




Using Claims Data to Predict Pre-Operative BMI Among Bariatric Surgery Patients: Development of the BMI Before Bariatric Surgery Scoring System (B3S3)

Jenna Wong ^{1,3}, Xiaojuan Li ^{1,2}, David E Arterburn⁴, Dongdong Li^{1,2}, Elizabeth Messenger-Jones¹, Rui Wang^{1,2}, Sengwee Toh ^{1,2}

¹Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA, USA; ²Department of Population Medicine, Harvard Medical School, Boston, MA, USA; ³Optum Labs Visiting Fellow, Eden Prairie, MN, USA; ⁴Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA

Correspondence: Jenna Wong, Department of Population Medicine, Harvard Medical School & Harvard Pilgrim Health Care Institute, 401 Park Drive, Suite 401 East, Boston, MA, 02215, USA, Tel +1 617 867-4513, Fax +1 617 867-427, Email jenna_wong@hphci.harvard.edu

Background: Lack of body mass index (BMI) measurements limits the utility of claims data for bariatric surgery research, but pre-operative BMI may be imputed due to existence of weight-related diagnosis codes and BMI-related reimbursement requirements. We used a machine learning pipeline to create a claims-based scoring system to predict pre-operative BMI, as documented in the electronic health record (EHR), among patients undergoing a new bariatric surgery.

Methods: Using the Optum Labs Data Warehouse, containing linked de-identified claims and EHR data for commercial or Medicare Advantage enrollees, we identified adults undergoing a new bariatric surgery between January 2011 and June 2018 with a BMI measurement in linked EHR data ≤ 30 days before the index surgery ($n=3226$). We constructed predictors from claims data and applied a machine learning pipeline to create a scoring system for pre-operative BMI, the B3S3. We evaluated the B3S3 and a simple linear regression model (benchmark) in test patients whose index surgery occurred concurrent (2011–2017) or prospective (2018) to the training data.

Results: The machine learning pipeline yielded a final scoring system that included weight-related diagnosis codes, age, and number of days hospitalized and distinct drugs dispensed in the past 6 months. In concurrent test data, the B3S3 had excellent performance (R^2 0.862, 95% confidence interval [CI] 0.815–0.898) and calibration. The benchmark algorithm had good performance (R^2 0.750, 95% CI 0.686–0.799) and calibration but both aspects were inferior to the B3S3. Findings in prospective test data were similar.

Conclusion: The B3S3 is an accessible tool that researchers can use with claims data to obtain granular and accurate predicted values of pre-operative BMI, which may enhance confounding control and investigation of effect modification by baseline obesity levels in bariatric surgery studies utilizing claims data.

Plain Language Summary:

- Pre-operative BMI is an important potential confounder in comparative effectiveness studies of bariatric surgeries.
- Claims data lack clinical measurements, but insurance reimbursement requirements for bariatric surgery often result in pre-operative BMI being coded in claims data.
- We used a machine learning pipeline to create a model, the B3S3, to predict pre-operative BMI, as documented in the EHR, among bariatric surgery patients based on the presence of certain weight-related diagnosis codes and other patient characteristics derived from claims data.
- Researchers can easily use the B3S3 with claims data to obtain granular and accurate predicted values of pre-operative BMI among bariatric surgery patients.

Keywords: bariatric surgery, body mass index, confounding variable, comparative effectiveness research, administrative claims, supervised machine learning

Background

With a steady rise in the prevalence of severe obesity, defined as a body mass index (BMI) of 40 kg/m² or greater, among US adults over the past two decades (age-adjusted prevalence of severe obesity increasing from 4.7% in 1999–2000 to 9.2% in 2017–2018),¹ identifying effective and safe treatment options for severe obesity is an important public health priority. Compared to conventional non-surgical weight-loss treatments (e.g., lifestyle intervention or behavioral modification), bariatric surgery yields greater reductions in body weight, greater remission of chronic conditions like type 2 diabetes and metabolic syndrome, less medication use, and greater improvements in quality of life.^{2,3}

Despite these benefits, more evidence on the comparative effectiveness and safety of different bariatric operations is needed, especially in the longer term. This need has been fueled by the rapid uptake of sleeve gastrectomy (SG) over the past decade, which has quickly displaced other previously common weight loss surgeries as the top bariatric operation in the United States, despite little supporting data.^{4–6} To address this gap, recent large-scale observational studies comparing post-surgical outcomes among different bariatric operations have largely utilized data from electronic health records (EHRs) or bariatric surgery registries.^{5,7–10} Administrative claims databases may be another viable and more cost-efficient data source for bariatric surgery research, offering the potential to identify larger and more representative patient populations and capture a more complete picture of patients' medical encounters and post-surgical events across multiple health systems and over longer follow-up horizons.^{6,11}

A concern of using claims databases for bariatric surgery research, however, relates to the ability to capture BMI data. This concern is warranted given that traditional claims databases lack clinical measurements,¹¹ and weight-related diagnosis codes under the International Classification of Diseases (ICD) system are severely underutilized in most patient populations.^{12–14} However, bariatric surgery patients, particularly in the pre-operative period, may be an exception because most US health insurers require documentation of a BMI ≥ 40 kg/m² or ≥ 35 kg/m² with an obesity-related comorbidity for patients to receive prior approval and coverage of their operations.¹⁵ This requirement incentivizes documentation of weight-related ICD codes for virtually all (>98%) bariatric surgery patients,¹⁶ creating a unique opportunity for claims data to contain valuable information on pre-operative BMI. An important potential confounder of bariatric surgery comparative effectiveness studies, pre-operative BMI has been found to both vary across patients undergoing different bariatric operations^{5,7,16,17} and impact the risk of numerous safety and effectiveness outcomes, including post-operative complications,^{17–20} mortality,^{20–22} weight loss,^{17,23,24} and type 2 diabetes remission.^{25,26}

A previous study¹⁶ found that weight codes in claims data during the pre-operative period had excellent concordance with pre-operative BMI in the EHR for identifying bariatric patients with a BMI of ≥ 35 kg/m² at baseline. However, as more granular measures of BMI are desired, such high concordance is increasingly challenging to attain due to differences in the precision of weight-related ICD codes (that can only denote BMI *ranges*, eg, 45.0–49.9 kg/m², or nonspecific weight categories, eg, obesity) versus numeric BMI values in the EHR (eg, 47.2 kg/m²), as well as potential inaccuracies in documentation of pre-operative weight codes due to reasons such as miscoding (eg, data entry errors), differential coding (eg, unequal documentation of BMI-specific versus nonspecific weight codes across operations), or weight loss before the operation (eg, as part of the patient's treatment plan).¹⁶

Rather than infer pre-operative BMI from the descriptions of weight-related ICD codes, two enhancements may allow for more precise and accurate prediction of pre-operative BMI from claims data: 1) use of additional predictors (ie, features) from claims data, besides weight-related diagnosis codes, and 2) use of flexible, data-driven modeling techniques to predict pre-operative BMI. However, for such prediction models to be of practical use in bariatric surgery research, they must also be easy to disseminate and for investigators to interpret and implement in claims databases. In this study, we applied a data-driven machine learning pipeline that considered a broad range of features from claims data to create and internally validate an accessible scoring system to predict pre-operative BMI, as documented in the EHR, among bariatric surgery patients.

Materials and Methods

Data Source

This study used data from the Optum Labs Data Warehouse (OLDW), containing linked de-identified administrative claims and EHR data for commercially insured and Medicare Advantage enrollees. The database contains longitudinal

health information on enrollees and patients, representing a diverse mixture of ages, races/ethnicities, and geographical regions across the United States.²⁷ The claims component includes physician, pharmacy, and facility claims submitted for covered enrollees, while the EHR component includes clinical diagnoses, procedures, prescriptions, clinical notes, laboratory results, and vital signs (including BMI) documented for enrollees as part of routine clinical practice.

This study was approved by the Harvard Pilgrim Health Care Institutional Review Board with a waiver of individual patient consent.

Study Population

Following the inclusion and exclusion criteria from a previous study,¹⁶ we used the claims portion of the OLDW to identify a retrospective cohort of patients aged ≥ 18 years who newly underwent SG, Roux-en-Y gastric bypass (RYGB), or adjusted gastric banding (AGB) for weight loss between January 1, 2011, and June 30, 2018. The index bariatric operation could occur in an inpatient or ambulatory care setting, and eligible patients required continuous medical and pharmacy coverage during the 6-month period before their index surgery (see [eFigure](#)). From this eligible cohort, we identified those patients with a BMI measurement recorded in their linked EHR data on or within 30 days prior to the day of index surgery, where we used this subset of patients to derive and validate the scoring system for pre-operative BMI.

Model Outcome

We used the BMI values recorded in the EHR (“EHR-based”) on or within 30 days prior to the date of the index surgery as the reference standard measure of pre-operative BMI and the model outcome. All EHR-based BMI measurements were recorded as a continuous value, which we expressed to the nearest tenth of a unit (eg, 47.2 kg/m²). If multiple BMI measurements were recorded for the patient within the 30-day pre-operative period, we used the measurement closest to the index surgery.

Candidate Predictors

From the claims portion of the OLDW, we used data in the 6 months prior to the index surgery to create various domains of candidate features for pre-operative BMI (“claims-based”). The first domain included indicators for 14 weight code categories, 12 of which captured ICD codes for specific BMI categories and 2 of which captured ICD codes for non-specific weight categories (see [eTable 1](#) for category descriptions and code mappings). These indicators were used to capture all weight-related codes documented for a given patient in the prior 6 months. Other domains of claims-based features included patient demographics (age, sex, race/ethnicity, geographical region within the US, and insurance type), comorbidities (33 indicators of conditions in the Charlson/Elixhauser combined comorbidity score^{28,29} and conditions strongly associated with obesity, such as gastroesophageal reflux disease, sleep apnea, dyslipidemia, and acquired hypothyroidism), hospitalizations in the past 6 months (number of hospitalizations and total days hospitalized), dispensed medications in the past 6 months (total number of distinct generic drugs, indicator for medications approved for management of overweight and obesity, and indicators for 6 medication classes with weight gain as a known side effect), and the bariatric surgery type (SG, RYBG, or AGB).

Machine Learning Pipeline

Using a randomly selected 90% of study patients whose index operation occurred in 2011–2017 (training set), we applied a machine learning pipeline designed to yield an accurate, yet also accessible tool to predict pre-operative BMI from claims data ([Figure 1A](#)).

The pipeline involved first training a random forest regression model using all candidate features as inputs to predict pre-operative BMI. A nonparametric ensemble machine learning technique, the random forest methodology³⁰ can naturally accommodate interactions and nonlinear associations in the data without any manual model specification (making it more robust to model misspecification) and is more resilient to problems of high variability and overfitting that often plague the use of single decision trees.^{31,32} Despite these advantages and the often superior performance of the random forest algorithm over other machine learning techniques,³² this algorithm is also more complex to interpret and cumbersome to disseminate, creating practical barriers for such models to be shared and implemented by researchers.

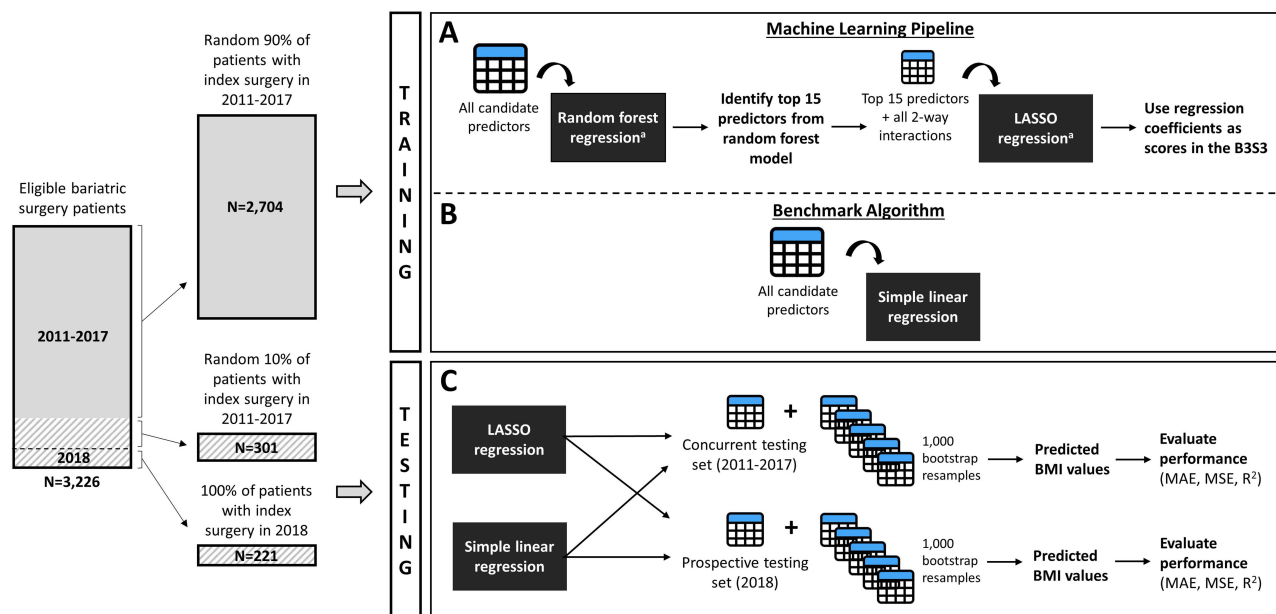


Figure 1 Derivation and validation of the B3S3 and benchmark algorithm. ^aHyperparameters were tuned using a 10-fold cross validation procedure in the training set. See [eTable 2](#) for the hyperparameters that were tuned for each algorithm and the final tuned values. The training data (solid grey box) were used to derive two prediction models for pre-operative BMI: 1) a LASSO regression model created through a machine learning pipeline, where a random forest regression model was first used to identify the 15 most important features among all candidate predictors, and then the top 15 features and all 2-way interactions between them were used as inputs into a LASSO regression model (A), and 2) a simple linear regression model utilizing all candidate predictors as inputs without any variable selection or interaction terms (B). The performance of the LASSO model and the simple linear regression model, fit on the full training data, was evaluated in two hold-out sets of the study data not used during training (the concurrent and prospective testing set; hatched grey boxes). The final fitted models were also evaluated in 1000 bootstrap resamples of the 2 testing sets to calculate empirical 95% confidence intervals for all performance metrics (C). The regression coefficients from the fitted LASSO model represent the scores in the B3S3. **Abbreviations:** B3S3, BMI Before Bariatric Surgery Scoring System; LASSO, least absolute shrinkage and selection operator; MAE, mean absolute error; MSE, mean squared error.

Thus, in the second part of the pipeline, we attempted to transfer the insights learned by the random forest model to a more familiar and accessible type of prediction algorithm – a linear regression model. We ranked the features in the random forest model in decreasing order of variable importance (measured by the total decrease in node impurities, or residual sum of squares, from splitting on a given variable, averaged over all trees in the random forest³³) and considered for inclusion in a linear regression model the top 15 predictors, along with all 2-way interactions between them to capture any 2-way interactions that may have been modeled in the random forest model. We selected a threshold of the top 15 predictors in light of the size of the training set to avoid overparameterization of the linear regression model after including all 2-way interactions. We offered these variables to a least absolute shrinkage and selection operator (LASSO) regression model, which applied shrinkage to the model coefficients (to reduce model variability) and performed automated variable selection (by shrinking some coefficients right to zero). The fitted LASSO model represented the final prediction model, and its coefficients represented the points in our scoring system, which we refer to as the *BMI Before Bariatric Surgery Scoring System* (B3S3). For both the random forest and LASSO algorithms, we tuned their hyperparameter values using a grid search procedure that iteratively assessed the algorithms' cross-validated performance over a range of plausible values (see [eTable 2](#) for details and the tuned values). Although it is typically recommended to first scale the value of continuous variables in a LASSO model, we found that in our case, scaling had no impact on the LASSO model's performance and therefore used the original values to facilitate the use and interpretation of the scoring system.

Benchmark Algorithm

We compared the performance of the LASSO model (ie, the B3S3), produced through our machine learning pipeline, to a simple linear regression model with all candidate predictors as inputs and no variable selection or interaction terms (Figure 1B). Thus, the simple linear regression model represented a benchmark algorithm that was offered the same set of

features as the machine learning pipeline but whose model specification was not determined through a flexible, machine-guided process.

Performance Evaluation

We evaluated (internally validated) the performance of the B3S3 and the benchmark algorithm by applying these models in two hold-out sets of the study data that were not used for training (Figure 1C). The first set included the remaining 10% of study patients whose index operation occurred in 2011–2017 (concurrent testing set). The second set included all study patients whose index operation occurred in 2018 (prospective testing set).

We assessed overall model performance using the following metrics: 1) mean absolute error (MAE), where the error was the absolute difference between the predicted and observed pre-operative BMI value, 2) mean squared error (MSE), where the absolute error was squared to amplify observations with larger errors and 3) R^2 , which quantified the proportion of variance in observed pre-operative BMI values explained by the model, where R^2 values closer to 1 indicated better overall performance.

We assessed model calibration (ie, accuracy of the model predictions) numerically and visually. We calculated the ratio of the mean predicted to observed pre-operative BMI, first overall (“weak calibration”) and then within three strata of predicted BMI: <40.0 , $40.0\text{--}50.0$, ≥ 50.0 kg/m² (“moderate calibration”),³⁴ where ratios closer to 1 indicated better calibration. We also constructed calibration plots of observed versus predicted BMI values, where well-calibrated observations fell along the diagonal (indicating perfect agreement between the predicted and observed BMI values). To maintain the de-identification nature of the database, we coarsened the calibration plot into 5-unit BMI categories and masked areas of the plot with <11 patients.

To measure variability in model performance due to the sampling procedure used to create the testing sets, we applied the final fitted models to 1000 bootstrap resamples of the testing sets and reported the values at the 2.5th and 97.5th percentiles of the resulting distribution.

All analyses were performed in the R environment for statistical computing, version 4.0.2. We used the glm (v4.0.2), glmnet (v4.0.2), and randomForest (v4.6.14) packages to fit the linear, LASSO, and random forest regression models, respectively.

Results

A total of 3226 bariatric surgery patients met the study inclusion criteria. Most patients were female (76.3%), middle-aged (median 48 years old), and had commercial health insurance (77.6%). The race/ethnicity of most patients was white (73.0%), followed by black (15.7%) and Hispanic (7.0%), and most patients were from the Midwest (54.0%) or South (33.6%). Hypertension, gastroesophageal reflux disease, sleep apnea, and dyslipidemia were the most common baseline comorbidities, with each affecting more than 50% of study patients (see eTable 3 for the prevalence of all 33 comorbidities considered for the scoring system). By far, the most common bariatric operation was SG (58.2%), followed by RYGB (35.2%) and AGB (6.6%). SG was performed more frequently in 2018 than in 2011–2017 (Table 1).

The median (interquartile range) pre-operative BMI in the EHR was 44.8 (40.5–50.6) kg/m², which was documented on the day of the index operation for over 50% of patients. In linked claims data during the 6-month period before the index operation, 98.3% of patients had a weight-related ICD code documented, and 96.0% had a BMI-specific code documented. Non-specific weight codes for severe obesity were more commonly documented than non-specific codes for obesity (97.4% versus 52.1%), while the top documented BMI-specific weight codes were for the ranges 40.0–44.9 kg/m² (39.5%) and 45.0–49.9 kg/m² (30.2%), which appeared more frequently for surgeries in 2018 than in 2011–2017 (Table 1).

Machine Learning Pipeline

The fitted random forest model identified the following 15 variables as most important: 12 indicators for weight-related code categories and 3 continuous variables for patient age (ranked 7th), number of distinct generic drugs dispensed in the past 6 months (ranked 12th), and number of days hospitalized in the past 6 months (ranked 15th). In ascending rank order, the top three predictors were indicators for the following weight code categories: “BMI 60.0–69.9 kg/m²”, “BMI 50.0–59.9 kg/m²”, and “Severely obese, non-specific”. Absent from the top 15 predictors were the weight code categories

Table I Characteristics of the Study Cohort

Characteristic	Overall (N=3226)	Training Set, 2011–2017 ^k (N=2704)	Concurrent Testing Set, 2011–2017 ^k (N=301)	Prospective Testing Set, 2018 ^k (N=221)
Model Outcome				
Pre-operative BMI recorded in the EHR, median (IQR)	44.8 (40.5–50.6)	44.8 (40.5–50.7)	44.8 (40.8–51.0)	44.4 (40.7–49.6)
Candidate Predictors				
Demographics				
Female, n (%)	2461 (76.3)	2057 (76.1)	232 (77.1)	172 (77.8)
Age in years, median (IQR)	48 (39–57)	47 (39–56)	48 (40–58)	50 (41–60)
Race/ethnicity, n (%)				
White	2355 (73.0)	<2006 (<74.2)	>203 (>67.4)	>146 (<66.1)
Black	505 (15.7)	423 (15.6)	50 (16.6)	32 (14.5)
Hispanic	226 (7.0)	186 (6.9)	24 (8.0)	16 (7.2)
Asian	48 (1.5)	>26 (>1.0)	<11 (<3.7)	<11 (<5.0)
Unknown	92 (2.9)	63 (2.3)	13 (4.3)	16 (7.2)
Region, n (%)				
Midwest	1742 (54.0)	<1480 (<54.7)	146 (48.5)	>116 (>52.5)
South	1084 (33.6)	886 (32.8)	120 (39.9)	78 (35.3)
Northeast	222 (6.9)	188 (7.0)	18 (6.0)	16 (7.2)
West	178 (5.5)	>150 (>5.5)	17 (5.7)	<11 (<5.0)
Commercial vs Medicare Advantage insurance, n (%)	2502 (77.6)	2128 (78.7)	231 (76.7)	143 (64.7)
Weight-related diagnosis codes^a, n (%)				
Underweight, normal weight, or overweight	99 (3.1)	>74 (>2.7)	14 (4.7)	<11 (<5.0)
BMI 30.0–34.9	73 (2.3)	>51 (>1.9)	<11 (<3.7)	<11 (<5.0)
BMI 35.0–35.9	102 (3.2)	>78 (>2.9)	13 (4.3)	<11 (<5.0)
BMI 36.0–36.9	142 (4.4)	115 (4.3)	15 (5.0)	12 (5.4)
BMI 37.0–37.9	171 (5.3)	136 (5.0)	19 (6.3)	16 (7.2)
BMI 38.0–38.9	192 (6.0)	162 (6.0)	11 (3.7)	19 (8.6)
BMI 39.0–39.9	246 (7.6)	208 (7.7)	22 (7.3)	16 (7.2)
BMI 40.0–44.9	1275 (39.5)	1060 (39.2)	116 (38.5)	99 (44.8)
BMI 45.0–49.9	973 (30.2)	799 (29.6)	89 (29.6)	85 (38.5)
BMI 50.0–59.9	886 (27.5)	745 (27.6)	83 (27.6)	58 (26.2)
BMI 60.0–69.9	253 (7.8)	210 (7.8)	24 (8.0)	19 (8.6)
BMI ≥70.0	49 (1.5)	>27 (>1.0)	<11 (<3.7)	<11 (<5.0)
Obese, nonspecific	1679 (52.1)	1397 (51.7)	167 (55.5)	115 (52.0)
Severely obese, nonspecific	3143 (97.4)	2630 (97.3)	297 (98.7)	216 (97.7)
Select comorbidities^b, n (%)				
Hypertension	2236 (69.3)	1871 (69.2)	219 (72.8)	146 (66.1)
Gastroesophageal reflux disease	1942 (60.2)	1614 (59.7)	191 (63.5)	137 (62.0)
Sleep apnea	1826 (56.6)	1516 (56.1)	181 (60.1)	129 (58.4)
Dyslipidemia	1825 (56.6)	1534 (56.7)	169 (56.2)	122 (55.2)
Diabetes	1159 (35.9)	952 (35.2)	118 (39.2)	89 (40.3)
Depression	1115 (34.6)	926 (34.3)	112 (37.2)	77 (34.8)
Anxiety	1080 (33.5)	905 (33.5)	95 (31.6)	80 (36.2)
Liver disease	875 (27.1)	718 (26.6)	93 (30.9)	64 (29.0)
Chronic pulmonary disease	841 (26.1)	675 (25.0)	97 (32.2)	69 (31.2)
Psychosis	830 (25.7)	680 (25.2)	67 (22.3)	83 (37.6)
Non-alcoholic fatty liver disease	825 (25.6)	681 (25.2)	87 (28.9)	57 (25.8)
Acquired hypothyroidism	635 (19.7)	519 (19.2)	68 (22.6)	48 (21.7)
Complicated diabetes	494 (15.3)	389 (14.4)	43 (14.3)	62 (28.1)
Hospitalizations in the past 6 months, n (%)				
At least 1 hospitalization ^c	160 (5.0)	136 (5.0)	11 (3.7)	13 (5.9)

(Continued)

Table 1 (Continued).

Characteristic	Overall (N=3226)	Training Set, 2011–2017 ^k (N=2704)	Concurrent Testing Set, 2011–2017 ^k (N=301)	Prospective Testing Set, 2018 ^k (N=221)
Total days in hospital ^d				
0	3066 (95.0)	<2588 (<95.7)	>279 (>92.7)	>199 (>90.0)
1–3	57 (1.8)	>35 (>1.3)	<11 (<3.7)	<11 (<5.0)
≥4	103 (3.2)	>81 (>3.0)	<11 (<3.7)	<11 (<5.0)
Dispensed medications in the past 6 months				
Approved drugs for overweight and obesity ^e , n (%)	126 (3.9)	>104 (>3.8)	<11 (<3.7)	<11 (<5.0)
Oral corticosteroids, n (%)	449 (13.9)	350 (12.9)	59 (19.6)	40 (18.1)
Diabetic agents ^f , n (%)	464 (14.4)	379 (14.0)	50 (16.6)	35 (15.8)
Beta blockers ^g , n (%)	493 (15.3)	405 (15.0)	54 (17.9)	34 (15.4)
Antidepressants and lithium ^h , n (%)	603 (18.7)	504 (18.6)	63 (20.9)	36 (16.3)
Antipsychotics ⁱ , n (%)	65 (2.0)	>43 (>1.6)	<11 (<3.7)	<11 (<5.0)
Anticonvulsants ^j , n (%)	361 (11.2)	283 (10.5)	41 (13.6)	37 (16.7)
Number of generic drugs dispensed, median (IQR)	8 (4–12)	7 (4–12)	8 (5–12)	7 (4–12)
Bariatric surgery type, n (%)				
Sleeve gastrectomy	1877 (58.2)	1554 (57.5)	169 (56.2)	154 (69.7)
Roux-en-Y gastric bypass	1136 (35.2)	971 (35.9)	105 (34.9)	60 (27.2)
Adjustable gastric banding	213 (6.6)	>175 (>6.5)	27 (9.0)	<11 (<5.0)

Notes: ^aConsiders all weight-related diagnosis codes recorded for the patient in the 6 months prior to the index surgery. [eTable 1](#) shows the code mappings used for these weight categories. ^bA total of 33 comorbidities (see [eTable 3](#)) were considered for inclusion in the prediction models. The 13 most common comorbidities are shown. ^cIn the models, this predictor was expressed as a continuous variable representing the number of hospitalizations in the past 6 months. ^dIn the models, the total days in hospital in the past 6 months was expressed as a continuous variable. ^eIncludes orlistat, phentermine-topiramate, bupropion-naltrexone, liraglutide, lorcaserin, diethylpropion, phendimetrazine, phentermine, and benzphetamine. ^fIncludes all insulins, sulfonylureas, and thiazolidinediones. ^gIncludes atenolol, metoprolol, nadolol, and propranolol. ^hAntidepressants include tricyclic antidepressants (amitriptyline, nortriptyline, doxepin, and imipramine), selective serotonin reuptake inhibitors (paroxetine and sertraline), mirtazapine, bupropion, and phenelzine. ⁱIncludes clozapine, olanzapine, quetiapine, risperidone, and haloperidol. ^jIncludes valproate, carbamazepine, pregabalin, and gabapentin. ^kCells with <11 patients have been masked to maintain the deidentification nature of the database.

Abbreviations: BMI, body mass index; EHR, electronic health record; IQR, interquartile range.

“Underweight, normal weight, or overweight” and “Obese, nonspecific”, as well as features related to comorbidities or previously dispensed medications.

When the top 15 predictors and all their 2-way interactions were offered to a LASSO regression model (a total of 120 terms), the fitted model retained 68 terms, of which 57 were 2-way interactions that largely included interactions between weight code categories, effectively adjusting the predicted BMI in cases where diagnosis codes for multiple weight categories (eg, conflicting categories or BMI-specific and non-specific weight categories) were documented for the patient in the previous 6 months. The coefficients of the LASSO model ranged in magnitude from -26.362 to 29.480 ([Table 2](#)) and were used as the scores in the B3S3. The [eSheet](#) contains step-by-step instructions on how to use the B3S3, along with an interactive example.

Overall Performance of the B3S3 and Benchmark Algorithm

In the concurrent testing set, the overall performance of the B3S3 was excellent, with a MAE of 2.309 (95% CI 2.045–2.575) (indicating that the prediction for a given patient was, on average, 2.309 BMI units different from the patient’s pre-operative BMI documented in the EHR), MSE of 10.682 (95% CI 7.854–13.584), and R^2 of 0.862 (95% CI 0.815–0.898). In comparison, the overall performance of the simple linear regression model (benchmark algorithm) was good, but inferior to the B3S3, with a higher MAE of 3.266 (95% CI 2.933–3.610), higher MSE of 19.306 (95% CI 15.274–23.983), and lower R^2 of 0.750 (95% CI 0.686–0.799), where all 95% CIs were non-overlapping with the B3S3. Findings were similar in the prospective testing set ([Table 3](#)).

Table 2 Coefficients of the LASSO Regression Model

Variable	Regression Coefficient ^a
Intercept	28.141
Individual predictors	
Code for BMI 36.0–36.9	−4.289
Code for BMI 38.0–38.9	−3.113
Code for BMI 40.0–44.9	−0.377
Code for BMI 45.0–49.9	3.329
Code for BMI 50.0–59.9	10.371
Code for BMI 60.0–69.9	17.795
Code for BMI 70.0+	29.480
Code for severely obese, nonspecific	15.750
Age in years	−0.025
Number of distinct generic drugs dispensed in the past 6 months	−0.010
Total days hospitalized in the past 6 months	−0.029
2-way interactions between weight-related codes	
Code for BMI 30.0–34.9 × Code for BMI 35.0–35.9	2.908
Code for BMI 30.0–34.9 × Code for BMI 36.0–36.9	1.814
Code for BMI 30.0–34.9 × Code for BMI 37.0–37.9	3.074
Code for BMI 30.0–34.9 × Code for BMI 38.0–38.9	2.324
Code for BMI 30.0–34.9 × Code for BMI 39.0–39.9	2.001
Code for BMI 30.0–34.9 × Code for BMI 40.0–44.9	2.471
Code for BMI 30.0–34.9 × Code for BMI 45.0–49.9	3.536
Code for BMI 30.0–34.9 × Code for BMI 50.0–59.9	2.935
Code for BMI 30.0–34.9 × Code for BMI 60.0–69.9	−19.947
Code for BMI 30.0–34.9 × Code for severely obese, nonspecific	−6.458
Code for BMI 35.0–35.9 × Code for BMI 36.0–36.9	1.719
Code for BMI 35.0–35.9 × Code for BMI 37.0–37.9	0.941
Code for BMI 35.0–35.9 × Code for BMI 38.0–38.9	1.131
Code for BMI 35.0–35.9 × Code for BMI 39.0–39.9	1.205
Code for BMI 35.0–35.9 × Code for BMI 40.0–44.9	1.171
Code for BMI 35.0–35.9 × Code for BMI 45.0–49.9	1.848
Code for BMI 35.0–35.9 × Code for severely obese, nonspecific	−4.299
Code for BMI 36.0–36.9 × Code for BMI 37.0–37.9	1.424
Code for BMI 36.0–36.9 × Code for BMI 38.0–38.9	1.637
Code for BMI 36.0–36.9 × Code for BMI 39.0–39.9	0.201
Code for BMI 36.0–36.9 × Code for BMI 50.0–59.9	−11.079
Code for BMI 37.0–37.9 × Code for BMI 38.0–38.9	1.135
Code for BMI 37.0–37.9 × Code for BMI 39.0–39.9	0.520
Code for BMI 37.0–37.9 × Code for BMI 40.0–44.9	0.728
Code for BMI 37.0–37.9 × Code for BMI 45.0–49.9	−1.560
Code for BMI 37.0–37.9 × Code for severely obese, nonspecific	−2.889
Code for BMI 38.0–38.9 × Code for BMI 39.0–39.9	1.911
Code for BMI 38.0–38.9 × Code for BMI 40.0–44.9	0.098
Code for BMI 38.0–38.9 × Code for BMI 45.0–49.9	−4.010
Code for BMI 38.0–38.9 × Code for BMI 50.0–59.9	−2.485
Code for BMI 39.0–39.9 × Code for BMI 45.0–49.9	0.063
Code for BMI 39.0–39.9 × Code for severely obese, nonspecific	−2.259
Code for BMI 40.0–44.9 × Code for BMI 45.0–49.9	−1.835
Code for BMI 40.0–44.9 × Code for BMI 50.0–59.9	−2.348
Code for BMI 40.0–44.9 × Code for BMI 70.0+	−26.362
Code for BMI 40.0–44.9 × Code for severely obese, nonspecific	−0.002

(Continued)

Table 2 (Continued).

Variable	Regression Coefficient ^a
Code for BMI 45.0–49.9 × Code for BMI 50.0–59.9	–8.077
Code for BMI 45.0–49.9 × Code for BMI 60.0–69.9	–14.917
Code for BMI 45.0–49.9 × Code for BMI 70.0+	–3.649
Code for BMI 45.0–49.9 × Code for severely obese, nonspecific	0.094
Code for BMI 50.0–59.9 × Code for BMI 60.0–69.9	–14.423
Code for BMI 50.0–59.9 × Code for BMI 70.0+	–15.892
Code for BMI 60.0–69.9 × Code for BMI 70.0+	–24.218
Code for BMI 60.0–69.9 × Code for severely obese, nonspecific	0.170
Code for BMI 70.0+ × Code for severely obese, nonspecific	0.402
2-way interactions between weight-related codes and age	
Code for BMI 45.0–49.9 × Age in years	0.014
Code for BMI 50.0–59.9 × Age in years	0.008
Code for BMI 60.0–69.9 × Age in years	0.035
2-way interactions between weight-related codes and number of distinct generic drugs dispensed in the past 6 months	
Code for BMI 37.0–37.9 × Number of generic drugs	–0.008
Code for BMI 45.0–49.9 × Number of generic drugs	0.005
2-way interactions between weight-related codes and total days hospitalized in the past 6 months	
Code for BMI 30.0–34.9 × Total days hospitalized	0.007
Code for BMI 39.0–39.9 × Total days hospitalized	0.079
Code for BMI 50.0–59.9 × Total days hospitalized	0.197
Code for BMI 60.0–69.9 × Total days hospitalized	0.062
Code for BMI 70.0+ × Total days hospitalized	–0.577
Code for severely obese, nonspecific × Total days hospitalized	0.005
2-way interactions between predictors besides weight-related codes	
Age in years × Total days hospitalized	–0.001

Notes: ^aRepresents the scores in the BMI Before Bariatric Surgery Scoring System (B3S3). See the [eSheet](#) for step-by-step instructions on how to use the B3S3, along with an interactive example.

Abbreviations: BMI, body mass index; LASSO, least absolute shrinkage and selection operator.

Table 3 Performance of the B3S3 and the Simple Linear Regression Model

Algorithm	Concurrent Testing Set, 2011–2017 (N=301)	Prospective Testing Set, 2018 (N=221)				
	Mean Absolute Error (95% CI)	Mean Squared Error (95% CI)	R ² (95% CI)	Mean Absolute Error (95% CI)	Mean Squared Error (95% CI)	R ² (95% CI)
B3S3 ^a	2.309 (2.045–2.575)	10.682 (7.854–13.584)	0.862 (0.815–0.898)	2.298 (1.974–2.637)	11.691 (8.056–15.804)	0.843 (0.780–0.885)
Simple linear regression model ^b	3.266 (2.933–3.610)	19.306 (15.274–23.983)	0.750 (0.686–0.799)	3.148 (2.744–3.561)	19.164 (14.276–24.366)	0.743 (0.658–0.801)

Notes: ^aUses the regression coefficients from the LASSO regression model offered the top 15 variables (and all 2-way interactions between them) from a random forest model offered all candidate predictors. ^bRepresents the benchmark algorithm offered all candidate predictors with no variable selection or interaction terms.

Abbreviations: B3S3, BMI Before Bariatric Surgery Scoring System; CI, confidence interval.

Calibration of the B3S3 and Benchmark Algorithm

In the concurrent testing set, the overall mean predicted pre-operative BMI from the B3S3 and simple linear regression model differed from the overall mean observed pre-operative BMI by only 0.3% (95% CI –0.5–1.1%) and 0.5% (95% CI –0.6–1.6%),

Table 4 Calibration of the B3S3 and the Simple Linear Regression Model

Algorithm, Overall And By Strata Of Predicted Pre-Operative BMI	Mean Pre-Operative BMI							
	Concurrent Testing Set, 2011–2017				Prospective Testing Set, 2018			
	N	P	O	P:O (95% CI)	N	P	O	P:O (95% CI)
B3S3 ^a								
Overall	301	46.347	46.222	1.003 (0.995–1.011)	221	44.878	44.489	1.004 (0.993–1.014)
BMI <40.0	53	36.759	36.089	1.019 (0.998–1.040)	40	36.227	36.185	1.001 (0.974–1.028)
BMI 40.0–<50.0	168	44.293	44.276	1.000 (0.990–1.010)	132	44.726	44.395	1.007 (0.997–1.018)
BMI ≥50.0	80	57.012	57.023	1.000 (0.985–1.016)	49	56.548	56.663	0.998 (0.972–1.024)
Simple linear regression model ^b								
Overall	301	46.434	46.222	1.005 (0.994–1.016)	221	45.086	44.489	1.010 (0.997–1.023)
BMI <40.0	38	36.467	36.029	1.012 (0.978–1.047)	35	35.350	36.091	0.979 (0.943–1.016)
BMI 40.0–<50.0	169	44.079	43.170	1.021 (1.009–1.033)	122	44.222	43.502	1.017 (1.004–1.030)
BMI ≥50.0	94	54.696	55.831	0.980 (0.960–1.001)	64	55.559	54.902	1.012 (0.983–1.041)

Notes: ^aUses the regression coefficients from the LASSO regression model offered the top 15 variables (and all 2-way interactions between them) from a random forest model offered all candidate predictors. ^bRepresents the benchmark algorithm offered all candidate predictors with no variable selection or interaction terms.

Abbreviations: B3S3, BMI Before Bariatric Surgery Scoring System; BMI, body mass index; CI, confidence interval; N, number of individuals; P, predicted mean pre-operative BMI from the fitted model; O, observed mean pre-operative BMI from the electronic health record; P:O, ratio of predicted to observed mean pre-operative BMI.

respectively. When patients were grouped into more homogeneous strata based on their predicted BMI (<40.0, 40.0–<50.0, ≥50.0 kg/m²), the calibration of the B3S3 was still excellent, with the mean predicted and observed BMI differing by <2% in the lowest stratum and <0.05% in the upper 2 strata, where the 95% CI around the predicted-to-observed ratio for all strata included 1, indicating no significant differences (Table 4). For the simple linear regression model, however, the mean predicted and observed values for the upper 2 BMI strata differed by >2%, with the 95% CI around the predicted-to-observed ratio for the middle (and most common) BMI strata being above 1, indicating that the model tended to slightly overestimate pre-operative BMI for predictions in this range. Findings were similar in the prospective testing set (Table 4). The calibration plot further illustrated the slightly better accuracy of predictions from B3S3, where patients in both testing sets were more tightly and evenly scattered along the diagonal for B3S3 (Figure 2, top row) than the simple linear regression model (Figure 2, bottom row).

Discussion

To our knowledge, this study is the first to create a claims-based model to predict pre-operative BMI among bariatric surgery patients. By applying flexible, machine-driven model specification techniques within a carefully crafted machine learning pipeline, we derived a model that could accurately predict pre-operative BMI, as documented in the EHR, and be easily implemented as a simple scoring system with claims data without requiring any knowledge or experience with more complex machine learning algorithms, thus representing an accessible tool to all bariatric surgery researchers using claims data. This scoring system, the B3S3, which we have made publicly available along with instructions and an example of how to implement it (see [eSheet](#)), uses only a handful of claims-based variables, and its predictions were found to be highly accurate, even when evaluated in prospective data. Moreover, the B3S3 outperformed a linear regression model that utilized considerably more claims-based predictors but was not specified using a flexible, machine-guided process.

An important potential baseline confounder of post-surgical outcomes, the ability to accurately measure pre-operative BMI is an important consideration when identifying suitable databases for comparative effectiveness and safety research on bariatric surgeries. Many studies to date have used data from EHRs and bariatric surgery registries^{5,7–10} for their capture of more granular, clinical data, but claims data also represent a valuable data source for their more complete capture of numerous effectiveness and safety outcomes for bariatric procedures (eg, cardiovascular events, abdominal operations, venous thromboembolism, and all-cause hospitalizations), particularly over longer follow-up horizons.⁶ However, as claims databases lack clinical measurements on BMI, our scoring system may enhance and promote the

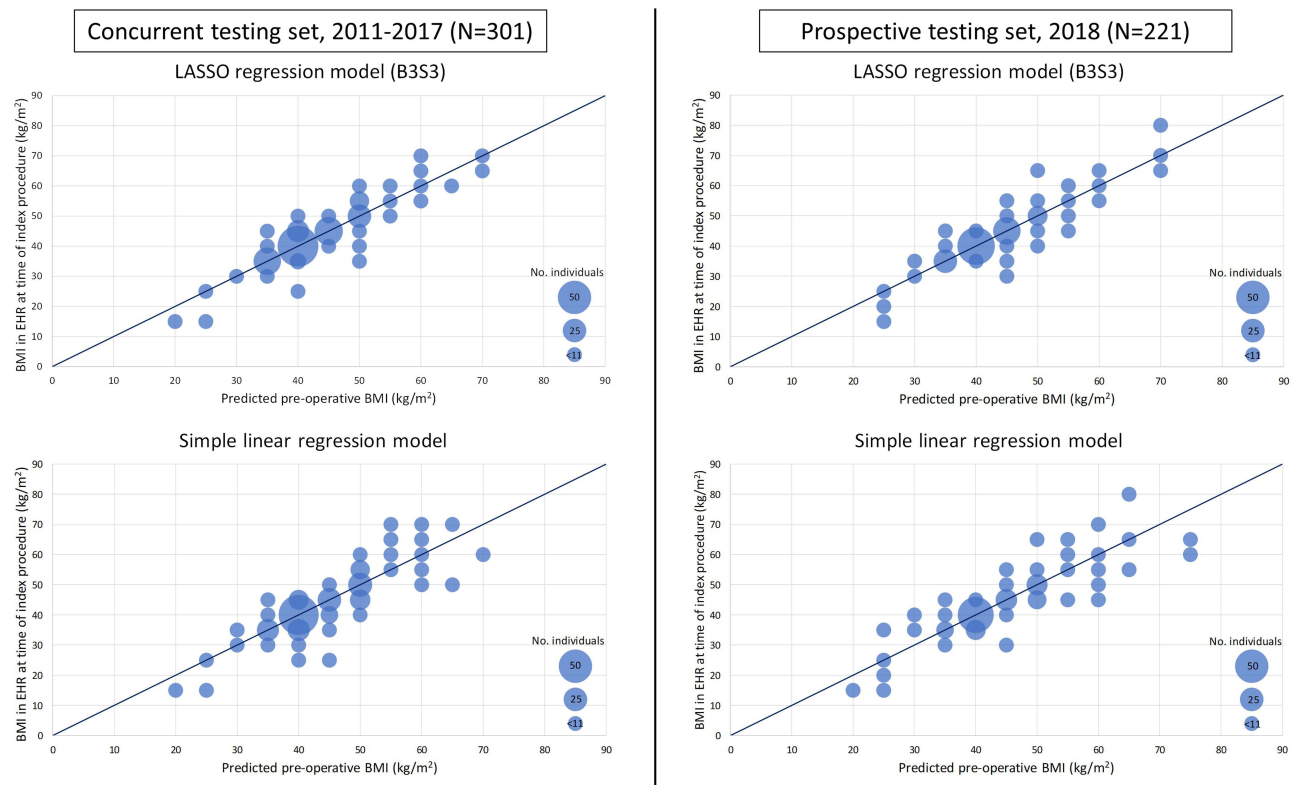


Figure 2 Calibration plots for the B3S3 and the simple linear regression model. The calibration plots show the number of individuals falling within each 5-unit category of observed versus predicted BMI, where the size of the dots is proportional to the number of individuals in the category. Areas of the plot with <11 patients are represented using the same sized dots to maintain the de-identification nature of the database. Perfect calibration occurs along the diagonal line, where the predicted and observed BMI values are equivalent.

Abbreviations: B3S3, BMI Before Bariatric Surgery Scoring System; BMI, body mass index; EHR, electronic health record; LASSO, least absolute shrinkage and selection operator.

utility of claims data for bariatric surgery research in several ways. First, compared to the approach of imputing pre-operative BMI solely based on the descriptions of weight-related diagnosis codes (which only denote non-specific weight categories, or at best, *ranges* of BMI), the B3S3 allows researchers to predict pre-operative BMI as a *continuous* value and with the same level of granularity as BMI measurements typically recorded in EHRs and bariatric surgery registries. Thus, our scoring system allows researchers to enjoy maximum flexibility in how pre-operative BMI is handled in claims-only studies (eg, as a categorical *or continuous* covariate). Second, as the B3S3 was found to be well calibrated along the full range of its predictions, these findings suggest that the B3S3 may be used to accurately identify cohorts of higher-risk bariatric surgery patients with extremely high pre-operative BMIs (eg, ≥ 50 or ≥ 60 kg/m²) or exclude patients with lower than expected pre-operative BMIs (eg, ≤ 35 kg/m²) whose operation may not be a true primary bariatric operation (eg, revisional surgery) but otherwise appears as such based on other claims-based inclusion criteria. Third, even in EHR-based studies, B3S3 may be useful as a remedy for missing data on pre-operative BMI when linked claims data are available. Indeed, a previous study⁵ had to exclude 16% of bariatric surgery patients due to a lack of baseline data on BMI in the EHR. Finally, as pre-operative BMI has been found to be a significant predictor of post-operative outcomes,^{19,23,35} the B3S3 may also be useful in claims-based studies aimed at forecasting outcomes after bariatric surgery by enabling the creation of pre-operative BMI as a candidate prognostic variable.

The capacity of the B3S3 to accurately predict pre-operative BMI is largely due to two factors: 1) the existence of weight-related codes in the ICD system, many of which denote BMI-specific ranges, and 2) reimbursement requirements by health insurers that strongly incentivize providers and institutions to document weight-related codes for nearly all bariatric patients in the months preceding their index surgery. However, it is important to note that the validity of the B3S3 is most likely specific to bariatric surgery patients in the pre-operative period and may not be generalizable to the

post-operative period (eg, to measure post-operative weight loss), when providers and institutions no longer have financial incentives to document weight-related codes. Indeed, a prior study¹⁶ found that in the first post-operative year, one-fifth of bariatric surgery patients did not have any weight-related codes in their claims data and less than one-half of patients had BMI-specific codes documented. Thus, in bariatric surgery research, the main utility of weight-related codes (and by extension, the B3S3) is to enable confounding control and investigation of effect modification by baseline severity of obesity for post-surgical outcomes that are well captured in claims data.

A strength of this study is the inclusion of a nationally representative sample of bariatric surgery patients with commercial or Medicare Advantage plans, who currently represent the majority of US individuals who undergo bariatric surgery.³⁶ By extension, a limitation of this study is that we could not include bariatric patients with other types of health insurance plans (eg, Medicaid) due to the nature of data available in the OLDW. Another limitation is the smaller size of the testing sets that were used to evaluate the B3S3. To address sampling variability in our testing sets, we bootstrapped the testing data to estimate uncertainty around all performance measures. To further evaluate the generalizability of B3S3, we strongly encourage other researchers to validate the performance of our scoring system in other claims databases, including data sources that capture other (eg, Medicaid) patient populations.

Conclusion

Pre-operative BMI is an important potential confounder in comparative effectiveness studies of bariatric surgeries. Although claims databases lack clinical measurements, the existence of weight-related diagnosis codes, combined with BMI-related reimbursement requirements by health insurers for bariatric surgeries, create valuable information on pre-operative BMI in claims data. We applied a machine learning pipeline to create a scoring system, the B3S3, that could accurately predict pre-operative BMI, as documented in the EHR, among bariatric surgery patients. Investigators can easily use the B3S3 with claims data to obtain granular predicted values of pre-operative BMI to enhance confounding control and investigation of effect modification by baseline severity of obesity in bariatric surgery studies.

Abbreviations

AGB, Adjusted gastric banding; BMI, Body mass index; B3S3, BMI Before Bariatric Surgery Scoring System; CI, Confidence interval; EHR, Electronic health record; ICD, International Classification of Diseases; LASSO, Least absolute shrinkage and selection operator; MAE, Mean absolute error; MSE, Mean squared error; OLDW, Optum Labs Data Warehouse; RYGB, Roux-en-Y gastric bypass; SG, Sleeve gastrectomy.

Data Sharing Statement

The raw data that support the findings of this study are available from Optum Labs, but restrictions apply to the availability of these data, which were used under license for the current study and so are not publicly available. The materials generated for the current study are included in this published article and its [Supplementary Files](#) as well as its [eSheet](#).

Ethics Approval and Informed Consent

This study was reviewed and approved by the Harvard Pilgrim Health Care Institutional Review Board with an exemption and waiver of individual patient consent. Exemption and waiver of individual patient consent were granted due to the use of a limited data set for this study and the execution of a data use agreement with Optum Labs to ensure specified safeguards for the information in the limited data set. All research in this study was conducted in accordance with the Declaration of Helsinki and in compliance with US data protection and privacy regulations to maintain confidentiality of patient data.

Acknowledgments

The authors would like to thank Dr. Susan Gruber for her helpful input on the analytic approaches used in this study.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising, or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study was funded by a grant from the Agency for Healthcare Research and Quality (R01HS026214).

Disclosure

Dr Xiaojuan Li reports grants from the National Institute on Aging (K01AG073651), outside the submitted work. Dr Sengwee Toh reports personal fees from Pfizer and TriNetX, outside the submitted work. In this work, the authors declare that they have no other financial or non-financial competing interests relevant to the interpretation or presentation of information.

References

- Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity and severe obesity among adults: United States, 2017-2018. *NCHS Data Brief*. 2020;360:1-8.
- Gloy VL, Briel M, Bhatt DL, et al. Bariatric surgery versus non-surgical treatment for obesity: a systematic review and meta-analysis of randomised controlled trials. *BMJ*. 2013;347:f5934. doi:10.1136/bmj.f5934
- Ribaric G, Buchwald JN, McGlennon TW. Diabetes and weight in comparative studies of bariatric surgery vs conventional medical therapy: a systematic review and meta-analysis. *Obes Surg*. 2014;24(3):437-455. doi:10.1007/s11695-013-1160-3
- Alalwan AA, Friedman J, Park H, Segal R, Brumback BA, Hartzema AG. US national trends in bariatric surgery: a decade of study. *Surgery*. 2021;170(1):13-17. doi:10.1016/j.surg.2021.02.002
- Arterburn D, Wellman R, Emiliano A, et al. Comparative effectiveness and safety of bariatric procedures for weight loss: a PCORnet cohort study. *Ann Intern Med*. 2018;169(11):741-750. doi:10.7326/M17-2786
- Ma Q, Mack M, Shambhu S, McTigue K, Haynes K. Characterization of bariatric surgery and outcomes using administrative claims data in the research network of a nationwide commercial health plan. *BMC Health Services Research*. 2021;21(1):116. doi:10.1186/s12913-021-06074-3
- Maciejewski ML, Arterburn DE, Van Scoyoc L, et al. Bariatric surgery and long-term durability of weight loss. *JAMA Surg*. 2016;151(11):1046-1055. doi:10.1001/jamasurg.2016.2317
- Casillas RA, Kim B, Fischer H, Zelada Getty JL, Um SS, Coleman KJ. Comparative effectiveness of sleeve gastrectomy versus Roux-en-Y gastric bypass for weight loss and safety outcomes in older adults. *Surg Obes Relat Dis*. 2017;13(9):1476-1483. doi:10.1016/j.soard.2017.03.011
- Benoit SC, Hunter TD, Francis DM, De La Cruz-Munoz N. Use of bariatric outcomes longitudinal database (BOLD) to study variability in patient success after bariatric surgery. *Obes Surg*. 2014;24(6):936-943. doi:10.1007/s11695-014-1197-y
- Ju Z, Anderson W, Istifan N, et al. Comparison of weight loss outcomes between Roux-en-Y gastric bypass and sleeve gastrectomy in a racially mixed urban patient population. *Surgery for Obesity and Related Diseases*. 2022;1:1. doi:10.1016/j.soard.2022.05.025
- Devine B. Sources of Data for Comparative Effectiveness Research. In: Levy A, Sobolev B editors. *Comparative Effectiveness Research in Health Services*. Health Services Research. Springer US; 2016:1-23. doi:10.1007/978-1-4899-7586-7_13-1
- Ammann EM, Kalsekar I, Yoo A, Johnston SS. Validation of body mass index (BMI)-related ICD-9-CM and ICD-10-CM administrative diagnosis codes recorded in US claims data. *Pharmacoepidemiology and Drug Safety*. 2018;27(10):1092-1100. doi:10.1002/pds.4617
- Lloyd JT, Blackwell SA, Wei II, Howell BL, Shrank WH. Validity of a claims-based diagnosis of obesity among medicare beneficiaries. *Eval Health Prof*. 2015;38(4):508-517. doi:10.1177/0163278714553661
- Martin BJ, Chen G, Graham M, Quan H. Coding of obesity in administrative hospital discharge abstract data: accuracy and impact for future research studies. *BMC Health Services Research*. 2014;14(1):70. doi:10.1186/1472-6963-14-70
- Gebran SG, Knighton B, Ngaage LM, et al. Insurance coverage criteria for bariatric surgery: a survey of policies. *Obes Surg*. 2020;30(2):707-713. doi:10.1007/s11695-019-04243-2
- Li X, Lewis KH, Callaway K, Wharam JF, Toh S. Suitability of administrative claims databases for bariatric surgery research – is the glass half-full or half-empty? *BMC Medical Research Methodology*. 2020;20(1):225. doi:10.1186/s12874-020-01106-8
- Nickel F, de la Garza JR, Werthmann FS, et al. Predictors of risk and success of obesity surgery. *Obes Facts*. 2019;12(4):427-439. doi:10.1159/000496939
- Wilkinson KH, Helm M, Lak K, Higgins RM, Gould JC, Kindel TL. The risk of post-operative complications in super-super obesity compared to super obesity in accredited bariatric surgery centers. *OBES SURG*. 2019;29(9):2964-2971. doi:10.1007/s11695-019-03942-0
- Gupta PK, Franck C, Miller WJ, Gupta H, Forse RA. Development and validation of a bariatric surgery morbidity risk calculator using the prospective, multicenter NSQIP dataset. *J Am Coll Surg*. 2011;212(3):301-309. doi:10.1016/j.jamcollsurg.2010.11.003
- Longitudinal Assessment of Bariatric Surgery (LABS) Consortium, Flum DR, Belle SH, et al.; Longitudinal Assessment of Bariatric Surgery. Perioperative safety in the longitudinal assessment of bariatric surgery. *N Engl J Med*. 2009;361(5):445-454. doi:10.1056/NEJMoa0901836
- DeMaria EJ, Portenier D, Wolfe L. Obesity surgery mortality risk score: proposal for a clinically useful score to predict mortality risk in patients undergoing gastric bypass. *Surg Obes Relat Dis*. 2007;3(2):134-140. doi:10.1016/j.soard.2007.01.005

22. Sun Y, Liu B, Smith JK, et al. Association of preoperative body weight and weight loss with risk of death after bariatric surgery. *JAMA Netw Open*. 2020;3(5):e204803. doi:10.1001/jamanetworkopen.2020.4803
23. Elbanna H, Ghnam W, Negm A, et al. Impact of preoperative body mass index on the final outcome after laparoscopic sleeve gastrectomy for morbid obesity. *Ulus Cerrahi Derg*. 2016;32(4):238–243. doi:10.5152/UCD.2016.3275
24. Kaplan U, Zohdy W, Gmora S, Hong D, Anvari M. What patient factors influence bariatric surgery outcomes? A multiple regression analysis of Ontario Bariatric Registry Data. *Canadian Journal of Surgery*. 2022;65(1):E66–E72. doi:10.1503/cjs.018319
25. AbdAlla Salman M, Rabiee A, Salman A, et al. Predictors of type-2 diabetes remission following bariatric surgery after a two-year follow up. *Asian Journal of Surgery*. 2022;45:2645–2650. doi:10.1016/j.asjsur.2021.12.070
26. Robert M, Ferrand-Gaillard C, Disse E, et al. Predictive factors of type 2 diabetes remission 1 year after bariatric surgery: impact of surgical techniques. *Obes Surg*. 2013;23(6):770–775. doi:10.1007/s11695-013-0868-4
27. Optum Labs. Optum Labs and Optum Labs Data Warehouse (OLDW) descriptions and citation. Eden Prairie, MN: Reproduced with permission from Optum Labs; 2022.
28. Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol*. 2011;64(7):749–759. doi:10.1016/j.jclinepi.2010.10.004
29. Sun JW, Rogers JR, Her Q, et al. Adaptation and validation of the combined comorbidity score for ICD-10-CM. *Med Care*. 2017;55(12):1046–1051. doi:10.1097/MLR.0000000000000824
30. Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5–32. doi:10.1023/A:1010933404324
31. McAlexander RJ, Mentch L. Predictive inference with random forests: a new perspective on classical analyses. *Research & Politics*. 2020;7(1):2053168020905487. doi:10.1177/2053168020905487
32. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*. 2019;19(1):281. doi:10.1186/s12911-019-1004-8
33. Liaw A, Wiener M. Classification and Regression by randomForest. 2002;2(3):18–22. Available from: <https://CRAN.R-project.org/doc/Rnews/>
34. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *Journal of Clinical Epidemiology*. 2016;74:167–176. doi:10.1016/j.jclinepi.2015.12.005
35. Padwal RS, Klarenbach SW, Wang X, et al. A simple prediction rule for all-cause mortality in a cohort eligible for bariatric surgery. *JAMA Surg*. 2013;148(12):1109–1115. doi:10.1001/jamasurg.2013.3953
36. Altieri MS, Yang J, Yin D, Talamini MA, Spaniolas K, Pryor AD. Patients insured by Medicare and Medicaid undergo lower rates of bariatric surgery. *Surg Obes Relat Dis*. 2019;15(12):2109–2114. doi:10.1016/j.soard.2019.05.001

Pragmatic and Observational Research

Dovepress

Publish your work in this journal

Pragmatic and Observational Research is an international, peer-reviewed, open access journal that publishes data from studies designed to reflect more closely medical interventions in real-world clinical practice compared with classical randomized controlled trials (RCTs). The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/pragmatic-and-observational-research-journal>