. W. & HOFER, T. P., 1999. Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care. *Medical Care*, 37:83-92.
- THOMAS, J. W.; HOLLOWAY, J. J. & GUIRE, K. E., 1993. Validating risk-adjusted mortality as an indicator for quality of care. *Inquiry*, 30:6-22.

---

Submitted on 1 November 2001

Final version resubmitted on 22 April 2002

Approved on 9 September 2002