FINAL PROGRESS REPORT:
ARE VOLUME STANDARDS ACCURATE MEASURES?

Principal Investigator:	Laurent G. Glance, M.D.
Co-Investigators:	Andrew W. Dick, Ph.D.
	Turner M. Osler, M.D.
	Dana B. Mukamel, Ph.D.

Organization:	University of Rochester School of Medicine and Dentistry
Project Period:	9/30/2003 - 9/29/2007
Project Officer:	Yen-Pin Chiang, Ph.D.

We gratefully acknowledge the support of the Agency for Healthcare and Quality Research.

Grant Award Number: R01 HS013617

This report represents the final progress report submitted on behalf of the Principal Investigators and the Subcontractors/Co-investigators.

Structured Abstract

Purpose: This project addresses the implications of regionalizing healthcare delivery using “volume” standards as opposed to “quality” criteria regionalization: (1) selective referral to high-quality hospitals; (2) selective referral to high-volume hospitals; (3) selective avoidance of low-quality hospitals; and (4) selective avoidance of low-volume hospitals. Administrative data were used to develop risk-adjusted measures of quality. Administrative data do not distinguish between complications and pre-existing conditions. We tested whether the absence of date stamping led to biased measures of hospital quality using the California State Inpatient Database (CA SID), which contains a Present-on-Admission (POA) indicator.

Scope and Methods: These studies were based on the California State Inpatient Database. Hierarchical logistic regression was used to produce the risk adjustment models. The impact of regionalization was assessed by simulating in-hospital mortality rate, the number of diverted patients, and the number of hospitals closed for each of the four referral strategies.

Results: There is a weak association between hospital quality and hospital volume. Most high-volume hospitals are not high quality, and most low-volume hospitals are not low quality. Selective referral to either high-volume centers or high-quality centers is moderately to highly effective but is extremely disruptive and unlikely to be feasible. Selective avoidance of low-volume centers does not lead to improved outcomes, whereas selective avoidance of low-quality hospitals yields minor improvements in outcome. The use of procedure volume as the basis for evidence-based hospital referrals should be re-evaluated by all stakeholders before undertaking more efforts to regionalize healthcare delivery using volume-based referral strategies. We also found that the use of routine administrative data without the POA indicator to construct hospital quality report cards may lead to inaccurate report cards.

key words: quality, risk-adjustment, administrative data, selective referral, regionalization, volume outcome association

Purpose

The Institute of Medicine (IOM) report, *To Err is Human,* states that between 44,000 and 98,00 deaths in the United States each year are due to medical errors (1). In response, the IOM has outlined a national agenda to improve patient safety (1). In order to accomplish this agenda, a report sponsored by the Agency for Healthcare Research and Quality (AHRQ) has designated "localizing specific surgeries and procedures to high-volume centers" as a high-priority area for patient safety research (2). There exists a large body of evidence suggesting that hospitals performing higher volumes of high-risk surgery have better outcomes than low-volume hospitals (3). In the absence of a systematic approach to measuring healthcare outcomes, selective referral of patients to high-volume hospitals has been advocated for high-risk surgery on the grounds that procedure volume is a quality proxy (4-6). With the launching of the Leapfrog Group safety initiative, sponsored by a coalition of some of the largest companies and third-party payers in the United States, these policy recommendations are now being implemented in the private sector (7). Nonetheless, the Leapfrog Group recognizes that volume is only an indirect measure for quality and supports working toward the goal of the "universal adoption of programs for measuring performance directly (7)."

In the absence of direct performance measures, the effort to regionalize healthcare using volume standards has a strong appeal. However, questions have been raised concerning the strength of the evidence supporting the existence of the volume-outcome relationship (8). Furthermore, it is not clear that volume standards should be substituted for quality standards given the option of constructing robust risk-adjusted outcome measures using widely available administrative data. This project was designed to address the implications of regionalizing healthcare delivery using "volume" standards as opposed to "quality" criteria. *Because a policy of "selective referral" that restricts care to a small number of centers of excellence may be very disruptive and may have a significantly negative impact on healthcare delivery (9), we also chose to contrast the potential benefits of a policy of "selective avoidance" (9) that seeks to restrict care at either "low-volume" or "low-quality" hospitals. The Main Hypothesis was to test whether:*

> *Regionalizing high-risk surgery by diverting patients from low-quality centers will lead to better population outcomes than does diverting patients from low-volume centers.*

To test this hypothesis, we used administrative data to develop risk-adjusted measures of quality. In most states and for all Medicare enrollees, administrative data are available for every hospitalization. These data could serve as an "information infrastructure" (10) for performance measurement. *The primary problem with using administrative data for benchmarking performance is that the data do not distinguish between complications (that occurred during the hospitalization) and conditions present prior to admission.* Clearly, the interpretation of secondary diagnoses as either pre-existing conditions or complications is critical to accurate risk adjustment. *Secondary diagnoses are used to stratify severity of disease (conditions that modify the severity of the primary diagnosis) and to specify comorbidities (conditions that are not directly related to the primary diagnosis) (11)*. Pre-existing conditions should serve as risk factors, whereas complications should be viewed as outcomes. The inability to separate the two in most administrative databases is likely to lead to biased risk adjustment and inaccurate quality measures. Despite this limitation, most abstract-based risk-adjustment measures "use discharge diagnoses codes representing all conditions treated during the hospitalization, regardless of when they occurred (12)."

Only two states, California and New York, include a date stamp on all secondary diagnoses that indicates whether a secondary diagnosis was present at the time of admission to the hospital. The Secondary Hypotheses to be tested were:

1. *The absence of date stamping leads to biased measures of severity of disease and comorbidity.*
2. *The absence of date stamping yields biased measures of hospital quality.*

We also explored the following additional Secondary Hypotheses:

1. Intensive Care Units with high patient volumes have lower mortality rates compared to low-volume ICUs.
2. Surgeons performing high-volume of off-pump coronary artery bypass surgery have lower mortality rates compared to low-volume surgeons.
3. Changing the statistical methodology used to identify quality outliers will change the relative quality rankings of hospitals.
4. High-quality surgeons are less likely to perform coronary artery bypass surgery on high-risk patients compared to low-quality surgeons.
5. Complications after coronary artery bypass surgery are important predictors of mortality.

The text of this report will divided into sections specific to each of the separate hypotheses tested in this project:

Scope

The absence of date stamping leads to biased measures of severity of disease and comorbidity.

Comorbidity measures are designed to exclude complications when they map ICD-9-CM codes to diagnostic categories. The use of data fields that indicates whether each secondary diagnosis was present on hospital admission may lead to the more accurate identification of pre-existing conditions. The objective was to examine the rate of misclassification of ICD-9-CM codes into diagnostic categories by the Dartmouth-Manitoba adaptation of the Charlson Index and by the Elixhauser comorbidity algorithm. This study was based on an analysis of 178,838 patients in the California State Inpatient Database (CA SID) admitted in 2000 for one of seven major medical and surgical conditions. The CA SID includes a present-on-admission (POA) indicator for each ICD-9-CM code that indicates whether an ICD-9-CM code represents a condition that was present on admission.

The absence of date stamping yields biased measures of hospital quality.

We conducted three separate studies to evaluate this hypothesis: the first based on our own set of condition-specific risk-adjustment models and the other two based on (1) the AHRQ Inpatient Quality Indicators and (2) the AHRQ Patient Quality Safety Indicators.

Many of the publicly available health quality report cards are based on administrative data. ICD-9-CM codes in administrative data are not date stamped to distinguish between medical conditions present on hospital admission and complications that occur during the hospital admission. Treating complications as pre-existing conditions gives poor-performing hospitals “credit” for their complications and may cause some low-quality hospitals to be misclassified as average- or high-performing hospitals. The purpose of these studies was to determine whether hospital quality assessment based on administrative data is impacted by the inclusion of the POA indicator in administrative data as a date stamp indicator.

The *first study* was based on 648,866 inpatient admissions between 1998 and 2000 for CABG surgery, coronary angioplasty, carotid endarterectomy (CEA), abdominal aortic aneurysm (AAA) repair, total hip replacement (THR), acute MI (AMI), and stroke using the California State Inpatient Database (SID), which includes CPAA modifiers. The *second study* was based on 2.07 million inpatient admissions between 1998 and 2000 in the California State Inpatient Database. The third study was based on 82,063 patient records for CABG surgery between 1998 and 2000 in the California State Inpatient Database.

Regionalizing high-risk surgery by diverting patients from low-quality centers will lead to better population outcomes than diverting patients from low-volume centers.

The objective of this study was to examine whether basing regionalization on risk-adjusted mortality would lead to better population outcomes than basing regionalization on procedure volume.

This study was based on 243,000 patients who underwent either abdominal aortic aneurysm surgery, coronary artery bypass surgery, or coronary angioplasty between 1998 and 2000 in California. Four regionalization strategies were compared: (1) selective referral to high-quality hospitals; (2) selective referral to high-volume hospitals; (3) selective avoidance of low-quality hospitals; and (4) selective avoidance of low-volume hospitals.

Intensive Care Units with high patient volumes have lower mortality rates compared to low-volume ICUs.

The objective of this study was to examine whether high-volume ICUs have superior mortality outcomes compared to low-volume ICUs. The study was based on 70,757 patients admitted to 92 ICUs between 2001 and 2003 using Project IMPACT (a clinical outcomes database created by the Society of Critical Care Medicine.

Surgeons performing high-volume of off-pump coronary artery bypass surgery have lower mortality rates compared to low-volume surgeons.

Off-pump CABG has been recently reintroduced into clinical practice. In light of the relatively low level of experience of most cardiac surgeons with off-pump CABG surgery and the exceptional technical challenge of working on a “beating heart,” off-pump CABG is a unique opportunity to analyze the impact of surgeon case volume on surgical outcome after controlling for the effects of patient case mix and hospital volume. The analyses were based on the New York State clinical CABG surgery registry. The study sample consisted of 36,930 patients undergoing isolated CABG surgery between 1998 and 1999 performed by 181 surgeons at 33 hospitals.

Changing the statistical methodology used to identify quality outliers will change the relative quality rankings of hospitals.

Study 1: Risk adjustment is central to the generation of health outcome report cards. It is unclear, however, whether risk adjustment should be based on standard logistic regression, fixed effects or random effects modeling. The purpose of this study was to determine how robust the New York State (NYS) Coronary Artery Bypass Graft (CABG) surgery report card is to changes in the underlying statistical methodology. This retrospective cohort study was based on data from the NYS Cardiac Surgery Reporting System on all patient undergoing isolated CABG surgery in NYS and who were discharged between 1997 and 1999 (51,750 patients).

Study 2: This study was conducted using the New York State (NYS) Coronary Artery Bypass Surgery Reporting System (CSRS) using all patients undergoing isolated CABG surgery in NYS who were discharged in 1999 (18,116 patients).

High-quality surgeons are less likely to perform coronary artery bypass surgery on high-risk patients compared to low-quality surgeons.

It is unknown whether high-risk cardiac surgical patients have less access to high-quality surgeons compared to lower-risk patients.

The objective of this study was to determine whether high-quality surgeons are less likely to perform CABG surgery on high-risk patients compared to low-quality surgeons. This retrospective cohort study was based on the New York State (NYS) Coronary Artery Bypass (CABG) Surgery Reporting System (CSRS) of all patients undergoing CABG surgery in NYS who were discharged between 1997 and 1999 (51,750 patients; 2.20% mortality).

Complications after coronary artery bypass surgery are important predictors of mortality.

Complications are associated with increased risk of death. The objective of this study was to quantify the increased odds of dying from complications following isolated coronary artery bypass grafting (CABG) surgery. This retrospective cohort study was conducted using the New York State CABG Surgery Reporting System for all patients undergoing isolated CABG surgery in NYS who were discharged between 1997 and 1999 (51,750 patients; 2.20% mortality).

Methods

The absence of date stamping leads to biased measures of severity of disease and comorbidity.

The Dartmouth/Charlson Index and the Elixhauser comorbidity measure were used to map the ICD-9-CM codes into diagnostic categories for patients in each study population. We calculated the misclassification rate for each mapping algorithm using information from the POA indicator as the “gold standard.”

The absence of date stamping yields biased measures of hospital quality.

Three studies were conducted to explore this hypothesis:

Study 1: Hierarchical logistic regression was used to create separate condition-specific risk adjustment models. For each study population, one model was constructed using only secondary diagnoses present at admission based on the POA indicator: “date stamp” model. The second model was constructed using all secondary diagnoses, ignoring the information present in the POA indicator: the “no date stamp model.” Each model adjusted for the severity of the principal diagnosis and for comorbid conditions. The severity of the principal diagnosis was coded using the Disease Staging classification system developed by Gonnella et al (13). Comorbid conditions were coded either using Disease Staging or the Elixhauser algorithm (11). Low- and high-performance hospitals were identified using the ratio of the observed mortality rate to the expected mortality rate (OE ratio) (14). Hospital quality was assessed separately using the “date stamp” and the “no date stamp” risk-adjustment models.

Study 2: The AHRQ IQI software was used to calculate risk-adjusted mortality rates using either (1) routine administrative data that included all ICD-9-CM codes or (2) enhanced administrative data that included only ICD-9-CM codes representing pre-existing conditions (based on the POA indicator).

Study 3: We calculated the positive predictive value of selected AHRQ Patient Safety Indicators (PSI) using information from the POA indicator as the gold standard and using the intra-class correlation coefficient to assess the level of agreement between the hospital risk-adjusted PSI rates with and without the information contained in the POA indicator.

Regionalizing high-risk surgery by diverting patients from low-quality centers will lead to better population outcomes than diverting patients from low-volume centers.

The goal of this study was to estimate the potential impact of four regionalization strategies for patients undergoing abdominal aortic aneurysm surgery, coronary artery bypass surgery, and coronary angioplasty: (1) selective referral to high-quality hospitals; (2) selective referral to high-volume hospitals; (3) selective avoidance of low-quality hospitals; and (4) selective avoidance of low-volume hospitals. The potential impact of regionalization was assessed by simulating in-hospital mortality rate, the number of diverted patients, and the number of hospital closed for each of the four referral strategies. We used hierarchical regression to identify high-quality and low-quality hospitals. High-volume hospitals were identified using the Leapfrog volume criteria.

Intensive Care Units with high patient volumes have lower mortality rates compared to low-volume ICUs.

The goal of this study was to examine the association between ICU patient volume and in-hospital mortality. Two separate analyses were performed. The first examined the volume outcome association controlling only for patient risk factors in order to determine if ICU volume is associated with better outcomes. The second examined the volume outcome association controlling for patient risk factors and ICU characteristics. The goal of this latter analysis was to determine if high ICU volume by itself, independent of other ICU characteristics, is associated with reduced mortality. Hierarchical logistic regression modeling was used to examine the volume-outcome association.

Surgeons performing high volumes of off-pump coronary artery bypass surgery have lower mortality rates compared to low-volume surgeons.

Random-effects models were constructed to model the effect of surgeon volume on outcome after adjusting for the effect of hospital volume and patient case mix.

Changing the statistical methodology used to identify quality outliers will change the relative quality rankings of hospitals.

Study 1: Using the same risk factors as in the NYS models, fixed-effects and random-effects models were fitted to the NYS data. Quality outliers were identified using (1) the ratio of observed to expected mortality rates (O/E ratio) and confidence intervals (CI) calculated using both parametric (Poisson distribution) and nonparametric (bootstrapping) techniques; and (2) shrinkage estimators.

Study 2: Patients from specific hospital were matched to a control group using the Mahalanobis distance. The hospitals' expected mortality rate was calculated in two ways: (1) as the mortality rate of the control group and (2) as the mortality rate predicted by the NYS CABG model.

Hospitals whose observed mortality rate was significantly different from their expected mortality rate (OE difference) were defined as quality outliers.

High-quality surgeons are less likely to perform coronary artery bypass surgery on high-risk patients compared to low-quality surgeons.

We used multivariate linear regression to examine the association between surgeon quality and patient risk. The analysis was performed at the level of the individual CABG patient. The surgeon O-to-E ratio was the dependent variable.

Complications after coronary artery bypass surgery are important predictors of mortality.

We estimated the independent effect of individual postoperative complications on in-hospital mortality after controlling for patient clinical risk factors and demographics.

Results

The absence of date stamping leads to biased measures of severity of disease and comorbidity.

The Dartmouth/Charlson Index underestimated the prevalence of hemiplegia/paraplegia by 70%, cerebrovascular disease by 70%, myocardial infarction by 65%, congestive heart failure (CHF) by 45%, and peptic ulcer disease by 34%. The Elixhauser algorithm misidentified as complications rather than as pre-existing conditions 43% of the coagulopathies, 25% of the fluid and electrolyte disorders, 18% of the cardiac arrhythmias, 18% of the cardiac arrhythmias, and 9% of the cases of CHF. Adding the POA indicator to administrative data would significantly enhance the ability of the Dartmouth/Charlson Index and of the Elixhauser algorithm to map ICD-9-CM codes to diagnostic categories accurately. This study provides significant insights into the potential risk of relying on conventional administrative data that are not date stamped for the construction of hospital report cards. It is possible that a relatively simple addition to administrative datasets – the POA indicator – may greatly improve risk adjustment. The findings of this study may prove useful to healthcare policymakers exploring the value and feasibility of instituting date stamping of ICD-9-CM codes in the other 48 states and in the Medicare/Medicaid programs.

The absence of date stamping yields biased measures of hospital quality.

Three studies were conducted to explore this hypothesis:

Study 1: Forty percent of the CABG hospitals, 33% of the PTCA hospitals, 40% of the THR hospitals, and 33% of the AMI hospitals identified as low-performance hospitals by the “date stamp” models were not classified as low-performance hospitals by the “no date stamp” models. Fifty percent of the CABG hospitals, 33% of the PTCA hospitals, 50% of the CEA hospitals, and 36% of the AMI hospitals identified as low-performance

hospitals by the “no date stamp” models were not identified as low-performance hospitals by the “date stamp” models. The inclusion of the POA indicator had a minor impact on hospital quality assessment for AAA repair, stroke, and carotid endarterectomy. This study supports the hypothesis that the use of routine administrative data without date stamp information to construct hospital quality report cards may result in the misidentification of quality outliers. Until report cards based on clinical data become more widely available, we recommend that POA indicator be added to all administrative data sets. However, the POA indicator will need to be further validated before date stamped administrative data can be used as the basis for health quality report cards.

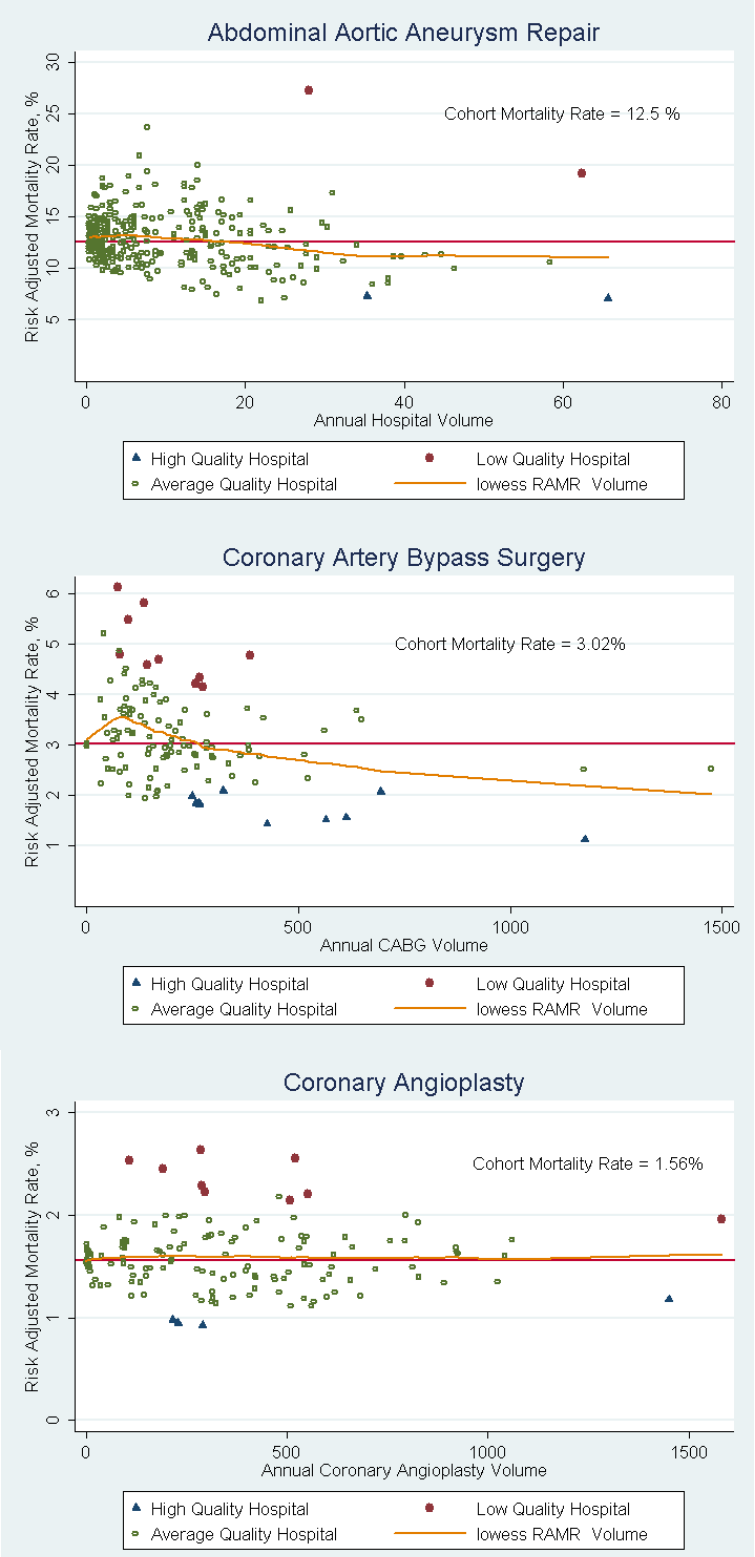
Study 2: The inclusion of the POA indicator frequently results in changes in the quality ranking of hospitals classified as high quality or low quality using routine administrative data based on the AHRQ IQIs. Twenty-seven percent (stroke) to 94% (CABG) of hospitals classified as high quality using routine administrative data were re-classified as intermediate- or low-quality hospitals using the enhanced administrative data (with the POA indicator). Twenty-five percent (congestive heart failure) to 76% (PCI) of hospitals classified as low-quality hospitals using enhanced administrative data were misclassified as intermediate-quality hospitals using routine administrative data. Despite the fact that the AHRQ IQIs were primarily intended to serve as a screening tool, they increasingly are being used to publicly report hospital quality. Our findings emphasize the need to improve the “quality” of administrative data by including a POA indicator if these data are to serve as the information infrastructure for quality reporting.

Study 3: The false positive error rate, defined as one minus the positive predictive value, was greater than or equal to 20% for four of the eight PSIs examined: decubitus ulcer, failure-to-rescue, postoperative physiologic and metabolic derangement, and postoperative pulmonary embolism or deep venous thrombosis. Pairwise comparison of the hospital risk-adjusted PSI rates, with and without the POA indicator information, demonstrated almost perfect agreement for five of the eight PSIs. For decubitus ulcer, failure-to-rescue, and postoperative pulmonary embolism or DVT, the intraclass-correlation coefficient ranged between 0.63 to 0.79. For some of the AHRQ Patient Safety Indicators, there are significant differences in the risk-adjusted rates of adverse events depending on whether the POA indicator is used to distinguish between pre-existing conditions and complications. The use of the POA indicator will increase the accuracy of the AHRQ PSIs as measures of adverse outcomes.

<u>Regionalizing high-risk surgery by diverting patients from low-quality centers will lead to better population outcomes than diverting patients from low-volume centers.</u>

There is a relatively weak association between hospital quality and hospital volume. Most high-volume hospitals are not high quality, and most low-volume hospitals are not low quality.

Figure. Association between annual hospital case volume and hospital risk-adjusted mortality rates for abdominal aortic aneurysm surgery, coronary artery bypass surgery, and coronary angioplasty. Abbreviations: RAMR – risk-adjusted mortality rate



Selective referral to high-volume centers would be only moderately effective (2-20% relative reduction in mortality) and extremely disruptive (70-99% reduction in the number of hospitals treating these conditions). Selective referral to high-quality centers was estimated to result in dramatic reduction in mortality (50%) but would also be highly disruptive, with greater than 80% of the patients re-directed to high quality centers. Selective avoidance of low-volume hospitals would not improve mortality, whereas selective avoidance of low-quality hospitals was estimated to result in a small improvement in overall mortality (2-6%) while causing relatively minor disruptions in patient referral patterns.

We found that selective referral to high-volume centers may, at best, yield only modest improvement in outcomes but would result in massive disruption in hospital services in California and possibly in other decentralized healthcare environments. Selective avoidance of low-volume hospitals would not lead to any improvement in outcomes. Selective avoidance of low-quality hospitals might yield some improvement in mortality rates without causing large-scale disruptions in referral patterns. The use of procedure volume as the basis for evidence-based hospital referrals should be re-evaluated by all stakeholders before undertaking further efforts to regionalize healthcare delivery using volume-based referral strategies.

Intensive Care Units with high patient volumes have lower mortality rates compared to low-volume ICUs.

This study may support the hypothesis that higher-volume ICUs have lower risk-adjusted mortality rates. The primary analysis found evidence of an association between “high-risk” patient volume and outcome, but only when controlling for other ICU characteristics. ICUs that treated higher volumes of very-high-risk patients (SAPS II > 41) have lower overall mortality rates after adjusting for patient risk factors. This “volume” effect was limited to high-risk (SAPS II between 30 and 41) and very-high-risk patients. There was no association between ICU volume and outcome when the volume calculation was based on all ICU admissions.

This study has potentially important policy implications. First, ICU volume alone may not be an accurate surrogate measure of ICU quality. We found that the volume of very-high-risk patients, but not overall volume, is associated with in-hospital mortality. Unfortunately, the need to quantify patient risk precludes the use of a simple patient count to identify potentially low-quality ICUs. However, if we are going to quantify patient risk, there is no rational basis for substituting volume measures for risk-adjusted measures of quality. Second, high-risk patient volume is but one of many organizational variables that affect outcomes. This association is no longer statistically significant after controlling for patient characteristics alone. High ICU volumes alone are not necessarily associated with better patient outcomes.

Surgeons performing high volumes of off-pump coronary artery bypass surgery have lower mortality rates compared to low-volume surgeons.

There is no association between the number of cases performed off pump by an individual surgeon and in-hospital mortality ($p=0.93$) after controlling for hospital CABG volume and patient-level risk factors. There is also no association between off-pump CABG mortality and the total number of both off-pump and on-pump cases ($p=0.78$). In the on-pump CABG cohort, high-volume surgeons had significantly lower risk-adjusted mortality rates compared to very-low-volume, low-volume, and medium-volume surgeons ($p<0.006$).

This study shows that high-volume surgeons performing off-pump CABG surgery do not have better mortality outcomes than low-volume surgeons. The absence of a volume-outcome association for off-pump surgery persisted when the procedure volume was based on the total number of off-pump and on-pump cases. However, high-volume surgeons have better outcomes than low-volume surgeons performing on-pump CABG. The absence of a volume-outcome relation for off-pump CABG is especially surprising in light of the more technically demanding nature of off-pump CABG surgery compared to on-pump CABG surgery.

Changing the statistical methodology used to identify quality outliers will change the relative quality rankings of hospitals.

Study 1: In this study, we examined whether different statistical methodologies affect which surgeons and hospitals are classified as high- and low-performance outliers in the publicly available NYS DOH CABG surgery report card. At the surgeon level, the standard logistic regression model, the fixed effects model, and the fixed effects component of the random effects model demonstrated near-perfect agreement on the identity of quality outliers using a quality indicator based on the O/E ratio and the Poisson distribution. Shrinkage estimators identified the fewest outliers, whereas the O/E ratios with bootstrap CI identified the greatest number of outliers. The results were similar for hospitals, except that the fixed effects model identified more outliers than either the NYS model or the fixed effects component of the random effects model.

Assigning outlier status to providers has potentially profound implications. Prior to publicly releasing a quality report cards, it is incumbent on the analyst to explore the extent to which their findings vary using different statistical approaches. More work is necessary to establish “best practices” for constructing report cards in order to ensure the validity of quality reporting.

Study 2: The purpose of this study was to assess hospital quality using a matching algorithm based on a generalized distance metric and to compare this approach to the more traditional regression-based approach. The two risk-adjustment methodologies disagreed on the outlier status of four of the 33 hospitals. Kappa analysis demonstrated substantial agreement between these two methods for identifying quality outliers: $\kappa=0.61$. There was excellent agreement between the point estimates of the OE difference obtained using these two risk adjustment methodologies.

Basing outcome assessment on either matching or regression modeling yielded similar findings on hospital ranking but only moderate level of agreement on hospital quality. The use of matching may enhance the transparency and acceptance of outcome report cards by hospitals and physicians.

High-quality surgeons are less likely to perform coronary artery bypass surgery on high-risk patients compared to low-quality surgeons.

In this study, we found that high-risk patients undergoing CABG surgery are more likely to be treated by high-quality surgeons than by low-quality surgeons. We found no evidence, in this study, that high-quality surgeons are selectively avoiding high-risk patients. Higher-risk patients are more likely to receive CABG surgery from higher-quality surgeons. For every 10 percentage point increase in patient risk of death (e.g., from 5% to 15%), there is an absolute reduction of 0.034 in the surgeon O-to-E ratio ($p<0.001$). This finding is important in light of prior evidence that some surgeons are more reluctant to care for the highest-risk patients in the aftermath of publicly released outcomes report cards. Our study does not, however, exclude the possibility that some of the highest-risk patients no longer have access to surgical revascularization.

Complications after coronary artery bypass surgery are important predictors of mortality.

There is a strong association between postoperative complications and in-hospital mortality. The mortality rate for patients without complication was 0.77% versus 16.1% for patients with complications ($p<0.001$). After adjusting for preoperative risk factors, transmural myocardial infarction (adj OR 7.90; $p<0.001$), respiratory failure (adj OR 6.02; $p<0.001$), renal failure (adj OR 7.15; $p<0.001$), and stroke within 24 hours (adj OR 4.09; $p<0.001$) were most strongly associated with mortality. This information may prove valuable to hospitals in their efforts to design quality improvement initiatives and care protocols to improve mortality after CABG surgery.

List of Publications

1. Glance LG, Dick AW, Olser TM, Mukamel DB. Does Date Stamping ICD-9-CM Codes Increase the Value of Clinical Information in Administrative Data? Health Services Research 2006; 41:231-251.

2. Glance LG, Dick AW, Osler TM, Mukamel DB. Accuracy of Hospital Report Cards Based on administrative Data. Health Services Research 2006; 41: 1413-1437.

3. Glance LG, Osler TM, Mukamel DB, Dick AW. Impact of Date Stamping on Hospital Quality Measurement: Experience with the AHRQ Inpatient Quality Indicators. Medical Care (in press).

4. Glance LG, Impact of Date Stamping on Patient Safety Measurement: Experience with the AHRQ Patient Safety Indicators. (submitted for publication)

5. Glance LG, Osler TM, Mukamel DB, Dick AW. Effect of Complications on Mortality after CABG Surgery: Evidence from New York State. Journal of Thoracic and Cardiovascular Surgery. 134(1):53-8, 2007 July.

6. Glance LG, Osler TM, Mukamel DB, Dick AW. Estimating the Impact of Regionalizing Health Care Delivery Based on Volume Standards Versus Based on Risk Adjusted Mortality Rate. International Journal for Quality in Health Care 19 (4): 195-202, 2007 August.

7. Glance LG, Li Y, Osler TM, Dick AW, Mukamel DB. Impact of Patient Volume on the Mortality of Adult Intensive Care Unit Patients. Crit Care Med 2006;34:1925-1934.
8. Glance LG, Dick AW, Olser TM, Mukamel DB. The Relation between Surgeon Volume and Outcome Following Off-Pump CABG versus On-Pump CABG. CHEST 2005; 128:820-837.

9. Glance LG, Dick AW, Osler TM, Li Y, Mukamel DB. Impact of Changing the Statistical Methodology on Hospital and Surgeon Ranking: The Case of the NYS Cardiac Surgery Report Card. Medical Care 2006; 44: 311-319.

10. Glance LG, Osler TM, Mukamel DB, Dick AW. Use of a Matching Algorithm to Evaluate Hospital CABG Performance as an Alternative to Conventional Risk Adjustment. Medical Care 2007; 45: 292-299.

11. Glance LG, Dick AW, Mukamel BD, Li Y, Osler TM. Are High Quality Cardiac Surgeons Less Likely to Operate on High Risk Patients Compared to Low-Quality Surgeons? Evidence from New York State. Health Services Research (in press)

Reference List

(1) To Err is Human: Building a Safer Health System. Washington, DC: National Academy Press, 2000.

(2) Making health care safer: a critical analysis of patient safety practices. Evid Rep Technol Assess (Summ) 2001;(43):i-668.

(3) Hewitt M. Interpreting the Volume-Outcome Relationship in the Context of Health Care Quality. 1 . 2000. Institute of Medicine.

(4) Dudley RA, Johansen KL, Brand R, Rennie DJ, Milstein A. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. JAMA 2000; 283(9):1159-1166.

(5) Birkmeyer JD. High-risk surgery--follow the crowd. JAMA 2000; 283(9):1191-1193.

(6) Birkmeyer JD, Lucas FL, Wennberg DE. Potential benefits of regionalizing major surgery in Medicare patients. Eff Clin Pract 1999; 2(6):277-283.

(7) Birkmeyer JD, Finlayson EV, Birkmeyer CM. Volume standards for high-risk surgical procedures: potential benefits of the Leapfrog initiative. Surgery 2001; 130(3):415-422.

(8) Sowden AJ, Sheldon TA. Does volume really affect outcome? Lessons from the evidence. J Health Serv Res Policy 1998; 3(3):187-190.

(9) Epstein AM. Volume and outcome--it is time to move ahead. N Engl J Med 2002; 346(15):1161-1164.

(10) Crossing the Quality Chasm. Washington, DC: National Academiy Press, 2001.

(11) Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998; 36(1):8-27.

(12) Iezzoni LI. The risks of risk adjustment. JAMA 1997; 278(19):1600-1607.

(13) Gonnella JS, Louis DZ, Gozum ME. Disease Staging: Classification and Severity Stratification. 1994. Ann Arbor, Michigan, The Medstat Group.

(14) Shwartz M, Ash AS, Iezzoni LI. Comparing Outcomes Across Providers. In: Iezzoni LI, editor. Risk Adjustment for Measuring Healthcare Outcomes. Chicago: AHSR, 1997: 471-516.