Published in final edited form as:

J Am Coll Surg. 2011 August ; 213(2): 231–235. doi:10.1016/j.jamcollsurg.2011.04.004.

Reliability of Superficial Surgical Site Infections as a Hospital Quality Measure

Lillian S. Kao, MD, MS, FACS¹, Amir A. Ghaferi, MD, MS², Clifford Y. Ko, MD, MS, MSHS, FACS³, and Justin B. Dimick, MD, MPH, FACS²

¹ Center for Surgical Trials and Evidence-based Practice (C-STEP), Department of Surgery, University of Texas Health Science Center at Houston, Houston, TX

²Michigan Surgical Collaborative for Outcomes Research and Evaluation (M-SCORE), Department of Surgery, University of Michigan, Ann Arbor, MI

³Division of Research and Optimal Patient Care, American College of Surgeons, Chicago, IL

Abstract

Background—Although rates of superficial surgical site infection (SSI) are increasingly used as measures of hospital quality, the statistical reliability of using SSI rates in this context is uncertain. We used the American College of Surgeons National Surgical Quality Improvement Project (ACS NSQIP) data to determine the reliability of SSI rates as a measure of hospital performance and to evaluate the effect of hospital caseload on reliability.

Study Design—We examined all patients who underwent colon resection in hospitals participating in ACS NSQIP in 2007 (n=18,455 patients, n= 181 hospitals). We first calculated the number of cases and the risk-adjusted rate of SSI at each hospital. We then used hierarchical modeling to estimate the reliability of this quality measure for each hospital. Finally, we quantified the proportion of hospital-level variation in SSI rates due to patient characteristics and measurement noise.

Results—The average number of colon resections per hospital was 102 (standard deviation, 65). The risk-adjusted rate of superficial SSI was 10.5%, but varied from 0 to 30% across hospitals. Approximately 35% of the variation in SSI rates was explained by noise, 7% could be attributed to patient characteristics, and the remaining 58% represented true differences in SSI rates. Just over half of the hospitals (54%) had a reliability greater 0.70, which is considered a minimum acceptable level. To achieve this level of reliability, 94 cases were required.

Conclusions—SSIs rates are a reliable measure of hospital quality when an adequate number of cases have been reported. For hospitals with inadequate caseloads, the NSQIP sampling strategy could be altered to provide enough cases to ensure reliability.

© 2011 American College of Surgeons. Published by Elsevier Inc. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosure information: Nothing to disclose.

Presented at the 6th Annual Academic Surgical Congress, Huntington Beach, CA, February 2011.

INTRODUCTION

Surgical site infections (SSI) are increasingly used to measure hospital quality with surgery. Hospital-specific rates of SSI are central to several value based purchasing, public reporting, and quality improvement initiatives. For example, SSI is the most common complication reported in the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP). In addition, The Center for Medicare and Medicaid Services will begin public reporting of SSI rates on their Hospital Compare website in 2012. To date, most of the controversy surrounding the use of outcome measures for these purposes focuses on methods for adjusting for patient risk.¹⁻³ The use of such measures also requires that they are highly reliable – i.e., high performing hospitals can be confidently distinguished from the low-performing hospitals.

However, the reliability of superficial SSI as a hospital quality measure is unknown. Reliability reflects the proportion of hospital-level variation attributable to differences in quality (e.g., “signal”), where the remaining variation is attributable to measurement error (e.g., “noise”). When assessing surgical outcomes, such as SSI, low hospital caseloads and low event rates conspire to reduce the reliability of these measurements. Low reliability increases the likelihood that extreme outcome (bad or good) are due to chance.⁴ The reliability of quality indicators is important to define since misclassification of hospitals can have significant impact on choice of quality improvement efforts, public perception, and reimbursement.

This study uses data on patients undergoing colon resections in the American College of Surgeons National Surgical Quality Improvement Project (ACS NSQIP) to assess the reliability of SSI rates as a measure of hospital performance and to evaluate the effect of hospital caseload on reliability.

METHODS

Data Source and Study Population

The 2007 American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) data file was used. ACS NSQIP is a national clinical registry used for quality improvement. ACS-NSQIP provides risk-adjusted 30-day mortality and morbidity rates to participating hospitals. Dedicated surgical clinical nurse reviewers collect data on over 135 variables including demographics and pre-operative risk factors, intra-operative variables, and post-operative complications using standardized definitions. Sampling of cases for inclusion is based on an eight-day cycle to minimize bias due to day of the week. High-volume hospitals participating in the general and vascular surgery program report the first 40 consecutive cases in the cycle, while reduced volume hospitals must enter the maximum number of eligible cases per cycle to meet a minimum requirement of 22 cases per cycle. On the 30th postoperative day, nurse reviewers obtained outcomes information by way of chart review, reports from morbidity and mortality conferences, and communication with each patient by letter or by telephone. Audit and feedback are performed to ensure the accuracy of the data, and the analytic methods for risk adjustment have been demonstrated to be robust and valid.⁵⁻⁶ For the study, all patients who underwent colon resection were identified by the relevant Current Procedure Terminology (CPT) Codes.

Risk-adjusted hospital infection rates

Surgical site infections (SSI) were ascertained from the medical record by clinical nurse reviewers according to standard definitions. We assessed hospital-specific, risk-adjusted rates of SSI using standard ACS-NSQIP techniques. In brief, we first fit a logistic regression model with SSI as the dependent variable and all potential patient risk factors as

independent variables. Patient risk factors in this model included functional status, ASA class, albumin, emergency surgery, laparoscopic approach, age, body mass index, race, gender, diabetes, and wound class. We then used this model to determine the expected probability of SSI for each patient. These expected values were summed at each hospital. We then calculated the ratio of observed to expected (“O/E ratio”) SSI and multiplied this by the average SSI rate to determine risk-adjusted SSI rates.

Estimating reliability

We next calculated the reliability of risk-adjusted SSI rates at each hospital. Reliability, measured from 0 to 1, can be thought of as the proportion of observed hospital variation that can be explained by true differences in quality.⁷ A reliability of 0 means that all of the variance in the outcome is due to measurement error, while a reliability of 1 means that all of the variance in outcome is due to true differences in performance. To perform this calculation, we used the following formula: Reliability = signal/(signal + noise). We estimated the “signal” using a hierarchical logistic regression model. In this model, the signal is the variance of the hospital random effect. We calculated “noise” using standard techniques for determining the standard error of a proportion. Commonly used cut-offs for acceptable reliability when comparing performance of groups and individuals are 0.70 and 0.90 respectively.⁷

Partitioning variation

We next determined the proportion of hospital variation attributable to patient factors, noise, and signal. The proportion of variation due to noise was calculated simply as 1-reliability, which was estimated as described above. To estimate the proportion of variation due to patient factors, we used two sequential random effects models. The first (“empty”) model was estimated with a random hospital effect but no patient characteristics. We then ran a second random effects model that included patient characteristics. Using standard techniques, we calculated the proportion of variation due to patient factors from the change in the variance of the random effect: (variance model1 – variance model2)/variance model1. To graphically demonstrate the proportion of variation due to each factor, we created hospital terciles (3 equal size groups). Statistical analyses were conducted using STATA 10 (College Station, TX).

RESULTS

A total of 18,455 patients from 181 ACS NSQIP participating hospitals underwent colon resections. The mean number of resections per hospital was 102 ± 65 . The mean risk-adjusted SSI rate per hospital was 10.5% with a range from 0 to 30%. Patient demographics are listed in **Table 1**.

The overall variation due to noise was 35%, while 7% was due to patient characteristics. 58% represented true differences in hospital SSI rates. When the caseloads were divided into low (< 65 cases), medium (65-115), and high (> 115) caseloads, the proportion of variation explained by patient factors was relatively constant across volume ranging from 3-5% (**Figure 1**). The number of patients in each group was 10,455 (low caseload), 5,850 (medium), and 2,150 (high).

The proportion of variation between hospitals attributable to patient risk factors remained relatively constant, ranging from 3-5%. The proportion of variation explained by noise decreased as caseload increased – from 57% to 28% to 18%. When hospital volume was graphed against reliability, there was increased reliability with increased hospital caseload (**Figure 2**). In order to achieve a cut-off of 70% reliability, a minimum of 94 cases had to be

reported. By this standard, only 54% of hospitals had enough cases for SSI rates to be considered a reliable quality indicator.

DISCUSSION

Surgical site infections (SSIs) are increasingly used as a measure of hospital quality.^{8,9} This study demonstrates that SSI rates are a reliable measure of hospital quality when an adequate number of cases have been reported. When the number of cases is low (< 65), more than 50% of variability between hospitals is due to statistical noise. When the number of cases reported is less than 94, reliability falls below the acceptable threshold of 70% reliability. Furthermore, for hospitals in the highest tertile by caseload, quality was the largest contributor to explaining variations in outcome. Although patient factors are important for explaining variation at the level of the individual, they contributed little overall to the variations in hospital outcomes.

Reliability is primarily driven by the number of cases and frequency of the outcome.^{7, 10} Previous studies have evaluated reliability of other outcome measures. Hofer et al evaluated the reliability of physician performance measures for diabetic care such as number of physician visits and hospitalizations, laboratory resource use, and adequacy of glycemic control as measured by hemoglobin A1c. They found that these performance measures, even after adjustment for case-mix, were only 40% reliable, meaning that 60% of the variation between physicians as due to noise.¹¹ Adams et al demonstrated that physician cost-profile scores, based on resource use for all episodes of care, were largely unreliable. Vascular surgery cost-profiles had the lower median reliability of 0.05 among the specialties, with reliabilities ranging from 0.05 to 0.79.¹⁰ Using 2007 ACS NSQIP data, Osborne et al demonstrated that as vascular surgery case volume increased across quartiles, the proportion of variation in mortality across hospitals due to statistical noise decreased from 94% in the lowest quartile to 64% in the highest quartile and the reliability of mortality as a quality indicator improved.¹²

Presently, about half of ACS NSQIP hospitals collecting data on colon resections submit enough cases to meet the threshold for 70% reliability. Despite the demonstrated validity of the ACS NSQIP methodology,⁵⁻⁶ reliability of its outcome measures is necessary to prevent misclassification of hospitals when ranking hospital performance. For example, Osborne et al demonstrated that 43% of hospitals participating in the ACS NSQIP vascular surgery program were misclassified into the wrong quartile when using standard regression methods, including 51% of the top quartile and 26% of the bottom quartile.¹² This misclassification can have significant implications for quality improvement efforts, public perception, and hospital finances in an era of pay for performance.

One potential solution might be to require ACS NSQIP participating hospitals to submit at least 94 colon cases so as to achieve at least 70% reliability. However, if other outcomes and types of surgeries are included, the number of cases that need to be reported to ensure reliability might be prohibitive, particularly for low volume hospitals. Furthermore, increasing reporting requirements would require more time and effort by the clinical nurse reviewer and could reduce the quality of other data collection efforts. The new generation of ACS NSQIP will address the problem of cost containment versus sufficient sampling to ensure reliability by using a 100% sampling strategy only for selected high risk procedures.¹³ An alternative solution would be to use a novel technique known as reliability adjustment.

Reliability adjustment is being increasingly used in quality measurement. This technique uses empirical Bayes methods to adjust for measurement error ("noise"), which is usually

due to low sample size or low event rates. As a result, unreliable outcomes from low volume hospitals will move closer to the mean, while more reliable estimates from higher volume hospitals will remain relatively stable. For example, low volume hospitals may be incorrectly classified as having extreme performance using standard analytic models, when the results were due to chance alone. Reliability adjustment would move those estimates closer to the mean and decrease the likelihood of classifying them as outliers. The disadvantages of reliability adjustment include the potential for overestimation of performance for low volume hospitals with high SSI rates; by shrinking their risk-adjusted outcomes towards the mean, we may be obscuring quality problems at low volume providers. Low volume hospitals with poor outcomes should be closely scrutinized and other additional methods for evaluating quality of care in these hospitals should be considered to avoid this problem. Lastly, there have not been any prospective studies demonstrating the superiority of reliability adjustment. Nonetheless, Dimick et al have demonstrated using cohort data that reliability adjustment for uncommon major surgical procedures such as abdominal aortic aneurysm repair or pancreatic resection significantly reduced variations in hospital mortality rates and improved the ability to predict future low mortality.¹⁴

This study has several limitations. First, the study uses ACS NSQIP data which selects a representative sample of cases to determine risk-adjusted outcomes. Use of only some rather than all cases may underestimate the reliability of SSIs and overestimate the percentage of low reliability hospitals. However, this methodology currently forms the basis for participating hospitals' quality improvement efforts. Second, only colon resection cases were included in the analysis. Because colon resections are a common and high-risk procedure, the reliability of superficial SSIs in this study may be higher than that for other procedures. Whether superficial SSIs are reliable across all surgical procedures or whether pooling rates across procedures to increase reliability will be appropriate to guide hospital quality improvement efforts are unclear.

In conclusion, superficial SSI rates after colon resections are a reliable indicator of hospital quality when the number of cases is adequate, likely due to the prevalence of both the procedure and the outcome. Consideration should be given to methods to increase the reliability of measured outcomes such as 100% sampling of targeted high risk procedures that will be used in the new generation of ACS NSQIP and/or reliability adjustment, particularly given the implications of misclassifying hospitals and surgeons based on performance.

Acknowledgments

This study was supported by a career development award to Dr. Dimick from the Agency for Healthcare Research and Quality (K08 HS017765), a research grant to Dr. Dimick from the National Institute of Diabetes and Digestive and Kidney Diseases (R21DK084397), and a career development award to Dr. Kao from the National Institutes of Health (K23 RR020020). The views expressed herein do not necessarily represent the views of Center for Medicare and Medicaid Services or the United States Government.

REFERENCES

1. Anderson DJ, Chen LF, Sexton DJ, Kaye KS. Complex surgical site infections and the devilish details of risk adjustment: important implications for public reporting. *Infect Control Hosp Epidemiol.* 2008; 29:941–6. [PubMed: 18808342]
2. Brandt C, Hansen S, Sohr D, Daschner F, Ruden H, Gastmeier P. Finding a method for optimizing risk adjustment when comparing surgical-site infection rates. *Infect Control Hosp Epidemiol.* 2004; 25:313–8. [PubMed: 15108729]

3. Nosocomial infection rates for interhospital comparison: limitations and possible solutions. A Report from the National Nosocomial Infections Surveillance (NNIS) System. *Infect Control Hosp Epidemiol.* 1991; 12:609–21. [PubMed: 1664844]
4. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. *Jama.* 2004; 292:847–51. [PubMed: 15315999]
5. Shiloach M, Frencher SK Jr, Steeger JE, et al. Toward robust information: data quality and inter-rater reliability in the American College of Surgeons National Surgical Quality Improvement Program. *J Am Coll Surg.* 2010; 210:6–16. [PubMed: 20123325]
6. Daley J, Forbes MG, Young GJ, et al. Validating risk-adjusted surgical outcomes: site visit assessment of process and structure. National VA Surgical Risk Study. *J Am Coll Surg.* 1997; 185:341–51. [PubMed: 9328382]
7. Adams, JL. The Reliability of Provider Profiling: A Tutorial. RAND Corporation; Santa Monica: 2009.
8. Smith RL, Bohl JK, McElearney ST, et al. Wound infection after elective colorectal resection. *Ann Surg.* 2004; 239:599–605. discussion -7. [PubMed: 15082963]
9. de Lissovoy G, Fraeman K, Hutchins V, Murphy D, Song D, Vaughn BB. Surgical site infection: incidence and impact on hospital utilization and treatment costs. *Am J Infect Control.* 2009; 37:387–97. [PubMed: 19398246]
10. Adams JL, Mehrotra A, Thomas JW, McGlynn EA. Physician cost profiling--reliability and risk of misclassification. *N Engl J Med.* 2010; 362:1014–21. [PubMed: 20237347]
11. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *Jama.* 1999; 281:2098–105. [PubMed: 10367820]
12. Osborne NH, Ko CY, Upchurch GR Jr, Dimick JB. The impact of adjusting for reliability on hospital quality rankings in vascular surgery. *J Vasc Surg.* 2011; 53:1–5. [PubMed: 21093202]
13. Birkmeyer JD, Shahian DM, Dimick JB, et al. Blueprint for a new American College of Surgeons: National Surgical Quality Improvement Program. *J Am Coll Surg.* 2008; 207:777–82. [PubMed: 18954793]
14. Dimick JB, Staiger DO, Birkmeyer JD. Ranking hospitals on surgical mortality: the importance of reliability adjustment. *Health Serv Res.* 2010; 45:1614–29. [PubMed: 20722747]

Figure 1. The proportion of superficial SSIs after colon resections that are attributable to patient factors, “noise” or measurement error, and hospital performance.

Figure 2.
Relationship between reliability and hospital caseload of colon resections based on the ACS NSQIP 2007 database.

Table 1

Demographics of patients who underwent colon resections included in the 2007 ACS NSQIP database.

Preoperative patient characteristics		<u>Risk-adjusted Hospital Rates of Superficial Surgical Site Infection</u>		
		Low (0% to 8%)	Medium (8% to 12%)	High (12% to 28%)
Number of patients, N (%)		6,204 (33%)	6,107 (33%)	6,144 (33%)
Age in years, mean		62	63	63
Non-white race, N (%)		1,451 (23%)	1,481 (24%)	1,352 (22%)
Female, N (%)		3,201 (52%)	3,232 (53%)	3,136 (51%)
Body mass index, mean		27.7	27.8	27.9
Emergency operation		932 (15%)	1,119 (18%)	976 (16%)
Dependent functional status		788 (13%)	827 (14%)	734 (12%)
Laparoscopic approach		1,838 (30%)	1,698 (28%)	1,856 (30%)
ASA Score 3 or higher		3,194 (51%)	3,243 (53%)	3,280 (53%)
Wound class contaminated	Clean	4,343 (70%)	4,160 (68%)	4,376 (71%)
		1,032 (17%)	1,048 (17%)	1,030 (17%)
	Contaminated Dirty/Infected	826 (13%)	897 (15%)	737 (12%)
Serum albumin level, mean (SD)		3.6 (0.8)	3.6 (0.8)	3.5 (0.8)
Diabetes mellitus dependant	Insulin	313 (5%)	301 (5%)	328 (5%)
	Non-insulin dependant	596 (10%)	530 (9%)	564 (9%)
Operative time, minutes		150 (79)	160 (84)	164 (88)
Work Relative Value Unit (RVU)		25 (5)	26 (5)	26 (5)