**Supplementary Material***

Arterburn D, Wellman R, Emiliano A, Smith SR, Odegaard AO, Murali S, et al. Comparative effectiveness of bariatric procedures for weight loss and safety. A PCORnet cohort study. Ann Intern Med. doi:10.7326/M17-2786

| **Contents** | **Page** |
|---|---|


* This supplementary material was provided by the authors to give readers further details on their article. The material was reviewed but not copyedited.

# Statistical Appendix

Below we provide additional detail regarding the propensity score model and variable selection using the LASSO procedure, as well as the functional form of the outcome model. All analyses were conducted in R version 3.4.3 Kite Eating Tree[1].

**Propensity Score Model and Variable Selection**

To select variables for the propensity score model for each pairwise comparison we utilized the LASSO procedure, or Least Absolute Shrinkage and Selection Operator, as implemented in the R package glmnet[2] to automate the process across a large number of analyses. The LASSO shrinks regression coefficients by maximizing a likelihood that is penalized by the absolute size of the coefficients[2]. This results in some coefficients being shrunk completely to zero and consequently being removed from the propensity score model. The degree of shrinkage can be controlled via a tuning parameter, lambda, which corresponds to the degree of penalty, or shrinkage. For variable selection in practice, a penalty factor which multiplies lambda is set to 0 for variables that are retained in all models and 1 for variables that are subject to selection. Cross validation is then used to find the value of lambda that minimizes the error of the model.

In this analysis, each potential propensity score model for each pairwise comparison included age category, sex, race/ethnicity, baseline weight category, plus continuous linear covariates Charlson/Elixhauser Comorbidity index, baseline weight (both continuous and categorical representations), number of days between baseline weight measurement and bariatric surgery as "fixed" covariates, that is, covariates forced to be included in the model and assigned a penalty factor of 0 (referred to as fixed_covs in code below). Site, procedure year, smoking status, inpatient hospitalization days in the year prior to surgery, and comorbidities at baseline were included subject to the variable selection process (referred to as select_covs in code below). Further, to account for differing effects of confounders on propensity scores by site, interactions between site and all other potential confounders mentioned above were made available for selection (referred to as site_interactions in code below). Site and procedure year were included as selection variables (rather than fixed variables forced into the propensity score model) to allow for pooling in the event of small numbers of procedures at certain sites and in certain years, e.g., the model does not include a parameter for all non-referent sites and thus several sites may be pooled together as the referent group. Code snippet 1 below shows the rough implementation of the LASSO that was used.

R code – Snippet 1. LASSO for Propensity Score Variable Selection

```
# Procedure is the outcome for propensity score models

x <- grp == intervention_procedure



# Matrix of fixed covariates, covariates for selection

# and site interactions

cov_mat <- cbind(fixed_covs, select_covs, site_interactions)
```

```
# Count of included variables and designated penalty for LASSO
n_fixed <- length(names(fixed_covs))
n_select <- length(names(select_covs))
n_interact <- length(names(site_interactions))

# Penalty is 0 or 1 depending on whether it is fixed or available for selection
penalty <- c(rep(0, n_fixed), rep(1, n_select), rep(1, n_interact))

# Overall pscore using glmnet::glmnet
fit <- glmnet(cov_mat, x, family = "binomial", penalty.factor = penalty)
```

To select a value for the parameter lambda, we used cross validation with the cv.glmnet[2] function which is included in the glmnet R package. In order to save computational time in performing cross validation, and because the results of cross-validation are random due to the sampling of the folds, we performed twenty separate replications of 5-fold cross validation instead of the perhaps more conventional 10-fold cross validation. Figure 1 below shows one example of a plot generated following a single 5-fold cross validation. The two dotted vertical lines represent the value of the log of lambda that minimizes the binomial deviance and the largest value for which the deviance is within one standard error of the minimizing value. For larger values of lambda, more shrinkage occurs and fewer variables are selected. For each cross validation, by selecting a value for lambda that was within one standard error of the minimizing value we sought to obtain a final propensity score model with comparable deviance and fewer variables in the final model. R code snippet number 2 shows the basic loop that performed cross validation and estimation of lambda. This procedure was performed separately for each pairwise comparison at each time-point analyzed (i.e., 3 pairwise procedure comparisons at each of year 1, 3, and 5 and for sensitivity analysis). Plots of the estimated propensity scores from the main analyses are shown in Figure 2 below. We note that the histogram of propensity score estimates for the SG vs. AGB comparison (right hand column) show less overlap between the two procedure groups because site and year were highly predictive of procedure type. However, both procedures are represented in all propensity score deciles included in the outcome model.

R code – Snippet 2. Cross Validation to Select LASSO Parameter

```
lambda = NULL
for (i in 1:nreps)
{
```

```
  temp <- NULL
  cvfit <- cv.glmnet(cov_mat, x, family = "binomial", type.measure = "deviance",
                     parallel=TRUE, nfolds = 5)
  temp = data.frame(cvfit$lambda.min, cvfit$lambda.1se)
  lambda <- rbind(lambda, temp)
}
stopCluster(cl)


# take mean value over 20 replications
bestlambda.1se <- mean(lambdas[, 2])
```

In addition to adjustment for deciles of the predicted propensity score, we included main effects for baseline weight, sex, age, and all other baseline covariates in the outcome model (Table 1 of the manuscript lists these covariates). This approach obtains an unbiased estimate of the effect of procedure type on weight if either the propensity score model or the outcome regression model is correctly specified, and increases precision by including covariates that are strongly related to the outcome. (42) To estimate the trajectory of weight over time, days from surgery was included in the outcome model as a b-spline (cubic polynomial with 9 degrees of freedom, or 6 internal knots). The knots were located at quantiles of the follow up times chosen automatically by the bs() function in R. The chosen quantiles were roughly the 14th, 28th, 43rd, 57th, 71st and 86th. The decision to use 9 degrees of freedom, while somewhat arbitrary, was based on balancing the need for a reasonable number of internal knots to allow a flexible curve without compromising the ability to fit all of the models that were run.

**Outcome Model**

Estimates of percent total weight loss over time were generated by fitting a linear mixed effects model with patient-level random intercepts and slopes[3] with weight as the outcome and main effects for all baseline covariates, including both those that were fixed and those that were available for selection in the propensity score model (site and procedure year included). Additionally, propensity score deciles were included. Time was included in the model as a cubic b-spline with six internal knots, and procedure group and an interaction between procedure group and follow up time were also included. The general form of the model is as follows

$$Y_{ij} = \beta_0 + \beta_T^{bs} T_{ij}^{bs} + \beta_X X_i + \beta_{XT}^{bs} XT_{ij}^{bs} + \beta_{BL} BL_i + \beta_{Site} Site_i + \beta_{PS} PS_i + b_{0i} + b_{1i} T_{ij} + \epsilon_{ij},$$

Where $Y_{ij}$ is the weight measurement for subject i at time j, $T_{ij}$ is the $j^{th}$ time of weight measurement for the $i^{th}$ subject, $X_i$ is the dichotomous procedure variable, $T_{ij}^{bs}$ is the vector of b-spline components for time, $XT_{ij}^{bs}$ is the vector of interactions between procedure group and the b-spline for time, $BL_i$ is a vector of all baseline

covariates, $\text{Site}_i$ is a vector of indicator variables for all sites and $\text{PS}_i$ is vector of indicators for all propensity score deciles. Further, $b_{0i}$and $b_{1i}$ are random effects for the intercept and slope, respectively, and were assumed normally distributed with a mean of 0. This model was fit in R using the lme function in the nlme package[4]. Residual plots from the main analyses are shown in Figure 3 below. Code snippet 3 shows an abbreviated version of the function used to fit the outcome model for each of the main analyses.

R code – Snippet 3. Linear mixed effects model fit with lme.

```
fit_lme <- do.call(lme, list(fixed=y ~ time_bs + x + x*time_bs + bl_covs + site + pscores,
                             random = ~ 1 + time | factor(id)))
```

After fitting the model, estimated mean weight by procedure group over time was computed by evaluating the b-spline at the appropriate time point, setting the baseline covariates and propensity score variables at their mean value (dichotomous variables are set at the proportion observed in the dataset). Specifically, baseline weight was set at 127 kg (280 lbs), where 127 kg represents the approximate average baseline weight for the cohort described in Table 1 of the manuscript. Once the estimated mean weight was obtained the percent weight change was computed by subtracting and dividing by the baseline weight of 127 kg. Standard errors for the estimated percent weight change were obtained via the delta method.

**Sensitivity Analyses**

Several sensitivity analyses were conducted to assess whether there was variation in observed results due to certain choices that were made as part of the analysis plan and modeling. To assess the implications of drop out and baseline missing data on our results, we compared our primary analysis results to a simple, covariate adjusted model (no propensity scores) run on a single data set pooling all longitudinal data among patients with at least one post-surgery measurement sample, while excluding the race, ethnicity and blood pressure variables, which were the primary sources of missing baseline data (see Table 1 of the manuscript). This was deemed appropriate given that race/ethnicity and blood pressure are not thought to be strong confounders of the relationship between the choice of bariatric procedure and weight loss outcome. The sample size for our sensitivity analysis was 56,156 patients, which includes the 46,510 in our primary analysis plus an additional 9,646 patients who were eligible at baseline and had at least 1 post-surgery BMI measure, but lacked a BMI measure within the 1, 3, and 5-year windows used for the primary analysis. This allowed the model to be fit to a much larger cohort and to include an assessment of the effect of missing baseline data. Additionally, this model was simplified by the removal of propensity score adjustment and instead included only direct adjustment of covariates other than race/ethnicity and blood pressure. Since propensity score adjustment in this case was used to provide a degree of redundancy in control of confounding, we do not think that the exclusion of propensity score adjustment limits the conclusions of the sensitivity analysis in any considerable way.

To address concerns about possible violations of the positivity assumption and lack of overlap of the propensity scores, we compared our primary results to those obtained after propensity score trimming, that is, identifying a region of the propensity with substantial representation from both comparison arms and reducing the analysis sample to include only patients with propensity scores in that region. We, next, re-estimated the propensity score regression model in this reduce sample and, finally, refit the outcome regression model. This procedure was followed for each pairwise comparison at 1, 3, and 5 years follow-up. Trimmed propensity score distributions for all comparisons are shown in Figure 4 below.

For the comparisons of SG and RYGB at 1, 3 and 5 years, where propensity scores had the best overlap in the original analyses, we trimmed the overall propensity scores at the 15th and 85th percentile of the combined propensity score distribution, refit the propensity score models, and refit the outcome model adjusting for covariates and propensity score deciles.

For the comparisons of RYGB and SG with AGB, complete covariate adjustment (as in the primary analysis) led to overfitting in the outcome regression model. Therefore, we dropped the adjustment for covariates and instead included only baseline weight and propensity score deciles in the outcome models. Further, it was necessary to customize the trimming for these comparisons to achieve subgroups with overlapping propensity scores.

For the 1- and 3-year analyses comparing the AGB and RYGB we trimmed the overall propensity scores at the 30th and 70th and the 25th and 75th percentiles, respectively. For the 5-year comparison of AGB and RYGB it was necessary to trim the propensity scores by procedure group due considerable skew in the distribution of propensity scores in the RYGB arm (see Figure 2 below) and a considerably smaller number of patients with 5 years of follow-up in the AGB arm than the RYGB arm. For this comparison, we used patients with propensity scores within the 15th and 90th percentiles for the RYGB arm and the 2.5th and 15th percentile for the AGB arm.

The 1-, 3-, and 5-year comparisons of AGB and SG were the most difficult to address, as these propensity scores had the least overlap (Figure 2 below) and there is large imbalance in the number of patients receiving each, requiring procedure-wise trimming for all three comparisons to obtain propensity scores with considerable overlap (Figure 4 below). For the 1-year comparison, propensity scores were trimmed at the 15th and 95th percentiles for SG and the 2.5th and 12th percentiles for AGB. For the 3-year comparison, the 2.5th and 80th percentiles were used for SG and the 2.5th and 12.5th percentiles were used for the AGB. Finally, for the 5-year comparison, the 40th and 85th percentiles were used for the SG and the 2.5th and 15th percentiles were used for the AGB.

Exploratory analyses were performed to identify propensity score cut-offs that maximized sample size and overlap. Outcome regression was also performed with other propensity score thresholds and comparisons between procedures were unaffected, that is, results were not sensitive to the thresholds selected.

Additional sensitivity analyses included assessment of sensitivity to censoring of individuals based on conversion to a second bariatric surgery and/or pregnancy, as well as an assessment of whether 1-year weight loss for SG patients differed by year of surgery. The approach to the sensitivity analysis that removed the follow-up censoring for the second procedure was the same as the primary model. The analysis approach for the percent weight loss for SG patients at 1 year used both an unadjusted model and a model adjusted for age, sex, comorbidity score and baseline BMI.

**Variable Construction**

- Categorical variables included: year of surgery; age at index procedure; sex; race/ethnicity; Charlson/Elixhauser comorbidity index score
- Continuous variables included: height, weight, body mass index (BMI), and blood pressure;
- Binary/indicator variables included: prevalence of relevant comorbidities (anxiety, deep vein thrombosis, depression, eating disorders, type 2 diabetes, dyslipidemia, gastroesophageal reflux disease,

hypertension, infertility, kidney disease, non-alcoholic fatty liver disease, obstructive sleep apnea, osteoarthritis of the lower extremity, polycystic ovarian syndrome, psychoses, pulmonary embolism, smoking, and substance use disorders), and all diagnoses and procedures related to pregnancy.

**Site-Level Variation in Bariatric Procedures and Outcomes**

Additional analyses were conducted to understand site-level variation in bariatric procedures and outcomes. First, we found that there was significant variation in bariatric procedure volume across data marts (Table 1 below). As noted in the footnote to Table 1, we report these results at the data mart level because 9 of our 41 data contributing sites provided data that were rolled up or aggregated into three data marts, which makes it impossible to separate the data from their individual data-contributing sites. As a result, the number of data marts in our study is 35, but this represents 41 data contributing sites. As previously noted, for each pairwise comparison, we restricted the analysis to those sites (data marts) that included at least 1 patient of each procedure type at each time point of interest.

To understand whether there was any evidence of site-level variations in outcomes, we conducted unadjusted analyses of observed changes in weight at 1-, 3-, and 5-years following surgery. At one year, the mean weight loss was larger with RYGB than with SG at 29 of 32 data marts and RYGB was larger than AGB at all 28 data marts performing both procedures. Similarly, mean weight loss at 1 year was larger with SG than with AGB at all 28 data marts performing both procedures. One of the three data marts where weight loss was larger for SG than for RYGB had only one RYGB procedure represented.

Based on unadjusted analyses of observed changes in weight at three-years following surgery, the mean weight loss was larger with RYGB than with SG at 25 of 29 data marts and RYGB was larger than AGB at all 27 data marts performing both procedures. Similarly, mean weight loss at 3 years was larger with SG than with AGB at all 27 data marts performing both procedures.

Based on unadjusted analyses of observed changes in weight at five-years following surgery, the mean weight loss was larger with RYGB than with SG at 14 of 17 data marts and RYGB was larger than AGB at all 20 data marts performing both procedures. Similarly, mean weight loss at 5 years was larger with SG than with AGB at 15 of 16 data marts performing both procedures. One of the three data marts where weight loss was larger for SG than for RYGB had only one SG procedure represented. Further, at the one data mart where weight loss was greater for AGB than for SG there was only one SG procedure represented.

These findings suggest there is little unadjusted evidence of important heterogeneity in treatment effects by site (data mart). Where some evidence of heterogeneity is present, the sample sizes are too small to make any reasonable inference.

**References**

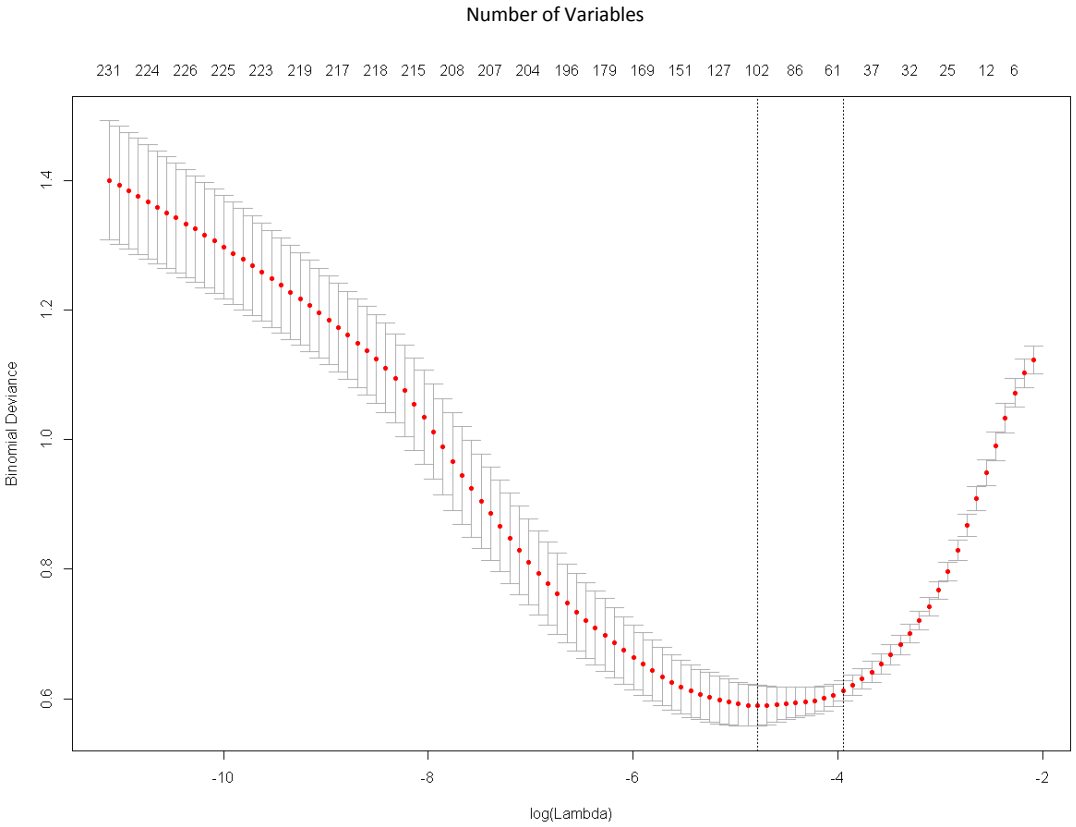
1. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

2. Friedman, J., Hastie, T. and Tibshirani, R. (2008) Regularization Paths for Generalized Linear Models via Coordinate Descent, Journal of Statistical Software, Vol. 33(1), 1-22 Feb 2010

3. Laird, N.M. and Ware, J.H. (1982) "Random-Effects Models for Longitudinal Data", Biometrics, 38, 963–974.

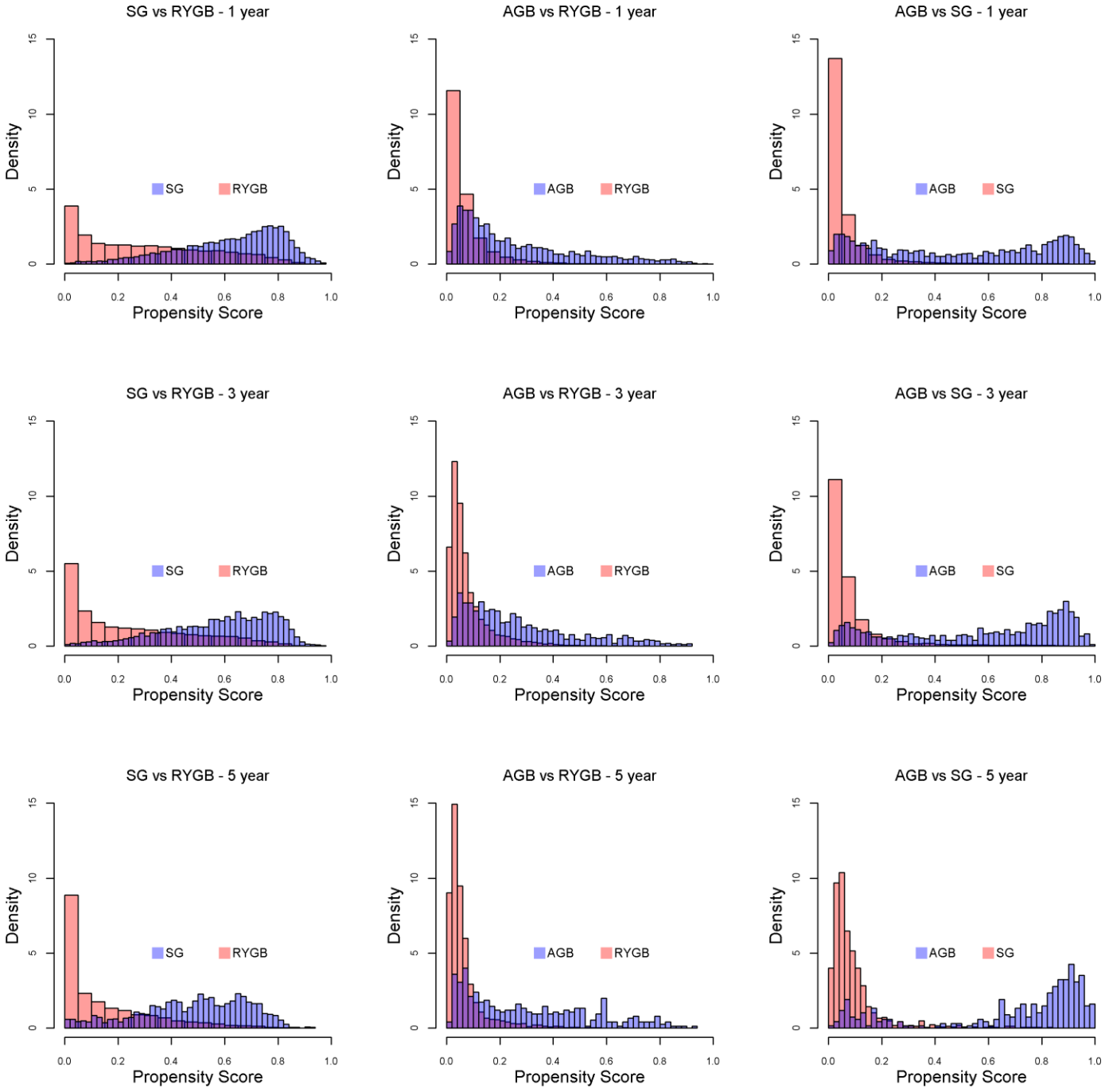
4. Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2017). _nlme: Linear and Nonlinear Mixed Effects Models_. R package version 3.1-131, <URL: https://CRAN.R-project.org/package=nlme>.

**Figure 1. Selecting the model with small binomial deviance and a reasonable number of covariates.**




* This figure shows one example of a plot generated following a single 5-fold cross validation. The two dotted vertical lines represent the value of the log of lambda that minimizes the binomial deviance and the largest value for which the deviance is within one standard error of the minimizing value. Bars represent standard errors.

**Figure 2. Plots of propensity score estimates from the main analyses.**




*Although it is difficult to see in the figure due to small numbers, AGB procedures are represented in all deciles of the propensity score for each pairwise comparison with RYGB and SG.

**Figure 3. Plots of residuals from the main analyses with a cubic smoothing spline.**

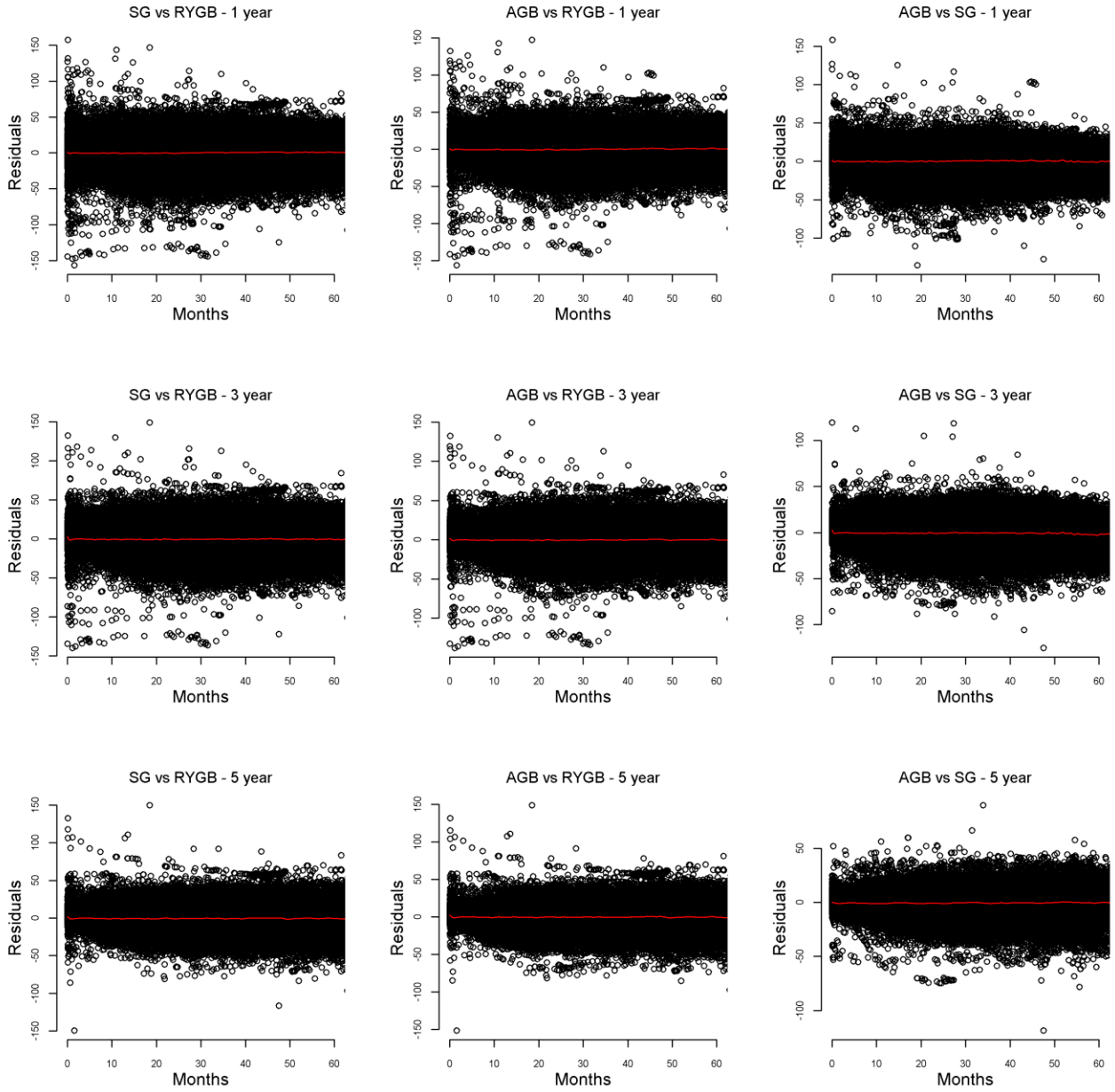


**Figure 4. Plots of propensity score estimates from the sensitivity analysis after trimming of propensity scores.**

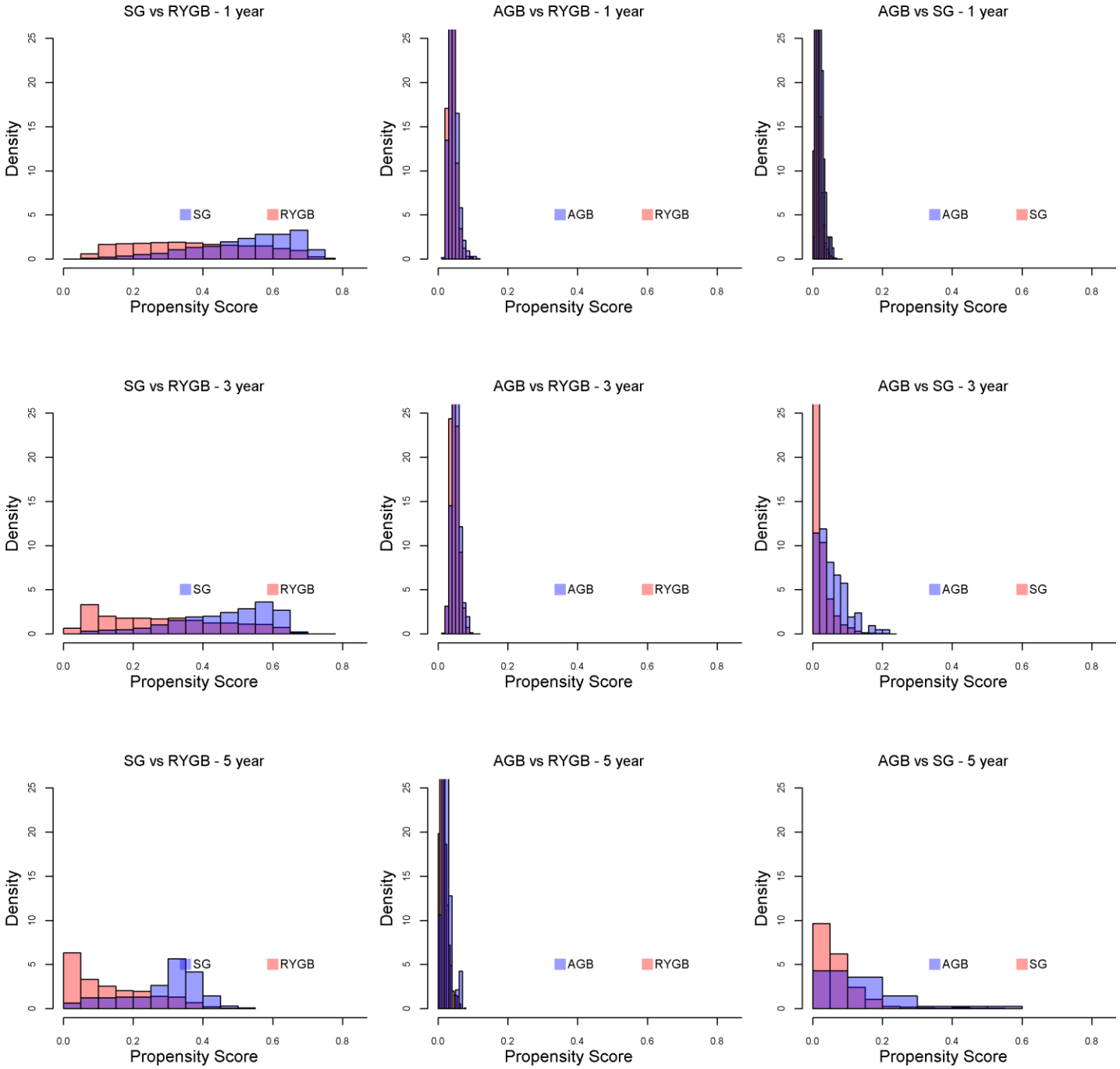


**Table 1. Procedures performed at each study Data Mart***

| | AGB | | RYGB | | SG | | Total |
|---|---|---|---|---|---|---|---|
| **Data Marts** | **n** | **%** | **n** | **%** | **n** | **%** | **n** |
| **Data Mart 1** | 119 | 3.0 | 3580 | 89.0 | 324 | 8.1 | 4023 |
| **Data Mart 2** | 2 | 0.2 | 427 | 48.8 | 446 | 51.0 | 875 |
| **Data Mart 3** | 7 | 1.8 | 191 | 50.4 | 181 | 47.8 | 379 |
| **Data Mart 4** | 361 | 7.7 | 3114 | 66.8 | 1186 | 25.4 | 4661 |
| **Data Mart 5** | 7 | 2.5 | 194 | 68.8 | 81 | 28.7 | 282 |
| **Data Mart 6** | 50 | 2.1 | 1035 | 43.9 | 1270 | 53.9 | 2355 |
| **Data Mart 7** | 152 | 26.3 | 144 | 24.9 | 282 | 48.8 | 578 |
| **Data Mart 8** | 219 | 5.2 | 1960 | 46.6 | 2030 | 48.2 | 4209 |
| **Data Mart 9** | 49 | 4.4 | 599 | 53.9 | 464 | 41.7 | 1112 |
| **Data Mart 10** | 216 | 55.4 | 174 | 44.6 | 0 | 0.0 | 390 |
| **Data Mart 11** | 148 | 6.4 | 1152 | 49.4 | 1030 | 44.2 | 2330 |
| **Data Mart 12** | 136 | 6.3 | 1440 | 66.9 | 578 | 26.8 | 2154 |
| **Data Mart 13** | 40 | 11.9 | 48 | 14.3 | 248 | 73.8 | 336 |
| **Data Mart 14** | 0 | 0.0 | 39 | 17.9 | 179 | 82.1 | 218 |
| **Data Mart 15** | 24 | 1.8 | 1009 | 74.3 | 325 | 23.9 | 1358 |
| **Data Mart 16** | 1 | 0.2 | 351 | 72.8 | 130 | 27.0 | 482 |
| **Data Mart 17** | 48 | 34.0 | 63 | 44.7 | 30 | 21.3 | 141 |
| **Data Mart 18** | 159 | 23.6 | 121 | 18.0 | 394 | 58.5 | 674 |
| **Data Mart 19** | 20 | 3.3 | 406 | 67.9 | 172 | 28.8 | 598 |
| **Data Mart 20** | 75 | 5.4 | 1159 | 84.1 | 144 | 10.4 | 1378 |
| **Data Mart 21** | 148 | 15.5 | 542 | 56.9 | 262 | 27.5 | 952 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Data Mart 22** | 167 | 7.7 | 1138 | 52.7 | 855 | 39.6 | 2160 |
| **Data Mart 23** | 77 | 6.1 | 534 | 42.5 | 646 | 51.4 | 1257 |
| **Data Mart 24** | 6 | 0.8 | 783 | 98.5 | 6 | 0.8 | 795 |
| **Data Mart 25** | 287 | 1.4 | 7847 | 39.0 | 12011 | 59.6 | 20145 |
| **Data Mart 26** | 62 | 3.5 | 767 | 42.8 | 961 | 53.7 | 1790 |
| **Data Mart 27** | 29 | 3.0 | 199 | 20.3 | 750 | 76.7 | 978 |
| **Data Mart 28** | 0 | 0.0 | 0 | 0.0 | 8 | 100.0 | 8 |
| **Data Mart 29** | 0 | 0.0 | 4 | 33.3 | 8 | 66.7 | 12 |
| **Data Mart 30** | 1 | 7.1 | 2 | 14.3 | 11 | 78.6 | 14 |
| **Data Mart 31** | 390 | 7.1 | 2242 | 40.9 | 2843 | 51.9 | 5475 |
| **Data Mart 32** | 68 | 13.9 | 219 | 44.9 | 201 | 41.2 | 488 |
| **Data Mart 33** | 69 | 6.3 | 412 | 37.9 | 606 | 55.7 | 1087 |
| **Data Mart 34** | 3 | 0.5 | 192 | 33.8 | 373 | 65.7 | 568 |
| **Data Mart 35** | 52 | 6.3 | 121 | 14.6 | 658 | 79.2 | 831 |

*The 35 Data Marts include a total of 41 data contributing sites because some sites roll up their data into one Data Mart.

**Table 2. Follow-up information in the PCORnet Bariatric Study cohort.**

| Cohort | Follow-up window of interest | | |
|---|---|---|---|
| | **1 year (measured at 6-18 months)** | **3 years (measured at 30-42 months)** | **5 years (measured at 54-66 months)** |
| Adult bariatric population with BMI ≥35 at baseline (n=65,093 total; 3192 AGB, 32208 RYGB, 29693 SG) (overall denominator population, Figure 1 of the manuscript) | | | |
| Estimated number of patients eligible for follow-up at each time point, overall and by procedure[a] (denominator for each time point) | | | |
| All | 53351 | 30521 | 10442 |
| AGB | 3098 | 2519 | 1111 |
| RYGB | 28039 | 18684 | 7824 |
| SG | 22214 | 9318 | 1507 |
| Adult bariatric population with BMI ≥35 at baseline and having a BMI measurement at either 1, 3, or 5 years follow-up (n=46,510) (overall numerator, Figure 1 of the manuscript) | | | |
| Observed number of patients having a BMI measurement during each follow-up time point, overall and by procedure (numerator for each time point) [b] | | | |
| All | 44978 | 20783 | 7159 |
| AGB | 2367 | 1507 | 609 |

| Cohort | Follow-up window of interest | | |
|---|---|---|---|
| | **1 year (measured at 6-18 months)** | **3 years (measured at 30-42 months)** | **5 years (measured at 54-66 months)** |
| RYGB | 24061 | 12429 | 5257 |
| SG | 18550 | 6847 | 1293 |
| **% of all patients eligible for follow-up who have a BMI measurement at that time point, overall and by procedure** | | | |
| **All** | **84%** | **68%** | **69%** |
| **AGB** | **76%** | **60%** | **55%** |
| **RYGB** | **86%** | **67%** | **67%** |
| **SG** | **84%** | **73%** | **86%** |

[a] This represents an estimated number of patients who can be followed for a certain follow-up window of interest based on the study timeframe, which ended on September 30, 2015. For example, only patients who had a bariatric procedure on October 1, 2014 or earlier would be eligible for having one complete year of follow-up information. However, the number of eligible patients was an estimate because we did not obtain actual dates for the analysis for privacy consideration. For example, all patients who had their procedure performed in 2013 or earlier and 3/4 of patients who had their procedure performed in 2014 will be eligible for at least one year of follow-up.

[b] These are the actual observed numbers of patients with measurements at each time point.

AGB = adjustable gastric banding; RYGB = Roux-en-Y gastric bypass; SG = sleeve gastrectomy

**Table 3. Characteristics of adult bariatric surgery patients with or without a body mass index measurement during baseline and follow-up***

| | All bariatric surgery patients in the PBS cohort* | | Patients with BMI ≥35 at baseline and BMI at 1, 3 or 5 years* | | Patients with BMI ≥35 at baseline but missing BMI at 1, 3 & 5 years ǂ | | Patients with missing BMI at baseline ǂ | |
|---|---|---|---|---|---|---|---|---|
| N (row %) | 65093 | (100.0) | 46510 | (100.0) | 18583 | (100.0) | 12510 | (100.0) |
| Bariatric procedure, N, % | | | | | | | | |
| AGB | 3192 | (4.9) | 2567 | (5.5) | 625 | (3.4) | 3049 | (24.4) |
| RYGB | 32208 | (49.5) | 24982 | (53.7) | 7226 | (38.9) | 6621 | (52.9) |
| SG | 29693 | (45.6) | 18961 | (40.8) | 10732 | (57.8) | 2840 | (22.7) |
| Age, mean (SD) | 45.0 | (11.7) | 45.5 | (11.6) | 43.6 | (11.9) | 45.8 | (12.1) |
| Age category, N (%) | | | | | | | | |
| 20-44 years | 32136 | (49.4) | 22075 | (47.5) | 10061 | (54.1) | 5866 | (46.9) |
| 45-64 years | 29622 | (45.5) | 22042 | (47.4) | 7580 | (40.8) | 5782 | (46.2) |
| 65-80 years | 3335 | (5.1) | 2393 | (5.2) | 942 | (5.1) | 862 | (6.9) |
| Female, N (%) | 51619 | (79.3) | 37315 | (80.2) | 14304 | (77.0) | 9535 | (76.2) |
| Max BMI, mean (SD) | 49.0 | (8.2) | 49.1 | (8.2) | 48.8 | (8.4) | NA | NA |
| Max BMI category, N (%) | | | | | | | | |
| 35-39 kg/m2 | 5828 | (9.0) | 3865 | (8.3) | 1963 | (10.6) | 0 | NA |
| 40-49 kg/m2 | 35010 | (53.8) | 25123 | (54.0) | 9887 | (53.2) | 0 | NA |
| 50-59 kg/m2 | 17699 | (27.2) | 12870 | (27.7) | 4829 | (26.0) | 0 | NA |
| 60+ kg/m2 | 6556 | (10.1) | 4652 | (10.0) | 1904 | (10.3) | 0 | NA |
| Year of surgery, N (%) | | | | | | | | |

| | All bariatric surgery patients in the PBS cohort* | | Patients with BMI ≥35 at baseline and BMI at 1, 3 or 5 years* | | Patients with BMI ≥35 at baseline but missing BMI at 1, 3 & 5 years ł | | Patients with missing BMI at baseline ł | |
|---|---|---|---|---|---|---|---|---|
| 2005 -2009 | 5572 | (8.6) | 5183 | (11.1) | 389 | (2.1) | 2532 | (20.2) |
| 2010 | 6494 | (10.0) | 5819 | (12.5) | 675 | (3.6) | 2721 | (21.8) |
| 2011 | 10292 | (15.8) | 9015 | (19.4) | 1277 | (6.9) | 2518 | (20.1) |
| 2012 | 10885 | (16.7) | 8977 | (19.3) | 1908 | (10.3) | 2041 | (16.3) |
| 2013 | 11121 | (17.1) | 8556 | (18.4) | 2565 | (13.8) | 1227 | (9.8) |
| 2014 | 11983 | (18.4) | 7951 | (17.1) | 4032 | (21.7) | 896 | (7.2) |
| 2015 | 8746 | (13.4) | 1009 | (2.2) | 7737 | (41.6) | 575 | (4.6) |
| Hispanic Ethnicity, N (%) | 13094 | (24.1) | 9612 | (24.5) | 3482 | (23.1) | 876 | (13.5) |
| Missing ethnicity, N (%) | 10778 | (16.6) | 7262 | (15.6) | 3516 | (18.9) | 6024 | (48.2) |
| Race, N (%) | | | | | | | | |
| Asian | 540 | (0.9) | 422 | (1.0) | 118 | (0.7) | 63 | (0.6) |
| African-American | 12094 | (21.1) | 8526 | (20.8) | 3568 | (22.1) | 1586 | (14.9) |
| Multiple | 802 | (1.4) | 465 | (1.1) | 337 | (2.1) | 90 | (0.8) |
| Caucasian | 41248 | (72.1) | 30371 | (74.0) | 10877 | (67.3) | 8523 | (79.9) |
| Pacific Islander | 188 | (0.3) | 143 | (0.4) | 45 | (0.3) | 9 | (0.1) |
| Native American | 354 | (0.6) | 279 | (0.7) | 75 | (0.5) | 45 | (0.4) |
| Other | 2001 | (3.5) | 866 | (2.1) | 1135 | 7.03 | 355 | (3.3) |
| Missing race, N (%) | 7866 | (12.1) | 5438 | (11.7) | 2428 | (13.1) | 1839 | (14.7) |
| Systolic BP, mean (SD) | 131.1 | (17.9) | 130.5 | (17.1) | 132.8 | (19.8) | 131.1 | (16.7) |
| Diastolic BP, mean (SD) | 75.8 | (12.1) | 75.6 | (11.6) | 76.1 | (13.4) | 76.1 | (11.4) |

| | All bariatric surgery patients in the PBS cohort* | | Patients with BMI ≥35 at baseline and BMI at 1, 3 or 5 years* | | Patients with BMI ≥35 at baseline but missing BMI at 1, 3 & 5 years ł | | Patients with missing BMI at baseline ł | |
|---|---|---|---|---|---|---|---|---|
| Missing BP, N (%) | 4006 | (6.2) | 1801 | (3.9) | 2205 | (11.9) | 10736 | (85.8) |
| Charlson-Elixhauser score category, N (%) | | | | | | | | |
| ≤-1 | 20243 | (31.1) | 15488 | (33.3) | 4755 | (25.6) | 2845 | (22.7) |
| 0 | 35152 | (54.0) | 23879 | (51.3) | 11273 | (60.7) | 8463 | (67.7) |
| ≥1 | 9698 | (14.9) | 7143 | (15.4) | 2555 | (13.8) | 1202 | (9.6) |
| Inpatient hospital days in year before surgery, mean (SD) | 0.46 | (4.4) | 0.43 | (3.7) | 0.53 | (5.7) | 0.79 | (9.5) |
| Health Conditions, N (%) | | | | | | | | |
| Anxiety | 13657 | (21.0) | 9917 | (21.3) | 3740 | (20.1) | 1837 | (14.7) |
| Depression | 19324 | (29.7) | 14339 | (30.8) | 4985 | (26.8) | 3370 | (26.9) |
| Diabetes | 23411 | (36.0) | 17320 | (37.2) | 6091 | (32.8) | 3835 | (30.7) |
| DVT | 454 | (0.7) | 336 | (0.7) | 118 | (0.6) | 94 | (0.8) |
| Dyslipidemia | 30377 | (46.7) | 22823 | (49.1) | 7554 | (40.7) | 5152 | (41.2) |
| Eating Disorder | 6321 | (9.7) | 4983 | (10.7) | 1338 | (7.2) | 238 | (1.9) |
| GERD | 26650 | (40.9) | 18995 | (40.8) | 7655 | (41.2) | 4764 | (38.1) |
| Hypertension | 38297 | (58.8) | 28017 | (60.2) | 10280 | (55.3) | 7089 | (56.7) |
| Infertility | 448 | (0.7) | 339 | (0.7) | 109 | (0.6) | 75 | (0.6) |
| Kidney Disease | 4897 | (7.5) | 3824 | (8.2) | 1073 | (5.8) | 491 | (3.9) |
| NAFLD | 13293 | (20.4) | 9646 | (20.7) | 3647 | (19.6) | 1380 | (11.0) |

| | All bariatric surgery patients in the PBS cohort* | | Patients with BMI ≥35 at baseline and BMI at 1, 3 or 5 years* | | Patients with BMI ≥35 at baseline but missing BMI at 1, 3 & 5 years ł | | Patients with missing BMI at baseline ł | |
|---|---|---|---|---|---|---|---|---|
| Osteoarthritis, lower limb | 1220 | (1.9) | 814 | (1.8) | 406 | (2.2) | 165 | (1.3) |
| PCOS | 3345 | (5.1) | 2345 | (5.0) | 1000 | (5.4) | 435 | (3.5) |
| Pulmonary Embolism | 717 | (1.1) | 547 | (1.2) | 170 | (0.9) | 96 | (0.8) |
| Psychotic Disorder | 1990 | (3.1) | 1471 | (3.2) | 519 | (2.8) | 235 | (1.9) |
| Sleep apnea | 31785 | (48.8) | 22894 | (49.2) | 8891 | (47.8) | 5254 | (42.0) |
| Smoker | 5552 | (8.5) | 4008 | (8.6) | 1544 | (8.3) | 618 | (4.9) |
| Substance Use Disorder | 1319 | (2.0) | 984 | (2.1) | 335 | (1.8) | 199 | (1.6) |

* The PBS cohort includes patients with BMI ≥35 at baseline, with and without follow-up BMI measurements, but not patients with missing baseline BMI. Baseline = measured in the year prior to surgery; Adult defined as age 20-79 years; AGB = adjustable gastric banding; RYGB = Roux-en-Y gastric bypass; SG = sleeve gastrectomy; BMI = body mass index ($kg/m^2$); BP = blood pressure in year prior to surgery; Health Conditions were identified by 1+ ICD-9 or SNOMED diagnosis code in the year prior to surgery; NAFLD = non-alcoholic fatty liver disease; GERD = gastroesophageal reflux disease; PBS = PCORnet Bariatric Study; PCOS = polycystic ovarian syndrome; DVT = deep vein thrombosis

ł These patients were excluded from the final PBS cohort. See Figure 1 of the manuscript for more details.

**Table 4. Major adverse events occurring in the first 30-days after bariatric surgery, by procedure type***

| 30-Day Adverse Event | AGB (n=3192) | | RYGB (n=32208) | | SG (n=29693) | | Overall (n=65093) | |
|---|---|---|---|---|---|---|---|---|
| | n | % | n | % | n | % | n | % |
| Death | 1 | 0.0 | 54 | 0.2 | 18 | 0.1 | 73 | 0.1 |
| Percutaneous, operative, or endoscopic intervention | 80 | 2.5 | 1309 | 4.1 | 576 | 1.9 | 1965 | 3.0 |
| Venous thromboembolism | 13 | 0.4 | 218 | 0.7 | 170 | 0.6 | 401 | 0.6 |
| Failure to discharge from hospital within 30 days | 1 | 0.0 | 75 | 0.2 | 39 | 0.1 | 115 | 0.2 |
| **Any Adverse Event** | **91** | **2.9** | **1605** | **5.0** | **783** | **2.6** | **2479** | **3.8** |

*AGB =adjustable gastric banding; RYGB = Roux-en-Y gastric bypass; SG = sleeve gastrectomy; We limited our adverse event analyses to 30 days because longer-term adverse events are more likely to occur outside of the original health system that performed the bariatric procedure, which would result in substantial underreporting.

**Table 5. Comparison of Percent Weight Loss at 1, 3, and 5 years Across Diabetes Subgroups**

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | % TWL | 95% CI | | N | %TWL | 95% CI | | N | %TWL | 95% CI | |
| **RYGB** | | | | | | | | | | | | |
| Diabetes | 8493 | -29.5% | (-29.6%, | -29.3%) | 4287 | -29.8% | (-30.0%, | -29.5%) | 1723 | -29.4% | (-29.8%, | -29.1%) |
| No Diabetes | 10536 | -32.5% | (-32.7%, | -32.3%) | 4938 | -32.5% | (-32.7%, | -32.3%) | 1953 | -31.9% | (-32.3%, | -31.6%) |
| Difference | | 3.0% | (2.8%, | 3.3%) | | 2.7% | (2.4%, | 3.0%) | | 2.5% | (2.0%, | 3.0%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| **SG** | | | | | | | | | | | | |
| Diabetes | 4387 | -23.7% | (-23.9%, | -23.4%) | 1487 | -23.7% | (-24.1%, | -23.4%) | 299 | -23.4% | (-24.2%, | -22.6%) |
| No Diabetes | 10542 | -25.9% | (-26.1%, | -25.8%) | 3817 | -25.7% | (-26.0%, | -25.5%) | 789 | -25.9% | (-26.5%, | -25.4%) |
| Difference | | 2.3% | (2.0%, | 2.5%) | | 2.0% | (1.6%, | 2.4%) | | 2.5% | (1.6%, | 3.4%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| | | | | | | | | | | | | |
| HTE Difference | | -0.8% | (-1.1%, | -0.4%) | | -0.7% | (-1.2%, | -0.2%) | | 0.0% | (-1.0%, | 1.0%) |
| Interaction P-value | | | | <0.001 ǂ | | | | 0.004 ǂ | | | | 0.99 |
| **RYGB** | | | | | | | | | | | | |

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | % TWL | 95% CI | | N | %TWL | 95% CI | | N | %TWL | 95% CI | |
| Diabetes | 8377 | -29.6% | (-29.7%, | -29.4%) | 4266 | -29.8% | (-30.0%, | -29.6%) | 1743 | -29.5% | (-29.8%, | -29.1%) |
| No Diabetes | 10307 | -32.6% | (-32.8%, | -32.5%) | 4886 | -32.6% | (-32.8%, | -32.4%) | 1990 | -32.1% | (-32.4%, | -31.7%) |
| Difference | | 3.1% | (2.8%, | 3.3%) | | 2.8% | (2.5%, | 3.1%) | | 2.6% | (2.1%, | 3.1%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| **AGB** | | | | | | | | | | | | |
| Diabetes | 512 | -12.3% | (-13.0%, | -11.7%) | 302 | -13.0% | (-13.8%, | -12.2%) | 103 | -13.0% | (-14.4%, | -11.6%) |
| No Diabetes | 1169 | -14.4% | (-14.8%, | -14.0%) | 641 | -15.2% | (-15.8%, | -14.6%) | 234 | -16.4% | (-17.4%, | -15.5%) |
| Difference | | 2.1% | (1.3%, | 2.8%) | | 2.2% | (1.3%, | 3.2%) | | 3.4% | (1.8%, | 5.1%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| HTE Difference | | -1.0% | (-1.8%, | -0.2%) | | -0.6% | (-1.5%, | 0.4%) | | 0.8% | (-0.9%, | 2.5%) |
| Interaction P-value | | | | 0.014 ł | | | | 0.27 | | | | 0.34 |
| **SG** | | | | | | | | | | | | |
| Diabetes | 4294 | -23.6% | (-23.8%, | -23.4%) | 1474 | -23.7% | (-24.1%, | -23.4%) | 299 | -23.6% | (-24.5%, | -22.7%) |
| No Diabetes | 10370 | -25.9% | (-26.0%, | -25.7%) | 3796 | -25.5% | (-25.7%, | -25.3%) | 789 | -25.8% | (-26.4%, | -25.2%) |

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | % TWL | 95% CI | | N | %TWL | 95% CI | | N | %TWL | 95% CI | |
| Difference | | 2.3% | (2.0%, | 2.6%) | | 1.8% | (1.3%, | 2.2%) | | 2.2% | (1.1%, | 3.2%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| **AGB** | | | | | | | | | | | | |
| Diabetes | 512 | -11.7% | (-12.4%, | -11.0%) | 298 | -12.1% | (-12.9%, | -11.2%) | 93 | -12.3% | (-14.1%, | -10.6%) |
| No Diabetes | 1169 | -13.9% | (-14.4%, | -13.4%) | 635 | -14.2% | (-14.8%, | -13.6%) | 213 | -15.4% | (-16.7%, | -14.2%) |
| Difference | | 2.2% | (1.4%, | 3.0%) | | 2.1% | (1.2%, | 3.1%) | | 3.1% | (1.3%, | 4.9%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | 0.001* |
| HTE Difference | | -0.1% | (-0.9%, | 0.8%) | | 0.3% | (-0.7%, | 1.4%) | | 0.9% | (-1.1%, | 2.9%) |
| Interaction P-value | | | | 0.88 | | | | 0.51 | | | | 0.38 |

%TWL = percent total weight loss; CI = confidence interval; HTE = heterogeneity of treatment effects

* These results show statistically significant subgroup effects comparing weight outcomes of patients with diabetes vs. no diabetes within procedures.

The magnitude of these between subgroup differences is ≤3.4 percentage points at all time points.

ł These results show statistically significant interactions (heterogeneity in treatment effects), where the results of the pairwise comparison between two procedures differs across subgroups of patients with diabetes vs. no diabetes at baseline. The magnitude of these differences between pairwise comparisons is clinically small and ≤1.0 percentage points of total weight loss.

**Table 6. Comparison of Percent Weight Loss at 1, 3, and 5 years Across BMI subgroups**

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% TWL** | **95% CI** | | **N** | **%TWL** | **95% CI** | | **N** | **%TWL** | **95% CI** | |
| **RYGB** | | | | | | | | | | | | |
| BMI 50+ | 5564 | -34.7% | (-35.0%, | -34.5%) | 2601 | -33.9% | (-34.2%, | -33.6%) | 1096 | -32.4% | (-32.9%, | -31.9%) |
| BMI<50 | 13465 | -29.7% | (-29.8%, | -29.6%) | 6624 | -30.3% | (-30.5%, | -30.1%) | 2580 | -30.2% | (-30.5%, | -29.9%) |
| Difference | | -5.0% | (-5.3%, | -4.8%) | | -3.6% | (-4.0%, | -3.2%) | | -2.3% | (-2.9%, | -1.6%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| **SG** | | | | | | | | | | | | |
| BMI 50+ | 3604 | -28.5% | (-28.7%, | -28.2%) | 1098 | -26.9% | (-27.3%, | -26.5%) | 221 | -26.6% | (-27.6%, | -25.6%) |
| BMI<50 | 11325 | -24.2% | (-24.4%, | -24.1%) | 4206 | -24.6% | (-24.8%, | -24.4%) | 867 | -24.5% | (-25.1%, | -24.0%) |
| Difference | | -4.2% | (-4.6%, | -3.9%) | | -2.3% | (-2.8%, | -1.8%) | | -2.1% | (-3.2%, | -1.0%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| HTE | | 0.8% | (0.4%, | 1.2%) | | 1.3% | (0.7%, | 1.8%) | | 0.2% | (-1.0%, | 1.3%) |
| Difference | | | | | | | | | | | | |
| Interaction P-value | | | | <0.001ǂ | | | | <0.001ǂ | | | | 0.76 |

| | 1-YEAR | | | 3-YEAR | | | 5-YEAR | | |
|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% TWL** | **95% CI** | **N** | **%TWL** | **95% CI** | **N** | **%TWL** | **95% CI** |
| **RYGB** | | | | | | | | | |
| BMI 50+ | 5487 | -34.5% | (-34.7%, -34.3%) | 2592 | -33.8% | (-34.1%, -33.4%) | 1117 | -32.6% | (-33.2%, -32.1%) |
| BMI<50 | 13197 | -29.9% | (-30.0%, -29.8%) | 6560 | -30.4% | (-30.6%, -30.2%) | 2616 | -30.1% | (-30.4%, -29.8%) |
| Difference | | -4.6% | (-4.9%, -4.3%) | | -3.4% | (-3.8%, -3.0%) | | -2.5% | (-3.2%, -1.9%) |
| P-value | | | <0.001* | | | <0.001* | | | <0.001* |
| **AGB** | | | | | | | | | |
| BMI 50+ | 266 | -12.7% | (-13.6%, -11.8%) | 138 | -12.1% | (-13.3%, -10.9%) | 53 | -14.8% | (-16.7%, -12.8%) |
| BMI<50 | 1415 | -13.8% | (-14.2%, -13.4%) | 805 | -14.7% | (-15.2%, -14.2%) | 284 | -15.1% | (-16.0%, -14.2%) |
| Difference | | 1.1% | (0.1%, 2.0%) | | 2.5% | (1.3%, 3.8%) | | 0.3% | (-1.8%, 2.5%) |
| P-value | | | 0.024* | | | <0.001* | | | 0.75 |
| HTE | | 5.7% | (4.7%, 6.6%) | | 5.9% | (4.6%, 7.2%) | | 2.9% | (0.7%, 5.0%) |
| Difference | | | | | | | | | |
| Interaction P-value | | | <0.001 ǂ | | | 0.001 ǂ | | | 0.009 ǂ |
| **SG** | | | | | | | | | |

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% TWL** | **95% CI** | | **N** | **%TWL** | **95% CI** | | **N** | **%TWL** | **95% CI** | |
| BMI 50+ | 3513 | -29.2% | (-29.5%, | -28.9%) | 1078 | -27.8% | (-28.2%, | -27.3%) | 221 | -26.9% | (-28.1%, | -25.8%) |
| BMI<50 | 11151 | -24.0% | (-24.1%, | -23.8%) | 4192 | -24.3% | (-24.6%, | -24.1%) | 867 | -24.6% | (-25.2%, | -24.0%) |
| Difference | | -5.2% | (-5.6%, | -4.9%) | | -3.4% | (-4.0%, | -2.9%) | | -2.3% | (-3.7%, | -1.0%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | 0.001* |
| **AGB** | | | | | | | | | | | | |
| BMI 50+ | 266 | -13.3% | (-14.3%, | -12.4%) | 137 | -12.2% | (-13.4%, | -11.0%) | 48 | -14.5% | (-16.7%, | -12.3%) |
| BMI<50 | 1415 | -13.2% | (-13.7%, | -12.8%) | 796 | -13.8% | (-14.4%, | -13.2%) | 258 | -15.0% | (-16.2%, | -13.8%) |
| Difference | | -0.1% | (-1.2%, | 0.9%) | | 1.6% | (0.3%, | 2.9%) | | 0.5% | (-1.9%, | 2.8%) |
| P-value | | | | 0.80 | | | | 0.014* | | | | 0.70 |
| HTE | | 5.1% | (4.0%, | 6.2%) | | 5.0% | (3.7%, | 6.3%) | | 2.8% | (0.3%, | 5.3%) |
| Difference | | | | | | | | | | | | |
| Interaction P-value | | | | <0.001 ł | | | | <0.001 ł | | | | 0.028 ł |

%TWL = percent total weight Loss; CI = confidence interval; HTE = heterogeneity of treatment effects

* These results show statistically significant subgroup effects comparing weight outcomes of patients with BMI 50+ vs. BMI <50 within procedures. The magnitude of these between subgroup differences is ≤5.2 percentage points at all time points.

ł These results show statistically significant interactions (heterogeneity in treatment effects), where the results of the pairwise comparison between two procedures differs across subgroups of patients with BMI 50+ vs. BMI <50 at baseline. The magnitude of these subgroup differences for the RYGB vs. SG pairwise comparisons are clinically small and ≤1.3 percentage points of total weight loss; however, the magnitude of these differences comparing AGB vs. either RYGB or SG is larger (5.0 to 5.9 percentage points total weight loss at 1 and 3 years; 2.8 to 2.9 percentage points at 5 years).

**Table 7. Comparison of Percent Weight Loss at 1, 3, and 5 years Across Age subgroups**

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | |
| **RYGB** | | | | | | | | | | | | |
| Age 65+ | 1077 | -28.0% | (-28.5%, | -27.6%) | 541 | -28.8% | (-29.4%, | -28.2%) | 218 | -28.7% | (-29.7%, | -27.7%) |
| Age <65 | 17952 | -31.4% | (-31.5%, | -31.2%) | 8684 | -31.4% | (-31.6%, | -31.3%) | 3458 | -31.0% | (-31.2%, | -30.7%) |
| Difference | | 3.3% | (2.8%, | 3.8%) | | 2.6% | (2.0%, | 3.2%) | | 2.3% | (1.3%, | 3.3%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| **SG** | | | | | | | | | | | | |
| Age 65+ | 756 | -20.9% | (-21.5%, | -20.3%) | 245 | -21.5% | (-22.4%, | -20.7%) | 38 | -21.9% | (-24.1%, | -19.7%) |
| Age <65 | 14173 | -25.4% | (-25.6%, | -25.3%) | 5059 | -25.2% | (-25.4%, | -25.0%) | 1050 | -25.1% | (-25.6%, | -24.6%) |
| Difference | | 4.5% | (4.0%, | 5.1%) | | 3.7% | (2.8%, | 4.6%) | | 3.2% | (1.0%, | 5.4%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | 0.005* |
| HTE | | 1.2% | (0.5%, | 1.9%) | | 1.1% | (0.0%, | 2.1%) | | 0.9% | (-1.5%, | 3.4%) |
| Difference | | | | | | | | | | | | |
| Interaction P-value | | | | 0.001 ǂ | | | | 0.045 ǂ | | | | 0.45 |
| **RYGB** | | | | | | | | | | | | |

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | |
| Age 65+ | 1057 | -28.2% | (-28.6%, | -27.7%) | 539 | -28.9% | (-29.5%, | -28.3%) | 221 | -28.7% | (-29.7%, | -27.8%) |
| Age <65 | 17627 | -31.4% | (-31.5%, | -31.3%) | 8613 | -31.5% | (-31.6%, | -31.3%) | 3512 | -31.0% | (-31.3%, | -30.8%) |
| Difference | | 3.2% | (2.8%, | 3.7%) | | 2.6% | (2.0%, | 3.2%) | | 2.3% | (1.3%, | 3.3%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| **AGB** | | | | | | | | | | | | |
| Age 65+ | 125 | -13.8% | (-15.1%, | -12.6%) | 73 | -14.8% | (-16.4%, | -13.3%) | 25 | -15.1% | (-17.8%, | -12.3%) |
| Age <65 | 1556 | -13.6% | (-14.0%, | -13.2%) | 870 | -14.2% | (-14.7%, | -13.7%) | 312 | -15.0% | (-15.9%, | -14.2%) |
| Difference | | -0.2% | (-1.6%, | 1.1%) | | -0.6% | (-2.3%, | 1.0%) | | -0.1% | (-2.9%, | 2.8%) |
| P-value | | | | 0.72 | | | | 0.44 | | | | 0.97 |
| HTE Difference | | -3.5% | (-4.9%, | -2.1%) | | -3.3% | (-5.0%, | -1.5%) | | -2.3% | (-5.3%, | 0.7%) |
| Interaction P-value | | | | <0.001 ǂ | | | | <0.001 ǂ | | | | 0.127 |
| **SG** | | | | | | | | | | | | |
| Age 65+ | 734 | -20.6% | (-21.2%, | -20.0%) | 243 | -21.4% | (-22.3%, | -20.5%) | 38 | -22.1% | (-24.4%, | -19.8%) |
| Age <65 | 13930 | -25.4% | (-25.6%, | -25.3%) | 5027 | -25.2% | (-25.4%, | -25.0%) | 1050 | -25.3% | (-25.8%, | -24.8%) |

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | |
| Difference | | 4.8% | (4.2%, | 5.5%) | | 3.8% | (2.9%, | 4.6%) | | 3.2% | (0.8%, | 5.5%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | 0.008* |
| **AGB** | | | | | | | | | | | | |
| Age 65+ | 125 | -13.2% | (-14.6%, | -11.8%) | 73 | -13.8% | (-15.4%, | -12.2%) | 25 | -14.9% | (-18.1%, | -11.8%) |
| Age <65 | 1556 | -13.1% | (-13.6%, | -12.7%) | 860 | -13.4% | (-14.0%, | -12.9%) | 281 | -14.7% | (-15.9%, | -13.6%) |
| Difference | | -0.1% | (-1.5%, | 1.3%) | | -0.3% | (-2.0%, | 1.3%) | | -0.2% | (-3.4%, | 2.9%) |
| P-value | | | | 0.89 | | | | 0.70 | | | | 0.89 |
| HTE Difference | | -5.0% | (-6.5%, | -3.4%) | | -4.1% | (-5.9%, | -2.2%) | | -3.4% | (-7.3%, | 0.5%) |
| Interaction P-value | | | | <0.001 ł | | | | <0.001 ł | | | | 0.089 |

%TWL = percent total weight Loss; CI = confidence interval; HTE = heterogeneity of treatment effects

* These results show statistically significant subgroup effects comparing weight outcomes of patients with age 65+ vs. <65 years within procedures. The magnitude of these between subgroup differences is ≤3.3 percentage points at all time points.

ł These results show statistically significant interactions (heterogeneity in treatment effects), where the results of the pairwise comparison between two procedures differs across subgroups of patients with age 65+ vs. <65 years at baseline. The magnitude of these subgroup differences for the

RYGB vs. SG pairwise comparisons are clinically small and ≤1.2 percentage points of total weight loss; however, the magnitude of these differences comparing AGB vs. either RYGB or SG is larger but always ≤5 percentage points.

**Table 8. Comparison of Percent Weight Loss at 1, 3, and 5 years Across Gender subgroups**

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | |
| **RYGB** | | | | | | | | | | | | |
| Female | 15230 | -30.9% | (-31.0%, | -30.8%) | 7486 | -31.3% | (-31.5%, | -31.1%) | 3021 | -31.0% | (-31.3%, | -30.8%) |
| Male | 3799 | -32.2% | (-32.5%, | -32.0%) | 1739 | -31.7% | (-32.0%, | -31.3%) | 655 | -30.7% | (-31.3%, | -30.2%) |
| Difference | | 1.3% | (1.0%, | 1.6%) | | 0.4% | (-0.0%, | 0.8%) | | -0.3% | (-0.9%, | 0.4%) |
| P-value | | | | <0.001* | | | | 0.057 | | | | 0.40 |
| **SG** | | | | | | | | | | | | |
| Female | 12160 | -24.9% | (-25.1%, | -24.8%) | 4402 | -25.1% | (-25.3%, | -24.9%) | 899 | -25.3% | (-25.8%, | -24.7%) |
| Male | 2769 | -26.4% | (-26.7%, | -26.1%) | 902 | -25.0% | (-25.5%, | -24.6%) | 189 | -24.4% | (-25.5%, | -23.4%) |
| Difference | | 1.5% | (1.2%, | 1.8%) | | -0.1% | (-0.6%, | 0.4%) | | -0.8% | (-2.0%, | 0.3%) |
| P-value | | | | <0.001* | | | | 0.76 | | | | 0.146 |
| | | | | | | | | | | | | |
| HTE Difference | | 0.2% | (-0.2%, | 0.6%) | | -0.5% | (-1.1%, | 0.2%) | | -0.6% | (-1.8%, | 0.7%) |
| Interaction P-value | | | | 0.34 | | | | 0.141 | | | | 0.38 |
| **RYGB** | | | | | | | | | | | | |

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | |
| Female | 14946 | -31.2% | (-31.3%, | -31.1%) | 7426 | -31.3% | (-31.5%, | -31.2%) | 3062 | -31.0% | (-31.3%, | -30.7%) |
| Male | 3738 | -32.2% | (-32.4%, | -31.9%) | 1726 | -31.7% | (-32.0%, | -31.3%) | 671 | -31.2% | (-31.8%, | -30.6%) |
| Difference | | 1.0% | (0.7%, | 1.3%) | | 0.3% | (-0.1%, | 0.8%) | | 0.2% | (-0.5%, | 0.9%) |
| P-value | | | | <0.001* | | | | 0.125 | | | | 0.56 |
| **AGB** | | | | | | | | | | | | |
| Female | 1331 | -14.1% | (-14.5%, | -13.7%) | 770 | -14.8% | (-15.3%, | -14.3%) | 274 | -15.7% | (-16.6%, | -14.8%) |
| Male | 350 | -12.3% | (-13.1%, | -11.6%) | 173 | -12.8% | (-13.9%, | -11.8%) | 63 | -12.7% | (-14.5%, | -10.9%) |
| Difference | | -1.8% | (-2.6%, | -0.9%) | | -1.9% | (-3.1%, | -0.8%) | | -3.0% | (-5.0%, | -1.0%) |
| P-value | | | | <0.001* | | | | 0.001* | | | | 0.003* |
| HTE Difference | | -2.7% | (-3.6%, | -1.8%) | | -2.3% | (-3.5%, | -1.1%) | | -3.2% | (-5.2%, | -1.0%) |
| Interaction P-value | | | | <0.001 ł | | | | <0.001 ł | | | | 0.002ł |
| **SG** | | | | | | | | | | | | |
| Female | 11963 | -24.7% | (-24.8%, | -24.5%) | 4377 | -24.9% | (-25.1%, | -24.7%) | 899 | -25.3% | (-25.9%, | -24.8%) |
| Male | 2701 | -27.0% | (-27.3%, | -26.6%) | 893 | -25.9% | (-26.4%, | -25.4%) | 189 | -24.6% | (-25.8%, | -23.5%) |

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | |
| Difference | | 2.3% | (1.9%, | 2.7%) | | 1.0% | (0.5%, | 1.6%) | | -0.7% | (-2.0%, | 0.6%) |
| P-value | | | | <0.001* | | | | <0.001* | | | | 0.306 |
| **AGB** | | | | | | | | | | | | |
| Female | 1331 | -13.4% | (-13.9%, | -12.9%) | 761 | -13.8% | (-14.4%, | -13.2%) | 249 | -14.8% | (-16.1%, | -13.6%) |
| Male | 350 | -12.7% | (-13.5%, | -11.8%) | 172 | -12.9% | (-14.0%, | -11.8%) | 57 | -13.1% | (-15.1%, | -11.1%) |
| Difference | | -0.7% | (-1.7%, | 0.2%) | | -0.9% | (-2.1%, | 0.3%) | | -1.7% | (-3.9%, | 0.5%) |
| P-value | | | | 0.134 | | | | 0.133 | | | | 0.127 |
| HTE Difference | | -3.0% | (-4.0%, | -2.0%) | | -1.9% | (-3.2%, | -0.7%) | | -1.0% | (-3.4%, | 1.4%) |
| Interaction P-value | | | | <0.001 ł | | | | 0.002 ł | | | | 0.40 |

*%TWL = percent total weight Loss; CI = confidence interval; HTE = heterogeneity of treatment effects

* These results show statistically significant subgroup effects comparing weight outcomes of male and female patients within procedures. The magnitude of these between subgroup differences is ≤3.0 percentage points at all time points.

ł These results show statistically significant interactions (heterogeneity in treatment effects), where the results of the pairwise comparison between two procedures differs across subgroups of male and female patients at baseline. The magnitude of these subgroup differences for the RYGB vs. SG

pairwise comparisons are clinically small and ≤0.6 percentage points of total weight loss; however, the magnitude of these differences comparing AGB vs. either RYGB or SG is larger but always ≤3.2 percentage points.

**Table 9. Comparison of Percent Weight Loss at 1, 3, and 5 years Across Racial/Ethnic subgroups**

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | |
| **SG** | | | | | | | | | | | | |
| Caucasian | 6163 | -25.5% | (-25.7%, | -25.3%) | 2182 | -25.0% | (-25.2%, | -24.8%) | 485 | -24.6% | (-25.0%, | -24.1%) |
| Hispanic | 2498 | -25.0% | (-25.3%, | -24.7%) | 1125 | -24.6% | (-24.9%, | -24.3%) | 265 | -24.0% | (-24.7%, | -23.4%) |
| African American | 3477 | -24.0% | (-24.2%, | -23.7%) | 1440 | -23.5% | (-23.7%, | -23.2%) | 289 | -22.9% | (-23.5%, | -22.3%) |
| **AA vs. Caucasian** | | 1.6% | (1.3%, | 1.9%) | | 1.6% | (1.2%, | 1.9%) | | 1.7% | (0.9%, | 2.4%) |
| P-Value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| **Hispanic vs. Caucasian** | | 0.01% | (0.2%, | 0.9%) | | 0.00% | (0.0%, | 0.7%) | | 0.01% | (-0.2%, | 1.3%) |
| P-Value | | | | <0.001* | | | | 0.041* | | | | 0.153 |
| SG Race P-Value | | | | <0.001 | | | | <0.001 | | | | <0.001 |
| **RYGB** | | | | | | | | | | | | |
| Caucasian | 11448 | -31.7% | (-31.8%, | -31.6%) | 5725 | -31.8% | (-32.0%, | -31.7%) | 2452 | -31.3% | (-31.5%, | -31.0%) |
| Hispanic | 2123 | -29.6% | (-29.9%, | -29.4%) | 1140 | -29.8% | (-30.1%, | -29.5%) | 488 | -29.3% | (-29.8%, | -28.9%) |
| African American | 2269 | -29.0% | (-29.3%, | -28.7%) | 1111 | -28.6% | (-28.9%, | -28.3%) | 396 | -28.1% | (-28.6%, | -27.6%) |
| **AA vs. Caucasian** | | 2.7% | (2.4%, | 3.0%) | | 3.2% | (2.9%, | 3.6%) | | 3.1% | (2.6%, | 3.7%) |
| P-Value | | | | <0.001* | | | | <0.001* | | | | <0.001* |

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | |
| **Hispanic vs. Caucasian** | | 2.1% | (1.7%, | 2.4%) | | 2.0% | (1.7%, | 2.4%) | | 1.9% | (1.4%, | 2.5%) |
| P-Value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| RYGB Race P-Value | | | | <0.001 | | | | <0.001 | | | | <0.001 |
| SG-RYGB Interaction P-Value | | | | <0.001 ǂ | | | | <0.001 ǂ | | | | <0.001 ǂ |
| **AGB** | | | | | | | | | | | | |
| Caucasian | 807 | -14.2% | (-14.7%, | -13.8%) | 445 | -14.8% | (-15.3%, | -14.3%) | 167 | -15.2% | (-15.9%, | -14.5%) |
| Hispanic | 296 | -12.8% | (-13.6%, | -12.1%) | 68 | -13.2% | (-14.4%, | -11.9%) | 21 | -14.8% | (-16.7%, | -12.8%) |
| African American | 120 | -12.7% | (-13.8%, | -11.5%) | 165 | -12.9% | (-13.7%, | -12.0%) | 57 | -14.5% | (-15.8%, | -13.3%) |
| **AA vs. Caucasian** | | 1.4% | (0.5%, | 2.2%) | | 1.9% | (1.0%, | 2.9%) | | 0.7% | (-0.7%, | 2.1%) |
| P-Value | | | | 0.002* | | | | <0.001* | | | | 0.349 |
| **Hispanic vs. Caucasian** | | 0.02% | (0.3%, | 2.8%) | | 0.02% | (0.3%, | 3.0%) | | 0.0% | (-1.7%, | 2.5%) |
| P-Value | | | | 0.013* | | | | 0.017* | | | | 0.70 |
| AGB Race P-Value | | | | 0.001 | | | | <0.001 | | | | 0.63 |

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | |
| **RYGB** | | | | | | | | | | | | |
| Caucasian | 11192 | -31.8% | (-31.9%, | -31.6%) | 5663 | -31.7% | (-31.8%, | -31.6%) | 2494 | -31.2% | (-31.4%, | -31.0%) |
| Hispanic | 2112 | -29.8% | (-30.1%, | -29.5%) | 1137 | -29.8% | (-30.1%, | -29.5%) | 491 | -29.3% | (-29.7%, | -28.9%) |
| African American | 2254 | -29.0% | (-29.3%, | -28.7%) | 1104 | -28.5% | (-28.8%, | -28.2%) | 406 | -28.2% | (-28.6%, | -27.7%) |
| **AA vs. Caucasian** | | 2.8% | (2.5%, | 3.1%) | | 3.2% | (2.8%, | 3.5%) | | 3.0% | (2.5%, | 3.5%) |
| P-Value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| **Hispanic vs. Caucasian** | | 2.0% | (1.6%, | 2.3%) | | 1.9% | (1.5%, | 2.2%) | | 1.9% | (1.4%, | 2.4%) |
| P-Value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| RYGB Race P-Value | | | | <0.001 | | | | <0.001 | | | | <0.001 |
| AGB-RYGB Interaction P-Value | | | | 0.010 ǂ | | | | 0.057 | | | | 0.006 ǂ |
| **AGB** | | | | | | | | | | | | |
| Caucasian | 807 | -13.8% | (-14.3%, | -13.3%) | 436 | -14.7% | (-15.2%, | -14.1%) | 148 | -15.9% | (-16.9%, | -15.0%) |

| | 1-YEAR | | | | 3-YEAR | | | | 5-YEAR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | | **N** | **% WC** | **95% CI** | |
| Hispanic | 120 | -12.7% | (-14.0%, | -11.4%) | 68 | -13.4% | (-14.6%, | -12.1%) | 19 | -16.1% | (-18.3%, | -14.0%) |
| African American | 296 | -12.9% | (-13.7%, | -12.0%) | 164 | -13.0% | (-13.8%, | -12.1%) | 49 | -13.5% | (-14.9%, | -12.1%) |
| **AA vs. Caucasian** | | 0.9% | (-0.0%, | 1.9%) | | 1.7% | (0.8%, | 2.7%) | | 2.4% | (0.8%, | 4.0%) |
| P-Value | | | | 0.061 | | | | <0.001* | | | | 0.004* |
| **Hispanic vs. Caucasian** | | 0.01% | (-0.3%, | 2.4%) | | 0.01% | (-0.0%, | 2.7%) | | 0.0% | (-2.5%, | 2.1%) |
| P-Value | | | | 0.129 | | | | 0.052 | | | | 0.87 |
| **AGB Race P-Value** | | | | 0.086 | | | | <0.001 | | | | 0.010 |
| **SG** | | | | | | | | | | | | |
| Caucasian | 5960 | -25.6% | (-25.8%, | -25.4%) | 2157 | -25.2% | (-25.4%, | -25.0%) | 485 | -25.1% | (-25.6%, | -24.7%) |
| Hispanic | 2477 | -25.0% | (-25.3%, | -24.7%) | 1125 | -24.7% | (-25.0%, | -24.4%) | 265 | -24.4% | (-25.0%, | -23.9%) |
| African American | 3443 | -24.1% | (-24.3%, | -23.8%) | 1432 | -23.7% | (-23.9%, | -23.4%) | 289 | -23.3% | (-23.8%, | -22.7%) |
| **AA vs. Caucasian** | | 1.5% | (1.2%, | 1.8%) | | 1.5% | (1.2%, | 1.9%) | | 1.9% | (1.2%, | 2.5%) |
| P-Value | | | | <0.001* | | | | <0.001* | | | | <0.001* |
| **Hispanic vs. Caucasian** | | 0.6% | (0.3%, | 0.9%) | | 0.5% | (0.1%, | 0.9%) | | 0.7% | (0.0%, | 1.4%) |
| P-Value | | | | <0.001* | | | | 0.011* | | | | 0.04* |
| SG Race P-Value | | | | <0.001 | | | | <0.001 | | | | <0.001 |

| | 1-YEAR | | | 3-YEAR | | | 5-YEAR | | |
|---|---|---|---|---|---|---|---|---|---|
| | **N** | **% WC** | **95% CI** | **N** | **% WC** | **95% CI** | **N** | **% WC** | **95% CI** |
| AGB-SG Interaction P-Value | | | 0.33 | | | 0.49 | | | 0.53 |

%TWL = percent total weight Loss; CI = confidence interval; HTE = heterogeneity of treatment effects; AA = African-American

* These results show statistically significant subgroup effects comparing weight outcomes of patients with different race/ethnicity within procedures. The magnitude of these between subgroup differences is ≤3.2 percentage points at all time points.

ł These results show statistically significant interactions (heterogeneity in treatment effects), where the results of the pairwise comparison between two procedures differs across subgroups defined by the race/ethnicity of patients at baseline.

**Table 10. Sensitivity Analysis Linear Fixed Effects Regression Model of the Comparative Effectiveness of Gastric Bypass, Sleeve Gastrectomy, and Adjustable Gastric Banding for Percent Total Weight Loss (%TWL) among Adults at 1, 3, and 5 Years Follow-up***

| | Time Since Bariatric Procedure | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 Year | | | 3 Years | | | 5 Years | | |
| **Procedure Comparison** | N | %TWL | 95% CI | N | %TWL | 95% CI | N | %TWL | 95% CI |
| **SG** | 25,530 | -25.0 | (-25.1, -24.9) | 25,530 | -21.7 | (-22.0, -21.4) | 25,530 | -18.7 | (-19.2, -18.2) |
| **RYGB** | 28,194 | -31.0 | (-31.2, -30.9) | 28,194 | -28.9 | (-29.2, -28.7) | 28,194 | -26.3 | (-26.7, -25.9) |
| **Difference** | | 6.0 | (5.9, 6.2) | | 7.2 | (6.9, 7.6) | | 7.6 | (7.0, 8.2) |
| **P-value** | | | <0.001 | | | <0.001 | | | <0.001 |
| **AGB** | 2,432 | -13.5 | (-13.8, -13.2) | 2,432 | -11.3 | (-12.0, -10.6) | 2,432 | -8.1 | (-9.2, -6.9) |
| **RYGB** | 27,885 | -31.1 | (-31.2, -31.0) | 27,885 | -29.1 | (-29.3, -28.9) | 27,885 | -26.6 | (-27.0, -26.3) |
| **Difference** | | 17.7 | (17.3, 18.0) | | 17.8 | (17.1, 18.5) | | 18.6 | (17.4, 19.7) |
| **P-value** | | | <0.001 | | | <0.001 | | | <0.001 |

| | Time Since Bariatric Procedure | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 Year | | | | 3 Years | | | | 5 Years | | | |
| **Procedure Comparison** | N | %TWL | 95% CI | | N | %TWL | 95% CI | | N | %TWL | 95% CI | |
| **AGB** | 2,432 | -13.1 | (-13.4, | -12.7) | 2,432 | -10.7 | (-11.6, | -9.9) | 2,432 | -7.4 | (-8.8, | -6.1) |
| **SG** | 25,353 | -25.1 | (-25.2, | -25.0) | 25,353 | -21.8 | (-22.1, | -21.5) | 25,353 | -18.9 | (-19.4, | -18.4) |
| **Difference** | | 12.0 | (11.7, | 12.4) | | 11.0 | (10.2, | 11.9) | | 11.5 | (10.0, | 12.9) |
| **P-value** | | | | <0.001 | | | | <0.001 | | | | <0.001 |

*%TWL = percent total weight loss; To assess the implications of drop out and baseline missing data on our results, we compared our primary analysis results to a simple, covariate adjusted model (no propensity scores) run on a single data set pooling all longitudinal data among patients with at least one post-surgery measurement while excluding the race, ethnicity and blood pressure variables, which were the primary sources of missing baseline data (see Table 1 of the manuscript). The sample size for this sensitivity analysis is 56,156 patients (25,530 SG + 28,194 RYGB + 2,432 AGB), which includes the 46,510 in our primary analysis plus an additional 9,646 patients who were eligible at baseline and had at least 1 post-surgery BMI measure, but lacked a BMI measure within the 1, 3, and 5-year windows used for the primary analysis. This approach assumes that loss to follow-up is associated with patients' covariate data (not outcome data), and mean estimates of percent change in BMI are weighted to reflect BMI trajectories of patients with any post-surgical weight measurements, rather than the subset with follow-up data available only in the 1, 3, or 5-year windows.

**Table 11. Sensitivity Analysis using Trimmed Propensity Scores for our Models of the Comparative Effectiveness of Gastric Bypass, Sleeve Gastrectomy, and Adjustable Gastric Banding for Percent Total Weight Loss (%TWL) among Adults at 1, 3, and 5 Years Follow-up***

| | Time Since Bariatric Procedure | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 Year | | | 3 Years | | | 5 Years | | | |
| **Procedure Comparison** | N | %TWL | 95% CI | N | %TWL | 95% CI | N | %TWL | 95% CI | |
| **SG** | 10,420 | -25.3 | (-25.5, -25.1) | 3,480 | -21.0 | (-21.4, -20.6) | 558 | -19.3 | (-20.4, | -18.2) |
| **RYGB** | 13,350 | -31.2 | (-31.3, -31.1) | 6,631 | -28.9 | (-29.2, -28.6) | 2,763 | -25.6 | (-26.0, | -25.1) |
| **Difference** | | 5.9 | (5.7, 6.1) | | 8.0 | (7.5, 8.4) | | 6.2 | (5.0, | 7.5) |
| **P-value** | | | <0.001 | | | <0.001 | | | | <0.001 |
| **AGB** | 327 | -13.8 | (-14.6, -13.0) | 245 | -12.6 | (-14.0, -11.3) | 44 | -13.6 | (-17.2, | -9.9) |
| **RYGB** | 7,818 | -31.4 | (-31.5, -31.2) | 4,871 | -28.7 | (-29.0, -28.4) | 2,826 | -25.4 | (-25.9, | -24.9) |
| **Difference** | | 17.6 | (16.8, 18.4) | | 16.1 | (14.7, 17.4) | | 11.8 | (8.1, | 15.5) |
| **P-value** | | | <0.001 | | | <0.001 | | | | <0.001 |
| **AGB** | 159 | -13.7 | (-14.9, -12.5) | 102 | -12.8 | (-15.0, -10.5) | 41 | -13.5 | (-17.3, | -9.7) |

| | Time Since Bariatric Procedure | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 Year | | | | 3 Years | | | | 5 Years | | | |
| **Procedure Comparison** | N | %TWL | 95% CI | | N | %TWL | 95% CI | | N | %TWL | 95% CI | |
| **SG** | 11,730 | -24.8 | (-24.9, | -24.6) | 4,119 | -20.5 | (-20.9, | -20.2) | 492 | -17.2 | (-18.3, | -16.1) |
| **Difference** | | 11.1 | (9.8, | 12.3) | | 7.7 | (5.5, | 10.0) | | 3.7 | (-0.2, | 7.7) |
| **P-value** | | | | <0.001 | | | | <0.001 | | | | 0.065 |

*%TWL = percent total weight loss. For each of our pairwise comparisons we trimmed the propensity scores, refit the propensity score models, and refit the outcome models. Further details on our approach to this sensitivity analysis can be found in the Statistical Appendix.