Risk stratification of serious adverse events after gastric bypass in the Bariatric Outcomes Longitudinal Database

Matthew L. Maciejewski, Ph.D.\textsuperscript{a,b,\ast}, Deborah A. Winegar, Ph.D.\textsuperscript{c}, Joel F. Farley, Ph.D.\textsuperscript{d}, Bruce M. Wolfe, M.D.\textsuperscript{e}, Eric J. DeMaria, M.D.\textsuperscript{f}

\textsuperscript{a}Center for Health Services Research in Primary Care, Durham VA Medical Center, Durham, North Carolina
\textsuperscript{b}Division of General Internal Medicine, Department of Medicine, Duke University, Durham, North Carolina
\textsuperscript{c}American Society for Metabolic and Bariatric Surgery, Raleigh, North Carolina
\textsuperscript{d}Division of Pharmaceutical Outcomes and Policy, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina
\textsuperscript{e}Oregon Health & Science University, Portland, Oregon
\textsuperscript{f}New Hope Wellness Center, Raleigh, North Carolina

Received April 23, 2012; accepted July 20, 2012

Abstract

Background: There is now sufficient demand for bariatric surgery to compare bariatric surgeons and bariatric centers according to their postsurgical outcomes, but few validated risk stratification measures are available to enable valid comparisons. The purpose of this study was to develop and validate a risk stratification model of composite adverse events related to Roux-en-Y gastric bypass (RYGB) surgery.

Methods: The study population included 36,254 patients from the Bariatric Outcomes Longitudinal Database (BOLD) registry who were 18–70 years old and had RYGB between June 11, 2007, and December 2, 2009. This population was randomly divided into a 50% testing sample and a 50% validation sample. The testing sample was used to identify significant predictors of 90-day composite adverse events and estimate odds ratios, while the validation sample was used to assess model calibration. After validating the fit of the risk stratification model, the testing and validation samples were combined to estimate the final odds ratios.

Results: The 90-day composite adverse event rate was 1.48%. The risk stratification model of 90-day composite adverse events included age (40–64, >65), indicators for male gender, body mass index (50–59.9, ≥60), obesity hypoventilation syndrome, back pain, diabetes, pulmonary hypertension, ischemic heart disease, functional status, and American Society of Anesthesiology classes 4 or 5. Our final gastric bypass model was predictive (c-statistic = 0.68) of serious adverse events 90 days after surgery.

Conclusions: With additional validation, this risk model can inform both the patient and surgeon about the risks of bariatric surgery and its different procedures, as well as enable valid outcomes comparisons between surgeons and surgical programs. (Surg Obes Relat Dis 2012;8: 671–678.) Published by Elsevier Inc. on behalf of American Society for Metabolic and Bariatric Surgery.

Keywords: Bariatric surgery; Roux-en-Y gastric bypass; Gastric bypass; Risk stratification; Adverse events; Mortality

Bariatric surgery is the most effective weight loss intervention for obese patients and enables many patients to discontinue diabetes, hypertensive, and lipid-lowering medications without significant risk of mortality [1]. Unsurprisingly, there has been great demand in the United States for bariatric surgery, which increased from 8,597 procedures in 1993 [2] to >200,000 surgeries in 2007 [3]. Medicare
coverage of bariatric surgery began in 2005, which was followed by improvements in complication rates and shorter lengths of stay [4].

With this growth in surgical volume, there is now sufficient information to compare postsurgical complications and mortality, as has been done by the National Surgical Quality Improvement Program (NSQIP) [5–7], the Michigan Bariatric Surgery Collaborative [8], and the Bariatric Outcomes Longitudinal Database (BOLD) registries of bariatric patients. As noted in a recent article [9], outcomes need to be risk stratified to support meaningful comparisons and avoid implicating centers as providing poor quality when they may simply be performing surgery on higher-risk patients. There are few validated risk stratification systems for characterizing bariatric surgery patients, with the Obesity Surgery Mortality Risk Score (OS-MRS) being the most widely used risk stratification measure to date [10–12].

The purpose of this study was to develop and validate a risk stratification model to predict composite adverse events occurring within 90 days after surgery for patients aged 18–70 undergoing a primary (i.e., not revisional) Roux-en-Y gastric bypass (RYGB) performed by a surgeon participating in the American Society for Metabolic and Bariatric Surgery (ASMBS) Bariatric Surgery Center of Excellence (BSCOE) and tracked in BOLD. This risk model considered a broader array of preoperative patient risk factors and a larger population than have been considered in previous risk stratification models. Identification of predictors of composite adverse events can be used to inform patients and surgeons about the potential for serious complications after RYGB and enables valid outcomes comparisons between surgeons and surgical programs.

**Methods**

**Patients**

Details of the preoperative, operative, and postoperative BOLD data have been previously reported [13]. In brief, BSCOE participants enter detailed information on all bariatric surgery patients prospectively for each preoperative, intraoperative, and postoperative patient encounter. BOLD data are used to ensure compliance with the requirements of the BSCOE program and for research purposes to develop general knowledge about optimal bariatric surgery practices; Online Appendix 1 provides a detailed description of variables used in this study.

This study included 101,030 research-consented patients aged 18–70 years who had an initial RYGB performed by a surgeon participating in the BSCOE program between June 11, 2007, and December 2, 2009. After excluding patients who underwent a procedure other than RYGB, had a previous bariatric surgery, or met other exclusion criteria, there were 36,254 patients who had undergone a primary RYGB, who were at least 90 days past the date of surgery when the database was created, and who had complete covariate and outcomes data to enable risk stratification model building (Online Appendix 2). Patients with both open and laparoscopic RYGB were included in the final sample to enable the broadest possible generalizability.

This sample of 36,254 was randomly divided into a 50% testing sample (n = 18,127) and a 50% validation sample (n = 18,127). The testing sample was used to identify significant predictors of 90-day composite adverse events and estimate odds ratios, while the validation sample was used to assess model fit, calibration, and discrimination consistent with methods employed for risk adjustment of cardiac surgery by the Society of Thoracic Surgeons [14–16]. After the specification of the risk stratification model was finalized, the testing and validation samples were combined to estimate final odds ratios from the pooled sample.

**Outcome definition**

The outcome of interest in the RYGB model was a composite endpoint of 90-day major adverse events, based in part on an algorithm from the LABS team [17]. This composite adverse event outcome included 17 endpoints measured during the index hospitalization and in the 90 days after discharge. These endpoints included intraoperative, postoperative, or postdischarge death, anastomotic leakage, cardiac arrest, venous thromboembolism, eversion, heart failure, liver failure, multisystem organ failure, myocardial infarction, pneumothorax, pulmonary embolism, renal failure, respiratory failure, sepsis, stroke, systemic inflammatory response syndrome, and intraoperative bleeding requiring blood transfusion. Deaths due to suicide or accidents were excluded, because bariatric surgery is unlikely to be the cause of these deaths. A patient was assigned an outcome value of 1 if any of these 17 endpoints were recorded within 90 days post-surgery, and zero otherwise. Composite adverse events at 90 days were used as the first outcome for risk adjustment using BOLD data, because risk stratification results were consistent between 30-day and 90-day event rates. Mortality, occurring much less frequently than composite adverse events, was difficult to model independently, and this sample was not powered for mortality.

**Explanatory variable identification and selection**

Before estimating the risk stratification model, a PubMed search of articles published in the English language before June 2010 was conducted to determine which patient risk factors were significant predictors of postsurgical mortality and complications. The search combined terms for bariatric surgery, gastric bypass, Roux-en-Y, and obesity surgery with terms for mortality, morbidity, complications, death, and survival, and generated 47 full-text eligible articles that were extracted from 406 titles and abstracts identified in the search. Predictor variables from 19 articles were systematically assessed, and the primary predictors of mortality or
post-surgical complications included BMI, male gender, age, hypertension, diabetes, previously developed risk scores (e.g., Charlson Co-morbidity Index), and surgical center.

We then examined the prevalence of each risk factor in the overall sample to determine which risk factors might be too rare to model. For example, pulmonary hypertension was rarely (4%) coded preoperatively but was retained for consideration as a known mortality risk factor [18]. Then, pair-wise correlations were examined to ensure that redundant or highly correlated variables were not included in the final model. By necessity, co-morbidities are largely based on self-report, as patients present for surgery with co-morbid conditions that are under active treatment. To confirm the presence of each co-morbid condition (e.g., pulmonary hypertension), records dating back to the initial diagnosis and initiation of treatment would be required and are generally not available at the time of clinical assessment.

Finally, race and open versus laparoscopic procedure were explicitly excluded from consideration. Race is highly correlated with socioeconomic, insurance, and contextual factors and may not relate well to meaningful clinical physiologic factors that might predict outcomes, which makes outcomes differences by race difficult to interpret [19–21]. In addition, patient outcomes should be similar across race and ethnicity when other preoperative characteristics are accounted for. Therefore, the inclusion of race in the risk stratification model might bias surgeons to operate on patients on the basis of race instead of appropriate risk factors or benefit considerations. The procedure access (open versus laparoscopic RYGB) was also excluded from consideration in the risk stratification model because the surgical procedure is a function of feasibility, patient preference, surgeon experience, surgeon preference, and clinical decision making made intraoperatively in the face of emergent issues at the time of surgery. Procedure access type was excluded to ensure that the risk model included only preoperative factors related to the patient that could not be controlled by the surgeon.

**Final model building and testing**

Given the large RYGB sample, we were not constrained to select the most parsimonious model that retained the greatest predictive power. However, it was important to ensure that the risk stratification model would be clinically useful to bariatric surgeons, and most of the predictive power was driven by a limited number of risk factors. Model building with the testing sample was conducted in 3 steps. First, bivariate regressions were conducted using logistic regression to generate unadjusted odds ratios with initial P values to identify covariates that were likely to remain highly significant after more complete adjustment. Second, subsets of related patient demographics (e.g., age categories) and preoperative co-morbidities were estimated in restricted logistic regressions to evaluate the sensitivity of ORs from bivariate regressions when clinically related vari-
ables were also adjusted. For example, a regression that included several age categories (<40 years, 40–49, 50–59, 60–64, and ≥65) generated similar odds ratios for the 40–49, 50–59, and 60–64 subgroups, and so a single age grouping (40–64 years) was carried forward. Covariates with P values ≤.10 were carried forward into the final multivariate model, and this generous P value threshold was chosen to account for the fact that the final model was based on the pooling of the testing and validation samples.

The final risk stratification model from the testing sample included a subset of risk factors that satisfied one of two criteria: (1) the risk factor was a statistically significant (no greater than P < .05) predictor of the composite adverse event outcome in multivariate regression or (2) the OR of the risk factor was 1.25 or ≥8, if not statistically significant. The second criterion was included with the expectation that statistically insignificant risk factors with sizable ORs in the current analysis may become statistically significant as the BOLD population grows over time. In addition, risk factors that were not statistically significant on a population level may be important predictors of patient-level outcomes [22]. Goodness of fit was assessed via the c-index (a measure of model discrimination) and the Hosmer-Lemeshow test, which assesses whether the observed event rates match the expected event rates in 5 equally sized subsets of the RYGB cohort.

The initial subset of risk factors was reviewed by a bariatric surgeon (E.J.D.) to ensure face validity, content validity, and construct validity. Odds ratios and 95% confidence intervals for all covariates except one were similar in the testing and validation samples. The obstructive sleep apnea co-morbidity indicator was statistically significant in the testing sample but insignificant in the validation sample, so was excluded from the final risk stratification model. After estimating and validating the risk stratification model using logistic regression, the final model coefficients on the pooled sample of testing and validation patients were estimated using generalized estimating equations (GEE) with a binomial distribution, logit link, independence working correlation matrix, and robust standard errors. A GEE model was estimated to account for correlation among patients treated in the same hospital. All analyses were conducted using Stata MP version 11.0 (College Station, TX), and approval to conduct human subjects research was obtained from the Copernicus Group Institutional Review Board.

**Results**

*Risk profile and event rate of patients having Roux-en-Y gastric bypass*

The 36,254 patients in the RYGB cohort had an average age of 43.8 (Table 1). The average body mass index (BMI) was 47.6, and 31.6% were super-obese (BMI ≥ 50). The cohort was predominantly female (79%) and Caucasian (79%). Three percent of
patients had an American Society of Anesthesiology (ASA) score of 1, 23% had an ASA score of 2, 69% had an ASA score of 3, and 5% had an ASA score of 4 or 5. Several preoperative comorbidities were highly prevalent, including hypertension (56%), diabetes (31%), hyperlipidemia (43%), gastroesophageal reflux disease (49%), and back pain (29%). Less prevalent preoperative patient factors included ischemic heart disease (3%), obesity hypoventilation syndrome (2%), self-reported alcohol use (1%), self-reported tobacco use (7%), functional impairment requiring an assistive device for ambulation (4%), and pulmonary hypertension (4%). Finally, 91.5% of the sample had their RYGB performed by laparoscopic access while 8.5% had open RYGB.

The rate of composite adverse events at 30 days was 1.38%, and the event rate at 90 days was 1.48% (Table 2). The most common serious adverse events at 90 days were anesthesia leak (.42%), renal failure (.31%), respiratory failure (.27%), and death (.12%).

Model performance and coefficients

The final risk stratification model for 90-day composite adverse events included 12 covariates (Table 3). These included age 40–64, age ≥ 65, male gender, BMI 50–59.9, BMI ≥ 60, obesity hypoventilation syndrome, back pain, diabetes, pulmonary hypertension, ischemic heart disease, functional impairment to ambu-
<table>
<thead>
<tr>
<th>Variable</th>
<th>Pooled sample coefficient (95% CI)</th>
<th>Testing sample coefficient (95% CI)</th>
<th>Validation sample coefficient (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age ≥65</td>
<td>2.44 (1.70, 3.50)§</td>
<td>2.13 (1.18, 3.81)</td>
<td>2.70 (1.71, 4.25)</td>
</tr>
<tr>
<td>Obesity hypoventilation syndrome</td>
<td>2.12 (1.50, 3.00)§</td>
<td>2.29 (1.38, 3.78)</td>
<td>1.96 (1.21, 3.17)</td>
</tr>
<tr>
<td>Functional status</td>
<td>2.01 (1.51, 2.68)§</td>
<td>1.98 (1.29, 3.04)</td>
<td>2.04 (1.37, 3.01)</td>
</tr>
<tr>
<td>Pulmonary hypertension</td>
<td>1.94 (1.00, 3.77)</td>
<td>2.43 (.96, 6.15)</td>
<td>1.56 (.59, 4.10)</td>
</tr>
<tr>
<td>Obese 50–59.9</td>
<td>1.91 (1.47, 2.50)§</td>
<td>1.94 (1.30, 2.88)</td>
<td>1.90 (1.33, 2.72)</td>
</tr>
<tr>
<td>Back pain</td>
<td>1.72 (1.37, 2.16)§</td>
<td>1.92 (1.38, 2.68)</td>
<td>1.56 (1.15, 2.14)</td>
</tr>
<tr>
<td>Age 40–64 yr</td>
<td>1.58 (1.26, 1.98)§</td>
<td>1.73 (1.21, 2.48)</td>
<td>1.49 (1.11, 2.00)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>1.57 (1.26, 1.94)§</td>
<td>1.68 (1.22, 2.32)</td>
<td>1.46 (1.09, 1.96)</td>
</tr>
<tr>
<td>ASA class 4/5</td>
<td>1.53 (1.16, 2.02)†</td>
<td>1.51 (1.00, 2.29)</td>
<td>1.56 (1.07, 2.27)</td>
</tr>
<tr>
<td>Male</td>
<td>1.43 (1.19, 1.73)§</td>
<td>1.48 (1.11, 1.96)</td>
<td>1.41 (1.09, 1.81)</td>
</tr>
<tr>
<td>Ischemic heart disease</td>
<td>1.42 (1.01, 2.01)§</td>
<td>1.73 (1.07, 2.79)</td>
<td>1.19 (.72, 1.96)</td>
</tr>
<tr>
<td>Obese 50–59.9</td>
<td>1.25 (1.02, 1.53)*</td>
<td>1.36 (1.00, 1.84)</td>
<td>1.16 (.88, 1.52)</td>
</tr>
<tr>
<td>Sample size</td>
<td>36,254</td>
<td>18,127</td>
<td>18,127</td>
</tr>
<tr>
<td>C-statistic</td>
<td>.6773</td>
<td>.6951</td>
<td>.6624</td>
</tr>
</tbody>
</table>

P < .001.  
* P < .05, † P < .01, § P < .0001.

Limitations

Several limitations must be acknowledged. First, the validity of these results is a function of the data quality, which is subject to continuous quality improvement within BOLD. Preoperative patient factors are likely of high quality, and preliminary auditing of BOLD data supports this conclusion. However, several factors (e.g., back pain) are based on patient self-report and may be surrogate measures of unmeasured factors that are more directly related to postsurgical complications. Significant predictors, such as obstructive sleep apnea and pulmonary hypertension, could not be validated with data from original sleep studies or other clinical tests, which may introduce measurement error.

Collection and auditing of all patient outcomes data in the months after surgery to ensure sufficient data capture remains an ongoing challenge. BOLD data are verified...
during site inspection, which each BSCOE must undergo initially and every 3 years thereafter to maintain designation. All surgeries reported in BOLD are compared with a hospital-generated surgery list, and an unbiased sampling of 10% of medical records are reviewed for accuracy. In addition, all complications and readmissions occurring within 30 days of surgery are verified. Any unreported reoperations, readmissions, deaths, transfers, or revisions found during chart review trigger a 100% chart review. Inconsistencies noted during site inspections are reported to the Bariatric Surgery Review Committee, which recommends whether the applicant qualifies for or maintains BSCOE designation status. The 90-day rate of composite adverse events reported here (1.48%) is 4 times lower than the composite adverse event rate reported at 30 days postsurgery in a report from LABS [17], although the LABS composite included reinterventions that are not always associated with a “serious complication.” Under-reporting of perioperative complications by surgeons or their designee could partially explain the lower 30-day and 90-day adverse event rates in BOLD, which could affect the specification of the risk stratification model in two ways if the patients whose events are not reported are systematically different than the patients who had events. First, the risk factors included in the model may have somewhat different odds ratios. Second, there may be risk factors not included in this model that would have been added to the model. Future work is needed with more current data to replicate these initial results.

Second, the risk factors that were significant in this RYGB cohort may not generalize to patients who have not undergone RYGB. As the Society of Thoracic Surgeons has reported [16], risk factors that are significant predictors of a given outcome in a given population may not be the same predictors that are significant in another outcome or another population. Thus, future work will need to be conducted to extend this risk stratification to non-RYGB procedures (e.g., LAGB, sleeve gastrectomy) that have sufficient volume to enable similar rigorous modeling.

Third, the risk stratification variable included two highly predictive risk factors that were based on self-report by the patient (functional impairment for ambulation) or the physician (ASA class), so this model cannot easily be implemented from administrative claims data. This risk stratification model represents a comprehensive set of preoperative risk factors and suggests that other registries should consider collecting functional status and ASA information.

Conclusion

This risk stratification analysis was based on the largest registry of bariatric surgery patients available today and detailed preoperative patient characteristics and risk factors, intraoperative clinical detail, and postoperative outcome data. With additional validation, this risk model can be used by patients, surgical programs, and payors to evaluate postsurgical outcomes and improve the quality of surgical care, which will demonstrate the value of extensive data collection associated with the bariatric COE program that has been the subject of considerable debate in recent years [35,36].

Acknowledgments

Dr. Maciejewski acknowledges VA HSR&D funding (IIR 05-201, SHP 08-137). We appreciate helpful comments from Richard A. Hansen, PhD, the editor and the reviewers.

Maciejewski: Literature search, figure creation, study design, data construction, data analysis, data interpretation, writing

Winegar: Data collection, data construction, data interpretation, writing

Farley: Literature search, figure creation, study design, data construction, data analysis, data interpretation, writing

Wolfe: Data interpretation, writing

DeMaria: Data collection, data construction, data interpretation, writing.

Conflicts of Interest

Funding was provided by Surgical Review Corporation. This work was supported by the Office of Research and Development, Health Services Research, and Development Service, Department of Veterans Affairs. Dr. Maciejewski was also supported by a Research Career Scientist award from the Department of Veterans Affairs (RCS 10-391). The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs, Duke University, Surgical Review Corporation, or the University of North Carolina at Chapel Hill. ASMBS acknowledges the surgeons and facilities that enter data into BOLD for their participation in this important initiative.

Disclosure

Drs. Maciejewski and Farley have received consultation funds from Novartis, Takeda Pharmaceuticals, and the Surgical Review Corporation.

Appendix

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.soard.2012.07.020.

References

Comment on: Risk stratification of serious adverse events after gastric bypass in the bariatric outcomes longitudinal database

Ref.: SOARD 12-84R2

Dr. Maciejewski et al. are to be commended for their thoughtful and well-written analysis of risk factors for serious complications following Roux-en-Y gastric bypass. The study benefits from its access to the clinical registry of the Bariatric Outcomes Longitudinal Database (BOLD). This clinically rich data set represents patient data from >600 hospitals nationwide and is the world’s largest regis-