Bariatric Surgery Outcomes at Designated Centers of Excellence vs Nondesignated Programs

Edward H. Livingston, MD

Objective: To compare outcomes of patients undergoing bariatric procedures in hospitals designated as centers of excellence compared with nondesignated hospitals.

Design: The 2005 National Inpatient Survey was used to compare outcomes at designated vs nondesignated hospitals. In addition to conventional null-hypothesis statistical testing to assess differences, effect sizes were calculated to estimate the clinical significance for observed differences.

Results: Centers of excellence performed substantially more operations than nondesignated centers. Despite this, outcomes were equivalent at centers of excellence and hospitals without this designation. Volume-outcome modeling attempting to identify the optimal number for a minimum volume threshold for bariatric operations revealed that annual procedure volume has a weak effect on outcomes. Similarly, many variables that were statistically significantly different between centers and noncenters proved to be clinically unimportant by effect size analysis. Risk adjustment was effectively achieved by using the Agency for Healthcare Research and Quality–supplied variables all-payer severity-adjusted diagnostic related group expected charges and deaths.

Conclusions: Designation as a bariatric surgery center of excellence does not ensure better outcomes. Neither does high annual procedure volume. Extra expenses associated with center of excellence designation may not be warranted.

Arch Surg. 2009;144(4):319-325

See Invited Critique at end of article

METHODS

Bariatric surgery centers of excellence (COE) were promoted by the Centers for Medicaid and Medicare Services (CMS) in 2006. A new Bariatric Surgery National Coverage Determination included the requirement that CMS-insured patients undergo bariatric procedures in a COE designated by either the American College of Surgeons or the American Society for Metabolic and Bariatric Surgery.1 This recommendation resulted from concern about the safety of bariatric procedures, especially for high-risk patients such as those typically covered by the CMS.

The CMS did not specify what would constitute a COE. The American College of Surgeons and the American Society for Metabolic and Bariatric Surgery established criteria for COE status, with each organization issuing nearly identical guidelines.2-4 Meeting both sets of criteria requires a certain amount of program support and coordination, and both require entry of outcomes data into proprietary databases and the performance of at least 125 operations per year. These criteria make intuitive sense but lack an evidence base for their application.5 Program characteristics thought to be related to optimal outcomes for bariatric surgery were delineated in a 2004 Betsy Lehman Center report on patient safety.6 Because there have been few studies in this area, this report primarily relied on expert opinion as evidence for recommendations. The relationship between the recommended structural elements for a bariatric surgery program and its outcomes are not known. I hypothesize that bariatric surgery programs fulfilling the requirements necessary to become a designated COE will have better outcomes than nondesignated centers.

DATABASE ACQUISITION AND CASE IDENTIFICATION

Data from the 2005 National Inpatient Survey (NIS) were obtained from the Agency for Healthcare Research and Quality.7 Most programs with COE status obtained it in 2006 based on their performance for the prior year. The NIS is a population-representative sampling of hospital discharge data that includes 20% of all the hospitalizations in the United States in any given year.
year. In contrast to the other large hospital discharge database (the National Hospital Discharge Survey), which samples a fraction of discharges from any given hospital, the NIS obtains information about all discharges from a select number of facilities in the United States. This method has the advantage of including the full spectrum of activity from hospitals in various regions. Although the NIS is statistically corrected to be population representative, data are only collected from hospitals in 29 states, and ethnicity data are often incomplete. Therefore, there are regions and populations that may not be well represented in the NIS. The major advantage of the NIS is the completeness of information from individual hospitals, which facilitates volume-outcome analysis. The NIS also includes discharge information from a variety of hospital types, including small community facilities and large academic teaching hospitals.

To best reflect the primary reason a patient was hospitalized, patients are assigned to a diagnosis related group (DRG) at the time of discharge based on diagnostic and procedure codes, as well as information from the patient’s medical record. Bariatric procedures were identified as discharges encoded with DRG 288. The code DRG 288 (operating room procedures for obesity) is only used when the primary reason for hospitalization is to undergo procedures related to morbid obesity. Therefore, any patient encoded with DRG 288 was admitted with the intent to perform an operation for a problem related to a patient’s morbid obesity. Procedures associated with plastic surgery, ie, those with ICD-9-CM procedure codes ranging from 83.8X to 86.3X, were excluded. Similar to a prior approach for case identification, operations identified with ICD-9-CM codes from 43.80 to 44.98 and DRG 288 were defined as bariatric surgical procedures. Poverty was defined as living in a zip code region with a median annual income of less than $35 000, which since 2003 has been the Agency for Healthcare Research and Quality income threshold for defining poverty. Elderly was defined as being older than 65 years.

The bariatric procedure case volume for each facility was counted for the year 2005. Deaths were identified from the NIS variable DIED (ie, died while in the hospital). Complications were counted for the year 2005. Deaths were identified from the NIS than 65 years. DRG288 were defined as bariatric surgical procedures. The code DRG 288 (operating room procedures for obesity) is only used when the primary reason for hospitalization is to undergo procedures related to morbid obesity. Therefore, any patient encoded with DRG 288 was admitted with the intent to perform an operation for a problem related to a patient’s morbid obesity. Procedures associated with plastic surgery, ie, those with ICD-9-CM procedure codes ranging from 83.8X to 86.3X, were excluded. Similar to a prior approach for case identification, operations identified with ICD-9-CM codes from 43.80 to 44.98 and DRG 288 were defined as bariatric surgical procedures. Poverty was defined as living in a zip code region with a median annual income of less than $35 000, which since 2003 has been the Agency for Healthcare Research and Quality income threshold for defining poverty. Elderly was defined as being older than 65 years.

The bariatric procedure case volume for each facility was counted for the year 2005. Deaths were identified from the NIS variable DIED (ie, died while in the hospital). Complications were counted for the year 2005. Deaths were identified from the NIS than 65 years. DRG288 were defined as bariatric surgical procedures. The code DRG 288 (operating room procedures for obesity) is only used when the primary reason for hospitalization is to undergo procedures related to morbid obesity. Therefore, any patient encoded with DRG 288 was admitted with the intent to perform an operation for a problem related to a patient’s morbid obesity. Procedures associated with plastic surgery, ie, those with ICD-9-CM procedure codes ranging from 83.8X to 86.3X, were excluded. Similar to a prior approach for case identification, operations identified with ICD-9-CM codes from 43.80 to 44.98 and DRG 288 were defined as bariatric surgical procedures. Poverty was defined as living in a zip code region with a median annual income of less than $35 000, which since 2003 has been the Agency for Healthcare Research and Quality income threshold for defining poverty. Elderly was defined as being older than 65 years.
The overall in-house mortality was 0.1%, and the complication rate was 6.4%. The mean and variance for these variables were equal, indicating that they followed a Poisson distribution. Patient and program characteristics are presented in Table 1. The mean number of cases performed per facility was substantially greater for COEs than for the hospitals that were not COEs. Although there were statistically significant differences between the types of hospitals in patient age, the proportion of elderly patients, male sex, patient Medicaid status, patient poverty status, and teaching hospital status, the absolute differences and effect sizes were small. Hospital costs at the COE facilities were significantly higher, but, as indicated by the small effect size, this difference is not clinically important.

Logistic regression for mortality and complication rates is presented in Table 2. With risk adjustment by inclusion of the APS-DRG expected mortality rate in the regression equation, none of the factors tested were predictive of mortality. The APS-DRG variable expected charge was used as a risk-adjustment variable. It was significantly correlated with morbidity, with an OR (95% CI) of 1.78 (1.72-1.84) (P < .001). Procedure volume was inversely related to complication rates, whereas patient age and teaching hospital status were positively associated with postoperative complications. Further exploration of the volume-complication rate was carried out with GAM regression, using spline smoothing functions; GAM revealed that there was a significantly nonlinear relationship between the APS-DRG expected mortality rate and complications (Figure 1). Figure 1 also reveals that, although there is a procedure volume-complication rate relationship, it is weak. When a logistic regression equation was solved using quadratic terms for the expected mortality rate, the c-index improved to 0.856, and age lost statistical significance in the model. The OR (95%


### Table 2. Logistic Regression of In-House Mortality and Risk Factors After Bariatric Procedures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Point Estimate</th>
<th>Odds Ratio (95% Confidence Interval)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COE</td>
<td>0.5660</td>
<td>1.76 (0.73-4.26)</td>
<td>.21</td>
</tr>
<tr>
<td>No. of bariatric procedures per year</td>
<td>0.1957</td>
<td>1.22 (0.77-1.92)</td>
<td>.40</td>
</tr>
<tr>
<td>Age, y</td>
<td>0.2610</td>
<td>1.30 (0.83-2.03)</td>
<td>.25</td>
</tr>
<tr>
<td>Male sex</td>
<td>0.0908</td>
<td>1.10 (0.41-2.90)</td>
<td>.85</td>
</tr>
<tr>
<td>Teaching hospital status</td>
<td>0.8044</td>
<td>2.24 (0.84-5.93)</td>
<td>.11</td>
</tr>
<tr>
<td>Medicaid enrollment</td>
<td>-1.1571</td>
<td>0.31 (0.04-2.46)</td>
<td>.27</td>
</tr>
<tr>
<td>Poverty status a</td>
<td>0.2105</td>
<td>1.23 (0.44-3.45)</td>
<td>.69</td>
</tr>
<tr>
<td>Expected mortality rate</td>
<td>0.6316</td>
<td>1.88 (1.64-2.16)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>c-Index</td>
<td>0.7990</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

**Morbidity**

| COE                       | 0.0031         | 1.00 (0.87-1.15)                    | .97     |
| No. of bariatric procedures per year | -0.2199       | 0.80 (0.75-0.86)                    | <.001   |
| Age, y                    | 0.0931         | 1.10 (1.03-1.17)                    | .003    |
| Male sex                  | -0.0352        | 0.95 (0.81-1.11)                    | .51     |
| Teaching hospital status  | 0.1260         | 1.14 (1.00-1.29)                    | .05     |
| Medicaid enrollment       | -0.0475        | 0.96 (0.76-1.21)                    | .67     |
| Poverty status a          | 0.0840         | 1.08 (0.84-1.16)                    | .84     |
| Expected charges          | 0.5773         | 1.78 (1.72-1.84)                    | <.001   |
| c-Index                   | 0.8460         | NA                                  | NA      |

Abbreviations: COE, center of excellence; NA, not applicable.

a Defined as living in a zip code region with a median annual income of less than $35 000.

CI) for procedure volume in relation to complication rates was 0.84 (0.78-0.90) (P < .001).

**Figure 2** shows the volume-outcome effect for complications and the expected complication rate and 95% CIs for any given hospital volume based on the Poisson distribution. As hospital volumes decrease, the expected complication rate increases with profound expansion of the CIs as the volumes become very low. There are approximately as many hospitals above the 95% CIs as there are below them, suggesting that hospitals may exhibit higher or lower than expected complication rates irrespective of their procedure volume status.

Given that there was a statistically significant hospital volume effect for complications, we examined the effect of arbitrarily assigning hospital volumes to predefined categories on the resultant ORs (**Table 3**). The minimal model shows that, as one considers the procedure volume-effect in single unit increments, the OR barely reaches significance. With aggregation into larger volume units, the OR appears larger. Bariatric procedures conducted at hospitals performing fewer than 125 cases per year would appear to have an 18% greater likelihood of being associated with complications. This is reduced to 10% when the model is improved by addition of the appropriate quadratic terms and risk adjusters in the regression equation.

If there is a volume effect, what volume should be used as a threshold for certifying bariatric surgery centers? This question was addressed by O/E ratio calculation. I used the APS-DRG expected charges as a proxy for prediction that complications might develop. Regression modeling demonstrated that inclusion of this variable in the complication regression model resulted in a very high c-index, suggesting that it serves as a good predictor for untoward events. The O/E ratio APS-DRG charges were calculated, as were the 95% CIs (**Figure 3**). Statistical significance was observed only for those facilities performing more than 500 procedures per year.

Substantially more bariatric operations were performed at COEs than hospitals not having that designation. Despite their higher volumes and COE designation, out-
comes were equivalent to those at facilities not designated as COEs. Patient care costs were statistically significantly higher at COEs, but effect size measurements suggested that these differences were not clinically significant. It has been shown that the minimal annual procedure volume required to be designated as a COE (125 cases per year) does not necessarily result in better outcomes and that the minimum volume requirement is not evidence based. Most important, this volume criterion significantly restricts access for bariatric surgery care. The number of bariatric operations performed each year was the most striking difference between bariatric surgery COEs and hospitals that were not COEs. Patient and facility characteristics were similar as were complication and death rates. These findings demonstrate that COE designation is not associated with better outcomes, despite the higher annual procedure rate associated with this designation.

Designation as a bariatric surgery COE requires a significant amount of personnel and infrastructure support. All COEs must have a bariatric surgery coordinator, personnel dedicated to data entry into proprietary databases, personnel devoted to following up patients long term, and subscription to one of the database services used to track bariatric surgery outcomes. No evidence exists that these program structural elements translate to better outcomes. Criteria such as entry of outcomes data into proprietary databases result in substantial program costs yet do not have a clear relationship to surgical outcomes. Neither the American Society for Metabolic and Bariatric Surgery bariatric outcomes longitudinal database nor the American College of Surgeons National Surgical Quality Improvement Program databases have been shown to improve bariatric surgery outcomes. Assumptions have been made that use of these databases will mimic the successes of the US Department of Veterans Affairs National Surgical Quality Improvement Program experience. This is not likely, because the Department of Veterans Affairs oversees a self-contained health care system that has a central authority with the ability to intervene in underperforming surgical programs, a process not possible in the private sector. In an era where most hospitals have operational deficits and physician reimbursement is falling, requiring additional costs in the name of improved qual-

Table 3. Effect of Variable Categorization on Logistic Regression Results

<table>
<thead>
<tr>
<th>Unit</th>
<th>Odds Ratio (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimal Model</td>
</tr>
<tr>
<td>1</td>
<td>1.001 (1.001-1.002)</td>
</tr>
<tr>
<td>25</td>
<td>1.033 (1.026-1.040)</td>
</tr>
<tr>
<td>50</td>
<td>1.067 (1.052-1.082)</td>
</tr>
<tr>
<td>125</td>
<td>1.175 (1.153-1.217)</td>
</tr>
<tr>
<td>c-Index</td>
<td>0.574</td>
</tr>
<tr>
<td></td>
<td>Fully Loaded Model</td>
</tr>
<tr>
<td>1</td>
<td>1.001 (1.000-1.001)</td>
</tr>
<tr>
<td>25</td>
<td>1.019 (1.012-1.027)</td>
</tr>
<tr>
<td>50</td>
<td>1.039 (1.025-1.037)</td>
</tr>
<tr>
<td>125</td>
<td>1.101 (1.059-1.145)</td>
</tr>
<tr>
<td>c-Index</td>
<td>0.856</td>
</tr>
</tbody>
</table>

As the units (number of bariatric procedures performed per year) are aggregated into larger categories, the apparent risk for complications between high- and low-volume facilities increases. This effect is still present even with the substantially improved risk adjustment, as is seen in the fully loaded model containing quadratic terms that account for the shape of the observed complication-volume outcome relationship and a number of patient and hospital factors that can independently influence complication rates. 
P<.001 for all comparisons.
ity should only be imposed if those expenses can be irre-

futably justified by their benefit in terms of improved out-

comes. The present study suggests that expenses related to

the structural elements required to achieve bariatric sur-

gery COE status may not be justified.

Because of the large numbers of patients studied, most

of the differences observed for between-groups variables

were statistically significant. This is to be expected with a

large database analysis in which, because of large num-

bers of patients, small absolute differences tend to be sta-

tistically significant.\textsuperscript{18-20} In its most general usage, the term

\textit{statistical significance} simply means that a difference ex-

ists between 2 populations given some a priori definition

of the degree of certainty that such a difference exists. Most

tests used to quantify differences (ie, the \textit{t} test) were de-

veloped to extrapolate results from groups with small sample

sizes to the entire population and establish whether statisti-

cally significant differences are likely. When applied to

groups that are big enough to approximate the entire popu-

lation, small absolute differences are frequently inter-

preted as being statistically significant. Effect sizes are cal-

culated to provide a sense for how big a difference exists

between groups when a statistically significant result is

found\textsuperscript{13,14,21} For the present study, effect sizes were small

for all measured variables except for annual procedure vol-

ume, suggesting that, although most of the differences found

between COEs and non-COEs were statistically signifi-

cant, they are not of much clinical significance.

The largest difference between the COEs and non-

COEs was procedure volume. Procedure volume was also

significantly related to complication rates. The question was

asked: If procedure volume is related to complication rates,

what should the annual procedure volume be for COE de-

signation? One method used to answer this question was

GAM. Most regression techniques treat the variable coef-

ficients as constants to be determined by the regression pro-

cess. In contrast, GAM uses a function instead of a con-

stant, giving the regression process added flexibility and

allowing further exploration of the relationships between

variables. Using this technique, I had anticipated an inflec-

tion in the curve describing the complication–procedure

volume relationship at the point where the annual proce-

dure volume would affect outcomes. Figure 1 shows that

inflections occurred in the complication–expected charges

but were not very distinct for the complication–procedure

volume curve. This latter curve implies that the complica-

tion–procedure volume relationship is weak. The

O/E ratios were calculated from each hospital’s observed

and expected complication rates. None of these reached sta-

tistical significance for low-volume facilities and only re-

gularly were less than 1.0 (implying better than expected out-

comes) for facilities exceeding 500 bariatric surgery cases

per year. Finally, an individual hospital’s complication rates

were plotted against procedure volume along with the ex-

pected observed rates and their 95\% CIs based on the Pois-

son distribution. Count data such as those used for assess-

ing complication rates follow the Poisson distribution, mea-

ning that at low procedure volumes substantial uncertain-

ity exists regarding what the true complication rate is

for a facility given the uncertainty related to sampling phe-

nomenon.\textsuperscript{3} As seen in Figure 2, the very wide CIs ob-

served for low-volume hospitals preclude definite conclu-

sions regarding the volume-outcome relationship. The

complication–procedure volume relationship was also in-

consistent with as many hospitals being high outliers as low

outliers. Had low volumes been consistently associated with

more complications, one would have expected more low-

volume facilities to have complication rates outside of the

Poisson 95\% CI.

There are important limitations to this study. I only ex-

amined inpatient complication rates, yet the very short

lengths of stay associated with bariatric surgery suggest that

when complications occur they may manifest as read-

missions. This effect should be equally distributed among

COEs and non-COEs. I did find a complication rate of 6.4\%,

a figure consistent with several recent studies of bariatric

procedure complication rates. Because bariatric surgery

COE criteria emphasize facility requirements directed at

patient safety during the operation, one would expect to

find fewer complications in COEs than in non-COEs, which

should have been apparent in the index hospitalization.

Another limitation was our sampling of 7.6\% of the total

number of COEs. Although a small sample, the centers

included in the study were responsible for 30.0\% of all the

bariatric procedures in the cohort studied. I found that

COEs perform substantially more bariatric procedures than

non-COEs, yet both had commensurate outcomes. These

findings suggest that the much larger number of hospi-

tals that perform low volumes of bariatric procedures have

outcomes similar to the high-volume COEs.

Accepted for Publication: March 3, 2008.

Correspondence: Edward H. Livingston, MD, Division of Gastrointestinal and Endocrine Surgery, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Rm E7-126, Dallas, TX 75390-9156 (edward .livingston@utsouthwestern.edu).

Financial Disclosure: None reported.

Funding/Support: This study was funded by the Hudson-

Penn endowment, grants P20RR020691 and PL1DK081183

from the National Institutes of Health, and grant VA HSRD

IIR-05-201 from the Department of Veterans Affairs.

Additional Contributions: Tracy Schiffner-Smith, MS, of the Department of Veterans Affairs National Surgical Quality Improvement Program supplied the O/E ratio algorithm.

REFERENCES

1. Centers for Medicaid and Medicare Services Bariatric Surgery National Cover-


2. Livingston EH. Can’t we all get along? the need to unify our efforts at bariatric


cessed December 26, 2007.


5. Livingston EH, Elliott AC, Hynan LS, Engel E. When policy meets statistics: the

very real effect that questionable statistical analysis has on limiting health care


gov/db/nation/nis/NIS_Introduction_2005.jsp. Published June 2007. Ac-

cessed December 26, 2007.
This study by Livingston uses the NIS administrative database to examine mortality outcomes in accredited vs nonaccredited bariatric surgery centers. As with any study, there are methodological limitations. However, we would like to focus this commentary on improving the quality of bariatric surgery and care.

Clearly, outcome is an important issue for improving most types of medical care. Studies using administrative data, such as the current one, frequently use inpatient mortality as the outcome because it is available and well coded in administrative discharge databases. However, if the bariatric surgeon and patient communities had their druthers, a number of equally, if not more, meaningful outcomes would probably be used to measure the quality of bariatric surgery. At the current time, bariatric procedures have become markedly safer, with substantially lower inpatient mortality rates compared with reports from years past. So, to evaluate the quality of a bariatric surgery program, other outcomes such as weight loss, patient satisfaction, quality of life, and functional status should probably be used in addition to hospital mortality.

A second issue is that quality improvement is an iterative process, and, therefore, striving to continually move forward with improving quality of care requires data—high quality and appropriate data. Although some may contend that data collection and feedback might not produce improved quality in the private sector as they do in the Department of Veterans Affairs health care system, there are a number of well-known private sector quality improvement models (eg, Lean/Toyota, Six Sigma, Northern New England Cardiac Project, etc) that have led to markedly improved care. The common thread of these and other quality improvement programs is data collection, feedback, intervention, and more data collection.

How do we obtain high-quality, standardized, valid data on clinically meaningful outcomes? Currently, one of the better ways is through the formal bariatric surgery quality programs. Two such programs are those of the American College of Surgeons and the American Society of Metabolic and Bariatric Surgery. Although the latter is designed as a COE program, the former is a bariatric surgery network that includes large and small facilities achieving quality. Regardless of philosophical constructs, both include databases that require the facilities to input data, not just mortality data but also data that are more appropriate measures of bariatric surgery quality.

This notion of “iterativeness” is also important, not only because it pertains to the process of clinical quality improvement but also because it relates to the quality improvement of the programs themselves. When current bariatric quality improvement programs began, procedure volume was used as a proxy for quality care because of the lack of known quality indicators. Using procedure volume is nothing new and, in fact, is being seen in other current quality initiatives, such as programs for esophagectomy and pancreatcetomy. However, as programs mature, better insight is gained into quality evaluation, improvement, and maintenance, and, increasingly, more process and outcome indicators other than procedure volume should be used. This iterative progression is how a number of quality programs are advancing, and it is likely how the bariatric programs will also progress.

Overall, the quality train has already left the station, and the field of surgery (bariatric or otherwise) needs to continue to recognize this. Although many believe the movement is going too far, too fast, the movement is indeed continuing—and to this point, it is not slowing. Rather than pushing back or ignoring the issue, it may be best to engage in the movement and to have surgeons determine what is best for our field and our patients. Evaluations, such as the one performed by Livingston, are important for us to continually evaluate, refine, and advance our programs.