



Published in final edited form as:

Am J Public Health. 2010 February ; 100(2): 264–269. doi:10.2105/AJPH.2008.153759.

Measuring quality for public reporting: Making it meaningful to patients

Dana B. Mukamel, PhD^{*}, Laurent G. Glance, MD, Andrew W. Dick, PhD, and Turner M. Olser, MD

Abstract

Public quality reports of hospitals, health plans and physicians are being used to promote efficiency and quality in the health care system. Shrinkage estimators have been proposed as superior measures of quality to be used in these reports because they offer more conservative and stable quality ranking of providers. In this paper we examine their advantages and disadvantages. Unlike previous studies, we adopt the perspective of a patient who is faced with choosing a provider in their local area of residence. We contrast the information made available by the traditional, non-shrinkage estimators and the shrinkage estimators. We demonstrate that two properties of shrinkage estimators make them less useful for patients making choices in their area of residence.

Keywords

Incentives in Health Care; Quality Improvement/Report Cards (Interventions); Quality of Care/
Patient Safety (Measurement)

Introduction

Measuring the performance of medical care providers has become an important facet of the American health care system. It is one of the four cornerstones of the 2006 Administration's "Value-Driven Health Care Initiative",¹ which calls for measuring and publishing information about quality, and using this information to improve quality and to promote the efficiency of medical care. Quality measures, based on either patient outcomes (e.g. risk-adjusted mortality rates) or process measures (e.g. percent of HMO enrollees with diabetes who received an eye examination), are reported in public report cards,^{2, 3} and are driving pay for performance (P4P) programs.^{4–6}

The accuracy of quality measures depends on a number of factors, including data quality,⁷ the impact of risk adjustment,^{8–12} sample size,¹³ and the specification of the quality measures themselves, i.e. defining them as the difference or the ratio of observed to expected outcome

^{*}Corresponding Author: Professor and Senior Fellow, Department of Medicine, Health Policy Research Institute, University of California, Irvine, 111 Academy Way, Suite 220 Irvine, CA 92697-5800; 949-824-8873; dmukamel@uci.edu.
Statement of author's affiliations: At the time the research was completed Dana B. Mukamel was with the University of California Irvine-Health Policy Research Institute; Laurent G. Glance was with University of Rochester School of Medicine and Dentistry; Andrew W. Dick was with RAND Corporation and Turner M. Olser was with The University of Vermont Medical College.
Contact information for reprint requests: Request for reprints should be sent to Dr. Dana Mukamel, Health Policy Research Institute, University of California-Irvine, 111 Academy Way, Suite 220 Irvine, CA 92697-5800 (dmukamel@uci.edu)

Author Contributions: All authors contributed to the conceptualization and design of the study. Dr. Mukamel wrote the paper and all edited the article.

Statement of Institutional Review Board Approval: There were no human participants involved in this study.

rates.^{14, 15} The use of shrinkage estimators rather than the traditional, non-shrinkage estimators (further defined below) has also been shown to result in different quality rankings.^{16–19}

In this paper we focus on the choice between specifying quality measures based on shrinkage versus the more traditional, non-shrinkage estimators, and discuss the merits and the implications of each approach. Unlike others, who argued in favor of shrinkage estimators because of their stability,^{16–18} we approach this issue from the perspective of a patient whose objective is to choose the best provider from among those available to him or her locally, and thus consider different criteria in evaluating the usefulness of the quality measures. We first describe the two approaches to the estimation of quality measures and define the shrinkage and the non-shrinkage based measures. We then discuss their advantages and disadvantages, and conclude by considering the options for best meeting the needs of patients.

Defining quality measures based on patient outcomes: shrinkage and nonshrinkage estimators

We focus our discussion on quality measures that compare patient outcomes, e.g. mortality, across providers. We recognize that unbiased measurement of quality requires risk-adjustment, but the issues we discuss in this paper apply equally to both risk-adjusted and unadjusted measures. Therefore, for simplicity of exposition we omit risk-adjustment from the discussion below, and note that the arguments we make and the conclusions we reach are not affected by this omission.

Denote by O_{ij} the health outcome for patient i treated by provider j . The quality measure we seek is based on the average outcomes experienced by all n patients treated by provider j . For example, if mortality is the outcome of interest, the average mortality rate for all patients treated in hospital j , or some function of it, can be defined as the quality measure for hospital j , and then used to compare and rank the performance of all hospitals on this outcome.

The unshrunk estimator of the quality of provider j is defined as the mean for all n patients treated by provider j and denoted as \bar{O}_j . This is an accurate, i.e. unbiased estimate of the provider's outcome rate. Its precision depends on the sample size used to calculate it. Providers treating a large number of patients will have more precise estimates of \bar{O}_j than those treating fewer patients.

Stein²⁰ and later James and Stein²¹ proposed a different measure, called the shrinkage estimator, \bar{SO}_j , and showed that it is more efficient (i.e. has a lower squared mean error) than the unshrunk estimator, \bar{O}_j . The shrinkage estimator is defined as a weighted average of the unshrunk estimator, \bar{O}_j , and the average outcome rate calculated over all providers, i.e. the grand mean, \bar{O} . Conceptually, the shrinkage estimator is designed to be close to the unshrunk estimator, \bar{O}_j , when provider j has a large sample and thus \bar{O}_j can be estimated with high precision, and to be close to the grand mean, \bar{O} , when provider j has a small sample and \bar{O}_j cannot be estimated precisely. In the latter case, it is assumed that the grand mean is a better reflection of the true outcome (see further discussion below for the rationale for this and alternative assumptions), and therefore, the estimator is pulled, or “shrunk” towards the grand mean. The name shrinkage estimator is derived from this property of the estimator.

Specifically, the shrinkage estimator is calculated as $\bar{SO}_j = (1 - \alpha_j)\bar{O} + (\alpha_j)\bar{O}_j$ where the weight α_j depends on the relative variance of the outcome within each provider, the sample size within each provider, and the variance across providers.²² When the variance within the provider is relatively small and the sample size is large, the weight will be close to 1 and the shrinkage

estimator will be dominated by the second term in the equation and will approximate the unshrunk estimator, \bar{O}_j . When the variance within the provider is large relative to the variance across providers and the sample size is small, α_j will be very small, the second term in the equation will converge to zero, and the shrinkage estimator will be dominated by the first term, i.e. the grand mean, \bar{O} .

This shrinkage estimator is often referred to as an empirical Bayesian estimator. In the context of empirical Bayesian estimation it can be viewed as follows. Prior to the measurement of the outcome we have beliefs about the distribution of outcomes across providers. We perform the measurement and obtain new information about these outcomes. Because of the stochastic nature of outcome data, the new information is imprecise. Therefore, instead of completely abandoning our prior beliefs, we only partially update them to incorporate the new information. The degree of updating depends on our confidence in the new information, which in turn depends on the sample size used to estimate the outcome rates. The larger the sample size, the higher the confidence in the new information and the more the estimator will be weighed towards the new information, which is the unshrunk mean. The smaller the sample size, the less confidence we have in the new data and the more we weigh the estimator towards the prior belief. Typically, in the context of quality measurement, the analyst adopts a prior belief that all providers have the same performance and thus all observed outcomes rates are shrunk towards the grand mean for all providers. Clearly, a different choice of prior belief, or in the parlance of statisticians “prior”, could lead to vastly different estimators.

An example

To illustrate how shrinkage estimators are calculated and how they differ from the unshrunk estimators we provide an example constructed to highlight their salient properties. Table 1 provides data for coronary artery bypass graft (CABG) procedures for ten hospitals, based on 2004 data from the New York State Cardiac surgery report. For each hospital we show the number of cases (sample size), number of deaths, the observed mortality rate, which is the unshrunk estimator, and the variance within the hospital. From these we calculated the grand mean and variance across all hospitals and based on these we calculated the shrinkage factors and the shrinkage estimators for each hospital. As the table shows, hospital 1, with the largest sample has the largest shrinkage factor, and, therefore, its shrinkage estimator is very similar to its observed, unshrunk estimator, at 2.31 compared with 2.38 respectively. On the other hand, hospital 6 has the smallest shrinkage factor because it has the smallest sample and the largest variance. For this hospital, due to its small sample, one has little confidence in the observed mortality rate, and therefore the shrinkage estimator is very close to the grand mean of 2.10, rather than to its observed mortality rate of 3.70. This reflects the belief, i.e. the “prior”, that this hospital’s “true” mortality rate is more likely to be similar to the average of all other hospitals, than to its actual mortality this year, which might be an aberration.

Figure 1 depicts the same information graphically. The unshrunk estimators are shown at the bottom. Their shrinkage counterparts are shown at the top. For example, the estimate for hospital 2, with 118 patients, is shrunk more towards the overall mean than the estimator for hospital 8, with 277 patients.

The use of shrinkage estimators in multivariate regressions

In the preceding discussion we abstracted from the issue of risk adjustment. However, clearly, risk adjustment is important in the context of quality measurement. Shrinkage estimators can be calculated for quality measures that are risk adjusted as well, using random effect models. These multivariate regression models, which predict patient outcomes (e.g. mortality) based on individual patient risks, assume that the intercept of the model is different for each provider. The provider specific intercept is equivalent to the shrinkage estimator and is calculated in a

manner analogous to the equation \overline{SO}_j above, such that it equals the provider specific outcome rate when the provider has a large sample and small variance, and it is shrunk towards the grand mean as its sample declines and the variance increases. These estimators can be calculated in standard statistical packages, such as SAS, which offers proc MIXED for linear regression models and proc GLIMMIX for models with discrete dependent variables.²³ For examples using random effects risk adjustment models see Glance et al.¹⁹ and Arling et al.¹⁸

We note that we focus our discussion on the use of shrinkage estimators for quality measurement, but random effect models can be used in other instances, like multicenter clinical trials or observational studies involving hierarchical data structures.²⁴

The advantages of the shrinkage estimators

Stable predictions

Stein²⁰ and James and Stein²¹ have argued that the shrinkage estimator is superior to the unshrunk estimator. They have shown that it always results in a lower expected total square errors for the group of providers as a whole. This result is achieved because the shrinkage estimator trades off the bias introduced when estimates are shrunk towards the grand mean with higher efficiency (lower mean squared error).

The intuition behind this result is that when samples are small there is a higher likelihood that any one summary measurement will result in an extreme value. Thus, if a small hospital that treats a small number of patients has a high mortality rate this year, it might be due to chance rather than *true* poor quality. In the following year the same hospital may have a much lower mortality rate. If we believe that all hospitals have the same average mortality rate, we expect that an observed extreme rate in a given year is a “fluke”, and that next year we will observe it “regressing to the mean”. Note that this result crucially depends on the assumption that the smaller hospital provides the same quality of care as the larger hospitals.

Adjustment for multiple comparisons

Another advantage of the shrinkage estimator is that it adjusts for multiple comparisons. It is defined in such a way that the degree of shrinkage depends on the number of groups (e.g. hospitals) that are compared. The more comparisons the larger the shrinkage.^{24, 25}

The multiple comparison problem arises when we want to answer the question of whether the outcome of a specific provider is a statistical outlier compared with the average outcomes of all providers. In other words, having observed an extreme outcome rate for this provider, can we conclude that it is due to a true difference in the quality of care of this provider or is it due to the stochastic nature of outcome measures, which lead us to expect that a certain percent will be flagged as outliers, even if in truth they are not – type I error in the language of statisticians. The multiple comparison problem can be stated as follows: if we compare 100 hospitals and use a p value of 0.05 to identify outliers, 5%, or 5 of these hospitals, will have a p value below 0.05 by chance alone and we will therefore conclude that they are outliers, even though their outcome rate is not truly different from all other providers.

The typical remedy to guard against such type I errors is to require a more conservative (lower p value) threshold for concluding that an observation is a statistical outlier. The Bonferroni correction²⁶ is a common approach that can be applied to the unshrunk estimators, if one chooses.

The shrinkage estimator incorporates the number of comparisons into the shrinkage formula, such that the larger the number of comparisons the larger the shrinkage of the measured value

towards the grand mean. Thus, the likelihood that one would consider a given provider an outlier diminishes as the number of comparisons increases and the shrinkage factor increases.

The disadvantages of shrinkage estimators

Is the shrinkage estimator always the best predictor?

The predictive efficiency of the shrinkage estimators derives from the assumption that all providers are similar and are likely to have the same average performance. If that assumption is correct, then shrinking extreme values towards the grand mean mimics the naturally observed “regression to the mean” phenomenon that will occur in the following period of measurement. Thus, for the group of providers as a whole, the shrinkage estimator is superior. However, for any specific provider, this may not be the case. For those providers whose performance truly deviates from the performance of others, the naturally occurring “regression to the mean” phenomenon will result in a regression to a value that is different from the grand mean. They will regress to the mean of their own and different distribution. Therefore, the shrinkage estimate will not provide a superior prediction for these providers.

In fact, the motivation for quality report cards is the notion that some providers perform at substantially different levels from their peers. For them, one would not anticipate regression to the grand mean, and as Efron and Morris²⁵ note, for such providers the shrinkage estimator would do substantially worse as a predictor.

Ranking providers – the “black box”

As we have shown above the degree of shrinkage increases as the sample size and the precision of the outcome measurement for each provider decreases. An artifact of this is that the rank order of providers changes due to the shrinkage. This is demonstrated in table 1 and figure 1 for hospitals 1 and 6. Hospital 1 is large with a mortality rate close to the grand mean. Hospital 6 is small with a very high mortality rate. The shrinkage estimator for hospital 1, because it has a large sample, is very close to the unshrunk estimator. The shrinkage estimator for hospital 6, however, because it is much smaller and its own average mortality rate is imprecise, will be shrunk substantially towards the mean. The end result is that the shrinkage estimators for both are the almost the same, suggesting that both hospitals offer the same quality care. Consumers reviewing a report card based on shrinkage estimators will conclude that they will do equally well with either.

This, however, is not an accurate interpretation of the data, because in reality the two are very different. For hospital 1 we know with a high degree of certainty that it performs at the average level. For 6, however, we do not really know. The observed high mortality rate may be due to chance because of the sample size or it may be due to the fact that it actually provides lower quality care. Given the extant literature that shows that higher volume is often associated with better patient outcomes,^{27–31} one might actually find the second hypothesis more reasonable.

For patients, whose objective it is to choose the best provider for them, the shrinkage estimator is misleading. It offers patients a black box which combines information about the estimated mortality rate with the precision of this estimate and does not allow them to weigh these two pieces of information separately, in ways consistent with their own preferences.

For sophisticated consumers who are statistically savvy, such as large employers or payers, this problem might be somewhat mitigated if the report card also provides a confidence interval around the shrinkage estimate. For the average patient, who may have difficulties understanding quality measures in general, the statistical significance information is likely to be ignored, and this does not provide a remedy.

The assumption about the “prior”

The shrinkage estimator is typically calculated as the weighted average of the unshrunk estimate and prior distribution with a mean equal to the grand mean across all providers. It is unclear, however, whether this is the best, or most believable, assumption about the prior distribution. As mentioned before, there is a body of literature that indicates an association between provider volume and outcomes – those hospitals and physicians treating larger patient populations tend to have better outcomes. Given this information it would be more reasonable to adopt a prior distribution with a mean that depends on each provider’s sample size. Using such a prior, the shrinkage estimator will no longer pull all estimates towards the grand mean. Rather, the smaller provider will have estimates that are pulled towards lower quality.

Furthermore, the relationship between volume and outcomes is not universal. It seems to be important for some conditions, such as CABG,²⁷ and not others, such as trauma.^{32–34} Therefore, one might consider using a prior specific to the medical condition being measured. Adopting such more informed priors would mitigate the problem we identified in the previous section, in which a large average hospital has the same quality estimate as a small hospital with extreme quality.

Multiple comparisons – are they all relevant?

We noted above that the shrinkage estimator also incorporates an adjustment for multiple comparisons. The larger the number of comparisons, the more the estimator is shrunk towards the grand mean. Unlike the impact of differential sample sizes, the number of comparisons is the same for all providers included in the analysis and hence the impact on the estimates of their quality is the same. Thus, the multiple comparison adjustment does not affect the rank order of providers. It does, however, reduce the variation in outcome rates. As shown in figure 1, the range of values of the shrinkage estimators is more limited than the range of the unshrunk estimator.

The number of comparisons is typically determined by the availability of data and the nature of the entity calculating the quality measures, and is not related to the number of relevant choices that the consumer faces. For example, the 2004 New York State (NYS) Cardiac Surgery report included over 150 cardiac surgeons. While in principle all might be relevant to patients considering cardiac surgery, in practice studies have found that patients tend to stay within their area of residence. An analysis of migration patterns in NYS identified 9 distinct referral areas in the state with most patients (about 95%) staying within these areas.³⁵ For a patient residing in the Rochester area, whose choice includes 7 surgeons, adjusting the shrinkage estimator for over 150 comparisons, most of whom are irrelevant, might result in inappropriate shrinking of the quality measure, to the point that no variation between the seven relevant providers remains. Similarly, if the Centers for Medicaid and Medicare Services (CMS) were to adopt this approach for its Nursing Home Compare report card, which includes over 16,000 nursing homes nationwide, in all likelihood there will be no discernible variation between these facilities, turning the report card uninformative and irrelevant.

Discussion

The choice between a shrinkage and non-shrinkage estimators for quality measurement is important, as it clearly changes the rank order of providers, the degree of variation among them and the identification of statistical outliers. Several studies have demonstrated these differences and argued in favor of adopting shrinkage estimators in quality reporting.^{18, 36} Our analysis of the properties of these two estimators suggests that while the shrinkage estimators may be preferred if the objective is to increase the accuracy of predicted mortality across all providers, it may not serve the needs of individual patients who are making a choice among the providers

available to them locally, and who may have different prior beliefs and different preferences over the “riskiness” of the quality measures than those of the analyst producing the information. In particular, shrinkage estimators tend to be the most biased for providers who are extreme quality outliers. These providers are exactly those that patients and third-party payers are the most interested in identifying.

Unfortunately, there does not seem to be one “correct” solution. The uncertainty in quality measures based on outcomes is inherent and can only be addressed by increasing sample size, often an impractical solution given the realities of medical care. A strategy proposed by Spiegelhalter et al.²⁴ is to perform sensitivity analysis and present shrinkage estimates based on several prior distributions, allowing the consumer of the information to choose the prior that is most consistent with his or her beliefs. For example, one might include in report cards measures based on uninformative priors, as is current practice, as well as priors related to provider volume. Report cards based on this strategy are likely to be complex and difficult to understand for most patients. They may also face political obstacles if the priors are unacceptable to strong stakeholders groups. For example, the most obvious prior to consider, as mentioned before, is one based on volume, in which low volume providers are assumed to have lower quality relative to higher quality providers. Would CMS be able to publish a hospital public report card with quality measures based on such a prior given the strong lobbying power of hospitals?

Another strategy, adopted by NYS in its Cardiac Surgery Reports and the CMS in the Nursing Home Compare report, is to present unshrunk estimators, but to include in the public report only information about providers that have met a minimum volume cut off. Unlike the shrinkage estimators, this strategy clearly identifies those cases where the precision of the measures has been judged to be insufficient. The disadvantage of this strategy is that the analyst imposes his or her own judgment of what is an acceptable level of accuracy for quality measures, and this also may lead to bias. For example, cardiac patients in NYS can obtain information only about the quality of surgeons who performed at least 200 procedures in the last 3 years. If they live in Pennsylvania, however, the report card patients can access will offer them measures on all surgeons who performed at least 30 procedures in the last year.

While none of these strategies offers a completely satisfying solution to the problems inherent in evaluating quality based on outcomes, in the spirit of transparency, which motivates the efforts to publicly reporting on providers’ quality, we prefer the unshrunk measures, which when accompanied by a measure of their statistical significance such as a p value or a confidence interval, do not present patients with a “black box”, but are explicit about the degree of uncertainty in the estimated quality measures. The challenge remains to present the information in such a way that would allow patients and those who help them make referral decisions (family members, physicians, social workers, payers, and others) to understand the information, its accuracy and precision, and to apply it to their specific choices in accordance with their own preferences.

Acknowledgments

The authors gratefully acknowledge funding from the National Institutes of Aging, Grant# AG027420 and from the AHRQ grant # HS016737.

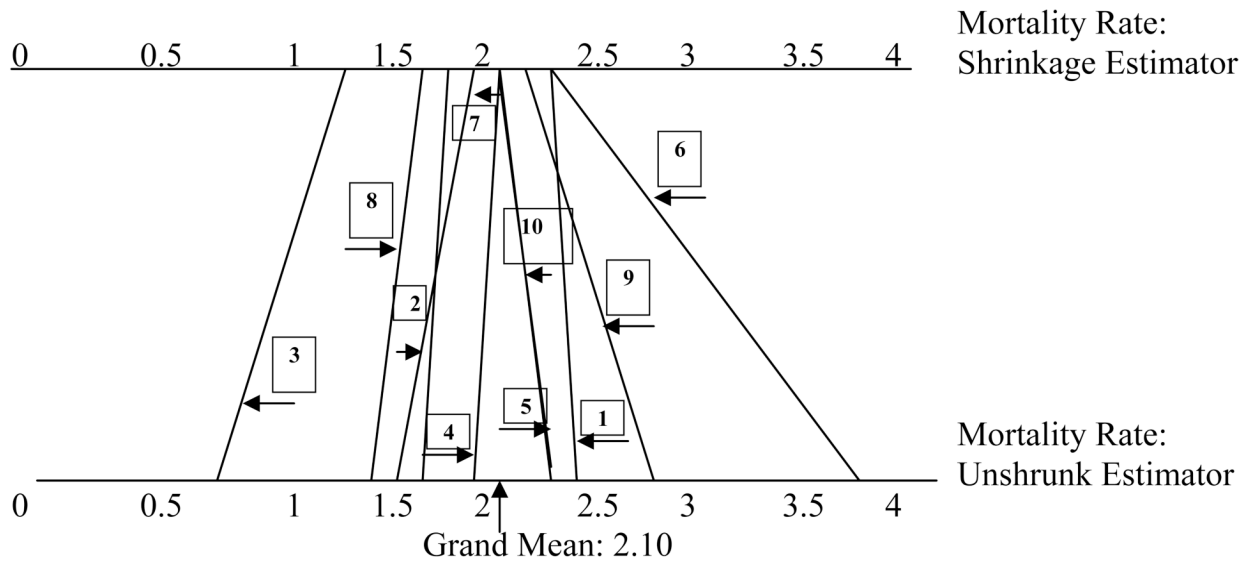
References

1. United States Department of Health and Human Services. Value-Driven Health Care. [Accessed 8/15/07]. <http://www.hhs.gov/valuedriven/fourcornerstones/index.html>
2. United States Department of Health and Human Services. Hospital Compare. [Accessed 8/15/07]. <http://www.hospitalcompare.hhs.gov/Hospital/Search/SearchMethod.asp?>

pagelist=Home&dest=NAV|Home|Search|SearchMethod|Welcome&search_dest=NAV|Home|Search|Welcome&version=default&browser=Firefox|3|WinXP&language=English&btnFindHosp=Find+and+Compare+Hospitals

3. Centers for Medicare & Medicaid Services. Nursing Home Compare Report Card. [Accessed 2007, 4/1/07]. <http://www.medicare.gov/NHCompare/>
4. Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. Pay for performance in commercial HMOs. *N Engl J Med* 2006 Nov 2;355(18):1895–1902. [PubMed: 17079763]
5. Centers for Medicare & Medicaid Services. Medicare "Pay for Performance (P4P)" Initiatives. [Accessed June 20, 2007]. <http://www.cms.hhs.gov/apps/media/press/release.asp?Counter=1343>
6. Silow-Carroll S, Alteras T. Value-Driven Health Care Purchasing: Four States That Are Ahead of the Curve. *The Commonwealth Fund*. Volume 67 http://www.commonwealthfund.org/publications/publications_show.htm?doc_id=515778&#doc515778.
7. Lawthers AG, McCarthy EP, Davis RB, Peterson LE, Palmer RH, Iezzoni LI. Identification of In-Hospital Complications From Claims Data Is It Valid? *Medical Care* 2000 Aug;38(8):785–795. [PubMed: 10929991]
8. Iezzoni LI, Shwartz M, Ash AS, Mackiernan Y, Hotchkin EK. Risk Adjustment Methods Can Affect Perceptions of Outcomes. *American Journal of Medical Quality* 1994 Summer;9(2):43–48. [PubMed: 8044051]
9. Iezzoni LI, Shwartz M, Ash AS, Hughes JS, Daley J, Mackiernan YD. Severity Measurement Methods and Judging Hospital Death Rates for Pneumonia. *Medical Care* 1996 Jan;34(1):11–28. [PubMed: 8551809]
10. Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Judging Hospitals by Severity-Adjusted Mortality Rates: The Influence of the Severity-Adjustment Method. *American Journal of Public Health* 1996 Oct;86(10):1379–1387. [PubMed: 8876505]
11. Mukamel DB, Brower CA. The Influence of Risk Adjustment Methods on Conclusions About Quality of Care in Nursing Homes Based on Outcome Measures. *The Gerontologist* 1998;38(6):695–703. [PubMed: 9868849]
12. Glance LG, Osler TM, Dick A. Rating the quality of intensive care units: is it a function of the intensive care unit scoring system? *Crit Care Med* 2002 Sep;30(9):1976–1982. [PubMed: 12352029]
13. Localio AR, Hamory BH, Fisher AC, TenHave TR. The Public Release of Hospital and Physician Mortality Data in Pennsylvania: A Case Study. *Medical Care* 1997 Mar;35(3):272–286. [PubMed: 9071258]
14. Mukamel DB, Dick A, Spector WD. Specification Issues in Measurement of Quality of Medical Care Using Risk Adjusted Outcomes. *Journal of Economic and Social Measurement* 2000;26:267–281.
15. Li Y, Dick A, Glance LG, Cai X, Mukamel D. Misspecification Issues in Risk Adjustment and Construction of Outcome Based Quality Indicators. *HSOR* 2007 Jun;7(1–2):39–56.
16. Shahian DM, Torchiana DF, Shemin RJ, Rawn JD, Normand SL. Massachusetts cardiac surgery report card: implications of statistical methodology. *Ann Thorac Surg* 2005 Dec;80(6):2106–2113. [PubMed: 16305853]
17. Normand SL, Glickman ME, Gatsonis CA. Statistical Methods for Profiling Providers of Medical Care: Issues and Applications. *Journal of the American Statistical Association* 1997;92:803.
18. Arling G, Lewis T, Kane RL, Mueller C, Flood S. Improving Quality Assessment through Multilevel Modeling: The Case of Nursing Home Compare. *Health Serv Res* 2007 Jun;42(3p1):1177–1199. [PubMed: 17489909]
19. Glance LG, Dick A, Osler TM, Li Y, Mukamel DB. Impact of changing the statistical methodology on hospital and surgeon ranking: the case of the New York State cardiac surgery report card. *Med Care* 2006 Apr;44(4):311–319. [PubMed: 16565631]
20. Stein, C. Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. Paper presented at: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability; Berkeley. 1955.
21. James, W.; Stein, C. Estimation with Quadratic Loss. Paper presented at: Proceedings of the Fourth Berkeley Symposium; Berkeley. 1960.
22. Hox, J. *Multilevel Analysis: Techniques and Applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers; 2002.

23. Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD.; Schabenberger, O. SAS for Mixed Models, Second Edition. Cary, NC: SAS Institute Inc.; 2006.
24. Spiegelhalter, DJ.; Abrams, KR.; Myles, JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Wiley & Sons, Ltd.; 2004.
25. Efron B, Morris C. Stein's Paradox in Statistics. *Sci Am* 1977;236:119–127.
26. Abdi, H. The Bonferroni and Sidak Corrections for Multiple Comparisons. *Encyclopedia of Measurement and Statistics*. 2008 [Accessed 1/28/08].
<http://www.utdallas.edu/~herve/Abdi-Bonferroni2007-pretty.pdf>
27. Hughes RG, Hunt SS, Luft HS. Effect of Surgeon Volume and Hospital Volume on Quality of Care in Hospitals. *Medical Care* 1987 Jun;25(6):489–503. [PubMed: 3695658]
28. Gordon TA, Bowman HM, Bass EB, et al. Complex gastrointestinal surgery: impact of provider experience on clinical and economic outcomes. *J Am Coll Surg* 1999 Jul;189(1):46–56. [PubMed: 10401740]
29. Long DM, Gordon T, Bowman H, et al. Outcome and cost of craniotomy performed to treat tumors in regional academic referral centers. *Neurosurgery* 2003 May;52(5):1056–1063. discussion 1063-1055. [PubMed: 12699547]
30. Barbieri CE, Lee B, Cookson MS, et al. Association of Procedure Volume With Radical Cystectomy Outcomes in a Nationwide Database. *J Urol* 2007 Oct;178(4 Pt 1):1418–1421. [PubMed: 17706712]
31. Nguyen NT, Paya M, Stevens CM, Mavandadi S, Zainabadi K, Wilson SE. The relationship between hospital volume and outcome in bariatric surgery at academic medical centers. *Ann Surg* 2004 Oct;240(4):586–593. discussion 593-584. [PubMed: 15383786]
32. Gance LG, Osler TM, Dick A, Mukamel D. The relation between trauma center outcome and volume in the National Trauma Databank. *J Trauma* 2004 Mar;56(3):682–690. [PubMed: 15128144]
33. Cooper A, Hannan EL, Bessey PQ, Farrell LS, Cayten CG, Mottley L. An examination of the volume-mortality relationship for New York State trauma centers. *J Trauma* 2000 Jan;48(1):16–23. discussion 23-14. [PubMed: 10647560]
34. Margulies DR, Cryer HG, McArthur DL, Lee SS, Bongard FS, Fleming AW. Patient volume per surgeon does not predict survival in adult level I trauma centers. *J Trauma* 2001 Apr;50(4):597–601. discussion 601-593. [PubMed: 11303152]
35. Mukamel DB, Mushlin AI, Weimer DL, Zwanziger J, Parker T, Indridason I. Do Quality Report Cards Play a Role in HMOs Contracting Practices? - Evidence from New York State. *Health Services Research* 2000 Apr;35(1 Part II):319–332. [PubMed: 10778818]
36. Shahian DM, Normand SL, Torchiana DF, et al. Cardiac surgery report cards: comprehensive review and statistical critique. *Annals of Thoracic Surgery* 2001 Dec;72(6):2155–2168. [PubMed: 11789828]



Legend: 1 = Hospital 1



Figure 1. The relationship between unshrunk and shrinkage estimators. Values based on the example in table 1.

Example: Data from the NYS 2004 Adult Cardiac Surgery Report Published in June 2006

Table 1

Hospital	1 n_j Number of cases: Sample Size	2 # of Deaths	3 \bar{O}_j Unshrunk Estimator: Mortality Rate per 100 Cases	4 σ_j^2 Variance within the Hospital	5 g_j Shrinkage Factor	6 \bar{SO}_j Shrunk Estimator: Mortality Rate per 100 Cases
6	81	3	3.70	0.0357	0.127	2.30
10	89	2	2.25	0.0220	0.205	2.13
2	118	2	1.69	0.0167	0.311	1.97
3	132	1	0.76	0.0075	0.528	1.39
4	256	5	1.95	0.0191	0.460	2.03
8	277	4	1.44	0.0142	0.554	1.73
9	286	8	2.80	0.0272	0.402	2.38
7	290	5	1.72	0.0169	0.522	1.90
5	488	11	2.25	0.0220	0.586	2.19
1	1050	25	2.38	0.0232	0.742	2.31

τ^2 = Variance across hospitals: 0.0000638. Calculated from column 3.

\bar{O} Overall mean: 2.10. Calculated from column 3.

$$\alpha_j = \frac{\sigma_j^2}{\sigma_j^2 + \tau^2}$$

Values for column 5 calculated as:

$$\bar{SO}_j = (1 - \alpha_j)\bar{O} + (\alpha_j)\bar{O}_j$$

Values for column 6 calculated as: