Evaluating mortality outlier hospitals to improve the quality of care in emergency general surgery

Robert D. Becher, MD, MS, Michael P. DeWane, MD, Nitin Sukumar, Marilyn J. Stolar, PhD, Thomas M. Gill, MD, Adrian A. Maung, MD, Kevin M. Schuster, MD, MPH, *and* **Kimberly A. Davis, MD, MBA***, New Haven, Connecticut*

AAST Continuing Medical Education Article

Accreditation Statement

This activity has been planned and implemented in accordance with the Essential Areas and Policies of the Accreditation Council for Continuing Medical Education through the joint providership of the American College of Surgeons and the American Association for the Surgery of Trauma. The American College Surgeons is accredited by the ACCME to provide continuing medical education for physicians.

AMA PRA Category 1 Credits™

The American College of Surgeons designates this journal-based CME activity for a maximum of 1 *AMA PRA Category 1 Credit*™. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Of the *AMA PRA Category 1 Credit*™ listed above, a maximum of 1 credit meets the requirements for self-assessment.

Credits can only be claimed online

AMERICAN COLLEGE OF SURGEONS

Inspiring Quality:
Highest Standards, Better Outcomes

100+*years*

Objectives

After reading the featured articles published in the *Journal of Trauma and Acute Care Surgery*, participants should be able to demonstrate increased understanding of the material specific to the article. Objectives for each article are featured at the beginning of each article and online. Test questions are at the end of the article, with a critique and specific location in the article referencing the question topic.

Claiming Credit

To claim credit, please visit the AAST website at http://www.aast.org/ and click on the "e-Learning/MOC" tab. You must read the article, successfully complete the post-test and evaluation. Your CME certificate will be available immediately upon receiving a passing score of 75% or higher on the post-test. Post-tests receiving a score of below 75% will require a retake of the test to receive credit.

Disclosure Information

In accordance with the ACCME Accreditation Criteria, the American College of Surgeons, as the accredited provider of this journal activity, must ensure that anyone in a position to control the content of *J Trauma Acute Care Surg* articles selected for CME credit has disclosed all relevant financial relationships with any commercial interest. Disclosure forms are completed by the editorial staff, associate editors, reviewers, and all authors. The ACCME defines a `commercial interest' as "any entity producing, marketing, re-selling, or distributing health care goods or services consumed by, or used on, patients." "Relevant" financial relationships are those (in any amount) that may create a conflict of interest and occur within the 12'months preceding and during the time that the individual is engaged in writing the article. All reported conflicts are thoroughly managed in order to ensure any potential bias within the content is eliminated. However, if you'perceive a bias within the article, please report the circumstances on the evaluation form.

Please note we have advised the authors that it is their responsibility to disclose within the article if they are describing the use of a device, product, or drug that is not FDA approved or the off-label use of an approved device, product, or drug or unapproved usage.

Disclosures of Significant Relationships with Relevant Commercial Companies/Organizations by the Editorial Staff

Ernest E. Moore, Editor: PI, research support and shared U.S. patents Haemonetics; PI, research support, TEM Systems, Inc. Ronald V. Maier, Associate editor: consultant, consulting fee, LFB Biotechnologies. Associate editors: David Hoyt and Steven Shackford have nothing to disclose. Editorial staff and Angela Sauaia have nothing to disclose.

Author Disclosures

The authors have nothing to disclose.

Reviewer Disclosures

The reviewers have nothing to disclose.

Cost

For AAST members and *Journal of Trauma and Acute Care Surgery* subscribers there is no charge to participate in this activity. For those who are not a member orsubscriber, the cost for each credit is $25.

System Requirements

The system requirements are as follows: Adobe® Reader 7.0 or above installed; Internet Explorer® 7 and above; Firefox® 3.0 and above, Chrome® 8.0 and above, or Safari™ 4.0 and above.

Questions

If you have any questions, please contact AAST at 800-789-4006. Paper test and evaluations will not be accepted.

BACKGROUND: Expected performance rates for various outcome metrics are a hallmark of hospital quality indicators used by Agency of Healthcare Research and Quality, Center for Medicare and Medicaid Services, and National Quality Forum. 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 three or more of each operation were included. An "outlier" hospital was defined as having an SMR with 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. Standardized mortality ratio 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 three times worse in the high SMR outliers compared with the low SMR outliers (1.7 vs. 0.5; $p < 0.001$). Hospital-, patient-, and operative-level characteristics were equivalent in each outlier group.

CONCLUSION: 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 underperforming and overperforming 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. (*J Trauma Acute Care Surg.* 2019;87: 297–306. Copyright © 2019 American Association for the Surgery of Trauma.)

LEVEL OF EVIDENCE: Epidemiologic Study, level III.

KEY WORDS: Standardized mortality ratio; EGS care.

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 postoperative 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] Also, the National Quality Forum (NQF) endorses hundreds of quality measures throughout health care, including many surgical quality metrics, such as operation- and specialty-specific standardized mortality rates and postoperative complication rates.[9]

All of these metrics fit into the Department of Health and Human Services' National Quality Strategy, which is a nationwide effort of over 300 stakeholder groups, organizations, and individuals designed to provide direction for improving the quality of health and health care 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 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

Data Sets and 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 the Section of General Surgery, Trauma, and Surgical Critical Care, Department of Surgery (R.D.B., M.P.D., A.A.M., K.M.S., K.A.D.), Yale School of Medicine; Yale Center for Analytical Sciences (N.S., M.J.S.), Yale School of Public Health; and Section of Geriatrics, Department of Internal Medicine (T.M.G.), Yale School of Medicine, New Haven, Connecticut.

Presented at: The 77th Annual Meeting of the American Association for the Surgery of Trauma San Diego, CA September 2018.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (www.jtrauma.com).

Address for reprints: Robert D. Becher, MD, Department of Surgery, Yale School of Medicine, 330 Cedar St, BB 310, New Haven, CT 06520; email: robert.becher@yale.edu.

DOI: 10.1097/TA.0000000000002271

from January 1, 2010, to December 31, 2011. The eight operations analyzed contribute to a majority of the morbidity and mortality in EGS; they are as follows: appendectomy, cholecystectomy, colectomy, inguinal, and 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 data sets were used. The first was the State Inpatient Database (SID) for California (data from 2010 and 2011). The state of California was chosen because it is the most populous state in the United States (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 data sets 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 data set 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 data set as having undergone one of the eight operations as a primary core operation were included. The 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 United States and have a nontrivial 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 three or more 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 2 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 and Outcome Measures

Our primary research question assessed the degree to which the hospital 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 2-year period. Standardized mortality ratio 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, sex, and Elixhauser-van Walraven comorbidity index. The inclusion of preadmission 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 sociodemographic 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 data set. While some may see this as a limitation, the evidence that case-mix differences (meaning things like differences in disease severity) explain 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 2-year period. Trauma center status was based on American College of Surgeons Committee on Trauma verified Level I or Level II 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]

The SMR-based caterpillar plots were then created to rank and compare standardized mortality performance among all hospitals studied; 80% confidence intervals (CI) were plotted around each SMR. We set an 80% CI *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% CI would lead to miscalculated conclusions of variance by type II error. An SMR of 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 CI 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 an SMR greater than 1.0 and an 80% CI lower bound that was above 1.0 (meaning CI did not cross 1.0) were considered high SMR outlier hospitals; these are below-average poor-performing institutions. Hospitals with an SMR less than 1.0 and an 80% CI upper bound that was below 1.0 (meaning CI 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.

Because there is no true fit statistic available for mixed-effects logistic regression models, we used Pearson χ^2 divided by degrees of freedom to assess whether the variability in the expected mortality data was modeled properly; this was performed 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% CIs 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 with 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 with 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. The χ^2 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 vs. ≥100 beds).

A two-sided *p* value less than 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 for biomedical research. The Human Investigation Committee is Yale's Institutional Review Board.

RESULTS

Over the course of the 2-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 I or Level II 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 two comorbidities based on the van Walraven comorbidity score, and 58% were female. Hospitals performed, on average, 638 EGS operations over the 2 years. Operative characteristics (Table 1) show that the most common operations performed were cholecystectomy (278 on average per hospital over the 2-year study), appendectomy (200 per 2 years on average per hospital), and colectomy (52 per 2 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. Standardized mortality ratio (Fig. 1) varied from a low of 0.2 (mortality, 80% lower than expected; CI, 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 three times worse in the high SMR outliers compared with 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-technology hospitals, teaching hospitals, average hospital volume, rural location, and small hospitals (<100 beds). Patient-level characteristics (age, sex, 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 2 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 2 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 χ^2 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 as follows: appendectomy, 0.27; cholecystectomy, 0.58; colectomy, 0.88; inguinal hernia repair, 0.53; LOAs, 0.76; repair of perforated peptic ulcer disease, 0.80; small-bowel resection, 0.88; ventral hernia repair, 0.95.

TABLE 1. Hospital, Patient, and 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-technology 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 and femoral hernia, hospital mean (SD); n = 3,313	15.1 (10.2)
LOA, hospital mean (SD); n = 8,403	38.2 (27.1)
Repair perforated PUD, hospital mean (SD); n = 2,079	9.5 (5.6)
Small-bowel resection, hospital mean (SD); n = 6,790	30.9 (19.8)
Ventral hernia, hospital mean (SD); n = 3,167	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 5,565 (3.8%) patients were excluded due to no sex specification.

** Operative characteristics are average number of operations performed at a hospital over 2 years.

SD, standard deviation; n, number; L1, Level I trauma center; L2, Level II trauma center; PUD, peptic ulcer disease (including gastric and duodenal perforations).

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 one in eight 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 underperforming and overperforming outlier institutions.

Developing a valid, national-level, hospital-specific, performance-based SMR metric requires many strategic decisions because there are multiple statistical issues encountered

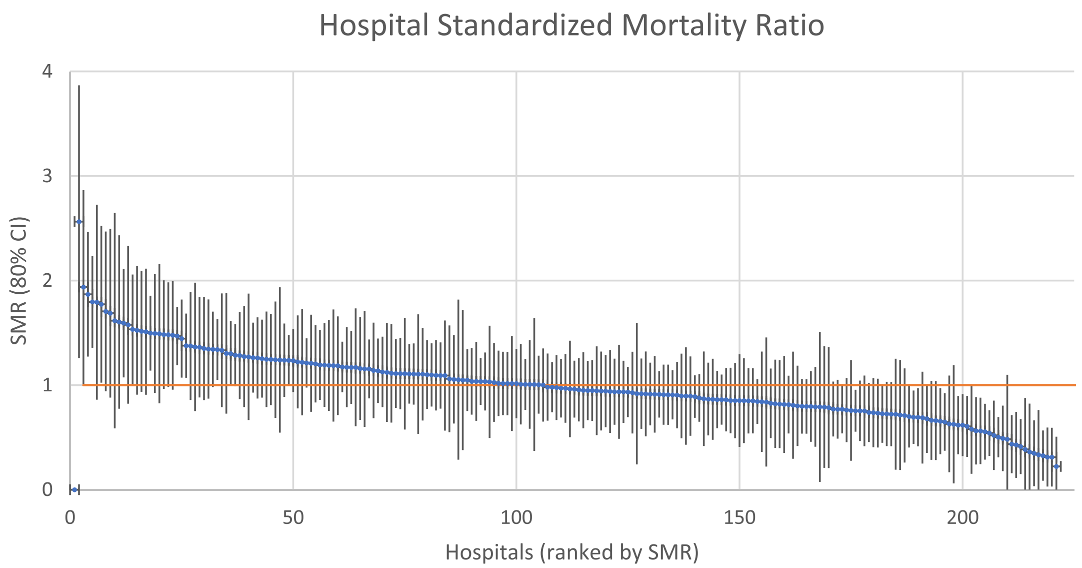

Figure 1. Hospital SMR. SMR-based caterpillar plot comparing standardized mortality performance among all acute care hospitals studied in California. Each line is one hospital; vertical bar represents 80% CI. SMR = 1.0 implies typical performance relative to the California standard for the types of patients 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.

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 use, 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

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-technology 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 hospital)			
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 and femoral hernia, hospital mean (SD)	12.8 (8.9)	18.9 (12.2)	0.08
LOA, 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

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

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 performed 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 socioeconomic 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 comorbid 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 performed 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 Patient Assessment and Outcomes Committee.[32] Emergency general surgery 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 CIs 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. Also, 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 toward 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, because it approaches health care 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 data set, 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] 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 because 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 data set, 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 underperforming and overperforming 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.

AUTHORSHIP

R.D.B., M.P.DW., N.S., M.J.S., T.M.G., A.A.M., K.M.S., K.A.D. participated in the study conception and design. R.D.B., M.P.DW., N.S., M.J.S. participated in the acquisition of data. R.D.B., M.P.DW., N.S., M.J.S., T.M.G., A.A.M., K.M.S., K.A.D. participated in the analysis and interpretation of data. R.D.B., M.P.DW., N.S., M.J.S., T.M.G., A.A.M., K.M.S., K.A.D. participated in the drafting of the article. R.D.B., M.P.DW., N.S., M.J.S., T.M.G., A.A.M., K.M.S., K.A.D. participated in the critical revisions.

DISCLOSURE

The authors have no financial conflicts of interest to disclose.
Dr. Becher acknowledges that this publication was made possible by the support of: the American Association for the Surgery of Trauma (AAST) Emergency General Surgery Research Scholarship Award; and the Yale Center for Clinical Investigation CTSA Grant Number KL2 TR001862 from the National Center for Advancing Translational Science (NCATS), a component of the National Institutes of Health (NIH). The contents are solely the responsibility of the authors and do not necessarily represent the official view of the AAST or the NIH. Dr. Gill acknowledges the support of the Academic Leadership Award (K07AG043587) and Claude D. Pepper Older Americans Independence Center (P30AG021342) from the National Institute on Aging.
Conflicts of Interest and Source of Funding: No conflicts of interest exist. Dr. Becher acknowledges that this publication was made possible by the

support of: the American Association for the Surgery of Trauma (AAST) Emergency General Surgery Research Scholarship Award; and the Yale Center for Clinical Investigation CTSA Grant Number KL2 TR001862 from the National Center for Advancing Translational Science (NCATS), a component of the National Institutes of Health (NIH). The contents are solely the responsibility of the authors and do not necessarily represent the official view of the AAST or the NIH. Dr. Gill acknowledges the support of the Academic Leadership Award (K07AG043587) and Claude D. Pepper Older Americans Independence Center (P30AG021342) from the National Institute on Aging.

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: https://www.ahrq.gov/workingforquality/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. *Am Surg*. 2011;77:951–959.
17. Shiloach M, Frencher SK Jr., 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.
18. Schilling PL, Dimick JB, Birkmeyer JD. Prioritizing quality improvement in general surgery. *J Am Coll Surg*. 2008;207:698–704.
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.
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.
21. George EI, Ročková 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-Shoshan O , Lorch SA, Volpp KG. The hospital compare mortality model and the volume-outcome relationship. *Health Serv Res*. 2010;45:1148–1167.
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.
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.
26. Birkmeyer JD, Dimick JB. Understanding and reducing variation in surgical mortality. *Annu Rev Med*. 2009;60:405–415.
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.
29. Association of American Medical Colleges (AAMC). 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.
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.
32. Shafi S, Aboutanos M, Brown CV, 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.
33. Dimick JB, Birkmeyer JD. Ranking hospitals on surgical quality: does risk-adjustment always matter? *J Am Coll Surg*. 2008;207:347–351.
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.

DISCUSSION

SHAHID SHAFI, M.D., M.P.H., M.B.A. (Southlake, Texas): Thanks for the opportunity, Dr. Rotondo, Dr. Reilly. And I want to thank the author for sending the manuscript in a timely fashion.

Great concept, great statistical methodology. We presented something similar last year. We called it EQIP – EGS Quality

Improvement Program, similar to NSQIP and TQIP. And I think it's time to start using that term and trying to figure out the best way to do it. Here are some specific questions for the authors:

Number one, why did you use 80 percent confidence intervals as opposed to 90 percent, which is typically used for NSQIP and TQIP. And obviously, if you use 80 percent confidence intervals, your intervals are narrower so you can get falsely positive outliers.

The second question, which I think is a major limitation of this study, is that the study was restricted to eight operative procedures only. A large number of EGS diseases are non-operative. Why did you not include all EGS patients in this analysis?

Third question is the lack of information on anatomic and physiologic data. We have done several validation studies on AAST grading systems, which shows that the grading systems and physiologic parameters are critical to predict the outcome of the patients.

Next question, how many patients were excluded from the analysis for lack of data?

Next question, how much of the variability was explained by your model? And how did you combine those eight SMRs to a single hospital level SMR? Were they all weighed equally?

Lastly, I would ask the authors to comment on the value of establishing an EGS verification program and regionalized systems of care.

Thank you very much.

HAYTHAM KAAFARANI, M.D. (Boston, Massachusetts): Great work and definitely a step in the right direction, congratulations.

We would never think of doing this same study in trauma patients without taking into account the injury severity or acuity.

You could have two patients with the same age, same comorbidities, and undergoing the same surgery that is needed from one of the eight procedures you have, but one of the two patients is sick and on death's doorstep and the other one is looking at you and chatting with you. The question I have is as follows: how does the SMR, which is based on an administrative database from HCUP, take the physiological arrangement of these patients, how sick they are, into account?

MITCHELL J. COHEN, M.D. (Denver, Colorado): A really nice study, but I wonder if we're at too much of a macro level to explain and truly understand differences you suggest. In keeping with this I wonder what you know about who actually did these surgeries? Especially presenting to this group of acute care surgeons who all have additional training and expertise above and beyond general surgery, we think that we have a unique skill set.

There is data from our group in San Francisco a few years ago and Pittsburgh last year suggesting we as acute care surgeons have some additional skill and training which potentially makes us better at these operations.

And so I'm wondering who did these surgeries and if we can get into a more micro or at the ground level of the skill set required to have good quality in difficult cases and patients.

WALTER L. BIFFL, M.D. (San Diego, California): This is a great study and an important issue to address. Were you able to identify which procedures have more variation in the outcomes? I wouldn't expect as much in appendicitis as I would in bowel obstruction, for example.

And, in looking at the outcomes, did time to the OR make a difference? We know in small bowel obstruction, there may be a relationship. And then, finally, is there any way to track the patients managed non-operatively and compare the outcomes? Thanks.

PRESTON R. MILLER, M.D. (Winston-Salem, North Carolina): Thanks very much. That was a fantastic paper. This is an issue that, like everybody else has said, we cannot begin to move forward in emergency general surgery until we can define what's good performance and what's bad performance.

Again, also to piggyback on what other folks have said, I think the level at which you were able to look at this is a 30,000-foot level.

And we don't have the answers based on these data, as to what does make a difference so, I'll just ask you to speculate, what do you think makes a difference? Thanks very much.

ROBERT D. BECHER, M.D., M.S. (New Haven, Connecticut): Thanks for everybody's comments.

The first question about why we use an 80 percent confidence interval. The reason is that this is a probabilistic exploratory study, and we were looking to identify differences in hospital-based standardized mortality rates and to see if we could identify why those differences exist.

The second question related to EGS-specific operations, and why we didn't include a broader cohort. This study looked at operation-specific SMR, and so we included only patients who had had a procedure. Perhaps we could broaden that in a future study.

The third question was about anatomic and physiologic severity of disease. And this gets to another question just about disease severity. The HCUP data that we employed does not include any physiologic parameters, so they can't be included in our case mix adjustment. With that said, there is good data to show that when you are comparing between hospitals, case mix adjustment is not always as important as people think it is, meaning you get very valid comparisons between hospitals with adjusted for case-mix differences such as anatomic and physiologic severity of disease. So I think that even though some may see this as a limitation, and we certainly understand that viewpoint, we strongly believe that the result that some hospitals were outliers is still valid.

The question about which patients were excluded and how much of the variability was explained in our model, I don't have that data here, but I will include it in the manuscript.

There was another question about weighting the variables. This gets to the idea that if you have a hospital, say, that does 400 appendectomies per year, how can you compare that hospital to a different hospital that does, say, 400 colectomies per year? Because these are different groups of patients. This is an excellent question that really relates to how we designed the analyses and methodology. And the answer gets to why, I think, this study is so robust. Our hospital-based SMR was calculated by aggregating the operation-specific SMRs at each hospital; the operation-specific SMRs were created by individual patient-level expected mortality. So to answer the question, the model itself weighs for differences in mortality risk by operation as the expected mortality rate is different for different operations. For example, if you have 400 colectomies at one hospital and 400 appendectomies at another, that difference gets accounted for based on how we defined the SMR variables, and so the hospital-level SMRs are directly comparable between these two hospitals. In other words, it's included in our analysis and is part of the data; we don't need to make or use statistical assumptions with weighted adjustments.

There was a question about the EGS verification process moving forward. And I think that hospital verification is the ultimate goal here, and I believe it is part of the AAST's goal of having an ACS EGS Quality Program. As per regionalization, the current study certainly hints at the benefits of potentially creating a system of EGS care. But before we go down that path way based on the current study, I think that we have to get at the bottom of the real issue which our study highlights: why are these profound hospital-based differences in EGS mortality occurring? Because maybe the answer is much simpler than regionalization; maybe it is getting the poor performers up to average mortality with a few key fixes.

The comment about the macro level and needing to look at the micro level I think is sort of a theme throughout this. And certainly that would be the next step, is to really dig in to answer the question of why. We know that we can't identify these differences based on the macro level, and so we may need to go to individual hospitals and really ask why.

Thanks so much.