Statistical Uncertainty of Mortality Rates and Rankings for Children's Hospitals

WHAT'S KNOWN ON THIS SUBJECT: Many hospital quality-of-care stakeholders wish to identify hospitals with leading or lagging performance on the basis of mortality rates, but the degree of statistical imprecision of comparative rates or rankings may not be fully appreciated.

WHAT THIS STUDY ADDS: Children's hospitals' mortality rates have substantial statistical imprecision, and hospital rankings based on those rates have even more. Stakeholders who seek to improve quality of care and patient outcomes may want to find more informative measures to guide action.

abstract

OBJECTIVES: Hospitals are being required to report publically their adjusted mortality rates, which are then being used to rank hospitals. Our objectives were to assess the statistical reliability of the determination of a hospital's adjusted mortality rate, of comparisons of that rate with the rates of other hospitals, and of the use of those rates to rank the hospitals.

METHODS: A cross-sectional study of 473 383 patients discharged from 42 US children's hospitals in 2008 was performed. Hospital-specific observed/expected (O/E) mortality rate ratios and corresponding hospital rankings, with 95% confidence intervals (CIs), were examined.

RESULTS: Hospitals' O/E mortality rate ratios exhibited wide 95% CIs, and no hospital was clearly distinguishable from the other hospitals' aggregated mean mortality performance. Only 2 hospitals' mortality performance fell outside the comparator hospitals' 95% CI. Those hospitals' 95% CIs overlapped with the overall comparator set's 95% CI, which suggests that there were no statistically significant hospital outliers. Fourteen (33.3%) of the 42 hospitals had O/E ratios that were not statistically different from being in the 95% CI of the top 10% of hospitals. Hospital-specific mortality rate rankings displayed even broader 95% CIs; the typical hospital had a 95% CI range that spanned 22 rank-order positions.

CONCLUSION: Children's hospital-specific measures of adjusted mortality rate ratios and rankings have substantial amounts of statistical imprecision, which limits the usefulness of such measures for comparisons of quality of care. *Pediatrics* 2011;128:e966–e972

AUTHORS: Chris Feudtner, MD, PhD, MPH,^{a,b} Jay G. Berry, MD, MPH,^c Gareth Parry, PhD,^d Paul Hain, MD,^e Rustin B. Morse, MD,^f Anthony D. Slonim, MD, PhD,^g Samir S. Shah, MD, MSCE,^h and Matt Hall, PhDⁱ

^aPolicyLab, ^bDepartment of Medical Ethics, and ^cDivision of Infectious Diseases, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; ^dProgram for Patient Safety and Quality, Division of General Pediatrics, Children's Hospital Boston and School of Medicine, Harvard University, Boston, Massachusetts; ^eInstitute for Healthcare Improvement, Boston, Massachusetts; ^fDepartment of Pediatrics, Monroe Carell Jr Children's Hospital at Vanderbilt, Nashville, Tennessee; ^gEmergency Department, Phoenix Children's Hospital, Phoenix, Arizona; ^hSaint Barnabas Health Care System, West Orange, New Jersey; and ⁱDepartment of Informatics, Child Health Corporation of America, Kansas City, Kansas

KEY WORDS

quality appraisal, quality improvement, mortality rates, hospital performance

ABBREVIATIONS

O/E—observed/expected

CI—confidence interval

PHIS—Pediatric Health Information System

APR-DRG—all-patient refined diagnosis-related group

IQR—interquartile range

www.pediatrics.org/cgi/doi/10.1542/peds.2010-3074

doi:10.1542/peds.2010-3074

Accepted for publication May 31, 2011

Address correspondence to Chris Feudtner, MD, PhD, MPH, PolicyLab and Department of Medical Ethics, Children's Hospital of Philadelphia, 3535 Market St, Room 1523, Philadelphia, PA 19104. E-mail: feudtner@email.chop.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2011 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

Funded by the National Institutes of Health (NIH)

Hospitals are increasingly being asked to report their mortality rates publicly.^{1–3} Advocates of public reporting assume that (1) we can measure a hospital's mortality rate with an adequate level of statistical precision, adjusting for clinical and demographic characteristics of the patients served by that hospital, (2) hospitals with higher mortality rates are delivering poorer quality of care, and (3) public reporting of this signal of poorer quality of care will spur hospitals to improve outcomes by combining patient/consumer market-based pressure with an innate desire to reduce mortality rates.⁴

Are those assumptions valid? Recently, some authors questioned the second assumption, arguing that hospital mortality rates are inappropriate quality-of-care indicators,^{5,6} and other authors questioned the third assumption, describing how public reporting of mortality rates has resulted in some hospitals shunning the care of the sickest patients.^{7,8} Most criticism of public reporting of hospital-specific mortality rates, however, has focused on one aspect of the first assumption, namely, the need to account for differences in hospitals' case mixtures, with some hospitals potentially caring for sicker patients, with more-severe illnesses, who are more likely to experience death. The response to this criticism has been to create severity-adjustment methods to allow, in theory, for useful comparisons of mortality rates across hospitals.⁹ One widely used metric used to incorporate severity-adjusted mortality rates focuses on the observed/expected (O/E) mortality rate ratio, which separates all patients into groups of similar patients, usually defined according to the all-patient refined diagnosis-related group (APR-DRG) system,¹⁰ calculates the expected mortality rate for each group with particular severity-adjustment methods, and compares

that expected mortality rate with the actual observed mortality rate, not just within each group but also across all groups of patients in a hospital.^{11,12} Hospitals with O/E mortality rate ratios of >1 are perceived to have higher mortality rates than would be expected on the basis of the individual hospitals' case mixtures. The severity-adjustment methods often are proprietary and are not published; a recent comparison of several methods applied to the same sample of hospitalized patients demonstrated striking inconsistencies.¹³

In this study, we focused on a different aspect of the first assumption, that is, can we measure a hospital's mortality rate with sufficient precision to compare it with rates of other hospitals? This question has not been posed regarding children's hospitals. Research on adult hospitals in the United Kingdom showed that performance rankings are statistically imprecise, even after adjustment for case mixtures,^{14–17} which suggests that where a hospital appears in a ranking from best and worst may have as much to do with random chance as with any underlying quality of care. Furthermore, smaller hospitals have been shown to have greater statistical instability in their rankings, which creates a potential bias for comparisons of hospitals of different sizes¹⁸ and which may be quite relevant in comparisons of children's hospitals.

To examine the degree of statistical precision for hospital mortality rates, we measured and compared mortality rates and rankings among 42 freestanding children's hospitals by using 3 different questions: (1) How much do particular hospitals differ from the overall mortality rate average or central tendency of all of the hospitals? (2) Is a particular hospital an outlier if it is significantly different from the other hospitals? and (3) If we focus on a par-

ticular level or quantile of performance, such as the top 10% of hospitals, does a particular hospital's performance fall within that range? Finally, we culminated our investigation by asking the following: if measurements of mortality rates are imprecise, how imprecise are rankings based on those rates?

METHODS

Institutional Review Board Oversight

In accordance with 45 CFR 46.102(f), this deidentified data-set study did not meet the definition of human subjects research.

Data Sources and Quality

Observed mortality data were obtained from the Pediatric Health Information System (PHIS), which is maintained by the Child Health Corporation of America (Shawnee Mission, KS). The PHIS contains discharge data regarding patient demographic features, diagnoses, and procedures from 42 freestanding children's hospitals. Systematic monitoring of data quality and reliability are ensured through a joint effort between the Child Health Corporation of America and participating hospitals.

To establish mortality expectations for case-mixture adjustment among a wide variety of pediatric inpatients, we used the 2006 Healthcare Cost and Utilization Project Kids' Inpatient Database of the Agency for Healthcare Research and Quality and generated mortality expectations for patients in the PHIS sample. The Kids' Inpatient Database is a nationally representative discharge database that provides the largest collection of pediatric inpatient data in the country, with discharge data including patient demographic features, diagnoses, and procedures.¹⁹

Patients

Patients discharged from the 42 PHIS participating hospitals between January 1, 2008, and December 31, 2008, were included in the study.

Variables

The dependent variable was the in-hospital mortality rate. The primary independent variables used for case-mixture adjustment were APR-DRG version 20 (3M Corp, Minneapolis, MN) codes and risk of death. Each APR-DRG-specific risk-of-death level (1, minor; 2, moderate; 3, major; and 4, extreme) appraises retrospectively the mortality risk for each patient on the basis of the patient's age, principle and secondary diagnoses, and certain procedures. The risk of death method is widely used for hospital performance monitoring.²⁰

Statistical Analyses

Overall Approach

Our primary units of analyses were each PHIS hospital's O/E ratio and associated mortality performance rankings. We report the ratios and rankings by using 3 statistical perspectives, that is, central tendency, outlier, and quantile performance level. Within each perspective, bootstrap resampling was used to assign variance (95% confidence intervals [CIs]) to the hospitals' O/E ratios and performance rankings. All statistical analyses were performed by using SAS 9.2 (SAS Institute, Cary, NC) and Stata 11.1 (Stata Corp, College Station, TX).

Central Tendency Perspective

To demonstrate the central tendency perspective of hospital comparisons, we used 3 different statistical approaches to determine whether each hospital performed better or worse than the overall mean O/E ratio for all of the hospitals. In the first approach, we used the most commonly used fre-

quentist technique, summed the observed and expected mortality risks for each patient within a hospital, and created a summed O/E mortality rate ratio with associated 95% Poisson CI. In the second approach, we used a Bayesian technique that has been used in the United Kingdom and calculated the O/E ratio with corresponding 95% CI by using a vague proper conjugate prior for the variance of the hospital random effects [$\tau^{-2} \sim \gamma(0.001, 0.001)$] and a vague but proper prior for their mean [$\mu \sim N(0, 1000)$], with the expected mortality rate as the only covariate. We implemented a burn-in of 1000 iterations and 5000 iterations for inference, checking for convergence of all parameters. We used Markov-chain Monte Carlo methods with an adaptive, blocked, random-walk, Metropolis algorithm. We used a logistic random-effects (hospital) model, and we report the posterior means with 95% posterior CIs.²¹

We used statistical resampling as the third analytic approach, because this technique is robust and enables a broader range of questions. Specifically, we used bootstrap resampling, creating 1000 bootstrap samples of each hospital's data, and calculated each hospital's O/E ratio from each sample. We then computed the CI around each hospital's ratio by using the bootstrap, bias-corrected, accelerated method.²² For each iteration of the bootstrap, we also determined the rankings for the hospitals on the basis of their O/E ratios. This provided us with 1000 rankings for each hospital, from which we computed a 95% CI around the hospital's ranking by using the bootstrap, bias-corrected, accelerated method.

Outlier Perspective

To demonstrate the group outlier perspective, we used a bootstrap resampling technique to determine whether

each hospital performed better or worse than the other hospitals. We randomly sampled O/E ratios from each of the other hospitals' sets of 1000 bootstrapped O/E ratios and calculated the 2.5th and 97.5th O/E ratio percentiles for those values. We iterated this procedure 1000 times and computed the median of the 2.5th percentiles and the median of the 97.5th percentiles. This provided a 95% performance interval in which we would expect a hospital's O/E ratio to lie if it was performing similarly to the other hospitals. If a given hospital's O/E ratio was below this interval, then the hospital performed better than other hospitals; conversely, if the hospital's O/E ratio was above this interval, then the hospital's performance was worse.

Quantile Performance Perspective

To demonstrate the quantile performance level perspective, we used a bootstrap resampling technique to categorize the performance of each hospital within the top 10%, 20%, and 50% of hospitals. To do this, we randomly sampled 1 O/E ratio value from each of the hospitals' sets of 1000 bootstrapped O/E ratios, and we calculated the top 10th, 20th, and 50th O/E ratio percentiles for those values. We iterated this procedure 1000 times and computed the 2.5th and 97.5th percentiles (which represents the 95% CI) of each performance level category. We then took the upper 95% CI boundary for each percentile as the upper 95% CI boundary for each performance level group, and we defined hospitals with O/E ratios below the upper 95% CI boundary as being members of that performance level group.

We evaluated the statistical precision of hospital rankings based on O/E ratios by using bootstrap techniques (bootstrap, bias-corrected, accelerated method) to estimate the 95% CIs of the rankings; for the quantile per-

TABLE 1 Characteristics of Study Hospitals and Patients in 2008

Hospitals	
No. of hospitals	42
No. of patients	473 383
Cases per hospital, median (IQR)	12 353 (9795–14 222)
Hospital mortality rate, median (IQR), deaths per 100 patients	1.0 (0.8–1.1)
Hospital O/E ratio, median (IQR)	0.9 (0.8–1)
Patients	
Gender, <i>n</i> (%)	
Male	255 726 (54)
Female	218 079 (46)
Race, <i>n</i> (%)	
White	222 618 (48.5)
Black	96 446 (21)
Hispanic	88 890 (19.4)
Asian	13 434 (2.9)
Other	37 943 (8.3)
Disposition, <i>n</i> (%)	
Died	4455 (0.9)
Did not die	469 383 (99.1)
Age, <i>n</i> (%)	
<2 y	179 583 (37.9)
2–4 y	67 926 (14.3)
5–9 y	74 007 (15.6)
10–14 y	76 173 (16.1)
15–18 y	58 355 (12.3)
>18 y	17 794 (3.8)
Length of stay, <i>n</i> (%)	
1–5 d	362 835 (76.6)
6–10 d	57 475 (12.1)
11–15 d	20 215 (4.3)
≥16 d	33 313 (7)
Payer, <i>n</i> (%)	
Government	221 762 (46.8)
Private	161 884 (34.2)
Other	90 192 (19)

formance level perspective top 10%, 20%, and 50% ranking hospitals, we used an approach analogous to the approach outlined above but using rankings instead of O/E ratios. All bootstrap samples were of equal size, compared with the original sample.

RESULTS

In 2008, the 42 children's hospitals that contributed to this study admitted 473 383 infants, children, and young adults (Table 1). Hospitals ranged between 9795 and 14 222 patients admitted that year. The median unadjusted mortality rate among the hospitals was 1.0 death per 100 patients (interquartile range [IQR]: 0.8–1.1 death per 100 patients); the median O/E mortality rate ratio was 0.9 (IQR: 0.8–1.0).

How much do particular hospitals differ from the overall mortality rate average or central tendency of all of the hospitals? The 95% CIs of each hospital's O/E ratio, calculated by using frequentist, Bayesian, and bootstrap resampling statistical methods, were broad (Fig 1). The frequentist method identified 4 hospitals to be below and 3 hospitals to be above the expected mortality rate range set by the overall group mean, to a statistically significant degree; the Bayesian method identified 5 hospitals above and 5 below and the bootstrap resampling 3 above and 2 below. Only 1 hospital was identified consistently by all 3 methods to be significantly below the overall group mean and only 1 hospital above. To the degree that such departure of a

hospital's O/E mortality rate ratio from the overall hospitals' group mean conveys important information, the central tendency perspective may be a relatively imprecise but useful perspective for comparison.

Is a particular hospital an outlier that is significantly different from the other hospitals? When we compared each hospital's O/E ratio point estimate and 95% CI with the overall 95% performance interval, as determined by the aggregated data from all of the other hospitals, we found that there was considerable statistical uncertainty for the O/E ratio point estimate for each hospital; the median width of the 95% CI for the O/E ratio for each hospital was 0.31 (IQR: 0.25–0.36) (Fig 2). Furthermore, only 2 hospitals had point estimates for O/E mortality rate ratios outside the comparator hospitals' 95% performance interval, and even those 2 hospitals had 95% CIs that overlapped with the overall comparator set's 95% performance interval, which suggests that there were no statistically significant hospital outliers.

For a particular level or quantile of performance, does a particular hospital's performance fall within that range? When we used the bootstrap resampling approach to determine the range of the performance interval for the top 10%, 20%, or 50% of the hospitals, we found that the statistical imprecision of each hospital's O/E point estimate precluded firm judgment regarding the relative performance interval for a given hospital; 14 (33.3%) of the 42 hospitals had O/E ratios that were not statistically different from being in the 95% performance interval of the top 10% of hospitals, whereas 36 (83.3%) of the 42 hospitals had O/E mortality rate ratio 95% CIs that included values in the 95% performance interval range of the top 50% of the PHIS hospitals (Fig 3).

FIGURE 1 Central tendency perspective of hospital O/E mortality rate ratios with 95% CIs calculated by using frequentist, Bayesian, and bootstrap methods.

If measurements of mortality rates are imprecise, then how imprecise are rankings based on those rates? The breadth of statistical imprecision in-

creased for the hospital rankings that were based on the O/E mortality rate ratio of each hospital (Fig 4). The median width of the 95% CI for hospital

FIGURE 2 Outlier perspective of each hospital's O/E mortality rate ratio and 95% CI plotted against the 95% CI for the set of comparison hospitals.

rankings was 22 positions (IQR: 18–25 positions in the rankings), of 42 possible ranking positions. In more-specific terms, the hospitals ranked first and last had the narrowest 95% CIs, but the range of even those narrow 95% CIs still indicated that those hospitals could have been ranked, given the degree of statistical uncertainty, 5 positions away from the first or last rank; the hospitals with the broadest 95% CIs were ranked 9th and 13th, but the 95% CIs for both ranged from the first rank position to the 36th rank position. Indeed, 36 (85.7%) of the 42 hospitals had rankings based on their O/E mortality rate ratios that were statistically consistent with the 95% performance interval range of the top 50% of all hospitals.

DISCUSSION

Our analysis of pediatric mortality data from 42 children's hospitals revealed the following: (1) the O/E mortality rate ratio for any given hospital is surrounded by a substantial zone of statistical imprecision (which is even greater for smaller hospitals; data not shown); (2) hospital rankings based on O/E ratios have a large degree of imprecision regarding where any particular hospital (except for those at the very top or bottom of the rankings) should truly fall in the rank ordering; and (3) with the use of methods that accounted for those uncertainties, only 2 (4.8%) of the 42 hospitals could be potentially identified as outliers having higher or lower mortality rates.

Our findings should be interpreted with several caveats in mind. First, our sample includes only freestanding children's hospitals, which may exhibit less variation in mortality than other hospitals; thus, our findings may not generalize to all hospitals. Second, we could not identify cases in which pediatric patients were admitted to hospitals with the specific purpose

of receiving end-of-life care; although the number of such cases is small, quality-of-care ranking systems should take great care not to penalize hospitals for providing this type of care.²³ Finally, in the absence of widely accepted statistical methods to compare the performance of a given hospital against the expected range of performance of a group of comparison hospitals (and not simply against the overall mean of that sample), we used bootstrap resampling techniques; although we consider this approach to be appropriate and analytically robust, more methodologic research is warranted.

Despite these caveats, our findings should sound a note of caution in a health care environment in which many have advocated public reporting of quality measures and patient outcomes as a means to provide patients and health plans with appropriate information for selection of “the best” providers and health care institutions and to stimulate quality improvement implementation by health care providers fearing market loss. Increasingly, public reports comparing and often rank-ordering hospitals are published by magazines, private companies, the Department of Health and Human Services, and even hospitals on their Web sites. Hospitals are facing mounting pressure to release data, including mortality rates, for use in such reports, and hospitals are concerned with their public image as based on such rankings.

Evidence suggests that publically releasing performance data stimulates quality improvement initiatives at the hospital level but is inconsistent in leading to direct improvements in quality of care.²⁴ Consumers and purchasers rarely seek performance data information, and many do not understand or trust it.²⁵ Furthermore, public performance reporting may have unintended consequences, such as leading physicians or hospitals

to avoid sicker patients in an attempt to improve their quality rankings.^{26,27} A limited number of studies did conclude that the publication of performance data was associated with improvement in health outcomes.²⁸ For example, New York State has mandated public reporting of risk-adjusted mortality rates for adult coronary intervention procedures for more than 1 decade and, during that period, unadjusted rates of death after such procedures have decreased significantly.⁸ Additional potential benefits of public reporting include accelerating the adoption of “best practices” and establishing data sets for critical outcomes research.^{8,28}

At first glance, a hospital’s unadjusted mortality rate may seem to be a straightforward measure of how well that hospital cares for patients, and rates adjusted for patient case mixture and severity of illness may be thought to offer an even more reliable basis for comparisons. As demonstrated, however, rankings based on adjusted mortality measures are statistically uncertain and are liable to overinterpretation; the vast majority of children’s hospitals in this study exist in a zone of essentially indistinguishable mortality rates. Therefore, although the use of this particular quality-of-care indicator either within or across hospitals may prompt quality im-

FIGURE 3

Performance group perspective of hospital O/E mortality rate ratios with 95% CIs.

FIGURE 4

Performance group perspective of hospital rankings based on O/E mortality rate ratios with 95% CIs.

provement efforts, the data seem to be an inexact guide; metaphorically, hospital mortality rates and rankings may supply a stiff wind but a poor compass.

CONCLUSIONS

Drawing attention to the statistical uncertainty associated with aggregate mortality data for hospitals in no way denies the importance of each hospital scrutinizing the circumstances surrounding the death of every patient, as a means to guide internal quality improvement efforts. Leaders within hos-

pitals should examine year-to-year changes in mortality rates but always in the context of the associated interval of statistical uncertainty, so that small increases or decreases in mortality rates are understood to be as likely attributable to chance as to any other factor. The public release of comparative mortality data for hospitals, if such data are reported at all, needs to be performed responsibly, with clear emphasis on the inherent uncertainty of what such data denote regarding comparative quality of care.

REFERENCES

- Rothberg MB, Morsi E, Benjamin EM, Pekow PS, Lindenauer PK. Choosing the best hospital: the limitations of public quality reporting. *Health Aff (Millwood)*. 2008;27(6):1680–1687
- Bonow RO, Masoudi FA, Rumsfeld JS, et al. ACC/AHA classification of care metrics: performance measures and quality metrics: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures. *Circulation*. 2008;118(24):2662–2666
- Comarow A. Best hospitals 2011–12: the methodology. *US News and World Report*. Available at: <http://health.usnews.com/health-news/best-hospitals/articles/2011/07/18/best-hospitals-2011-12-the-methodology>. Accessed August 18, 2011
- Hannan EL, Kumar D, Racz M, Siu AL, Chassin MR. New York State's Cardiac Surgery Reporting System: four years later. *Ann Thorac Surg*. 1994;58(6):1852–1857
- Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ*. 2010;340:c2016
- Black N. Assessing the quality of hospitals. *BMJ*. 2010;340:c2066
- Apolito RA, Greenberg MA, Menegus MA, et al. Impact of the New York State Cardiac Surgery and Percutaneous Coronary Intervention Reporting System on the management of patients with acute myocardial infarction complicated by cardiogenic shock. *Am Heart J*. 2008;155(2):267–273
- Resnic FS, Welt FG. The public health hazards of risk avoidance associated with public reporting of risk-adjusted outcomes in coronary intervention. *J Am Coll Cardiol*. 2009;53(10):825–830
- Steinberg SM, Popa MR, Michalek JA, Bethel MJ, Ellison EC. Comparison of risk adjustment methodologies in surgical quality improvement. *Surgery*. 2008;144(4):662–667
- Hughes J. 3M HIS APR DRG classification software: overview. Available at: www.ahrq.gov/qual/mortality/Hughessumm.htm. Accessed March 11, 2011
- Iezzoni LI. *Risk Adjustment for Measuring Health Care Outcomes*. 3rd ed. Chicago, IL: Health Administration Press; 2003
- Frisch L, Anscombe L, Bamford M. How can we know whether short term trends in a hospital's HSMR are significant? *Stud Health Technol Inform*. 2009;143:149–154
- Shahian DM, Wolf RE, Iezzoni LI, Kirle L, Normand SL. Variability in the measurement of hospital-wide mortality rates. *N Engl J Med*. 2010;363(26):2530–2539
- Leyland AH, Boddy FA. League tables and acute myocardial infarction. *Lancet*. 1998;351(9102):555–558
- Marshall EC, Spiegelhalter DJ. Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates. *BMJ*. 1998;316(7146):1701–1704
- Rabilloud M, Ecochard R, Esteve J. Maternity hospitals ranking on prophylactic caesarean section rates: uncertainty associated with ranks. *Eur J Obstet Gynecol Reprod Biol*. 2001;94(1):139–144
- Ranstrom J, Wagner P, Robertsson O, Lidgren L. Health-care quality registers: outcome-orientated ranking of hospitals is unreliable. *J Bone Joint Surg Br*. 2008;90(12):1558–1561
- Davidson G, Moscovice I, Remus D. Hospital size, uncertainty, and pay-for-performance. *Health Care Financ Rev*. 2007;29(1):45–57
- Agency for Healthcare Research and Quality. *Introduction to the HCUP Kids' Inpatient Database (KID) 2006*. Rockville, MD: Agency for Healthcare Research and Quality; 2008
- Meurer SJ. Mortality risk adjustment methodology for University Health System's clinical data base. Available at: www.ahrq.gov/qual/mortality/Meurer.htm. Accessed June 25, 2011
- Normand S-LT, Shahian DM. Statistical and clinical aspects of hospital outcome profiling. *Stat Sci*. 2007;22(2):206–226
- Efron B. Better bootstrap confidence intervals. *J Am Stat Assoc*. 1987;82(397):171–185
- Holloway RG, Quill TE. Mortality as a measure of quality: implications for palliative and end-of-life care. *JAMA*. 2007;298(7):802–804
- Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med*. 2008;148(2):111–123
- Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. *JAMA*. 2000;283(14):1866–1874
- Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA*. 2005;293(10):1239–1244
- Chien AT, Chin MH, Davis AM, Casalino LP. Pay for performance, public reporting, and racial disparities in health care: how are programs being designed? *Med Care Res Rev*. 2007;64(5 suppl):283S–304S
- O'Connor GT, Malenka DJ, Quinton H, et al. Multivariate prediction of in-hospital mortality after percutaneous coronary interventions in 1994–1996. *J Am Coll Cardiol*. 1999;34(3):681–691

ACKNOWLEDGMENTS

Dr Feudtner was supported in part by funding from the Pew Charitable Trusts. Dr Berry was supported in part by the Eunice Kennedy Shriver National Institute of Child Health and Human Development career development award K23 HD058092. Dr Shah received support from the National Institute of Allergy and Infectious Diseases (grant K01 AI73729) and the Robert Wood Johnson Foundation through its Physician Faculty Scholar Program.