Evaluating Mortality Outlier Hospitals to Improve the Quality of Care in Emergency General Surgery

Robert D. Becher, MD, MS[1], **Michael P. DeWane, MD**[1], **Nitin Sukumar, MS**[2], **Marilyn J. Stolar, PhD**[2], **Thomas M. Gill, MD**[3], **Adrian A Maung, MD**[1], **Kevin M. Schuster, MD, MPH**[1], **Kimberly A. Davis, MD, MBA**[1]

[1]Section of General Surgery, Trauma, and Surgical Critical Care; Department of Surgery; Yale School of Medicine, New Haven, Connecticut

[2]Yale Center for Analytical Sciences; Yale School of Public Health, New Haven, Connecticut

[3]Section of Geriatrics, Department of Internal Medicine; Yale School of Medicine, New Haven, Connecticut

Abstract

Background: Expected performance rates for various outcome metrics are a hallmark of hospital quality indicators used by AHRQ, CMS, and NQF. The identification of outlier hospitals with above- and below-expected mortality for emergency general surgery (EGS) operations is therefore of great value for EGS quality improvement initiatives. The aim of this study was to determine hospital variation in mortality after EGS operations, and compare characteristics between outlier hospitals.

Methods: Using data from the California State Inpatient Database (2010–2011), we identified patients who underwent one of eight common EGS operations. Expected mortality was obtained from a Bayesian model, adjusting for both patient- and hospital-level variables. A hospital-level standardized mortality ratio (SMR) was constructed (ratio of observed to expected deaths). Only hospitals performing 3 of each operation were included. An "outlier" hospital was defined as having a SMR 80% confidence interval that did not cross 1.0. High- and low-mortality SMR outliers were compared.

Results: There were 140,333 patients included from 220 hospitals. SMR varied from a high of 2.6 (mortality 160% higher than expected) to a low of 0.2 (mortality 80% lower than expected); 12 hospitals were high-SMR outliers, and 28 were low-SMR outliers. Standardized mortality was over 3 times worse in the high-SMR outliers compared to the low-SMR outliers (1.7 vs 0.5;

Corresponding Author: Robert D. Becher, MD, MS, Department of Surgery, Yale School of Medicine, 330 Cedar Street, BB 310, New Haven, CT 06520, robert.becher@yale.edu, Phone: 1 (203) 785-2572.

p<0.001). Hospital-, patient-, and operative-level characteristics were equivalent in each outlier group.

Conclusions: There exists significant hospital variation in standardized mortality after EGS operations. High-SMR outliers have significant excess mortality, while low-SMR outliers have superior EGS survival. Common hospital-level characteristics do not explain the wide gap between under- and over-performing outlier institutions. These findings suggest that SMR can help guide assessment of EGS performance across hospitals; further research is essential to identify and define the hospital processes of care which translate into optimal EGS outcomes.

Level of Evidence: Level III Epidemiologic Study

BACKGROUND:

Expected performance rates for various outcome metrics in the perioperative period are a hallmark of hospital quality indicators. The Agency of Healthcare Research and Quality (AHRQ) has defined a spectrum of inpatient hospital "Quality Indicators"[1] – including multiple post-operative mortality rates[2] – which are standardized, evidence-based measures meant to provide a perspective on hospital quality, to measure clinical performance, and to track and improve outcomes. The Center for Medicare and Medicaid Services (CMS) has a longstanding "Hospital Quality Initiative"[3] which measures, tracks, and publicly reports procedure-specific metrics such as risk-standardized mortality and readmission rates after elective operations, with a goal of driving quality improvement through measurement and transparency;[4–6] CMS links their overall Quality Strategy to reimbursements through its Quality Payment Program.[7,8] And the National Quality Forum (NQF) endorses hundreds of quality measures throughout healthcare, including many surgical quality metrics such as operation- and specialty-specific standardized mortality rates and post-operative complication rates.[9]

All of these metrics fit into the Department of Health and Human Services' (HHS) National Quality Strategy (NQS), which is a nationwide effort of over 300 stakeholder groups, organizations, and individuals designed to provide direction for improving the quality of health and healthcare in the United States.[10–12] In the field of emergency general surgery (EGS), there are currently no standardized, evidence-based, widely-accepted quality metrics to measure, compare, and track clinical performance across hospitals with the goal of improving outcomes. To move the field of EGS forward, quality metrics must be identified, investigated, validated, and endorsed. With these metrics as a foundation, the next mechanism for accomplishing this goal would be to establish an American College of Surgeons (ACS) Quality Program for EGS (which does not currently exist).

Selecting appropriate EGS quality of care measures is a crucial yet challenging task.[7,13] To start this process, one common approach in the field of quality improvement is to first determine the degree to which health services increase or decrease the likelihood of a

desired health outcome for patients. Reflecting its predominance as an outcome metric, mortality after an emergency operation is a logical place to start. Measuring the variation in hospital-level standardized mortality after EGS operations would quantify hospital performance, and allow for an analysis of why such variations exist between the high- and low-mortality hospitals. This would in turn guide quality improvement initiatives across the spectrum of EGS hospitals, and would be of great value for advancing the field of EGS.

To test the concept that there is significant variation in hospital standardized mortality after surgical emergencies, we sought to answer two questions. First, our primary research question asked: to what degree does the hospital standardized mortality rate (SMR) vary for adult patients undergoing common EGS operations? Second, our secondary research question asked: what are the differences in hospital-level characteristics between high- and low-performing outlier hospitals which help explain the variation in SMR?

METHODS:

Datasets & Variables:

This is a population-based, retrospective cohort study of all adult patients (18 years) who underwent one of eight EGS operations in the state of California over a 24-month period, from January 1, 2010 to December 31, 2011. The eight operations analyzed contribute to a majority of the morbidity and mortality in EGS; they were: appendectomy; cholecystectomy; colectomy; inguinal & femoral hernia repair (analyzed together as one type of operation); lysis of adhesions (LOA; note by our definition no bowel resections were performed in the LOA group); repair of perforated peptic ulcers (either gastric or duodenal ulcers); small bowel resection; and ventral hernia repair. Both laparoscopic and open operations were included; trauma operations were excluded.

Two datasets were used. The first was the State Inpatient Database (SID) for California (data from 2010 and 2011). The state of California was chosen as it is the most populous state in the US (population of 37 million in 2011), with a diverse population and varied geography, with both urban and rural areas. The SID are part of a family of datasets developed by the Healthcare Cost and Utilization Project, and sponsored by AHRQ.[14] Data abstracted included patient demographics, chronic health conditions, hospital-based metrics, and in-hospital mortality. The second dataset was the American Hospital Association (AHA) Annual Survey of Hospitals Database for 2010 and 2011.[15] The same California acute care hospitals in the SID and the AHA were paired, thus enabling risk-adjustment at the hospital level.

For the current analyses, only patients undergoing urgent/emergency operations with specific EGS diagnoses were included. Patients were identified using International Classification of Disease, 9th Edition (ICD-9), procedural codes (Supplemental Digital Content 1, Appendix A, http://links.lww.com/TA/B338); only patients who were listed in the SID dataset as having undergone one of the eight operations as a primary core operation were included. ICD-9 diagnosis codes (Supplemental Digital Content 2, Appendix B, http://links.lww.com/TA/B339) identified patients with a specific diagnosis of an EGS condition.

Given the ability to longitudinally track patients within SID, patients were not included more than once, meaning only their first EGS operation during the study was assessed.

The patient populations were chosen as they are among most prevalent emergent surgical diagnoses requiring operative intervention in the US, and have a non-trivial risk of postoperative morbidity and mortality.[16–18] An operation was defined as being performed urgently/emergently if it was associated with an admission not scheduled, as defined by the SID unscheduled admission variable. Death was measured as an in-hospital mortality.

Transfer status of the patient to and from another acute care hospital was incorporated into the inclusion/exclusion criteria. For patients who emergently underwent an operation at one hospital and later transferred out to a second hospital (such as for bleeding or other complications), mortality was attributed to the transferring/primary hospital; this is in keeping with the public reporting of mortality rates.[19]

Acute care hospitals were the only hospital-type included in the analyses. Dedicated pediatric hospitals, rehabilitation hospitals, and government hospitals such as Veteran's Affairs Hospitals were excluded. Only hospitals performing 3 of each of the eight operation types were included, meaning that the absolute minimum number of EGS operations that any given hospital could perform over the two years was 24. We rationalized that this would give us a more consistent, less heterogeneous group of hospitals for comparison, as each hospital would be doing a wide- range of EGS operations and would therefore contribute more reliable information regarding mortality.

Statistical Analyses & Outcome Measures:

Our primary research question assessed the degree to which the hospital standardized mortality rate (SMR) varies for adult patients undergoing common EGS operations. Our approach to defining SMR was based on the methodology used by CMS.[20–23] Our SMR measure estimates a hospital-level, all-cause, in-hospital, risk-adjusted, standardized mortality for acute care hospitals in California.

The SMR was constructed for each individual hospital: first, operation-level SMRs were calculated for each of the eight unique EGS operation types; second, these eight unique SMRs were pooled to create the aggregated hospital SMR; we combined data over the two-year period. SMR was defined as the ratio of the observed in-hospital deaths to expected deaths which occur postoperatively at a given hospital. Observed operative mortality was defined as a death during the index inpatient hospitalization which occurred after one of the eight EGS operations. Expected mortality was then calculated for each operation type at each hospital, based on individual patients' expected deaths (explanation begins in next paragraph). The observed to expected mortality ratio was then calculated to define the hospital SMR for a given operation at a given hospital; at this point, each hospital had eight unique SMRs. These eight SMRs were then pooled for each hospital to create a composite hospital-wide SMR; weighting the operation-specific SMRs based on differences in operation-specific volume at a given hospital was not necessary as the operation-specific hospital SMRs are comprised of individual patients' expected deaths. This methodology ensured that we would make appropriately calibrated predictions of hospital-level SMR.

This approach is an indirect standardization method; the only valid comparisons of SMR are between hospitals that contributed data to the study.

To define expected mortality, which is the SMR denominator, we created hierarchical, Bayesian mixed-effects logistic regression models for each operation separately. A mixed-effects model with hospital-specific intercepts has advantages over the more basic random effects model or a purely fixed effects model.[22] First, it includes an adjustment for both patient-level and hospital-level effects; the inclusion of hospital-level attributes reduces the potential confounding of the patient attribute-risk relation.[22] Second, a mixed effects model allows for the accurate inclusion of smaller hospitals into the overall SMR analysis by more properly calibrating the estimate of expected death.

Smaller hospitals may have unstable estimates of mortality, due to either a small number of operations or a small number of events/deaths.[21] Several statistical models have been proposed in an effort to overcome this limitation. A random effects type model shrinks all variability at these small hospitals to the mean via a reliability adjustment, thereby often incorrectly defining their mortality as average – which is not useful for quality improvement initiatives. The fixed effects models, on the other hand, cannot incorporate the stochastic effects of clustering, thereby making SMR estimates somewhat unreliable. The mixed effects approach, used in the present analysis, replaces the hospital-specific fixed effects by an assumption that these effects are random variables drawn from a distribution of such effects.[22] With the hospital-level risk adjustment, such models allow for both between- and within-hospital variation.

We first adjusted our operation-specific Bayesian models of expected mortality for the individual patient-level case-mix characteristics: age, gender, and Elixhauser-van Walraven comorbidity index. The inclusion of pre-admission medical conditions is fundamental to creating accurate measures of SMR. The Elixhauser-van Walraven is a validated, weighted measure of a person's chronic disease burden.[24] Coexisting conditions were identified using ICD-9 diagnosis codes, which were then compiled into an Elixhuaser-van Walraven comorbidity index. We specifically did not include socio-demographic characteristics in our case-mix adjustment (such as race, ethnicity, or payor status), as some argue that their inclusion may mask disparities and inequities in quality of care.[25] Furthermore, we were not able to include anatomic and/or physiologic severity of disease since this information was not available in the HCUP SID dataset. While some may see this as a limitation, the evidence that case-mix differences (meaning things like differences in disease severity) explains variations in operative mortality rates is mixed.[26]

We then added hospital-level effects to the Bayesian models to get a revised expected mortality prediction for each operation type, at each hospital. While the incorporation of hospital-level effects enhances the calibration of the expected mortality prediction,[21,22] the question as to whether the shrinkage target should depend on hospital-level attributes remains a key issue of contention in creating SMR models.[23] Based on CMS estimates, we know that hospital volume, for example, plays a crucial role in hospital quality assessments because the amount of information to assess hospital quality depends on the number of patients treated and, with event data, more particularly the number of observed events.

Therefore, unless the analysis includes some form of stabilization, such as including hospital volume in the model, hospital performance estimates of SMR associated with low-volume hospitals will be unreliable.[22]

The hospital-level characteristics included were: hospital operative volume; trauma center status; high technology capability; and medical school affiliation. Hospital volume was defined as the total number of patients having one of the eight types of urgent/emergent EGS operations at each acute care hospital over the two-year period. Trauma center status was based on American College of Surgeons Committee on Trauma verified Level 1 or Level 2 trauma centers.[27] High technology capability was based on the AHA database definition, defined as hospitals which perform adult open heart surgery and/or major organ transplantation such as heart or liver transplant.[15,28] Medical school affiliation was based on the hospital being a teaching hospital for an accredited medical school, with either medical students and/or residents.[29]

SMR-based caterpillar plots were then created to rank and compare standardized mortality performance among all hospitals studied; 80% confidence intervals were plotted around each SMR. We set an 80% confidence interval *a priori* as this is a probabilistic exploratory study, and we were most interested in defining if variation exists. Because of this, we further assumed there would be small numbers of operations and/or deaths at some hospitals in the high and low outlier groups, and we were concerned that a 95% confidence interval would lead to miscalculated conclusions of variance by type II error. SMR=1.0 means the observed mortality was equal to the expected mortality, and this defined the average performing hospitals. Statistically speaking, average performing hospitals are also those for which the SMR confidence interval overlapped with 1.0, meaning they were not significantly different from average. This implies typical performance relative to the California standard for the types of patients treated at that hospital. Hospitals with a SMR >1.0 *and* an 80% confidence interval lower bound that was above 1.0 (meaning confidence interval did not cross 1.0) were considered high-SMR outlier hospitals; these are below-average poor-performing institutions. Hospitals with a SMR <1.0 *and* an 80% confidence interval upper bound that was below 1.0 (meaning confidence interval did not cross 1.0) were considered low-SMR outlier hospitals; these are above-average high-performing institutions. In this manner, the high mortality outliers were statistically significantly worse than average, and the low mortality outliers were statistically significantly better than average.

As there is no true fit statistic available for mixed effects logistic regression models, we used Pearson Chi-Square divided by degrees of freedom to assess whether the variability in the expected mortality data was modeled properly; this was done for the operation-specific models. If the value greatly exceeds 1.0, the model is a poor fit.

Our hospital-level SMR uses a multiple model approach, with results for each of the individual operation type models pooled to create the overall hospital-wide SMR mortality measure. We therefore sought to increase the practical utility of the measure by assessing the differences in SMR performance by each operation type within hospitals. For each of the eight EGS operation types, operation-specific SMR-based caterpillar plots were then created to compare standardized mortality performance among all hospitals studied; 80% confidence

intervals were plotted for each SMR. This could potentially allow hospitals to better target quality improvement efforts, if certain operations were found to have much higher SMRs compared to other operation types within the same institution.

Our secondary research question was to identify differences in hospital-level characteristics between high- and low-performing outlier institutions to help explain any variation in SMR. To do this, hospital-level characteristics were compared using bivariate techniques between those in the high-outlier group compared to those in the low-outlier group. Patient-level and operation- type characteristics were also compared; these were reported at the hospital-level. For example, a comparison of mean age between the high and low outliers compared the mean age at each individual hospital; it did not compare the mean age of all individual patients in the high and low outlier groups. Chi-squared ($\chi2$) tests were used to compare differences in proportions of categorical variables; such data were summarized by frequencies with percentages. Group means were compared using t-tests for normally distributed continuous variables; such data were summarized by mean values with standard deviations (±SD). The hospital-level characteristics assessed were: hospital operative volume; trauma center status; high technology capability; medical school affiliation; hospital location (rural vs urban); and hospital size (<100 beds versus 100 beds).

A two-sided p-value of <0.05 was defined as significant. All statistical analyses were conducting using SAS 9.4 (SAS Institute Inc., Cary, NC). This study was approved by the Human Investigation Committee (HIC) of the Yale University Human Research Protection Program (HRPP) for biomedical research. The HIC is Yale's Institutional Review Board (IRB).

RESULTS:

Over the course of the two-year study, 140,333 patients were included from 220 acute care hospitals in California. Hospital-level characteristics can be found in Table 1. Overall, 22% of hospitals were level 1 or level 2 trauma centers, 9% were in a rural location, and 30% were teaching hospitals with a medical school affiliation. Patient-level characteristics (Table 1) demonstrate that the mean age was 50 years, most had over 2 comorbidities based on the van Walraven comorbidity score, and 58% were female. Hospitals performed, on average, 638 EGS operations over the two years. Operative characteristics (Table 1) show that the most common operations performed were cholecystectomy (278 on average per hospital over the two-year study), appendectomy (200 per two years on average per hospital), and colectomy (52 per two years on average per hospital).

In comparing hospital-level variation in standardized mortality rates, significant differences were found between the low and high SMR outliers. SMR (Figure 1) varied from a low of 0.2 (mortality 80% lower than expected; Confidence Interval 0.0–0.5) to a high of 2.6 (mortality 160% higher than expected; CI 1.3–3.9). A total of 28 hospitals (12.7%) were low-SMR outliers, indicating better than expected mortality rates, while 12 hospitals (5.5%) were high-SMR outliers, indicating poorer than expected mortality rates. Average SMR was over 3 times worse in the high-SMR outliers compared to the low-SMR outliers (1.7 vs 0.5;

p<0.001). For the SMR data by individual hospital, please see Supplemental Digital Content 3, Appendix C, http://links.lww.com/TA/B340.

In trying to define why the significant variation in SMR outcomes exist between the low and high SMR outliers, differences in hospital, patient, and operative level characteristics were assessed (Table 2). Hospital-level characteristics were equivalent in each outlier group, including percentage of verified trauma centers, high-tech hospitals, teaching hospitals, average hospital volume, rural location, and small hospitals (<100 beds). Patient-level characteristics (age, gender, comorbidities, hospital length of stay) were also similar in both outlier groups. Operative characteristics showed that the high and low outliers performed, on average, the same numbers of all types of EGS operations over the two years; the only exception was for repair of perforated peptic ulcer disease, with slightly higher numbers on average in the low performing cohort (7.6 vs 12.6 operations over two years; p = 0.01).

Differences across hospitals in operation-specific SMRs for each of the eight operation types were then assessed. Unlike in the aggregated SMR hospital data, the operation-specific hospital SMRs had no clear distinction between a group of well-performing, low-SMR hospitals and a group of poorly-performing, high-SMR hospitals. For six of the eight operation types, we found no true statistically significant outlier hospitals. The only exceptions were with the colectomy and small bowel resection groups, for which 14 hospitals were low-SMR outliers. However, no hospitals were high-SMR outliers in these groups.

Values for the Pearson Chi-Square divided by degrees of freedom were below 1.0 for each of the eight models, indicating variability in the expected mortality data was modeled properly. Exact values were: Appendectomy, 0.27; Cholecystectomy, 0.58; Colectomy, 0.88; Inguinal hernia repair, 0.53; Lysis of adhesions, 0.76; Repair of perforated peptic ulcer disease, 0.80; Small bowel resection, 0.88; Ventral hernia repair, 0.95.

DISCUSSION:

The current study documented significant hospital variation in standardized mortality rates at acute care hospitals in California that perform a wide spectrum of EGS operations. Nearly 1 in 8 hospitals are low-SMR outliers with superior EGS survival, while over 1 in 20 have significant excess mortality and are high-SMR outliers. On average, the high performing institutions have mortality rates three times lower than the poorly performing institutions. Common hospital-, patient-, and operation-level characteristics do not explain the wide gap between these under- and over-performing outlier institutions.

Developing a valid, national-level, hospital-specific, performance-based SMR metric requires many strategic decisions as there are multiple statistical issues encountered in modeling hospital quality based on outcomes.[22] These include the types of covariates to include in the model, the type of statistical modeling approach to employ, and the calculation of the metric itself. With these issues in mind, the aim of SMR modeling is to develop a parsimonious model that includes clinically relevant variables strongly associated with the

risk of mortality following an EGS operation, producing a level playing field to evaluate hospitals.

Guided by AHRQ and CMS standards, our hospital-level SMR is a rigorously created metric that can help guide assessment of hospital EGS performance by identifying both high and low SMR outlier hospitals. By providing a context to compare EGS performance across institutions, this will allow hospitals to better target quality improvement efforts. Risk-adjustment for our SMR measure was done for both case-mix differences and hospital service-mix differences. We did not adjust for patients' admission source or discharge disposition because these factors are associated with structure of the health care system, and may reflect the quality of care delivered by the system. We also did not adjust by socio-economic status, race, or ethnicity because hospitals should not be held to different standards of care based on the demographics of their patients. Lastly, complications occurring during a hospitalization are distinct from co-morbid illnesses and may reflect hospital quality of care, and therefore were not used for risk adjustment in our models.

The measurement and analyses of EGS-specific metrics has been endorsed as a key area of investigation to move the field of EGS forward.[30] Much of this is currently done at the patient level. A good example is the validated grading system for acute colonic diverticulitis[31] created by the American Association for the Surgery of Trauma's (AAST) Patient Assessment & Outcomes Committee.[32] EGS will equally benefit from hospital-level metrics. One such metric is SMR; similar standardized mortality measures are endorsed by AHRQ[2], CMS[4–6], and the NQF.[9] An NQF-endorsed, EGS-specific, national-level SMR would help to define the optimal care of the EGS patient, establish a benchmark for EGS hospital performance, and set achievable goals for hospital quality improvement initiatives at both a local and national level. It may also be possible to define and validate operation-specific SMR performance metrics. In the present study, due to the wide SMR confidence intervals for the operation-specific hospital SMRs, there was no clear distinction between a group of low- versus high-SMR hospitals for any operation.

One of the goals of surgical quality improvement is to measure the perioperative quality of hospital care through metrics which accurately discriminate across institutions; the standard outcome to measure is surgical mortality. Surgical mortality is inherently variable, and this variation can be attributed to three general contributing factors: chance, case-mix (referring to patient characteristics), and quality of care.[26] The present study highlights that standardized EGS mortality varies significantly across hospitals in California, and that these differences appear to be due mostly to quality of care. We minimized the risks of chance by building rigorous hierarchical, Bayesian mixed-effects logistic regression models. And we adjusted for case-mix in these models – though contrary to popularly accepted statistical wisdom, the evidence that case-mix differences are fundamental to explaining variations in surgical mortality *across hospitals* is conflicting.[26,33] Differences in quality of care therefore appear to be driving the SMR variations across institutions in California.

The quality of surgical care is a multidimensional construct.[26] Answering the simple question "why do patients die after EGS operations?" turns out to be quite complex. The current study, along with previous EGS research, provides only preliminary answers to this

basic question. To build on our present finding, the next steps are twofold. The first is to work towards validating SMR as a quality metric for EGS, a component of which will be to go through the rigorous steps of NQF-endorsement.[34] With this context in mind, it is clear that a national registry of EGS patients is needed to estimate reliable outcome rates for this complex patient-population. The second is to consider the root causes of this mortality variation, which will likely require both quantitative and qualitative research methodologies. Qualitative research can often answer questions that quantitative research cannot, as it approaches healthcare quality through lived experiences and relational processes that are extremely hard to capture quantitatively.

The present study has limitations. First, we used a retrospective administrative dataset, and our conclusions are thus constrained by their inherent limitations and biases, including errors in coding for procedures and diagnoses, selection biases, and inability to infer causation. Second, the ability to risk-adjust the data did not include admission physiologic parameters or condition-specific indicators such as admission level of inflammation; inclusion of these may help to explain some of the SMR variation. Third, hospital volume and possibly other hospital attributes are both predictors of and consequences of hospital quality.[22] In order to stabilize the regression models, it is widely accepted that hospital-level determinants be included in the SMR models to account for service-mix (as discussed at length). However, this is a tradeoff as there is the potential, mainly in the context of low volume institutions, to adjust away differences related to the quality of the hospital. Fourth, an "emergency" patient is a construct of the SID dataset, and generalizing to all patients requiring an urgent/ emergency operation may not be valid. And fifth, the data are from the state of California, and generalizations to a national level may not be valid.

In conclusion, there exists significant hospital variation in SMR at acute care hospitals in California that perform a wide spectrum of EGS operations. High-SMR outliers have significant excess mortality, while low-SMR outliers have superior EGS survival. Hospital-, patient-, and operation-level characteristics do not explain the wide gap between these under- and over- performing outlier institutions. These results suggest that SMR can help guide assessment of EGS performance across hospitals; further research is essential to identify and define the hospital processes of care which translate into optimal EGS outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

SOURCE OF FUNDING:

REFERENCES:

1. AHRQ (Agency for Healthcare Research and Quality) -- Quality Indicators homepage Available from: http://www.qualityindicators.ahrq.gov/. Accessed August 30, 2018.

2. AHRQ (Agency for Healthcare Research and Quality) -- Inpatient Quality Indicators Available from: http://www.qualityindicators.ahrq.gov/Modules/IQI_TechSpec_ICD10_v2018.aspx. Accessed August 30, 2018.

3. CMS (Centers for Medicare and Medicaid Services) -- Hopsital Quality Initiative Available from: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/index.html. 2013 Accessed August 30, 2018.

4. CMS (Centers for Medicare and Medicaid Services) -- Quality Outcome Measures Available from: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/OutcomeMeasures.html. 2017 Accessed August 30, 2018.

5. CMS (Centers for Medicare and Medicaid Services) -- Hospital Inpatient Quality Reporting Program Available from: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html. 2017 Accessed August 30, 2018.

6. Hospital Compare homepage Available from: https://www.medicare.gov/hospitalcompare/search.html. Accessed August 30, 2018.

7. CMS (Centers for Medicare and Medicaid Services) -- Quality Measure and Quality Improvement Available from: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Quality-Measure-and-Quality-Improvement-.html. 2017 Accessed August 30, 2018.

8. CMS (Centers for Medicare and Medicaid Services) -- The Quality Payment Program Available from: https://qpp.cms.gov/. Accessed August 30, 2018.

9. National Quality Forum (NQF): Quality Positioning System (QPS) homepage Available from: http://www.qualityforum.org/QPS/QPSTool.aspx. Accessed August 30, 2018.

10. NQS (National Quality Strategy) & Working for Quality --by AHRQ Available from: https://www.ahrq.gov/workingforquality/index.html. Accessed August 30, 2018.

11. NQS (National Quality Strategy) -- about the NQS, by AHRQ Available from: /workingforquality/about/index.html. 2017 Accessed August 30, 2018.

12. NQF (National Quality Forum) and the NQS (National Quality Strategy) Available from: http://www.qualityforum.org/Field_Guide/What_is_the_National_Quality_Strategy.aspx. Accessed August 30, 2018.

13. AHRQ (Agency for Healthcare Research and Quality) -- Selecting Quality and Resource Use Measures: A Decision Guide for Community Quality Collaboratives Available from: /professionals/quality-patient-safety/quality-resources/tools/perfmeasguide/index.html. 2010 Accessed August 30, 2018.

14. Healthcare Cost and Utilization Project (HCUP). Healthcare Cost and Utilization Project (HCUP) Databases. State Inpatient Databases (SID) overview State Inpatient Databases (SID) homepage. Available from: https://www.hcup-us.ahrq.gov/sidoverview.jsp. Accessed January 5, 2019.

15. American Hospital Association (AHA). American Hospital Association (AHA) Annual Survey of Hospitals Database overview AHA Annual Survey of Hospitals Database homepage. Available from: https://www.ahadataviewer.com/additional-data-products/AHA-Survey/. Accessed January 5, 2019.

16. Becher RD, Hoth JJ, Miller PR, Mowery NT, Chang MC, Meredith JW. A Critical Assessment of Outcomes in Emergency versus Nonemergency General Surgery Using the American College of

Surgeons National Surgical Quality Improvement Program Database. The American Surgeon. 2011;77:951–959. [PubMed: 21944366]

17. Shiloach M, Frencher SK, Steeger JE, Rowell KS, Bartzokis K, Tomeh MG, Richards KE, Ko CY, Hall BL. 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]

18. Schilling PL, Dimick JB, Birkmeyer JD. Prioritizing quality improvement in general surgery. J Am Coll Surg. 2008;207:698–704. [PubMed: 18954782]

19. Krumholz HM, Lin Z, Keenan PS, Chen J, Ross JS, Drye EE, Bernheim SM, Wang Y, Bradley EH, Han LF, et al. Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. JAMA. 2013;309:587–593. [PubMed: 23403683]

20. Silber JH, Satopää VA, Mukherjee N, Rockova V, Wang W, Hill AS, Even-Shoshan O, Rosenbaum PR, George EI. Improving Medicare's Hospital Compare Mortality Model. Health Serv Res. 2016;51 Suppl 2:1229–1247. [PubMed: 26987446]

21. George EI, Ro ková V, Rosenbaum PR, Satopää VA, Silber JH. Mortality Rate Estimation and Standardization for Public Reporting: Medicare's Hospital Compare. J Am Stat Assoc. 2017;112:933–947.

22. Ash A, Fienberg E, Louis T, Normand SL, Stukel TA, Utts J, The COPSS-CMS White Paper Committee. STATISTICAL ISSUES IN ASSESSING HOSPITAL PERFORMANCE Commissioned by the Committee of Presidents of Statistical Societies. 2012;70.

23. Silber JH, Rosenbaum PR, Brachet TJ, Ross RN, Bressler LJ, Even-Shashan O, Lorch SA, Volpp KG. The Hospital Compare mortality model and the volume-outcome relationship. Health Serv Res. 2010;45:1148–1167. [PubMed: 20579125]

24. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. Med Care. 2009;47:626–633. [PubMed: 19433995]

25. Krumholz HM, Brindis RG, Brush JE, Cohen DJ, Epstein AJ, Furie K, Howard G, Peterson ED, Rathore SS, Smith SC, et al. Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: cosponsored by the Council on Epidemiology and Prevention and the Stroke Council. Endorsed by the American College of Cardiology Foundation. Circulation. 2006;113:456–462. [PubMed: 16365198]

26. Birkmeyer JD, Dimick JB. Understanding and reducing variation in surgical mortality. Annu Rev Med. 2009;60:405–415. [PubMed: 18759554]

27. American College of Surgeons Committee on Trauma. American College of Surgeons Committee on Trauma (ACS-COT) ACS-COT Homepage. Available from: https://www.facs.org/quality-programs/trauma. Accessed January 5, 2019.

28. McHugh MD, Kelly LA, Smith HL, Wu ES, Vanak JM, Aiken LH. Lower mortality in magnet hospitals. Med Care. 2013;51:382–388. [PubMed: 23047129]

29. Association of American Medical Colleges (AAMC) Organization Directory Search Result. AAMC Organization Directory homeage. Available from: https://members.aamc.org/eweb/DynamicPage.aspx?webcode=AAMCOrgSearchResult&orgtype=Medical%20School. Accessed January 5, 2019.

30. Becher RD, Davis KA, Rotondo MF, Coimbra R. Ongoing Evolution of Emergency General Surgery as a Surgical Subspecialty. J Am Coll Surg. 2018;226:194–200. [PubMed: 29111417]

31. Shafi S, Priest EL, Crandall ML, Klekar CS, Nazim A, Aboutanos M, Agarwal S, Bhattacharya B, Byrge N, Dhillon TS, et al. Multicenter validation of American Association for the Surgery of Trauma grading system for acute colonic diverticulitis and its use for emergency general surgery quality improvement program. J Trauma Acute Care Surg. 2016;80:405–410; discussion 410–411. [PubMed: 26670116]

32. Shafi S, Aboutanos M, Brown CV-R, Ciesla D, Cohen MJ, Crandall ML, Inaba K, Miller PR, Mowery NT. Measuring anatomic severity of disease in emergency general surgery. J Trauma Acute Care Surg. 2014;76:884–887. [PubMed: 24553565]

33. Dimick JB, Birkmeyer JD. Ranking hospitals on surgical quality: does risk-adjustment always matter? J Am Coll Surg. 2008;207:347–351. [PubMed: 18722939]

34. National Quality Forum (NQF): How Endorsement Happens Available from: https://www.qualityforum.org/Measuring_Performance/ABCs/How_Endorsement_Happens.aspx Accessed September 7, 2018.
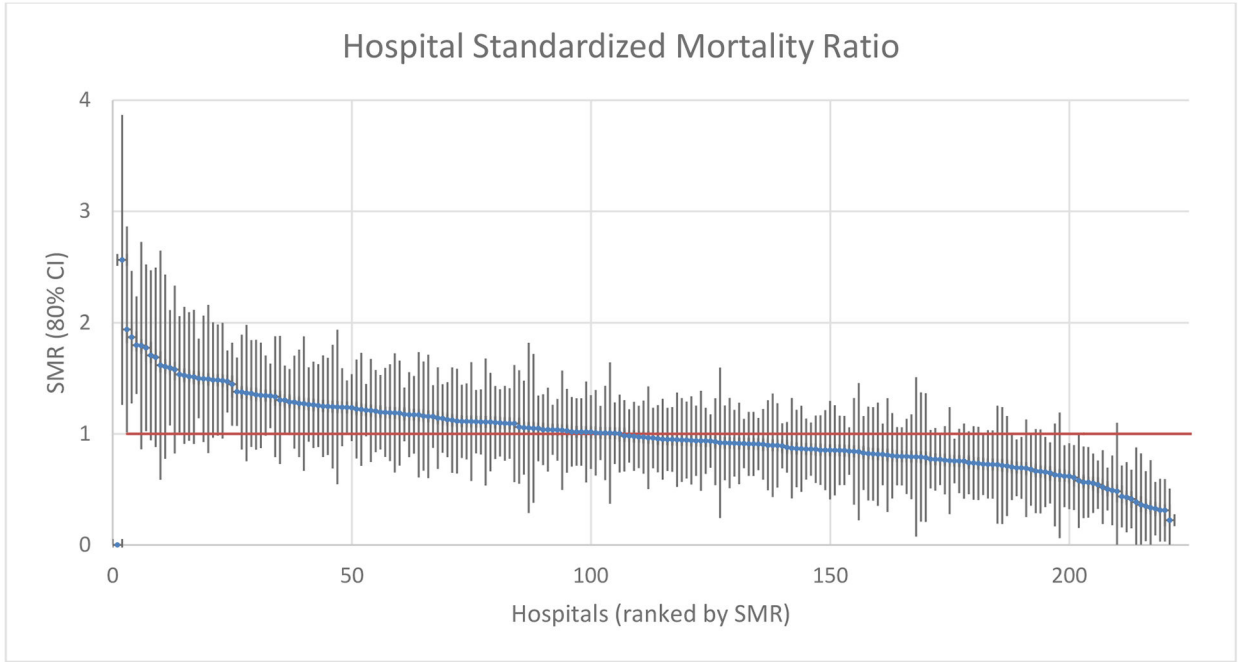
Figure 1: Hospital Standardized Mortality Ratio (SMR).
SMR-based caterpillar plot comparing standardized mortality performance among all acute care hospitals studied in California. Each line is one hospital; vertical bar represents 80% confidence interval. SMR=1.0 implies typical performance relative to the California standard for the types of patients treated at that hospital. SMR >1.0 with 80% CI lower bound above 1.0 is a high-mortality outlier (12 hospitals). SMR <1.0 with 80% CI upper bound below 1.0 is a low-mortality outlier (28 hospitals). For the SMR data that went into creating this chart, please see Supplemental Digital Content 3, Appendix C, http://links.lww.com/TA/B340.

Table 1:

Hospital, Patient, & Operation Characteristics

Hospital Characteristics (n=220)	
SMR, mean (SD)	1.0 (0.3)
Teaching Hospital, n (%)	65 (29.6%)
Trauma L1 or L2, n (%)	48 (21.8%)
High tech capable, n (%)	152 (69.1%)
Rural Location, n (%)	20 (9.1%)
>=100 beds, n (%)	200 (90.9%)
Hospital EGS operative volume, mean (SD)	637.9 (368.8)
Patient Characteristics (n=140,333) [++]	
Age in years, mean (SD) entire cohort	50.4 (19.8)
Age in years, mean (SD) by hospital	51.1 (4.9)
van Walraven score, mean (SD) entire cohort	2.3 (5.9)
van Walraven score, mean (SD) by hospital	2.5 (1.1)
Hospital length of stay in days, mean (SD) entire cohort	5.4 (7.9)
Hospital length of stay in days, mean (SD) by hospital	5.5 (1.1)
Male, n (%) entire cohort	58,960 (42.0%)
Proportion of male patients, mean (SD) by hospital	42.1% (4.4)
Operation Characteristics [*]	
Appendectomy, hospital mean (SD); n=44,031	200.1 (125.2)
Cholecystectomy, hospital mean (SD); n=61,081	277.6 (177.5)
Colectomy, hospital mean (SD); n=11,469	52.1 (33.9)
Inguinal & Femoral Hernia, hospital mean (SD); n=3313	15.1 (10.2)
Lysis of Adhesion, hospital mean (SD); n=8403	38.2 (27.1)
Repair Perforated PUD, hospital mean (SD); n=2079	9.5 (5.6)
Small Bowel Resection, hospital mean (SD); n=6790	30.9 (19.8)
Ventral Hernia, hospital mean (SD); n=3167	14.4 (9.8)

[++] Patient characteristics are listed two ways: First, for entire cohort, meaning at the individual level. Second, at the hospital level, which amounts to a mean of all hospital means. Of the 145,901 possible patients, 3 patients (0.002%) were excluded due to no hospital length of stay, and 5565 patients (3.8%) were excluded due to no gender specification.

[*] Operative characteristics are average number of operations performed at a hospital over two years SMR, risk-standardized mortality ratio; SD, standard deviation; n, number; L1, level 1 trauma center; L2, level 2 trauma center; EGS, emergency general surgery; PUD, peptic ulcer disease (including gastric and duodenal perforations)

Table 2

Comparison of Good and Poor Outlier Hospitals

	Good (n=28)	Poor (n=12)	p-value
Hospital Characteristics			
SMR, mean (SD)	0.5 (0.2)	1.7 (0.3)	<0.001
Teaching Hospital, n (%)	6 (21.4%)	3 (25.0%)	1.00
Trauma L1 or L2, n (%)	5 (17.9%)	3 (25.0%)	0.68
High tech capable, n (%)	19 (67.9%)	8 (66.7%)	1.00
Rural Location, n (%)	1(3.6%)	3 (25.0%)	0.07
>= 100 beds, n (%)	26 (92.9%)	10 (83.3%)	0.57
Hospital EGS operative volume, mean (SD)	641.0 (399.2)	788.3 (504.2)	0.33
Patient Characteristics (by hospital)			
Age, mean (SD)	51.0 (4.1)	48.7 (4.5)	0.12
van Walraven score, mean (SD)	2.6 (1.3)	2.0 (1.1)	0.17
Hospital length of stay, mean (SD)	5.3 (1.1)	5.7 (1.5)	0.35
Proportion male, mean (SD)	41.8% (3.9)	41.4% (5.0)	0.78
Operative Characteristics (by hospita·)			
Appendectomy, hospital mean (SD)	198.6 (140.1)	228.9 (158.4)	0.55
Cholecystectomy, hospital mean (SD)	263.6 (159.1)	366.3 (263.3)	0.23
Colectomy, hospital mean (SD)	52.1 (41.0)	53.8 (35.7)	0.90
Inguinal & Femoral Hernia, hospital mean (SD)	12.8 (8.9)	18.9 (12.2)	0.08
Lysis of Adhesion, hospital mean (SD)	40.5 (35.4)	34.8 (23.9)	0.61
Repair Perforated PUD, hospital mean (SD)	7.6 (4.6)	12.6 (6.4)	0.01
Small Bowel Resection, hospital mean (SD)	28.7 (20.2)	32.8 (22.0)	0.57
Ventral Hernia, hospital mean (SD)	12.7 (8.4)	18.2 (10.3)	0.09

NOTE: "Good" outlier hospitals have a low SMR <1.0 with an 80% confidence interval upper bound that does not cross 1.0; "Poor" outlier hospitals have a high SMR > 1.0 with an 80% confidence interval lower bound that does not cross 1.0. For our Bayesian models of expected mortality, age, van Walraven score, gender, hospital volume, trauma center status, tech capability, and teaching hospital status were included in the model

SMR, risk-standardized mortality ratio; SD, standard deviation; n, number; L1, level 1 trauma center; L2, level 2 trauma center; EGS, emergency general surgery; PUD, peptic ulcer disease (including gastric and duodenal perforations)