Surgical Site Infections

Volume-Outcome Relationship and Year-to-Year Stability of Performance Rankings

Michael S. Calderwood, MD, MPH,*† Ken Kleinman, ScD,‡ Susan S. Huang, MD, MPH,§
Michael V. Murphy, BA,† Deborah S. Yokoe, MD, MPH,* and Richard Platt, MD, MSc†

Background: Surgical site infection (SSI) rates are publicly reported as quality metrics and increasingly used to determine financial reimbursement.

Objective: To evaluate the volume-outcome relationship as well as the year-to-year stability of performance rankings following coronary artery bypass graft (CABG) surgery and hip arthroplasty.

Research Design: We performed a retrospective cohort study of Medicare beneficiaries who underwent CABG surgery or hip arthroplasty at US hospitals from 2005 to 2011, with outcomes analyzed through March 2012. Nationally validated claims-based surveillance methods were used to assess for SSI within 90 days of surgery. The relationship between procedure volume and SSI rate was assessed using logistic regression and generalized additive modeling. Year-to-year stability of SSI rates was evaluated using logistic regression to assess hospitals' movement in and out of performance rankings linked to financial penalties.

Results: Case-mix adjusted SSI risk based on claims was highest in hospitals performing <50 CABG/year and <200 hip arthroplasty/year compared with hospitals performing ≥ 200 procedures/year.

At that same time, hospitals in the worst quartile in a given year based on claims had a low probability of remaining in that quartile the following year. This probability increased with volume, and when using 2 years' experience, but the highest probabilities were only 0.59 for CABG (95% confidence interval, 0.52–0.66) and 0.48 for hip arthroplasty (95% confidence interval, 0.42–0.55).

Conclusions: Aggregate SSI risk is highest in hospitals with low annual procedure volumes, yet these hospitals are currently excluded from quality reporting. Even for higher volume hospitals, year-to-year random variation makes past experience an unreliable estimator of current performance.

Key Words: hospital quality, public reporting, performance measurement, health care–associated infection, Medicare, health care reform, hospital reimbursement, policy

Quality metrics are increasingly used to determine reimbursement and to guide consumer choice. As an example, the Centers for Medicare & Medicaid Services (CMS) has linked payment to specific quality measures through both the Hospital-Acquired Condition (HAC) Reduction Program and the Hospital Value-Based Purchasing Program.[1–3] These 2 programs put hospitals at risk of losing over $1.9 billion in annual revenue through a reduction in payment to those deemed to be providing lower quality care and a redistribution of payment to others deemed to be providing higher quality care.[1] Surgical site infection (SSI) rates are one of the major hospital quality metrics being tracked by these programs, with SSI rates for specific procedures independently contributing to the aggregate quality scores used to determine hospital reimbursement. SSI rates for various procedures are also publicly reported by both CMS and state health departments.[4,5]

There is an assumption that public reporting of quality metrics can be used by patients to select hospitals providing higher quality care, and that financial disincentives will improve poorly performing hospitals.[2,3] This assumption, however, is predicated on reporting metrics that are sufficiently meaningful to judge the quality of care.

One concern is that hospitals performing a low volume of procedures are often excluded from public reporting due to insufficient data,[4,6] despite evidence of worse outcomes in

these hospitals for postoperative complications, costs, and mortality.[7–11] Although state health departments typically exclude hospitals performing <20 procedures per year, CMS now excludes hospitals with <1 expected SSI in a year based on procedure volume.[4,6] This means that hospitals performing <50 procedures per year get excluded when the average SSI rate is 2%.

At the same time, for the hospitals that are deemed to have sufficient volume to allow performance ranking, it is unknown whether hospitals that are identified as outliers in any given year have a high likelihood for similar performance in the next year. Prior studies have raised concern that much of the year-to-year variability in performance metrics is due to random variation.[12–15] Although unproven to date, public reporting of the past year's performance leads patients, regulatory agencies, and payers to assume that patients will experience similar care in the coming year.

We sought to evaluate the volume-outcome relationship as well as the year-to-year stability of performance rankings following coronary artery bypass graft (CABG) surgery and hip arthroplasty, 2 procedures which have been a focus in state quality reports comparing hospital performance.[5] Our hypotheses were that low volume hospitals would have higher SSI rates and that past performance would not necessarily be a good predictor of future performance.

METHODS

Study Population

We used 2005–2011 Medicare Provider Analysis and Review Research Identifiable Files to identify all short-stay acute care US hospitals performing CABG and primary hip arthroplasty on fee-for-service Medicare patients between January 1, 2005 and December 31, 2011, based on Medicare Part A inpatient claims data.[16] We refer to these below as Medicare patients and Medicare procedures.

We identified CABG cases using International Classification of Diseases, Ninth Revision (ICD-9) codes 36.10–36.17, 36.19, and 36.2, and primary hip arthroplasty cases using ICD-9 codes 81.51 and 81.52, with multiple codes for the same procedure type on the same day counted as a single case.[17] We then assessed claims within 90 days of the surgical procedure for ICD-9 codes suggestive of a deep and organ/space SSI, including ICD-9 diagnosis codes 513.1, 682.2, 730.08, 996.61, 996.62, 998.31, 998.32, 998.51, and 998.59 following CABG, and 996.60, 996.66, 996.67, 996.69, 998.51, and 998.59 following primary hip arthroplasty.[18] Prior national studies showed that these codes accurately rank hospitals when compared with expert adjudication of full text medical records, often identifying cases missed by hospital surveillance.[19–21] Therefore, we feel this is an acceptable methodology for comparing SSI outcomes across hospitals. In fact, similar methodology is currently being used by the CMS Hospital Inpatient Quality Reporting Program to validate hospital SSI reporting.[22]

To focus on surgical complications rather than preexisting infections, we excluded all codes listed as present on admission at the time of surgery, while including present on admission codes during any readmissions to a hospital within 90 days of the surgical procedure. For patients who underwent another major surgery in the 90-day postoperative surveillance window, we censored our surveillance at the time of the subsequent surgery.[17] We used 2012 Medicare Provider Analysis and Review Research Identifiable Files data to capture coding and readmissions for procedures performed in the last 90 days of 2011.

For each patient in our CABG and primary hip arthroplasty cohorts, we collected data on age, sex, and comorbidities at the time of the surgical admission for individual-level risk adjustment. Comorbidities were assessed using publicly available comorbidity software from the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project, based on methods previously described by Elixhauser et al.[23,24]

SSI Risk by Procedure Volume

Medicare procedure volume was categorized into 1–24, 25–49, 50–99, 100–199, or 200+ surgical cases on Medicare patients each year, and hospitals could change volume categories across years. These categories correspond to hospitals that would be excluded from public reporting based on expected SSI rates of 4%, 2%, 1%, 0.05%, respectively, using the current CMS methodology of excluding hospitals with <1 expected SSI in a given year based on procedure volume.[4]

We used logistic regression to calculate the odds of having an SSI code by the annual Medicare surgical volume. Generalized estimating equations were used to control for repeated measures across years within individual hospitals, and the model was adjusted for the age, sex, and comorbidities of each Medicare patient. We also examined the relationship between continuous procedure volume and the probability of a patient having an SSI code graphically using generalized additive models.[25]

A more in depth discussion of the statistical methods and models is included in a supplementary appendix, Supplemental Digital Content 1 (http://links.lww.com/MLR/B247).

SSI Risk Over Time

Within each year, we fit a logistic regression mixed effects model with random intercepts for each hospital, including each individual's age, sex, and coded comorbidities as fixed effects. The predicted random intercepts from the models for each year indicate case mix-adjusted relative performance. Hospitals were ranked based on these predicted random intercepts and divided into case mix-adjusted quartiles of performance. We then used logistic regression, with generalized estimating equations to account for repeated measures across years within hospital, to model the outcome of being in the worst quartile next year based on: (1) worst, middle 2, and best quartile status this year, for all hospitals combined and by volume category; (2) worst, middle 2, and best quartile status this year and the year prior, by volume category. Our modeling of quartile status was based on the fact that the CMS HAC Reduction Program uses quartiles to determine financial penalties.

We also plotted the performance rank of randomly selected individual hospitals from 2005 through 2011. This allowed us to examine the stability of SSI performance visually.

RESULTS

Study Population

From 2005 through 2011, an average of 1165 US hospitals performed CABG procedures and 3117 US hospitals performed hip arthroplasty procedures for fee-for-service Medicare patients each year. An average of 135,882 CABG procedures and 217,223 hip arthroplasty procedures were performed each year. The median number of annual fee-for-service Medicare surgical cases per hospital was 89 for CABG (interquartile range, 46–150) and 46 for hip arthroplasty (interquartile range, 19–95). There was a 17% decline in the number of CABG procedures and a 17% increase in the number of hip arthroplasty procedures performed per year from 2005 through 2011.

SSI Risk by Procedure Volume

Table 1 shows SSI coding rates in US hospitals by the annual volume of surgical cases performed on Medicare patients. The table also shows the unadjusted and adjusted odds of SSI for each group compared with the largest volume hospitals. For both CABG and hip arthroplasty, coded SSI rates were highest at hospitals that performed the fewest procedures per year.

For CABG, patients at hospitals performing <50 Medicare procedures per year had significantly higher odds of an SSI code: 30% higher for patients in hospitals performing fewer than 25 procedures and 21% higher in hospitals performing 25–49 procedures. The risk in hospitals that performed 50–199 procedures per year was not different from those performing ≥ 200 procedures. Hospitals performing fewer than 50 procedures per year accounted for more than a quarter of all US hospitals performing CABG

procedures from 2005 to 2011, while only accounting for 7%–8% of the surgical volume.

For hip arthroplasty, there was an inverse correlation between procedure volume and patients' SSI risk, with patients at hospitals in each of the 4 lower volume categories having significantly higher risk than patients at hospitals performing ≥ 200 procedures per year. The excess odds of an SSI code ranged from 58% in hospitals performing fewer than 25 procedures to 14% in hospitals performing 100–199 procedures. Over 90% of US hospitals performing hip arthroplasty procedures from 2005 to 2011 had SSI risks that were significantly higher than those in the largest volume hospitals.

Figure 1 plots the probability of a patient having an SSI code by volume of Medicare procedures. The data shown is for 2011, but the trends were similar for each year. For CABG, coded SSI rates decreased with increasing volume up to approximately 100 Medicare fee-for-service procedures per year, with similar risk above 100 procedures per year. For hip arthroplasty, SSI risk continued to decline as hospital volume increased.

Stability of SSI Performance over Time

Table 2 shows the probability that a hospital ranked in the worst quartile in 1 year remained in that quartile the following year based on claims data suggestive of SSI.

For CABG, hospitals ranked in the worst quartile in the current year had a 44% chance of remaining in that quartile the following year. For comparison, the chance of remaining in the worst quartile would be 25% if quartile rankings were solely due to random chance. The probability of remaining in the worst quartile was directly related to a hospital's annual procedure volume. For example, the probability of remaining in the worst quartile for hospitals performing fewer than 25 CABG procedures annually on Medicare patients was 0.20, and statistically no different from chance [95% confidence interval (CI), 0.14–0.29], while this probability was 0.52 in the highest volume hospitals (95% CI, 0.45–0.60).

TABLE 1. Procedure Volume and Rates of Assignment of Surgical Site Infection Codes in Medicare Fee-for Service Patients

Annual Procedure Volume (Medicare)	Hospitals Performing Procedure* [N (%)]	Annual No. Surgical Cases* [N (%)]	Annual Coded Infections*	Coded SSI Rate (%)	Crude OR (95% Confidence Interval)[†]	Adjusted OR (95% Confidence Interval)[‡]
2005–2011 CABG procedures in US hospitals						
1–24	121 (10)	1581 (1.2)	120	7.57	1.29 (1.16–1.43)	**1.30 (1.17–1.44)**
25–49	202 (17)	7522 (5.5)	529	7.04	1.19 (1.11–1.28)	**1.21 (1.12–1.30)**
50–99	323 (28)	23,714 (17.5)	1448	6.11	1.02 (0.96–1.09)	1.04 (0.98–1.11)
100–199	343 (29)	48,013 (35.3)	2811	5.85	0.98 (0.92–1.04)	1.01 (0.95–1.07)
200+	176 (15)	55,052 (40.5)	3290	5.98	Ref.	Ref.
2005–2011 hip arthroplasty procedures in US hospitals						
1–24	984 (32)	11,250 (5.2)	348	3.10	1.63 (1.51–1.75)	**1.58 (1.47–1.69)**
25–49	657 (21)	23,809 (11.0)	628	2.64	1.38 (1.29–1.48)	**1.34 (1.26–1.44)**
50–99	744 (24)	53,234 (24.5)	1274	2.39	1.25 (1.17–1.33)	**1.22 (1.15–1.30)**
100–199	535 (17)	73,201 (33.7)	1631	2.23	1.16 (1.09–1.23)	**1.14 (1.07–1.21)**
200+	195 (6)	55,723 (25.7)	1074	1.93	Ref.	Ref.

Adjusted odds ratios in bold text were significantly higher than the null value of 1 ($P < 0.05$).

*Average across 2005–2011.

[†]Logistic regression model with generalized estimating equations to account for clustering within hospitals, robust standard errors used for confidence interval.

[‡]Logistic regression model adjusted for age, sex, and coded co-morbidities, with generalized estimating equations to account for clustering within hospitals, robust standard errors used for confidence interval.

CABG indicates coronary artery bypass graft; OR, odds ratio; Ref., reference; SSI, surgical site infection.
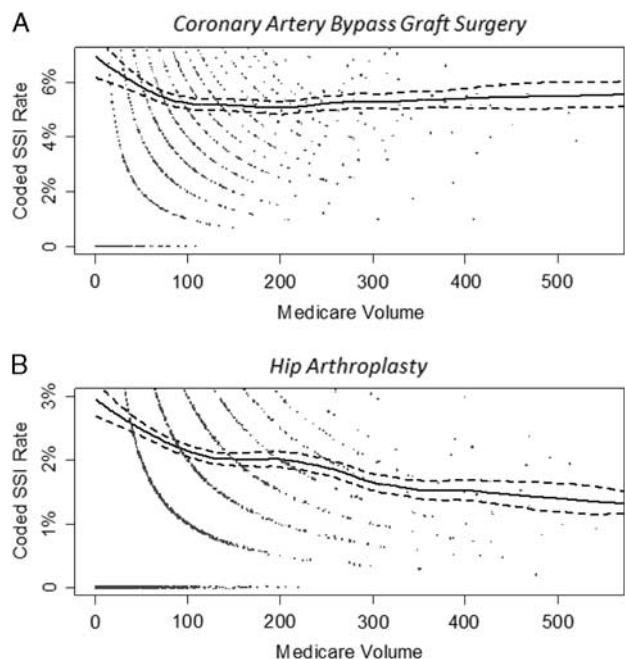
FIGURE 1. These are plots of the coding for surgical site infection by volume of procedures performed on fee-for-service Medicare patients who underwent coronary artery bypass graft surgery (A) and primary hip arthroplasty (B) in US hospitals in 2011. The dots represent the coding percentages from individual hospitals, jittered slightly to prevent overplotting. The solid line represents 100 times the predicted probability of surgical site infection coding obtained from a generalized additive model, and the dashed lines represent the 95% confidence interval for the predicted probability line. SSI indicates surgical site infection.

For hip arthroplasty, hospitals ranked in the worst quartile in the current year had a 34% chance of remaining in that quartile the following year. Similar to CABG, the probability of remaining in the worst quartile for hospitals performing fewer than 25 hip arthroplasty procedures annually on Medicare patients was no different from chance (0.26, 95% CI, 0.23–0.29). In the highest volume hospitals, the probability was 0.44 (95% CI, 0.37–0.50).

The modeled probability of moving from the best quartile to the worst quartile for CABG was 0.11 and for hip was 0.18. In other words, 1 of 9 hospitals ranked in the best quartile in the current year based on coding suggestive of SSI following CABG and 1 of 6 hospitals ranked in the best quartile in the current year based on coding suggestive of SSI

following hip arthroplasty would be expected to be ranked in the worst quartile the following year. This movement in annual performance ranking is demonstrated graphically in Figure 2, which plots the yearly case mix-adjusted ranks for 5 randomly selected hospitals in each volume category. Regardless of annual procedure volume, there was great variability in the year-to-year rank for each hospital from 2005 to 2011. In fact, we observed consecutive year movement from the worst quartile to the best quartile and then back to the worst quartile in a number of hospitals for both procedures.

The effect of using 2 years of data to predict the next year's performance is shown in Table 3. For both CABG and hip arthroplasty, hospitals ranked in the worst quartile 2 years in a row had a slightly higher probability of being ranked in the worst quartile the following year. However, even for the largest hospitals, the probabilities were still only 0.59 for CABG (95% CI, 0.52–0.66) and 0.48 for hip arthroplasty (95% CI, 0.42–0.55).

CONCLUSIONS

Public reporting of SSI rates and their role in determining hospital reimbursement have led to an increasing national focus on these potentially preventable health care–associated infections.[26] However, data from hospitals with low procedure volume are often ignored due to well founded concerns about instability in their individual rates. This exclusion is important, as our data suggest that procedure volume is a strong predictor of outcome following CABG and hip arthroplasty, with the highest SSI risk found in those hospitals currently excluded from public scrutiny. In addition, while SSI rates have declined nationally over time,[27] our modeling of longitudinal data from individual hospitals suggests that annual SSI performance rankings may be highly unstable from year-to-year, even for hospitals performing many procedures. This raises important concerns about the use of SSI data for both quality reporting and reimbursement programs.

Our finding that hospitals that perform a lower volume of surgical procedures have a higher SSI risk adds to prior literature showing volume-outcome associations for post-operative complications, costs, and mortality.[7–11] In one study, lower surgical procedure volume was associated with longer operative durations, a risk factor for SSI.[28] It is also possible that infection prevention practices are more standardized at hospitals with higher surgical volumes.[9] Whatever the reason,

TABLE 2. Probability of Worst Quartile Ranking Next Year for a Hospital With a Worst Quartile Ranking this Year (2005–2011)

Overall Probability (Ignoring Volume)	1–24 Procedures/y	25–49 Procedures/y	50–99 Procedures/y	100–199 Procedures/y	200+ Procedures/y
		Coronary artery bypass graft surgery			
0.44 (0.41–0.47)	0.20 (0.14–0.29)	**0.42 (0.36–0.48)**	**0.42 (0.37–0.47)**	**0.49 (0.44–0.54)**	**0.52 (0.45–0.60)**
		Hip arthroplasty			
0.34 (0.32–0.35)	0.26 (0.23–0.29)	**0.32 (0.29–0.35)**	**0.38 (0.36–0.41)**	**0.43 (0.39–0.46)**	**0.44 (0.37–0.50)**

Probabilities presented with 95% confidence intervals. Hospital quartile rankings in each year determined by a logistic regression model adjusted for age, sex, and co-morbidities, with generalized estimating equations to account for clustering within hospitals. Volume categories determined by surgical procedures performed on fee-for-service Medicare patients. The probabilities in bold text have confidence intervals that exclude the null value of 0.25, the probability of being in the worst quartile by chance.
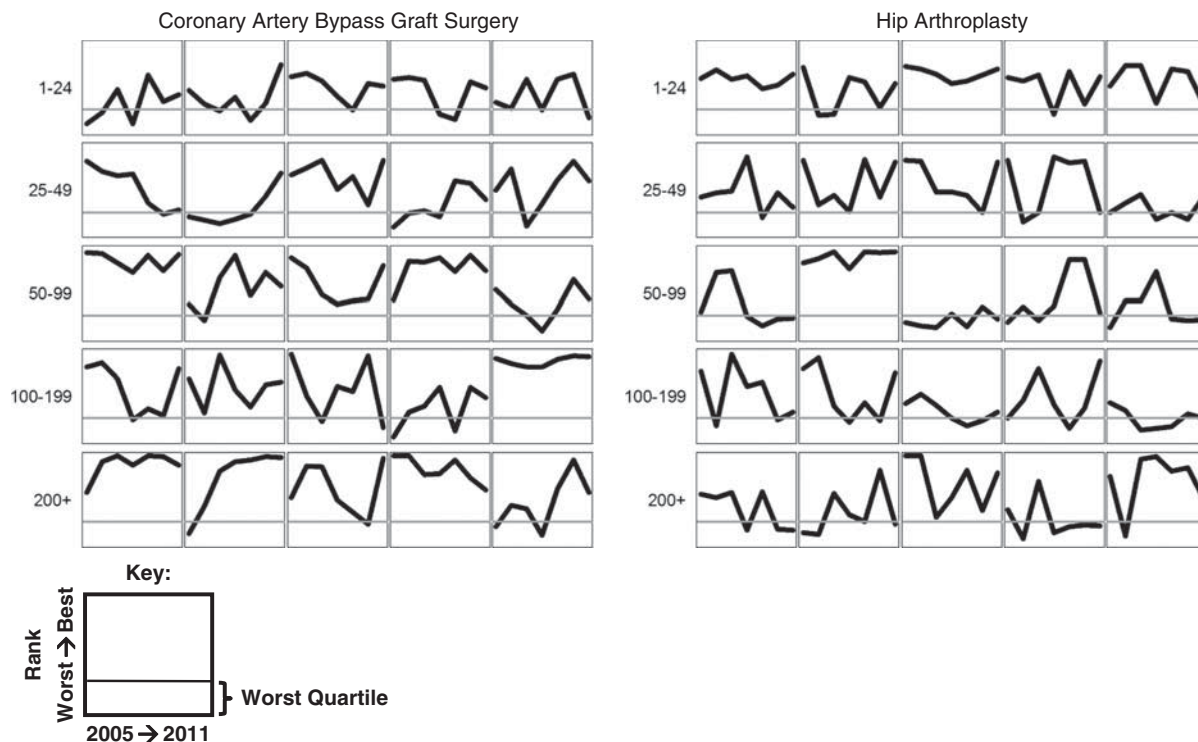
FIGURE 2. For each volume category, 5 hospitals were selected at random to show the case mix-adjusted rank from 2005 to 2011 relative to other hospitals performing coronary artery bypass graft surgery and hip arthroplasty for fee-for-service Medicare patients. This shows the variability in performance ranking over time.

the association is robust, and procedure volume should be considered a strong determinant of SSI risk. Instead, current policies ignore hospitals with lower procedure volume.

As for why coded SSI rates declined with increasing surgical volume only up to around 100 annual CABG cases on Medicare fee-for-service patients, while the volume-outcome relationship continued beyond this surgical case volume for hip arthroplasty, we do not know of existing literature highlighting this finding. It is not clear why this should differ between CABG and hip arthroplasty, and it is worth further investigation.

Two recent papers assessed the hospitals being penalized by the HAC Reduction Program and the Hospital Value-Based Purchasing Program.[29,30] These programs appear to disproportionately penalize larger volume hospitals, and major teaching hospitals in particular. Exclusion of hospitals with low surgical volume is an important contributor to this phenomenon.

Even among the included hospitals, there is a high degree of variability in the year-to-year performance ranking of individual hospitals. It is not clear that outcome from a single year, or even 2 years, are truly indicative of better or worse performance. Similar instability was previously reported for metrics comparing postoperative mortality. This prior work showed a failure to predict future outcomes based on past performance, with an argument that this was due to an insufficient number of procedures to adequately assess outcomes in relation to other hospitals.[12,14,31] For this reason, some have

advocated for changes in sampling strategies to capture additional cases in an effort to improve reliability.[13,15]

In exploring the impact of a broader sampling strategy, we did find a modest improvement in prediction by using quartile ranking over 2 consecutive years to predict whether a hospital would remain in the worst quartile. Even with this extra year of data, however, prediction remained poor. This was true even in the highest volume group, where 41% of hospitals performing CABG and 52% of hospitals performing hip arthroplasty were no longer in the worst quartile in the subsequent year despite having been in the worst quartile for 2 consecutive years. In addition, broader sampling strategies may compromise interpretability by combining data from different periods and the possibility for direct improvement based on early indicators of worse outcomes.

One interesting article suggested a statistic to measure rankability. Unfortunately, since the statistic depends on fitting a fixed effects model, we were unable to estimate it in our full sample of hospitals, due to the relative rareness of the outcome and small number of procedures in some hospitals.[14,32] Instead, our analyses used empirical Bayes methods in an effort to improve reliability when analyzing year-to-year performance ranking.[19,20] These methods have been advocated by others comparing quality outcomes between hospitals.[33,34] Although there are benefits to this approach, these methods pull hospitals with low procedure volumes toward the middle of the distribution, meaning they

TABLE 3. Probability of Worst Quartile Ranking Next Year Based on 2 Years of Performance Data (2005–2011)

	1–24 Procedures/y	25–49 Procedures/y	50–99 Procedures/y	100–199 Procedures/y	200+ Procedures/y
Probability of worst quartile ranking next year if					
Coronary artery bypass graft surgery					
Worst quartile this year and					
Best quartile year prior	**0.14 (0.08-0.21)***	0.29 (0.23-0.35)	0.26 (0.22-0.32)	**0.36 (0.30-0.42)**	**0.35 (0.28-0.42)**
Middle quartiles year prior	0.18 (0.12-0.27)	**0.37 (0.31-0.43)**	**0.34 (0.29-0.38)**	**0.44 (0.39-0.49)**	**0.43 (0.36-0.51)**
Worst quartile year prior	0.30 (0.20-0.42)	**0.52 (0.45-0.59)**	**0.49 (0.45-0.55)**	**0.60 (0.55-0.65)**	**0.59 (0.52-0.66)**
Hip arthroplasty					
Worst quartile this year and					
Best quartile year prior	**0.19 (0.17-0.22)***	0.26 (0.23-0.29)	**0.31 (0.28-0.34)**	**0.34 (0.31-0.38)**	**0.34 (0.28-0.40)**
Middle quartiles year prior	0.24 (0.21-0.27)	**0.31 (0.28-0.35)**	**0.37 (0.34-0.40)**	**0.40 (0.37-0.44)**	**0.40 (0.34-0.46)**
Worst quartile year prior	**0.30 (0.27-0.34)**	**0.39 (0.35-0.43)**	**0.45 (0.42-0.49)**	**0.49 (0.45-0.53)**	**0.48 (0.42-0.55)**

Probabilities presented with 95% confidence intervals. Hospital quartile rankings in each year determined by a logistic regression model adjusted for age, sex, and comorbidities, with generalized estimating equations to account for clustering within hospitals. Volume categories determined by surgical procedures performed on fee-for-service Medicare patients. The probabilities in bold text have confidence intervals (CI) that excludes the null value of 0.25, the probability of being in the worst quartile by chance.
*Cells where the hospitals have CI that excludes the null value and a reduced probability of a worst quartile ranking. Otherwise, all bolded entries have an increased probability.

are less likely to appear in the extreme quartiles. In seeking greater certainty, the outcomes in many small volume hospitals are ignored through these empirical Bayes methods, despite the fact that in the aggregate these hospitals have the highest complication rates. At the same time, the variability that we noted in year-to-year performance ranking was based on hospitals with the most robust data.

Our study does have some limitations. First, our analyses used claims data to identify SSI outcomes rather than data from hospital-based prospective surveillance systems. For both CABG and hip arthroplasty, we previously published national validation work showing that billing codes that are suggestive of SSI successfully identify reportable infections with similar confirmation rates among patients with an SSI code in hospitals with high versus low SSI coding. This suggests that hospitals with higher rates of SSI coding also have higher rates of chart confirmed SSIs.[18–21] We believe claims-based ranking is both more complete because it includes the substantial number of SSIs that become manifest after discharge and do not return to the hospital that performed the surgery, and also superior in that this method minimizes variable implementation of surveillance between hospitals.

Second, our analyses were limited to Medicare fee-for-service patients. In general, though, Medicare surgical volume tracks well with overall surgical volume when looking at data from State Inpatient Databases.[35] In addition, there is no reason to think that the selected diagnosis codes would be used differently in Medicare fee-for-service patients versus other patients in a way that would influence our results.

Third, we only included inpatient claims data, and did not look at coding for potential SSIs that may have been treated in the outpatient setting. In prior work, however, we showed that our SSI codes identified 87% of deep incisional primary and organ/space SSIs following CABG and 92% following hip arthroplasty when limiting to Part A inpatient claims data.[18]

Fourth, it is possible that surgical procedures with higher SSI rates might have different year-to-year variability. An example is colon surgery that has national SSI rates of 5.6%, compared with 1.3% for primary hip replacement and 2.8% for CABG.[36]

Fifth, we only used age, sex, and comorbidity data coded in inpatient claims for case-mix adjustment. It is possible that other factors not available in claims could further improve our case-mix adjustment, but a hospital's case-mix typically does not change significantly from year-to-year, so we do not believe that this significantly impacts our study findings.

Although we recognize that any grading system is subject to year-to-year reclassification, it is important to consider when this variation is so great that the process being used to compare hospital performance loses its value. The high degree of year-to-year variation in the performance rank of individual hospitals, shown in our data (Fig. 2), raises concerns about the value of a hospital's rank in any given year. Instead, there should be a greater focus on sustained performance trends which are more likely to be a better measure of quality. One alternative would be to financially reward hospitals with sustained longitudinal improvement, rather than focusing on performance in a single year or 2 years.

This is important, as hospital SSI rates following colon surgery and abdominal hysterectomy are now being publicly reported and used to determine CMS reimbursement.[1,4] If the problems we identified for CABG and hip arthroplasty apply for these other 2 procedures, hospitals may incur financial penalties and patients may make choices on the basis of information that is a poor measure of the hospital's current performance.

SSI risk for CABG and hip arthroplasty is greatest in the large number of US hospitals performing too few procedures to be included in public reporting. At the same time, for those hospitals that are included in public reports, past performance is a poor predictor of future performance. Assigning financial penalties and providing guidance to the public based on current SSI performance metrics risks ignoring the large group of low volume hospitals that have the poorest outcomes, and targeting hospitals whose outlier status in a given year may be due to random variation.

REFERENCES

1. Department of Health and Human Services, Centers for Medicare & Medicaid Services. Medicare program; hospital inpatient prospective payment systems for acute care hospitals and the long-term care hospital prospective payment system and fiscal year 2015 rates; quality reporting requirements for specific providers; reasonable compensation equivalents for physician services in excluded hospitals and certain teaching hospitals; provider administrative appeals and judicial review; enforcement provisions for organ transplant centers; and electronic health record (EHR) incentive program. Final rule. *Fed Register*. 2014;79: 49853–50536. Available at: www.gpo.gov/fdsys/pkg/FR-2014-08-22/pdf/2014-18545.pdf.
2. VanLare JM, Conway PH. Value-based purchasing—national programs to move from volume to value. *N Engl J Med*. 2012;367:292–295.
3. Blumenthal D, Jena AB. Hospital value-based purchasing. *J Hosp Med*. 2013;8:271–277.
4. Medicare.gov. Hospital compare. Available at: www.medicare.gov/hospitalcompare/search.html. Accessed May 24, 2016.
5. Makary MA, Aswani MS, Ibrahim AM, et al. Variation in surgical site infection monitoring and reporting state. *J Healthc Qual*. 2013;35:41–46.
6. McKibben L, Horan TC, Tokars JI, et al. Guidance on public reporting of healthcare-associated infections: recommendations of the Healthcare Infection Control Practices Advisory Committee. *Infect Control Hosp Epidemiol*. 2005;26:580–587.
7. Katz JN, Losina E, Barrett J, et al. Association between hospital and surgeon procedure volume and outcomes of total hip replacement in the United States medicare population. *J Bone Joint Surg Am*. 2001;83-A:1622–1629.
8. Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002;346:1128–1137.
9. Geubbels EL, Wille JC, Nagelkerke NJ, et al. Hospital-related determinants for surgical-site infection following hip arthroplasty. *Infect Control Hosp Epidemiol*. 2005;26:435–441.
10. Browne JA, Pietrobon R, Olson SA. Hip fracture outcomes: does surgeon or hospital volume really matter? *J Trauma*. 2009;66:809–814.
11. Post PN, Kuijpers M, Ebels T, et al. The relation between volume and outcome of coronary interventions: a systematic review and meta-analysis. *Eur Heart J*. 2010;31:1985–1992.
12. Birkmeyer JD, Dimick JB. Understanding and reducing variation in surgical mortality. *Annu Rev Med*. 2009;60:405–415.
13. van Dishoeck AM, Koek MB, Steyerberg EW, et al. Use of surgical-site infection rates to rank hospital performance across several types of surgery. *Br J Surg*. 2013;100:628–636.
14. Henneman D, van Bommel AC, Snijders A, et al. Ranking and rankability of hospital postoperative mortality rates in colorectal cancer surgery. *Ann Surg*. 2014;259:844–849.
15. Kao LS, Ghaferi AA, Ko CY, et al. Reliability of superficial surgical site infections as a hospital quality measure. *J Am Coll Surg*. 2011;213:231–235.
16. Research Data Assistance Center. Medicare provider analysis and review research identifiable files (Research Data Assistance Center web site). Available at: www.resdac.org/cms-data/files/medpar-rif. Accessed May 24, 2016.
17. The Joint Commission. Surgeries targeted by the Surgical Care Improvement Project, as outlined in The Specifications Manual for National Hospital Inpatient Quality Measures (Joint Commission web site). Available at: www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx. Accessed May 24, 2016.
18. Calderwood MS, Kleinman K, Murphy MV, et al. Improving public reporting and data validation for complex surgical site infections following coronary artery bypass graft surgery and hip arthroplasty. *Open Forum Infect Dis*. 2014;1:ofu106.
19. Huang SS, Placzek H, Livingston J, et al. Use of Medicare claims to rank hospitals by surgical site infection (SSI) risk following coronary artery bypass graft surgery. *Infect Control Hosp Epidemiol*. 2011;32:775–783.
20. Calderwood MS, Kleinman K, Bratzler DW, et al. Use of Medicare claims to identify US hospitals with a high rate of surgical site infection after hip arthroplasty. *Infect Control Hosp Epidemiol*. 2013;34:31–39.
21. Calderwood MS, Ma A, Khan YM, et al. Use of Medicare diagnosis and procedure codes to improve detection of surgical site infections following hip arthroplasty, knee arthroplasty, and vascular surgery. *Infect Control Hosp Epidemiol*. 2012;33:40–49.
22. Centers for Medicare & Medicaid Services, Department of Health and Human Services. HAI measures in the Hospital Inpatient Quality Reporting (IQR) Program to be validated for the FY 2015 payment determination and subsequent years, (ii) targeting SSI for validation. *Fed Register*. 2012;77:53545–53547. Available at: www.gpo.gov/fdsys/pkg/FR-2012-08-31/pdf/2012-19079.pdf.
23. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. *Med Care*. 1998;36:8–27.
24. Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project, comorbidity software, version 3.7 (Agency for Healthcare Research and Quality web site). Available at: www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp. Accessed May 24, 2016.
25. Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. Boca Raton, FL: Chapman and Hal/CRC; 1990.
26. Lee GM, Hartmann CW, Graham D, et al. Perceived impact of the Medicare policy to adjust payment for healthcare-associated infections. *Am J Infect Control*. 2012;40:314–319.
27. Centers for Disease Control and Prevention. Healthcare-Associated Infections (HAI) Progress Report (Centers for Disease Control and Prevention web site). 2015. Available at: www.cdc.gov/hai/progress-report/. Accessed September 21, 2015.
28. Muilwijk J, van den Hof S, Wille JC. Associations between surgical site infection risk and hospital operation volume and surgeon operation volume among hospitals in the Dutch nosocomial infection surveillance network. *Infect Control Hosp Epidemiol*. 2007;28:557–563.
29. Rajaram R, Chung JW, Kinnier CV, et al. Hospital characteristics associated with penalties in the Centers for Medicare & Medicaid Services Hospital-Acquired Condition Reduction Program. *JAMA*. 2015;314:375–383.
30. Kahn CN 3rd, Ault T, Potetz L, et al. Assessing Medicare's Hospital Pay-For-Performance Programs and whether they are achieving their goals. *Health Affairs (Millwood)*. 2015;34:1281–1288.
31. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. *JAMA*. 2004;292:847–851.
32. Lingsma HF, Steyerberg EW, Eijkemans MJ, et al. Comparing and ranking hospitals based on outcome: results from The Netherlands Stroke Survey. *QJM*. 2010;103:99–108.
33. Dimick JB, Ghaferi AA, Osborne NH, et al. Reliability adjustment for reporting hospital outcomes with surgery. *Ann Surg*. 2012;255:703–707.
34. Krell RW, Hozain A, Kao LS, et al. Reliability of risk-adjusted outcomes for profiling hospital surgical quality. *JAMA Surg*. 2014;149:467–474.
35. Agency for Healthcare Research and Quality. Overview of the state inpatient databases (Healthcare Cost and Utilization Project web site). Available at: www.hcup-us.ahrq.gov/sidoverview.jsp. Accessed on May 24, 2016.
36. Edwards JR, Peterson KD, Mu Y, et al. National Healthcare Safety Network (NHSN) report: data summary for 2006 through 2008, issued December 2009. *Am J Infect Control*. 2009;37:783–805.