The Relationship between Case Volume, Care Quality, and Outcomes of Complex Cancer Surgery

Andrew D Auerbach, MD, MPH, Judith Maselli, MSPH, Jonathan Carter, MD, Penelope S Pekow, PhD, Peter K Lindenauer, MD, MSC

BACKGROUND:
How case volume and quality of care relate to each other and to results of complex cancer surgery is not well-understood.

STUDY DESIGN:
Observational cohort of 14,170 patients 18 years or older who underwent pneumonectomy, esophagectomy, pancreatectomy, or pelvic surgery for cancer between October 1, 2003 and September 1, 2005 at a US hospital participating in a large benchmarking database. Case volumes were estimated within our dataset. Quality was measured by determining whether ideal patients did not receive appropriate perioperative medications (such as antibiotics to prevent surgical site infections), both as individual "missed" measures and as overall number missed. We used hierarchical models to estimate effects of volume and quality on 30-day readmission, in-hospital mortality, length of stay, and costs.

RESULTS:
After adjustment, we noted no consistent associations between higher hospital or surgeon volume and mortality, readmission, length of stay, or costs. Adherence to individual measures was not consistently associated with improvement in readmission, mortality, or other outcomes. For example, continuing antimicrobials past 24 hours was associated with longer length of stay (21.5% higher, 95% CI, 19.5–23.6%) and higher costs (17% higher, 95% CI, 16–19%). In contrast, overall adherence, although not associated with differences in mortality or readmission, was consistently associated with longer length of stay (7.4% longer with 1 missed measure and 16.4% longer with ≥2) and higher costs (5% higher with 1 missed measure, and 11% higher with ≥2).

CONCLUSIONS:
Although hospital and surgeon volume were not associated with outcomes, lower overall adherence to quality measures is associated with higher costs, but not improved outcomes. This finding might provide a rationale for improving care systems by maximizing care consistency, even if outcomes are not affected. (J Am Coll Surg 2010;211:601–608. © 2010 by the American College of Surgeons)

The volume to outcomes relationship—the association between improved surgical outcomes at sites that perform a procedure more often—has become the focus of payor-driven proposals to regionalize care to high-volume centers. This relationship has been of particular interest in complex cancer surgery, where evidence suggests that care from a more experienced surgeon and hospital produce better outcomes.

However, little is known about the specific mechanisms that explain variation in outcomes between high- and low-volume centers or surgeons, and to what extent these can be attributed to differences in quality as measured by adherence to recommended care processes. If care quality is the primary factor explaining differences in outcomes between high- and low-volume centers, then patients in need of cancer surgery could expect similar results at high- and low-volume centers with similar quality measure performance. Conversely, if high-volume centers are better regardless of adherence to recommended practices, then travel to a regional referral center would be the wisest course of action.
When contrasted with the impact of case volume on outcomes, associations between individual quality measures and outcomes have been small or absent. \textsuperscript{11-14} Recent data from our group confirm inconsistent associations between individual quality measures and outcomes in coronary artery bypass surgery. \textsuperscript{15} However, increasing overall performance on quality measures can have a powerful impact on mortality \textsuperscript{15} and is also associated with lower costs. \textsuperscript{16}

We hypothesized that for patients undergoing complex cancer surgery, advantages seen at high-volume systems would be related to greater adherence to recommended care practices. To explore this hypothesis, we analyzed data collected from adults undergoing cancer surgery (eg, pelvic exenteration, esophageal resection, pancreatic resection, or pneumonectomy) in a nationally representative sample of US hospitals. Using these data, we first examined the relationship between patient outcomes, hospital case volume, physician case volume, and care quality measures. We then examined the degree to which overall quality (an all-or-none measure of system reliability) influenced mortality in relationship to volume measures.

\section*{Data}

Our data were collected on 14,170 patients cared for by 1,629 physicians at 266 hospitals participating in Perspective (Premier Inc.), a database developed for measuring quality and health care use, which we have used in previous research. \textsuperscript{17}

In addition to standard hospital discharge file data, Perspective contains a date-stamped log of all materials (eg, serial compression devices used to prevent venous thromboembolism) and medications (eg, β-blockers) charged for during hospitalization. Perspective charge data are collected electronically and undergo comprehensive auditing as part of Premier efforts to ensure data validity. Previous research suggests that comorbidity indices collected using Premier data correspond closely to those collected from charts. \textsuperscript{17}

Located in all regions of the United States, Perspective sites are representative of the US hospital population \textsuperscript{18-20} because they are predominantly small to mid-size nonteaching facilities and serve a largely urban patient population. Perspective sites also have performance on publicly reported quality measures similar to non-Perspective sites. The Institutional Review Board at University of California San Francisco approved our study, and our funder (California HealthCare Foundation) had no role in the development or execution of the study, or preparation of the article.

Patients were initially eligible for our analysis if they were admitted between October 1, 2003 and September 30, 2005 and were 18 years of age or older. Patients in this cohort who underwent complex cancer surgery were then identified using ICD-9-CM procedure codes and diagnosis codes by replicating methods used by Begg and colleagues. \textsuperscript{2} Specifically, patients had to have a principal diagnosis of cancer and to have undergone one of the following operations as their principal procedure during hospitalization: esophageal resection (ICD-9 42.40-42.42, 42.51-42.56, 42.58-42.59, 42.61-42.66, 42.68-42.69), pancreatic resection (ICD9 52.51, 52.33, 52.59, 52.6, 52.7), liver resection (ICD-9 50.22, 50.3, 50.4), pelvic exenteration (ICD-9 57.71, 68.8, 48.4-48.6), or pneumonectomy (ICD9 32.5, 32.3, 32.4).

\section*{Methods}

Because some hospitals in our cohort did not contribute data for the entire study period, we estimated the annual case volume by dividing each hospital’s or physician’s observed patient count by the total number of months that the hospital or physician contributed patients to the data set. These “annualized” volumes were then divided into quartiles so that one quarter of the patient cohort was included in each quartile of volume, as done in previous work. \textsuperscript{4,25-27}
was received during hospitalization. These measures included whether antimicrobials were used to prevent surgical site infection on the operative day, whether an antimicrobial was continued inappropriately past the first postoperative day, and whether appropriate strategies were used to prevent venous thromboembolism on the operative day.

Because inpatient diagnosis codes cannot reliably distinguish between complications and pre-existing conditions, we measured the proportion of ideal candidates for each care process who failed to receive them—a missed quality measure. For example, we considered the opportunity for β-blocker use “missed” if a patient did not receive the drug and did not have ICD-9 coded principal or secondary diagnosis of hypotension, heart block, or congestive heart failure recorded in their hospital record. To provide a more sensitive measure of system-level ability to provide reliable care,4,25–27 we also counted the total number of quality measures missed during hospitalization.

**Outcomes variance**

Cost, length of stay, readmission, discharge status, and mortality outcomes were obtained from Perspective discharge abstract data as described. Length of stay and costs were log-transformed to account for skew and to stabilize variance of residuals in multivariable models. β-estimates and 95% confidence intervals for log-transformed outcomes were converted to percent differences using the formula 100 × (EXP (estimate) − 1).

We first described study patients and hospitals using univariable methods. Multivariable alternating logistic models (SAS PROC GENMOD) were used to account for clustering of patients within physicians and physicians within hospitals for dichotomous outcomes and calculate adjusted odds ratios and adjusted estimates. Mixed effect models (SAS PROC MIXED) were used to account for clustering of patients within physicians and within hospitals for continuous variables. Models were constructed using manual variable selection methods. Volume and quality measures were entered manually, and additional covariates (confounding factors) were selected for inclusion if they were associated with the outcomes at p < 0.05, if including them changed estimates for the primary predictors by >10%, or for face validity. All analyses were carried out using SAS version 9.1 (SAS Institute, Inc.).

**Patient characteristics**

A total of 14,170 patients underwent one of our target operations at one of our study sites between October 1, 2003 and September 30, 2005. Mean age of patients was 66.2 years (SD 11.0 years), and 56% were men. Most were white and had Medicare insurance. The most common Elixhauser-defined comorbidities in our cohort were hypertension (50.2%), metastatic cancer (23.8%), and chronic obstructive pulmonary disease (40.2%). Three percent (427 patients) died during the initial hospitalization or a subsequent admission to the same hospital, 11% were readmitted in 30 days (Table 1).

The proportion of patients who did not receive our target medications varied. Few did not receive a β-blocker (15%) or had no antimicrobial charges on the operative day (9%); two thirds had no venous thromboembolism preventative measures and 62% had antimicrobials continued after the first postoperative day. Few patients (9%) had no missed quality measures, 35% missed 1, and 55% missed ≥2.

**Hospital and physician volume** and rate of missed quality measures

Most hospitals (174 hospitals, 65%) and physicians (913 physicians, 56%) were lowest-volume (eg, 1st quartile of volume) providers. Hospital volume ranged from 13 (interquartile range [IQR] 8–19) in the lowest quartile to 110 per year (IQR 105 to 148) in the highest. Physician volume ranged from 4 patients per year (IQR 3 to 5) in the lowest quartile, to 29 (IQR 24 to 41) in the highest. The mean number of quality measures missed was similar across physician and hospital volume quartiles (Table 2).
reveal any underlying associations between volume and any of our study outcomes (Table 4).

In this cohort of patients undergoing complex cancer surgery, we observed no statistically significant associations between higher volume and improved outcomes, or between individual quality measures and improved outcomes. When quality was measured as an overall count, worse overall quality (indicated by the number of measures missed during hospitalization) was not associated with
clinical outcomes, but was strongly associated with higher costs and length of stay. These findings suggest that quality improvement efforts aimed at improving the reliability of systems that provide care of cancer surgery patients can have a substantial impact on costs of care.

A large literature describes the relationship between higher volume and better outcomes in cancer surgery. This observation has led to endorsement of case volume as a way to identify preferred sites and improve patient outcomes—an approach aptly termed follow the crowd. However, regionalization of services poses practical problems, and the evidence for the ability of volume benchmarks to accurately identify “best” sites has limitations. We did not see a striking association between higher volume and better outcomes. This might be because we had a relatively small sample size compared with previous work, or because previous studies were able to include longer-term outcomes at fixed time periods. Longer periods of follow-up accrue more events, further increasing statistical power to compare events across volume strata, as well as potentially increasing sensitivity to the effects of high-quality postsurgical care for cancer patients provided at more specialized high-volume centers, which take place long after hospitalization. Although our methods selected cases using diagnosis codes used previously, pooling a number of fairly disparate surgical procedures in our study might have also limited our ability to detect a volume effect by attenuating our ability to discern hospitals’ or surgeons’ procedure-specific

### Table 2. Number of Hospitals or Providers and Missed Quality Measures per Quartile of Patient Volume

<table>
<thead>
<tr>
<th>Hospital annual volume</th>
<th>n</th>
<th>No. of hospitals or providers</th>
<th>Median volume</th>
<th>Mean no. of missed quality measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st quartile</td>
<td>3,560</td>
<td>174</td>
<td>13 (8,19)</td>
<td>1.71 (0.75)</td>
</tr>
<tr>
<td>2nd quartile</td>
<td>3,550</td>
<td>49</td>
<td>39 (35,44)</td>
<td>1.63 (0.79)</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>3,590</td>
<td>29</td>
<td>63 (58,71)</td>
<td>1.44 (0.78)</td>
</tr>
<tr>
<td>4th quartile</td>
<td>3,470</td>
<td>14</td>
<td>110 (105,148)</td>
<td>1.57 (0.91)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Physician annual volume</th>
<th>n</th>
<th>No. of hospitals or providers</th>
<th>Median volume</th>
<th>Mean no. of missed quality measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st quartile</td>
<td>3,532</td>
<td>913</td>
<td>4 (3,5)</td>
<td>1.68 (0.78)</td>
</tr>
<tr>
<td>2nd quartile</td>
<td>3,537</td>
<td>404</td>
<td>8 (7,9)</td>
<td>1.58 (0.77)</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>3,592</td>
<td>212</td>
<td>15 (13,17)</td>
<td>1.49 (0.81)</td>
</tr>
<tr>
<td>4th quartile</td>
<td>3,509</td>
<td>100</td>
<td>29 (24,41)</td>
<td>1.61 (0.89)</td>
</tr>
</tbody>
</table>

IQR, interquartile range.

### Table 3. Volume, Individual Quality Measures, and Outcomes of Major Cancer Surgery

<table>
<thead>
<tr>
<th></th>
<th>Mortality, adjusted</th>
<th>Readmission, adjusted</th>
<th>Length of stay, adjusted</th>
<th>Cost, adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital annual volume</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st quartile</td>
<td>1.43 (0.91–2.32)</td>
<td>1.18 (0.92–1.52)</td>
<td>−3.40 (−9.83 to 3.48)</td>
<td>−11 (−20 to 0.1)</td>
</tr>
<tr>
<td>2nd quartile</td>
<td>1.18 (0.76–1.82)</td>
<td>1.01 (0.81–1.27)</td>
<td>−0.69 (−7.70 to 6.85)</td>
<td>−3 (−14 to 9)</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>1.09 (0.72–1.66)</td>
<td>1.15 (0.96–1.39)</td>
<td>−0.86 (−8.24 to 7.11)</td>
<td>−6 (−17 to 7)</td>
</tr>
<tr>
<td>4th quartile</td>
<td>Referent</td>
<td>Referent</td>
<td>Referent</td>
<td>Referent</td>
</tr>
</tbody>
</table>

| Physician annual volume |                     |                       |                          |               |
| 1st quartile            | 1.32 (0.92–1.89)    | 0.99 (0.81–1.22)      | 12.19 (7.46–17.13)       | −1 (−5 to 3)  |
| 2nd quartile            | 1.25 (0.87–1.81)    | 0.97 (0.82–1.15)      | 9.92 (5.13–14.92)        | 3 (−1 to 7)   |
| 3rd quartile            | 1.06 (0.76–1.48)    | 0.93 (0.77–1.12)      | 3.52 (−1.25 to 8.53)     | −1 (−5 to 3)  |
| 4th quartile            | Referent            | Referent              | Referent                 | Referent      |

| Missed quality measures |                     |                       |                          |               |
| No β-blockers first 2 days | 0.56 (0.39–0.79) | 1.04 (0.87–1.24) | −10.53 (−13.03 to −7.96) | −11 (−13 to −9) |
| No prophylactic antibiotics on day of surgery | 1.18 (0.78–1.78) | 1.05 (0.84–1.32) | 0.12 (−3.19 to 3.55) | −1 (−3 to 2) |
| Antibiotics received after first postoperative day | 1.05 (0.79–1.39) | 1.08 (0.96–1.21) | 21.54 (19.49–23.61) | 17 (16–19) |
| No VTEP first 2 days | 0.81 (0.62–1.06) | 0.88 (0.77–1.00) | −1.20 (−3.20 to 0.84) | −3 (−4 to −1) |

VTEP, venous thromboembolism prophylaxis.
experience. With these limitations, it is important to note that others have found that the volume to outcomes relationship in cancer surgery might be weaker than previously described.8,36 This might be because secular trends in surgical outcomes are disproportionately affecting high-mortality centers,32 or because the pressure to contract based on volume has already made substantial progress toward moving cases away from lower-volume centers and toward higher-volume ones. Although we used all available diagnosis code data for risk adjustment in our models and attempted to avoid pitfalls described by others,37 we might have been unable to fully adjust for shifting of higher-complexity patients to higher-volume centers. This possibility is suggested by the observation that higher-volume centers were more costly, although it seems equally possible that larger centers also tend to provide more costly care.

Few of our individual process measures were associated with improvements in outcomes or resource use. Although this might be because we used measures that parallel but do not entirely replicate chart-abstracted process measures, our data are consistent with previous evidence suggesting that performance on publicly reported quality measures explains only a small portion of differences in patient outcomes.38 In fact, early experience with Surgical Care Improvement Project measures—on which our measures were based—suggests no relationship between better performance on individual quality measures and improved outcomes11–14 in colorectal surgery. Like the Surgical Care Improvement Project measures, our quality measures address a few key processes in perioperative care and do not capture other key elements of operative or perioperative care.

In contrast, overall quality is thought to be a measure of a system’s ability to deliver care reliably7,34,39–41; reliability and consistency form the rationale for the growing use of checklists in clinical care.42–44 In our study, overall quality represents the proportion of patients who did not “miss” an opportunity to receive appropriate care. Our measure was developed (to the greatest extent possible with our data) using information that might represent appropriately withholding a medication and as such our overall quality measure represents the cumulative impact of multiple appropriate clinical decisions, in addition to the reliability of the system of care. This study did not demonstrate the strong impact of overall care quality on mortality observed in our previous work15 (for reasons described earlier), although it is important to note that maximizing overall quality in this study would save approximately $3,400 per patient. When applied to the >80% of patients in our study who missed at least 1 quality measure, such cost savings would have an enormous economic impact.

Our study has a number of limitations. Because we used administrative data, we cannot easily distinguish complications from pre-existing disease and cannot replicate chart-based Surgical Care Improvement Project measures exactly. However, we constructed our quality measures to focus on patients with no documented contraindications and we did not use comorbidities to define outcomes. Our quality measures focus primarily on inpatient medications and cannot distinguish between continuation of home medications and initiation of medications in hospital. This factor might be influencing the associations seen between β-blocker use and outcomes, but is less likely to affect antimicrobial or serial compression device use. In addition,
our quality measures were collected from electronic billing systems rather than chart abstraction and have not been validated in a scientific study. However, because Premier’s business model focuses on provision of accurate benchmarking data to their members, all charge and diagnosis data are regularly audited for accuracy. As an observational study, results are subject to biases related to nonrandom assignment of patients to receive medications or devices, as well as the documentation biases described. However, secondary analyses including adjustment for hospital-level likelihood of receipt of quality measures did not suggest this was a substantial threat (data not presented). Although Premier hospitals are similar to other US centers in terms of size, teaching status, and location, they can differ from non-Premier hospitals. Although we constructed our volume measures to be consistent with those used in previous work, they might not adequately represent expertise accrued if low-volume surgeons were performing other complex operations. Although we selected our surgery types according to previous studies, it is possible that our approach missed some procedures performed less frequently. In addition, our study had a somewhat smaller number of high-volume hospitals than other studies, a trend that might have further limited our ability to see strong volume effects. Our volume measures do not take into account cases that surgeons performed outside hospitals participating in Premier or cases of similar complexity within target hospitals and, as such, might underestimate surgeon volume and experience. Finally, it is likely that some operations in our dataset were at least partially performed by fellows or residents. However, whether the surgery was performed at a teaching hospital was not a major predictor of outcomes in our models.

Our study represents a first look at the important relationship between case volume, care quality, and outcomes of care. Borderline associations between improved outcomes and higher volume in our data can be countered by higher costs at high-volume sites. Quality of care as measured in our data has, at best, little association with patient outcomes, but worse quality care was far costlier. Efforts to simultaneously encourage patients to “follow the crowd” and increase the quality of health care have strong face validity, but might have a heterogeneous impact on the value of health care.

Study conception and design: Auerbach, Maselli, Carter, Pekow, Lindenauer

Acquisition of data: Auerbach, Pekow, Lindenauer

Analysis and interpretation of data: Auerbach, Maselli, Carter, Pekow, Lindenauer

Drafting of manuscript: Auerbach, Maselli, Carter, Pekow, Lindenauer

Critical revision: Auerbach, Maselli, Carter, Pekow, Lindenauer

Acknowledgment: We would like to acknowledge Erin Hartman, MS, for her expert editorial assistance, as well as Denise Remus, MD and Kathy Belk for their work in assembling the dataset used for this analysis.