

THE LANCET

Digital Health

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Saux P, Bauvin P, Raverdy V, et al. Development and validation of an interpretable machine learning-based calculator for predicting 5-year weight trajectories after bariatric surgery: a multinational retrospective cohort SOPHIA study. *Lancet Digit Health* 2023; published online Aug 29. [https://doi.org/10.1016/S2589-7500\(23\)00135-8](https://doi.org/10.1016/S2589-7500(23)00135-8).

Development and validation of an interpretable machine learning based calculator for predicting 5 year-weight trajectories after bariatric surgery: a multinational retrospective cohort SOPHIA study

Appendix 1: Preprocessing and visit date adjustment.....	3
Appendix 2: Smoothing of TWL trajectories.....	4
Appendix 3: Cohorts.....	5
Appendix 4: Model development.....	8
Appendix 5: Comparison with other methods.....	9
Appendix 6: Comparison with previously published models.....	10
Appendix Figure 1: ABOS patient flowchart.....	11
Appendix Figure 2: Smoothed observed median BMI and TWL trajectories and corresponding IQR, for A) ABOS cohort, B) BAREVAL cohort, C) NOK cohort, D) SGH cohort, E) SOS cohort, F) PRECOS cohort, G) Roma cohort, H) Lyon cohort, I) COD cohort, J) Mexico cohort, K) SleevePass RCT cohort and L) SM-BOSS RCT cohort.....	12
Appendix Figure 3: Machine learning pipeline diagram in A), and all decision trees for prediction of %TWL at B) month 12, C) month 24 and D) month 60.....	24
Appendix Figure 4: Bland-Altman plots at A) M12, B) M24 and C) M60 for all interventions.....	28
Appendix Figure 5: Predicted trajectory and IQR of prediction error, for a 40 years old patient with height 1.70m, non-smoker, with prediabetes, undergoing RYGB with preoperative weight 118kg (blue) and 138kg (green), corresponding to the 1st and 2nd tertile of weights in the ABOS cohort.....	31
Appendix Figure 6: Accuracy of BMI prediction, in terms of MAD error, on internal test data as well as (available) external test data, comparing decision trees with A) linear regression, B) longitudinal model (mixed effect model), C) random forest, and D) CART decision tree with pruning (without LASSO).....	32
Appendix Table 1: Validation of existing prediction models on the ABOS cohort for corresponding type of intervention and date of prediction.....	36
Appendix Table 2: Comparison of literature models versus our model on the ABOS testing subset.....	37
Appendix Table 3: Model-based means for primary and secondary outcomes after sleeve gastrectomy or Roux-en-Y gastric bypass at 1, 3 and 5 years (similar format as Table 2 from PMID: 33640917).....	38
Appendix Table 4: Baseline attributes by categories, before feature selection.....	40
Appendix table 5: TRIPOD checklist for model development and validation.....	62
Appendix Table 6: Baseline characteristics, among 633 expected at 5 years, comparing the complete cases with those with missing data at one or more follow-up visits.....	65
Appendix Table 7: List of authors in charge of the access and verification of each cohort.....	66
References.....	67

Appendix 1: Preprocessing and visit date adjustment

We first performed a preprocessing of all patients' features: features containing more than 50% of missing values were excluded, as well as categorical features with high-class imbalance (i.e. one class accounting for less than 2% of all patients) and features consisting of free text input. Categorical variables were converted into binary variables.

Regarding the visit date, we performed a linear adjustment to improve data quality and downstream prediction accuracy. Although visits were planned at 1 month, 3 months, 1 year, 2 years, and 5 years after the intervention, actual visits may occur earlier or later than the scheduled date. A direct consequence is that the weight losses were measured with a drift compared to the model outcomes (weight losses at scheduled dates), introducing an additional source of outcome variability, a priori unrelated to the baseline characteristics.

Firstly, patients with a previous history of bariatric surgery were removed (n=286). Moreover, patients with large delays between scheduled and actual visits related to postoperative complications were also removed (n=59). We defined significant delays as more than 10 days early or 1 month late for month 1; 25 days early or 2 months late for month 3; 2 months early or 6 months late for 1 year; 3 months early or 6 months late for 2 years; 6 months early or 24 months late for 5 years. In case of missing follow-up visits, patients were kept in the analysis but censored at the corresponding dates. Patients not expected at a given time (recent interventions) were also censored after the last completed visit.

Secondly, we linearly interpolated the weights at 1 month, 3 months, 1 year, 2 years, and 5 years between the previous and current visits and computed the corresponding TWL from the intervention. Indeed, weight predictions are made at conventional dates corresponding to fixed numbers of months after surgery. However, the actual visit dates in the training cohort (when weights are measured) differ from the conventional dates and vary across patients due to early or late follow-up visits. This introduces a bias between the prediction targets and the training data. For instance, if a patient had an M3 visit on schedule (i.e. exactly 3 months after surgery) and an M12 visit delayed by a month (i.e. 13 months after surgery rather than 12 months), we would expect the true weight 12 months after surgery to have been between the reported M3 and M12 weights, and larger than the latter since the patient was likely on a downward weight trajectory.

In order to account for this bias, we linearly interpolated the weights at M 1, 3, 12, 24, 60 between the previous and current visits and computed the corresponding TWL from intervention. In the case of the example patient above, we would estimate the scheduled M12 weight as $9/10 \times$ reported M12 weight $+1/10 \times$ reported M3 weight.

Appendix 2: Smoothing of TWL trajectories

In the online tool displayed in Figure 3 and Appendix Figure 5, predicted trajectories of BMI and TWL are represented by smooth, continuous curves. However, the predicted outcomes of the tree-based model correspond to a set of postoperative dates, namely 1, 3, 12, 24 and 60 months after surgery. In order to smoothly generalize to arbitrary dates, we use the following nonlinear parametric model:

$$TWL_{SG,RYGB}(t) = \alpha / \zeta,$$

where α , β , γ , t_0 , τ and *intercept* are coefficients fitted by nonlinear least-squares regression (Levenberg-Marquardt algorithm, available in the *scipy.optimize* library in the Python programming language). We designed this particular model to match qualitative features of the typical weight loss evolution observed after RYGB and SG. For $\alpha=0$, these curves

exhibit an initial weight loss until $t_{nadir} = t_0 + \tau \log \frac{1}{\gamma}$, followed by weight regain and long-term

flattening. Moreover, the first few months after surgery are often associated with rapid weight loss, which is modeled by the term α / ζ , the importance of which decays with time and thus mostly impacts the front-end of the curve; in particular, the curve still exhibits a nadir at

$t_{nadir} \approx t_0 + \tau \log \frac{1}{\gamma}$ followed by a weight regain even when $\alpha \neq 0$.

For AGB patients, as described in the discussion on patients' characteristics, the restriction to first-time interventions induces a downward weight trend on average, without weight regain after nadir. To represent this, we opted the following monotonic model:

$$TWL_{AGB}(t) = \alpha / \zeta,$$

where α , β , γ and *intercept* are coefficients fitted by least-squares regression. The rationale for the decaying factor α is the same as for RYGB and SG, that is to account for the initial rapid weight loss. The factors β and γ respectively account for the medium and long term trend of the TWL curve.

Appendix 3: Cohorts

Training cohorts

The ABOS cohort (Atlas Biologique de l'Obésité Sévère, NCT01129297), a prospective study conducted in Lille, France, aiming to identify the determinants of the outcome of bariatric surgery. All ABOS participants who underwent their first bariatric surgery at Lille University Hospital were enrolled in the present study (Appendix Figure 1). Patients with gastric bypass, sleeve gastrectomy and gastric banding recruited from 2006 through 2020 corresponding to a follow-up data duration of 60 months were included in this analysis. Informed consent was obtained from all participants. Demographic characteristics, anthropomorphic measurements, medical history, medication use, and clinical laboratory tests were prospectively collected before surgery. A 75 g oral glucose tolerance test (OGTT) was performed after overnight fasting at baseline. Diabetes status was defined at baseline, based upon the previous history of diabetes, use of antidiabetic medication, and/or fasting plasma glucose (FBG) ≥ 126 mg/dL (7.0 mmol/L) and/or a 2-hour plasma glucose ≥ 200 mg/dL (11.1 mmol/L) during the OGTT and/or HbA1c $\geq 6.5\%$ (48 mmol/L).¹ Homeostasis model assessment (HOMA) 2 estimates of β -cell function (HOMA2-B) and insulin resistance (HOMA2-IR) based on C-peptide concentrations (which performs better than insulin in patients with diabetes) was calculated with the HOMA calculator (University of Oxford, Oxford, UK).

Since ABOS enrolled a majority of patients with RYGB and AGB, we combined the data with those from the BAREVAL cohort from Montpellier (NCT02310178), a sleeve specific cohort to constitute the training set. The BAREVAL cohort is a prospective cohort designed to evaluate the long-term effectiveness of sleeve gastrectomy on weight loss. Demographic characteristics, anthropomorphic measurements, medical history, medication use, and clinical laboratory tests were prospectively collected before surgery. Diabetes status was defined at baseline, based upon the previous history of diabetes, use of antidiabetic medication, and/or fasting plasma glucose (FBG) ≥ 126 mg/dL (7.0 mmol/L) and/or HbA1c $\geq 6.5\%$ (48 mmol/L).¹ Data was available at baseline and at 3, 12, 24, and 60 months follow-up. All patients that underwent primary bariatric surgery between 2014 and 2020 were selected.

External testing cohorts

Datasets from the validation centers were prospectively collected as part of registered prospective cohort studies, and stored in the centers' databases in compliance with local and national regulatory requirements. Analyses were done locally (NOK and SOS study cohorts) and aggregated results were sent to Lille University Hospital. For other cohorts, pseudonymized data were received and analyzed in Lille, in accordance with the General Data Protection Regulation (GDPR) recommendations. The validation cohorts followed the legal and ethical rules applied in each country. Informed consent was obtained from all participants.

The Dutch Obesity Clinic (Nederlandse Obesitas Kliniek, NOK) is the largest outpatient clinic for the surgery of obesity in the Netherlands with currently ten locations over the country and is treating circa 5000 patients yearly. All patients attend an extensive pre- and postoperative counseling program, which is led by a multidisciplinary team, consisting of a

physician, dietician, psychologist and physiotherapist. Current cohort is a retrospective cohort with prospectively collected data. In order to provide the five-year follow-up data, all patients that underwent primary bariatric surgery in 2015 and 2016 were selected. Patients with interventions other than gastric bypass, gastric sleeve or gastric band were excluded, as well as patients with missing type of intervention and missing baseline weight. Demographic characteristics, anthropometric measurements and medical history were collected at baseline. Diabetes was scored as present or absent by the treating physician, based on the medical history and use of antidiabetic drugs. Smoking at baseline was also scored as either present or absent. Weight was collected before surgery and at 3, 6, 12, 18, 24, 36, 48, and 60 months at regular postoperative follow-up visits.

The SOS study is a non-randomized, controlled intervention study investigating the long-term effects of bariatric surgery. The study has been described in detail elsewhere.² In brief, 2007 participants who underwent bariatric surgery and 2040 matched controls were recruited between September 1, 1987 and January 31, 2001. The inclusion criteria were age 37 to 60 years and body mass index (BMI) of 34 or more for men and 38 or more for women. Exclusion criteria were identical in the surgery and control groups and were selected to ensure all participants were eligible for surgery. Oral or written informed consent was obtained from all study participants. Participants in the surgery group underwent gastric banding, vertical banded gastroplasty or gastric bypass. Participants in the control group received conventional obesity treatment. Participants in the surgery group underwent gastric banding, vertical banded gastroplasty or gastric bypass. Participants in the control group received conventional obesity treatment at their primary health care centers. The current analysis included only patients which underwent gastric banding (n=376) or gastric bypass (n=266). Participants that had a change in treatment type (i.e. band removal or conversion to a different surgery type) were censored on the corresponding date. Data was available at baseline and at 3, 12, 24, 48, and 72 months follow-up. Since no information about weight at 60 months (5 years) after surgery was available, we used the mean of the weights at 48 and 72 months as a proxy. Type 2 diabetes status at baseline was defined based on previous diagnosis of diabetes and/or use of antidiabetic medication and/or fasting plasma glucose ≥ 7 mmol/L and/or HbA1c ≥ 48 mmol/L.

The PRECOS (NCT03517072) study is a multicenter prospective cohort study, conducted in Arras, Boulogne sur/ Mer, Lille and Valenciennes to explore the determinants of 5 year-weight loss following bariatric surgery. Demographic characteristics, anthropomorphic measurements, medical history, medication use, and clinical laboratory tests were prospectively collected before surgery. Diabetes status was defined at baseline, based upon the previous history of diabetes, use of antidiabetic medication, and/or fasting plasma glucose (FBG) ≥ 126 mg/dL (7.0 mmol/L) and/or HbA1c $\geq 6.5\%$ (48 mmol/L).¹ Data was available at baseline and at 3, 12, 24, and 60 months follow-up.

The Roma cohort was collected from “Randomized clinical study comparing the effect of Roux-en-Y gastric bypass and Sleeve Gastrectomy on reactive hypoglycemia” (NCT01581801) and from “Diet and Medical Therapy Versus Bariatric Surgery in Type 2 Diabetes (DIBASY)” (NCT00888836). All subjects from these cohorts were enrolled for metabolic-bariatric surgery between 2009 and 2016 at Fondazione Policlinico Universitario A. Gemelli IRCCS - Catholic University of the Sacred Heart in Rome. The studies were approved by the Ethical Committee of the Fondazione Policlinico Universitario A. Gemelli IRCCS in Rome. Written informed consent was obtained from all participants. The weight data were obtained from 2009 to 2020 corresponding to a follow-up duration of 60 months.

Demographic characteristics, anthropometric measurements, medical history, medication use and clinical laboratory tests were prospectively collected before and after surgery. A 75 g oral glucose tolerance test (OGTT) was performed after overnight fasting at baseline in all subjects and also one year after surgery in participants at NCT01581801. Body composition was assessed by dual energy X-ray absorptiometry (Lunar-iDXA). Diagnosis of Type 2 Diabetes (T2D) was made at baseline taking into account the previous history of diabetes, use of antidiabetic medication, and/or fasting plasma glucose (FPG) ≥ 126 mg/dL (7.0 mmol/L) and/or a 2-hour plasma glucose ≥ 200 mg/dL (11.1 mmol/L) during the OGTT and/or HbA1c $\geq 6.5\%$ (48 mmol/L). Remission of T2D was defined as a FPG ≤ 100 mg/dl (5.6 mmol/L) and a HbA1c $\leq 6.5\%$ (48 mmol/L) for at least 1 year without active pharmacologic therapy.

The Sao Paulo cohort (COD) was comprised of 121 patients with T2D and obesity who received RYGB at the Oswaldo Cruz German Hospital, in Sao Paulo (SP; Brazil) between 2008 and 2016, and attended clinical and biological assessment at baseline and after one year of follow-up, as previously reported.^{3,4} All patients provided written informed consent prior to inclusion. Informed consent was obtained from all participants.

The Mexico cohort was collected from “Long versus short biliopancreatic limb in Roux-en-Y gastric bypass: short-term results of a randomized clinical trial” (clinical trial identifier NTC04609449), a randomized study with patients undergoing Roux-en-Y gastric bypass at a single academic center in Mexico city from 2016 to 2018.⁵ The study was approved by National and Institutional ethical Committees.

Randomized control trials

In 2018, two RCTs comparing laparoscopic SG and RYGB with a long-term follow-up rate of more than 80 per cent at 5 years were published. The Finnish SleevePass (Sleeve versus Bypass) equivalence study 10 enrolled 240 patients with percentage excess weight loss as the primary endpoint. The Swiss SM-BOSS (Swiss Multicentre Bypass or Sleeve Study) trial 11 included 217 patients with percentage excess BMI loss (%EBMIL) as the primary endpoint. Both trials showed better weight loss for RYGB. In the SleevePass trial the difference was not statistically significant based on the prespecified equivalence margins. In the SM-BOSS trial the difference was also not statistically significant after adjustments for multiple comparisons. The trials were designed as two-group, randomized, controlled, multicentre studies comparing SG with RYGB, involving 240 patients with severe obesity from Finland and 225 from Switzerland.¹⁴ Study protocols were reviewed and approved by the local ethical committees of each participating hospital. Both RCTs were conducted in accordance with the principles of the Declaration of Helsinki and registered at the clinical trials registry of the National Institutes of Health (NCT00356213, NCT00793143). Additional patient-level data were included, either prospectively collected unpublished 5-year data or information retrieved retrospectively from patient files. To ensure standardization of outcomes between the studies, all parameters and variables were translated into English to ensure identical definitions in both trials. T2DM remission was defined according to the American Diabetes Association criteria 14: complete remission by haemoglobin (Hb) A1c level below 6.0 per cent and fasting glucose concentration less than 100 mg/dl (below 5.6 mmol/l) for at least 1 year in the absence of active pharmacological therapy, and partial remission by HbA1c level below 6.5 per cent and fasting glucose concentration 100–125 mg/dl (5.6–6.9 mmol/l) for at least 1 year in the absence of active pharmacological therapy.

Appendix 4: Model development

To derive the model, the training cohort was divided into two subsets: a training subset consisting of 80% of randomly selected patients, and an internal testing subset comprising 20% of patients.

We first performed a preprocessing of all patients' features (Appendix P3). As the ABOS cohort had a large number of preoperative attributes per patient (Appendix Table 4), we ran a feature selection algorithm on this patient subgroup to extract the most statistically relevant ones concerning outcome prediction. We used the *Least Absolute Shrinkage and Selection Operator* (LASSO) with a cross-validated regularization parameter to limit overfitting, as well as the hierarchical group LASSO to account for interactions. The LASSOs were trained on each outcome visit separately (months 1, 3, 12, 24, 60).

To develop the model, we further leveraged a class of ML algorithms called *decision trees* to learn meaningful subgroups of patients that share statistical similarities in their baseline characteristics, and second, to fit a TWL prediction model for each subgroup. In opposition to classical regression analysis, such algorithms are valuable for learning hierarchical stratification of patients from baseline characteristics and observed outcomes without supervision. For instance, *decision trees* can predict weight loss when they are trained on a heterogeneous cohort of different bariatric interventions such as RYGB, SG and AGB, according to the type of intervention as well as using other variables such as the age at intervention, BMI and other clinical features.

For calibration, we used LASSO-extracted features as input for the Classification and Regression Trees (CART) algorithm. CART is a non-parametric statistical method used for prediction analysis of categorical (classification) and continuous (regression) response variables and for explanatory variables which may consist of nominal, ordinal, or continuous features. It is designed to create a sequence of binary choices according to attribute categories for categorical features, or based on a threshold value for numerical features. The CART algorithm recursively decides which attribute, category, or threshold value to use, and in which order, by maximizing the remaining statistical dispersion (variance) between the 2 subtrees that emerge from this choice. CART identifies groups of patients who share the same binary choices, and fits a single outcome prediction for each cluster. As weight loss outcomes are continuous, CART produced here a regression tree. A workflow diagram of the machine learning process is displayed in Appendix Figure 3A.

The algorithm was calibrated on the training subset of the training cohort. We further compared the predicted TWL to the observed outcomes of patients in the testing subset of the training cohort (internal validation).

Appendix 5: Comparison with other methods

In parallel with our LASSO+CART-based model, we also explored other methods for prediction: first we tried simple linear regression, based on the 7 selected attributes.

Furthermore, we tried to take into account the longitudinal nature of the data by using mixed effect model taking into account time with different functions (power functions, restricted cubic splines,...), with random intercept and/or slopes, to model the predicted weight loss after bariatric surgery from the preoperative data. The best mixed effect model uses the seven selected attributes, a restricted cubic splines to model time after surgery, as well as a random intercept and slope.

CARTs can achieve variable selection by themselves using a pruning mechanism (e.g. cost complexity pruning). Thus, we tried to train CART with all baseline variables and then pruning the resulting trees; it produced similar trees to the LASSO+CART approach (in particular all trees started by splitting on the type of intervention) and selected a total of six variables: weight, height, age, T2D status and duration and type of intervention. In terms of selected features, the only difference between LASSO and CART was the smoking history.

Lastly, because decision trees are typically weak predictors and that bagging/random forests (RFs) are a natural step to try and improve accuracy, at the cost of less interpretability, we tried RF on our training set, using the RandomForestRegressor from the *scikit-learn* library in Python and tuning hyperparameters using 3-fold cross-validation. These hyperparameters included the maximum depth, maximum number of features, minimum number of samples at a leaf node, number of tree predictors and cost-complexity pruning parameter.

Appendix Figure 6 displays the results of these comparisons.

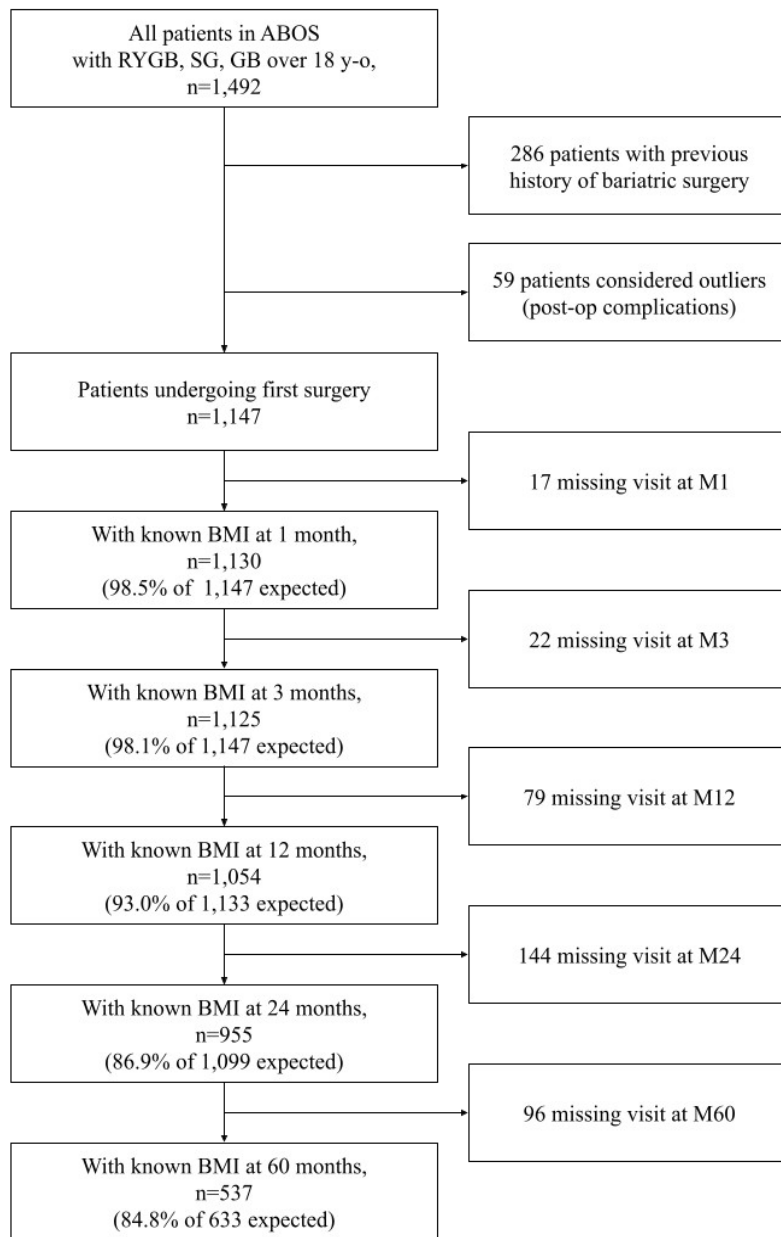
As results, linear models, mixed effect models and CART with pruning are underperforming compared to CART decision trees, in terms of MAD error, whereas RFs resulted in MAD slightly smaller than CARTs, but not enough to justify the loss of interpretability (in the sense that confidence intervals largely overlapped).

Appendix 6: Comparison with previously published models

We analyzed the results from previously published models^{6, 7, 8, 9, 10, 11, 12} between January 2000 and December 2021, identified using the following search terms: bariatric surgery, postoperative weight loss, weight loss prediction, and prediction model. We included English-language studies that investigated RYGB, SG, and AGB and used a prospective or retrospective design.

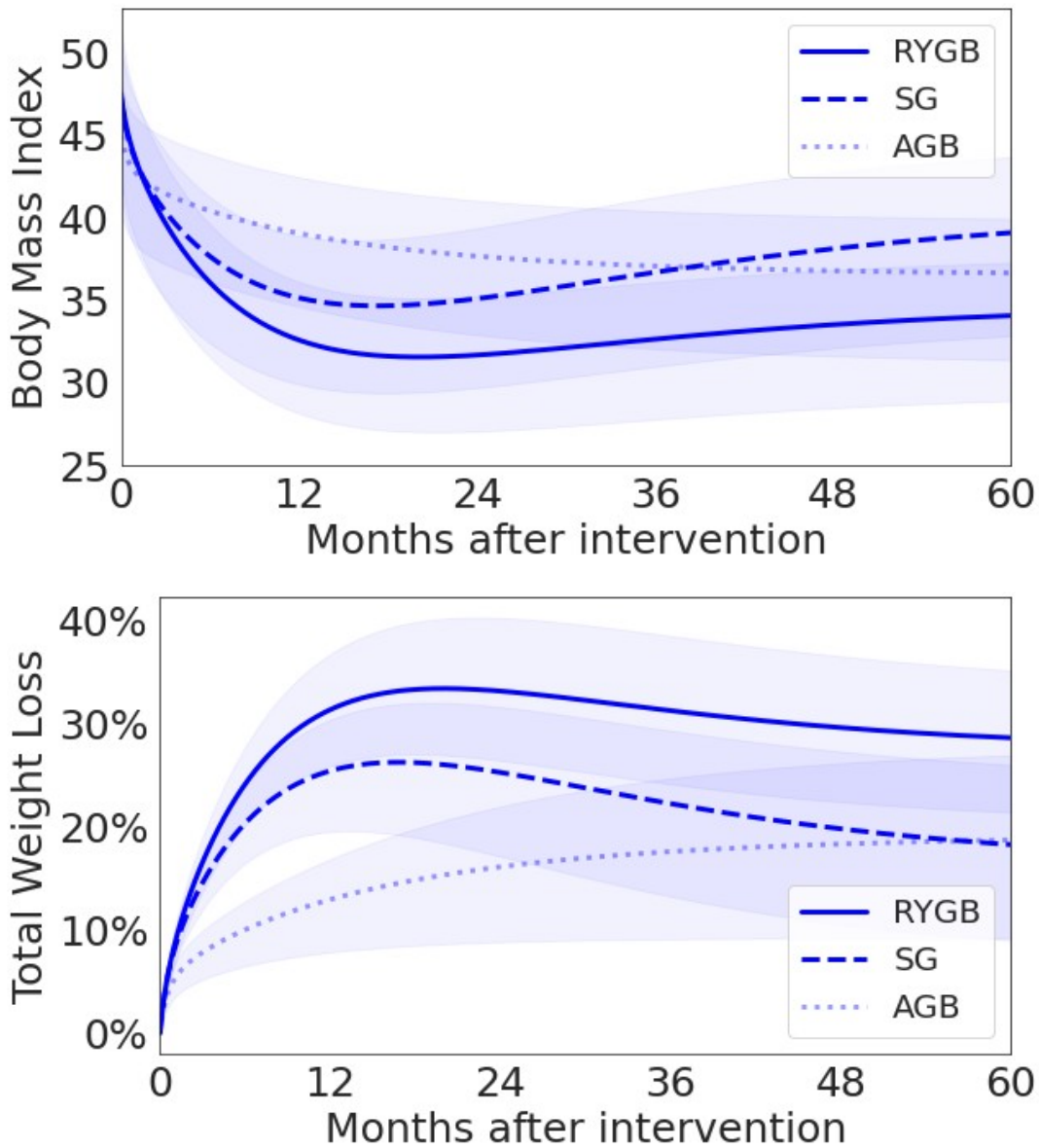
The identified models were implemented at the appropriate postoperative times to predict the TWL of corresponding patients from the training cohort, allowing the calculation of MAD and RMSE as previously described. In contrast with our tree-based model, which can simultaneously predict weight loss at different dates and across multiple types of bariatric surgery, the existing models were restricted to a single outcome date and a single type of intervention. For a fair comparison, these existing models were compared with corresponding subgroup of ABOS cohort according to type of surgery and outcome date.

Appendix Figure 1: ABOS patient flowchart

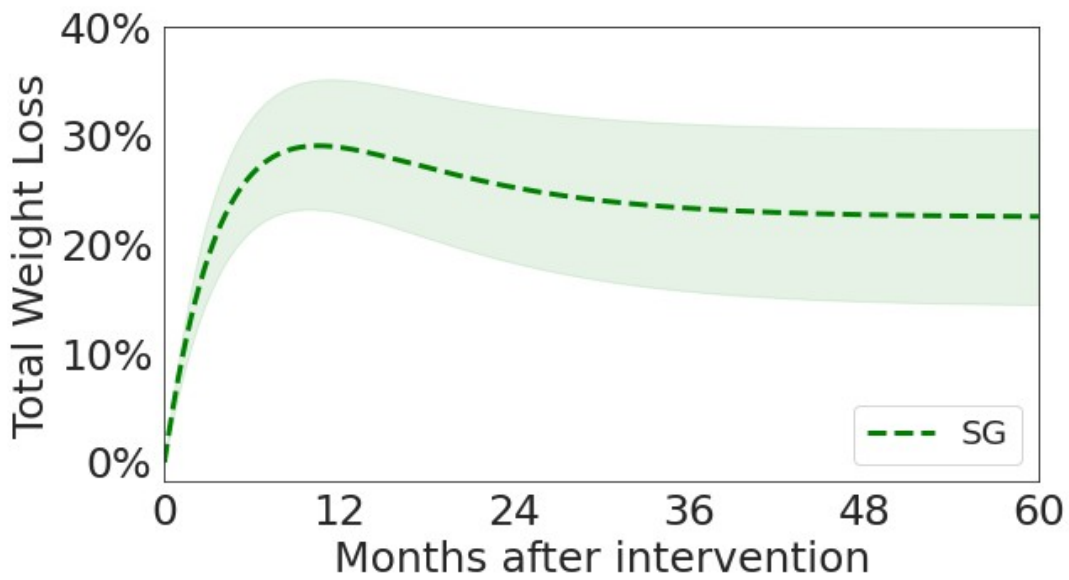
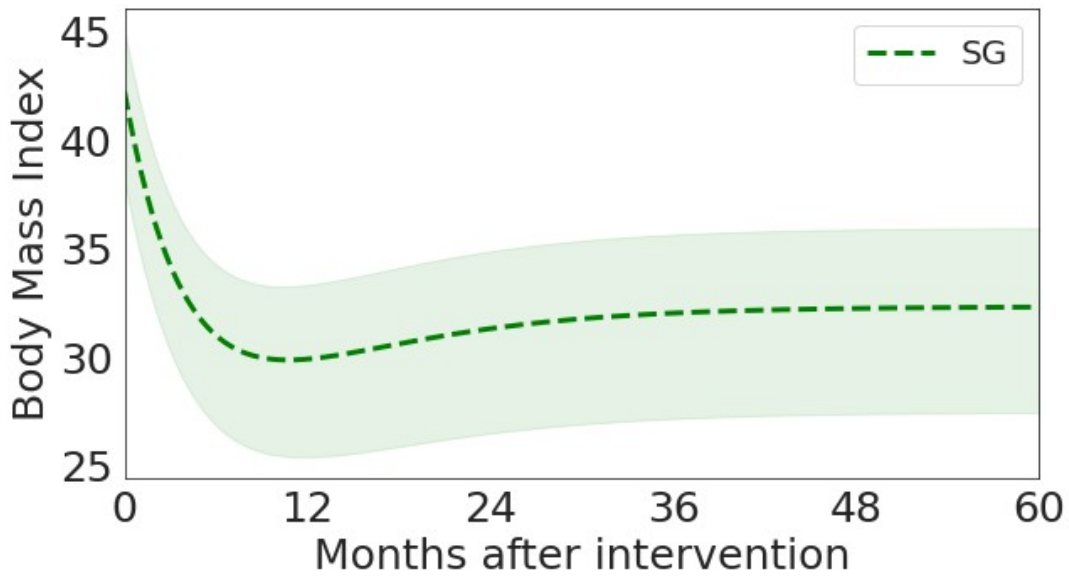


Expected: number of ABOS patients for which a visit is expected to be recorded in ABOS. For instance, patients who underwent surgery six months prior to the ABOS extraction date cannot have a M12 visit recorded in the database and thus do not count as expected. We used the following cutoff durations after intervention to compute the expected: 20 days at M1, 65 days at M3, 300 days at M12, 630 days at M24. For M60, we extended the cutoff to 6.5 years to reflect the postponed visits during the COVID-19 pandemic.

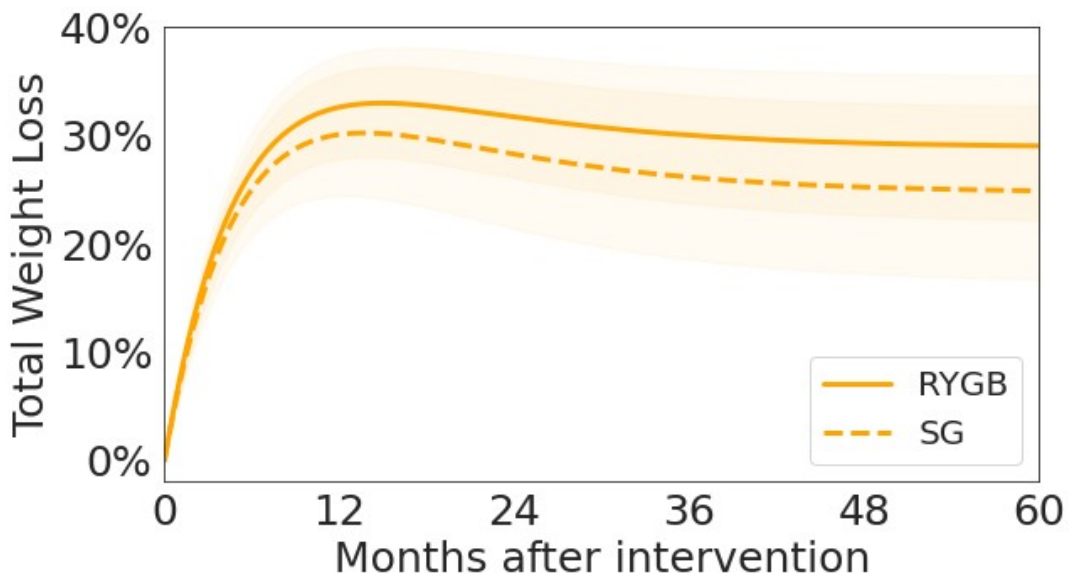
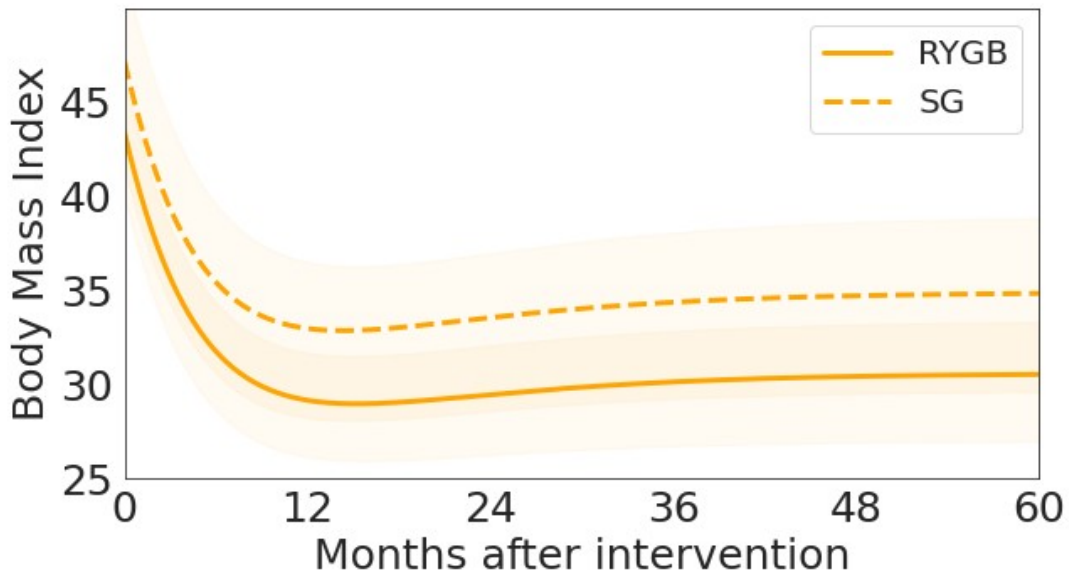
Appendix Figure 2: Smoothed observed median BMI and TWL trajectories and corresponding IQR, for A) ABOS cohort, B) BAREVAL cohort, C) NOK cohort, D) SGH cohort, E) SOS cohort, F) PRECOS cohort, G) Roma cohort, H) Lyon cohort, I) COD cohort, J) Mexico cohort, K) SleevePass RCT cohort and L) SM-BOSS RCT cohort



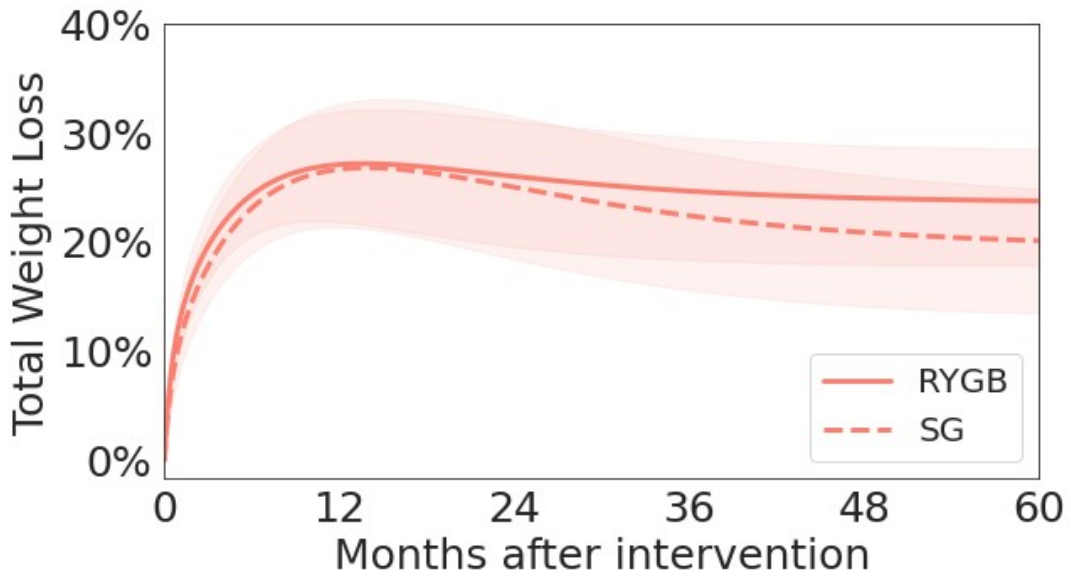
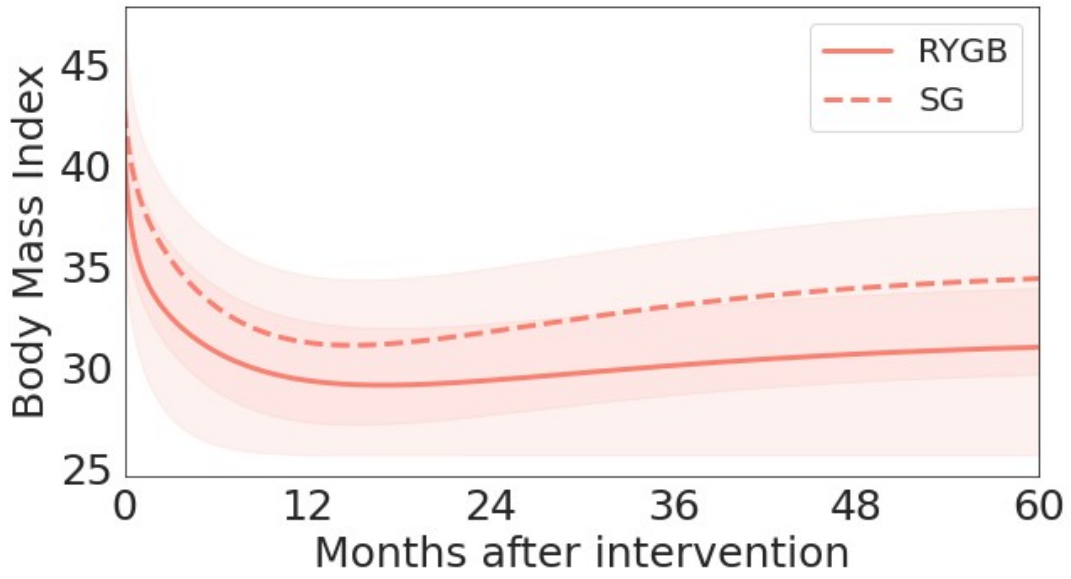
A. ABOS cohort, all interventions.



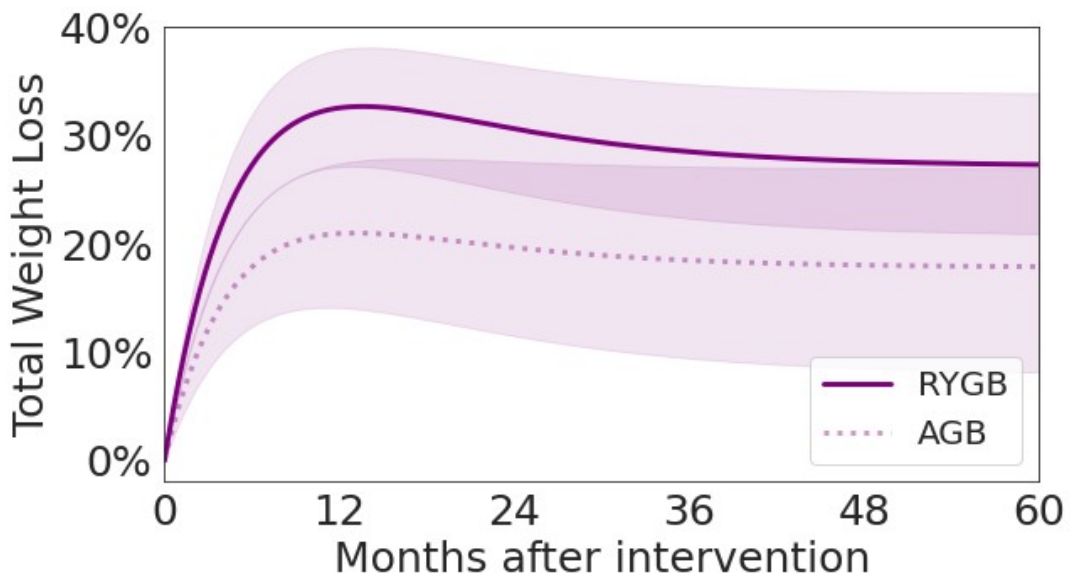
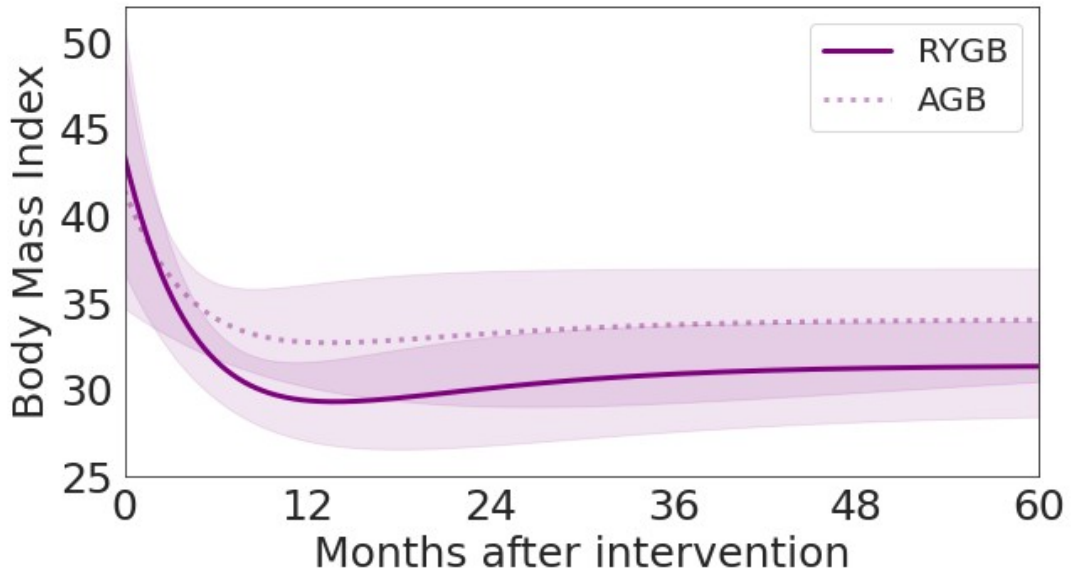
B. BAREVAL cohort, sleeve gastrectomy.



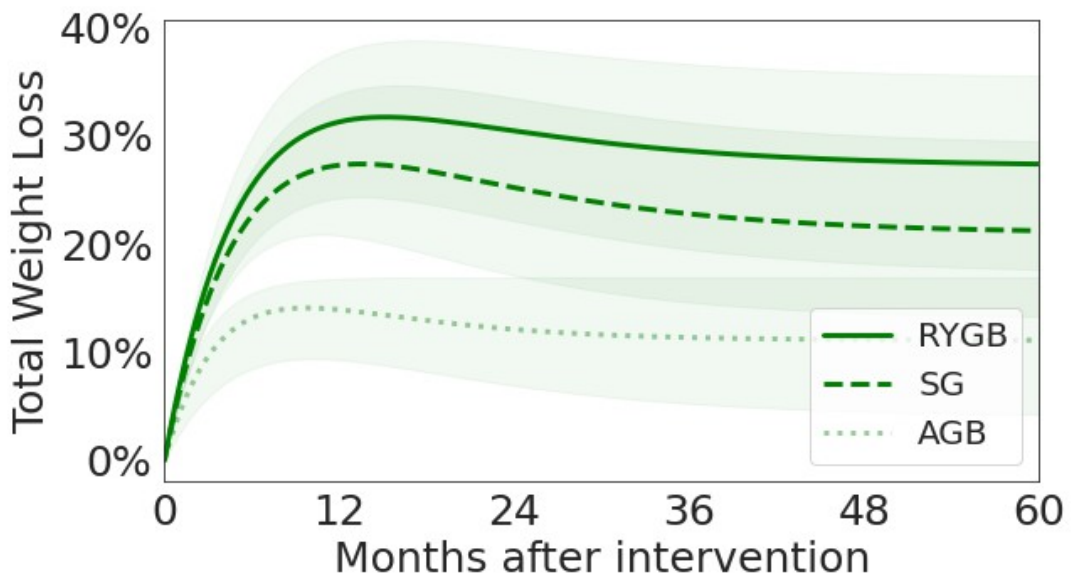
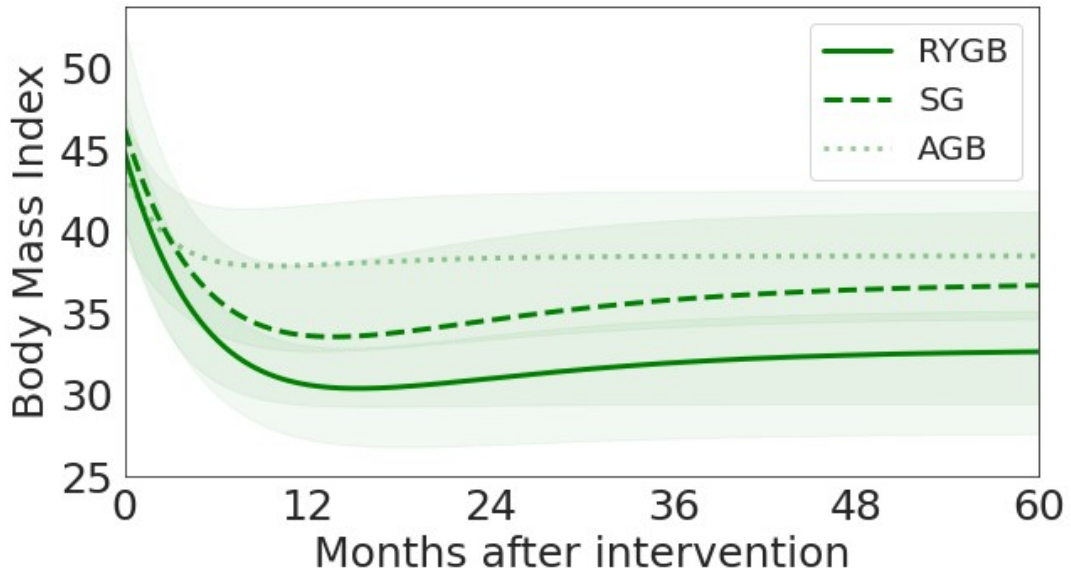
C. NOK cohort, all interventions.



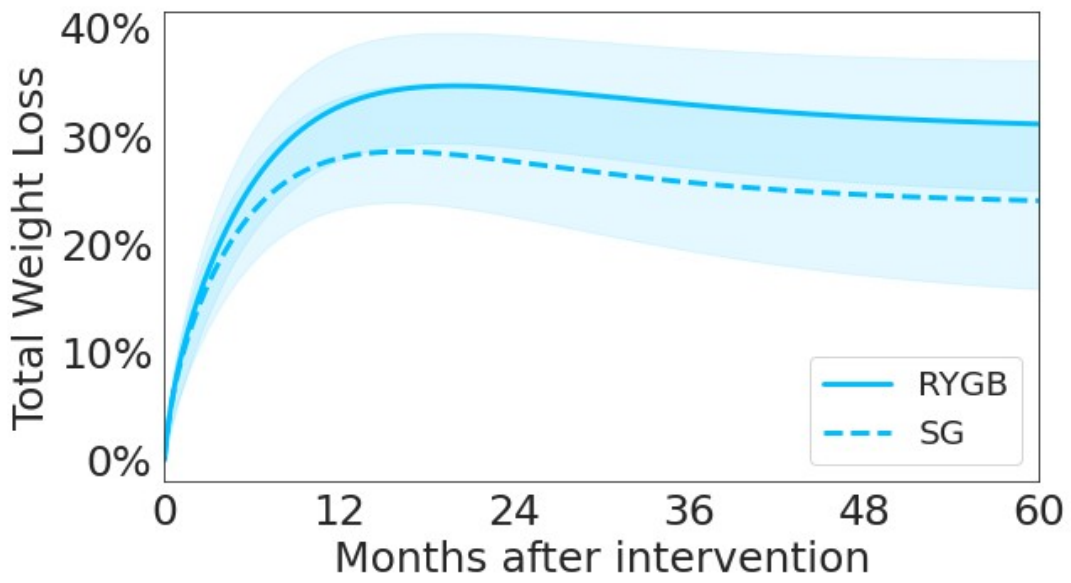
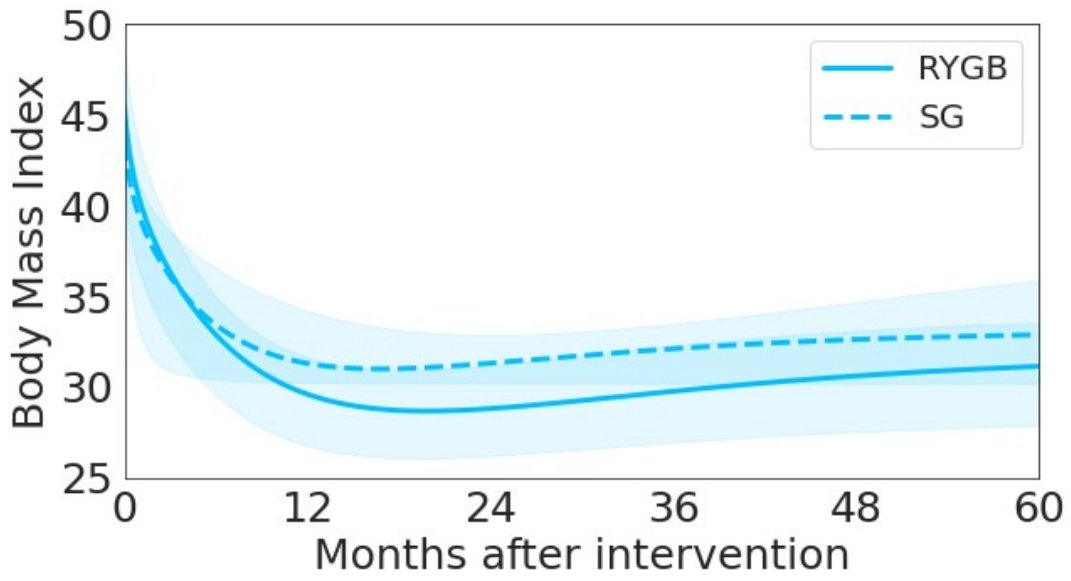
D. SGH, Roux-en-Y gastric bypass and sleeve gastrectomy.



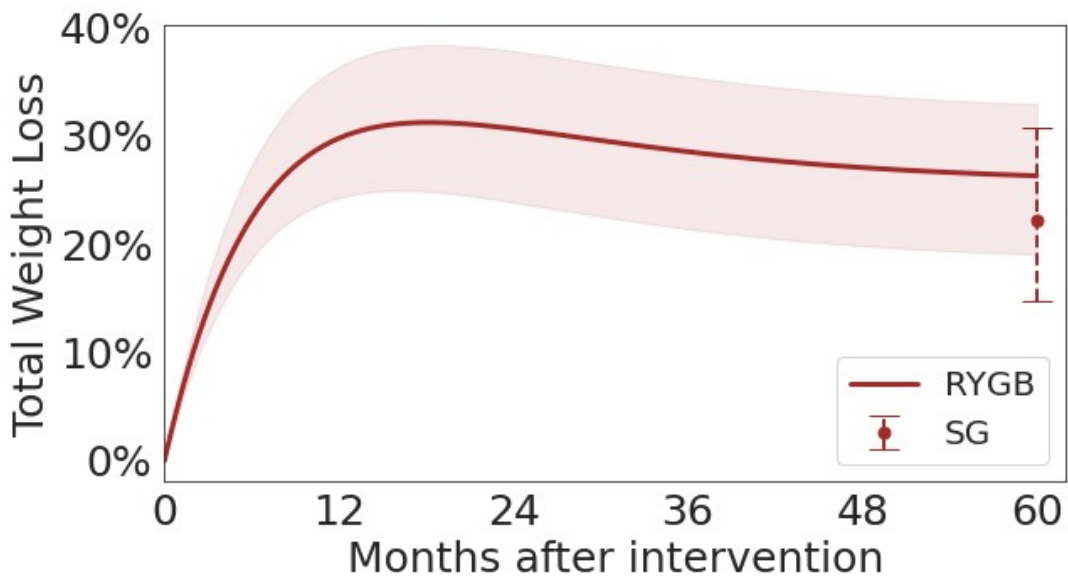
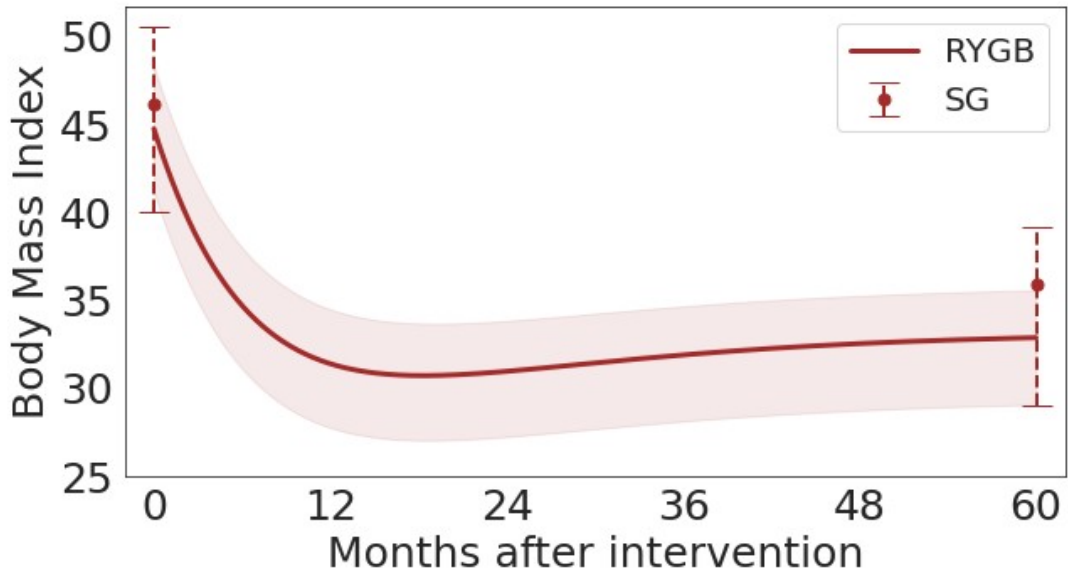
E. SOS cohort, Roux-en-Y gastric bypass and gastric banding.



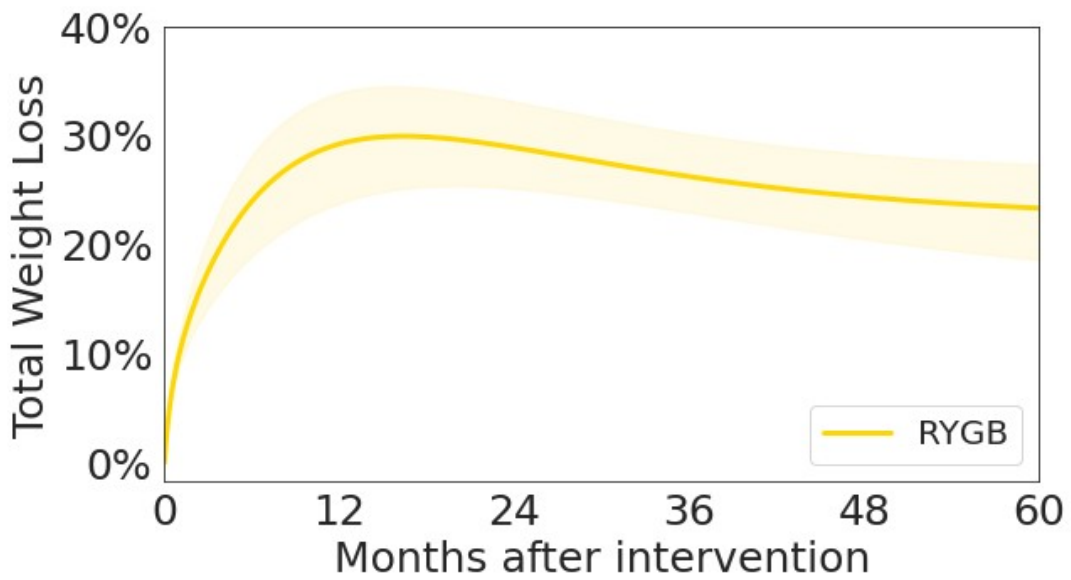
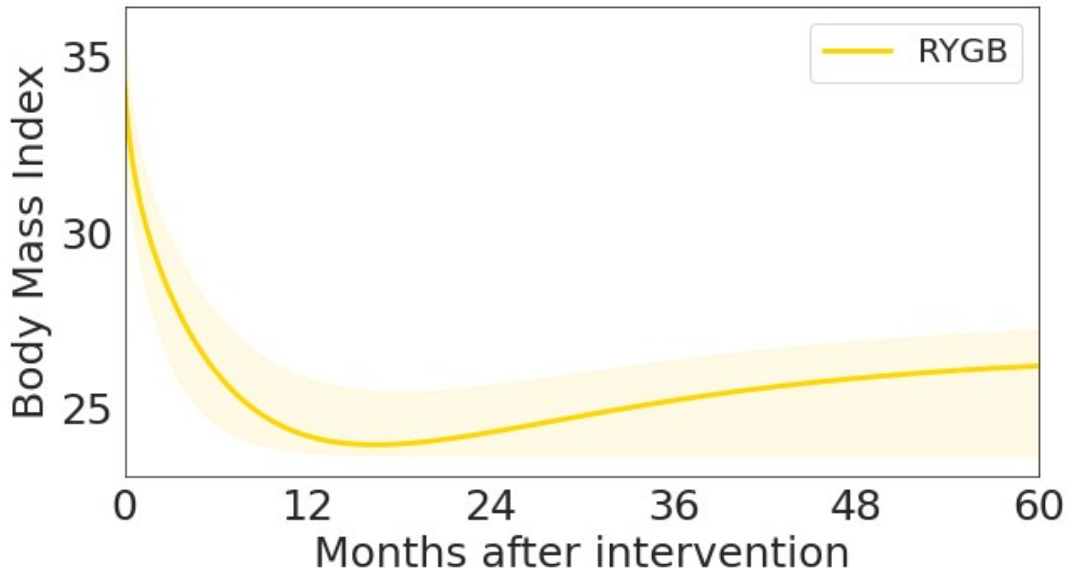
F. PRECOS cohort, all interventions.



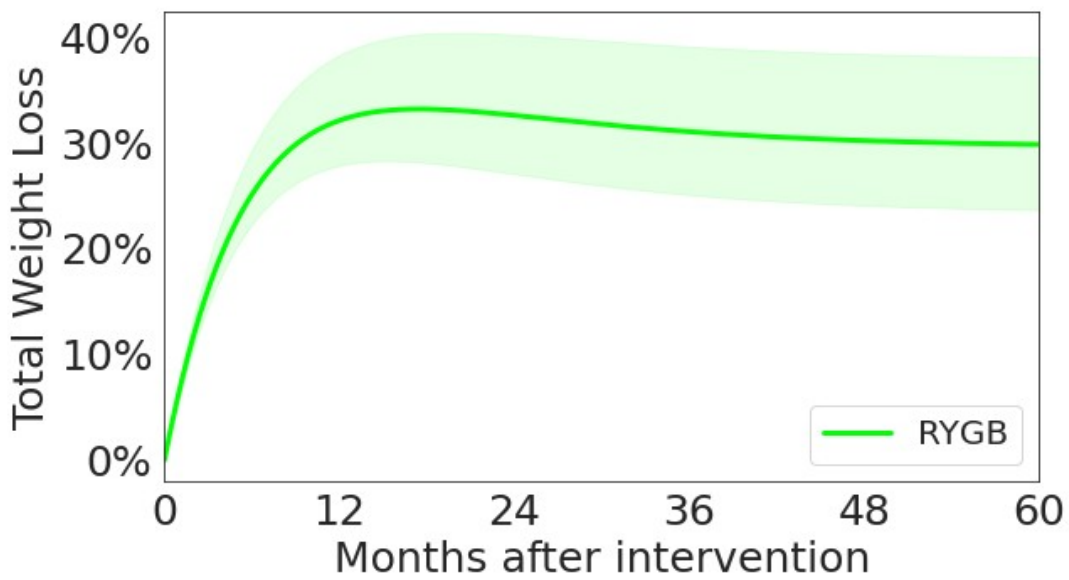
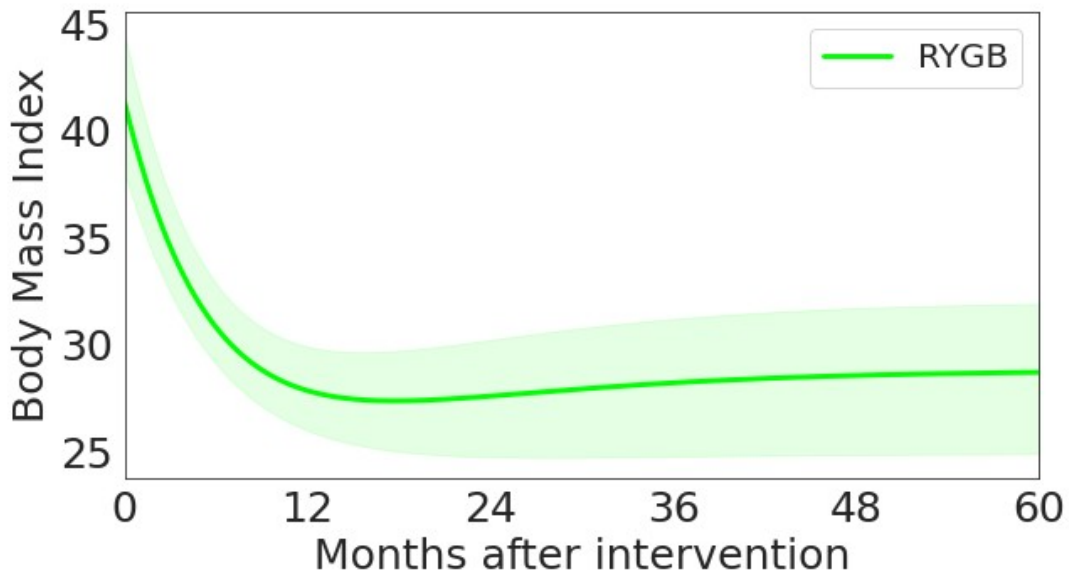
G. Roma cohort, Roux-en-Y gastric bypass and sleeve gastrectomy.



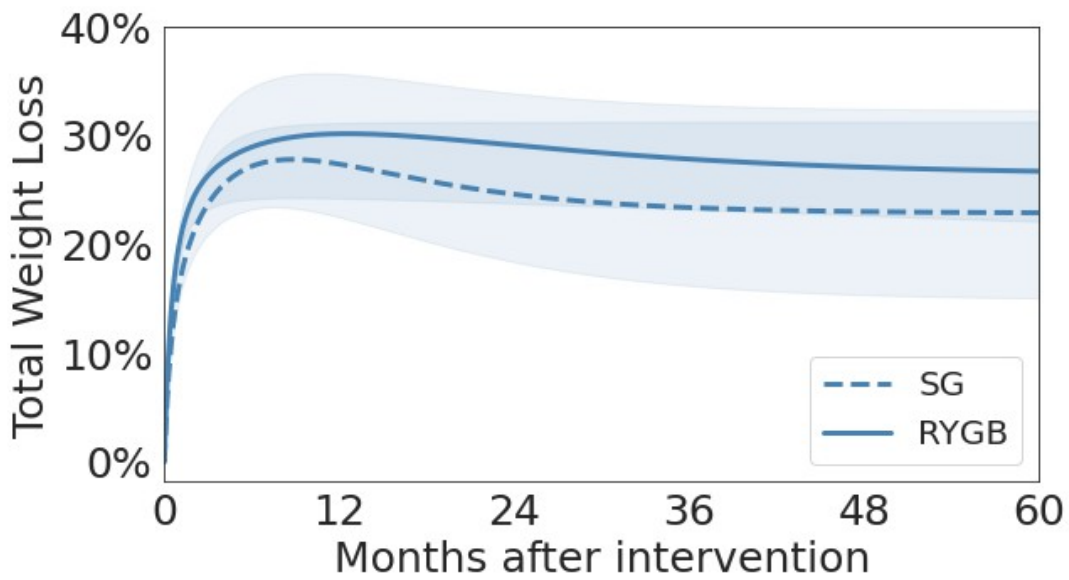
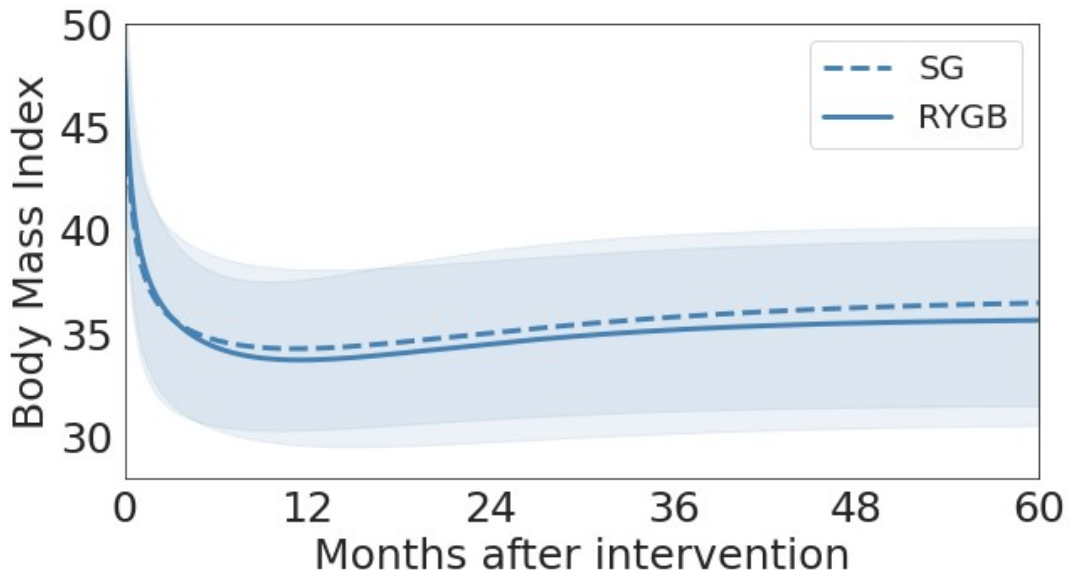
H. Lyon cohort, Roux-en-Y gastric bypass and sleeve gastrectomy (only baseline and 5 years sleeve data).



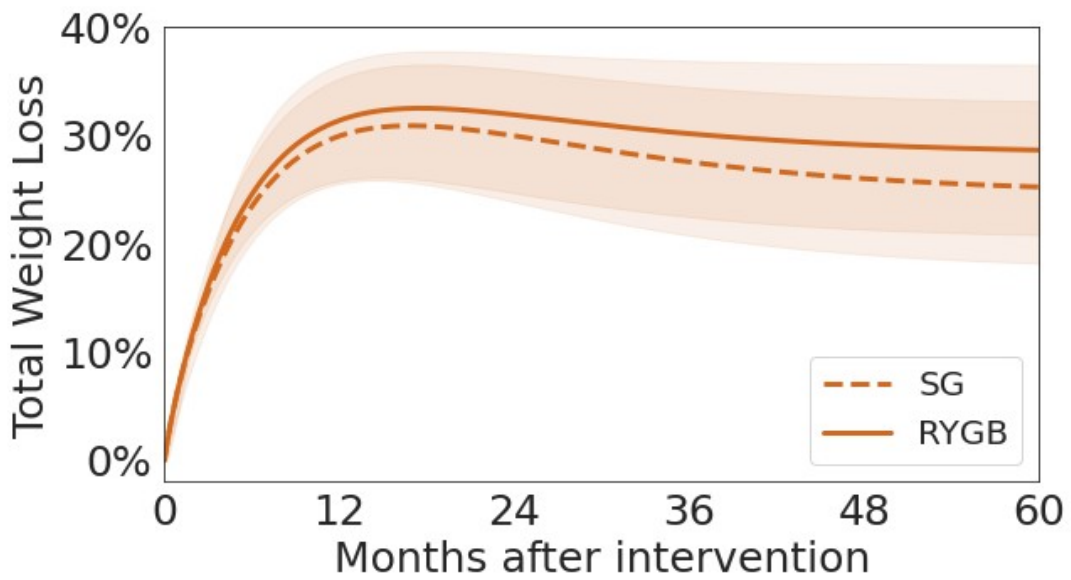
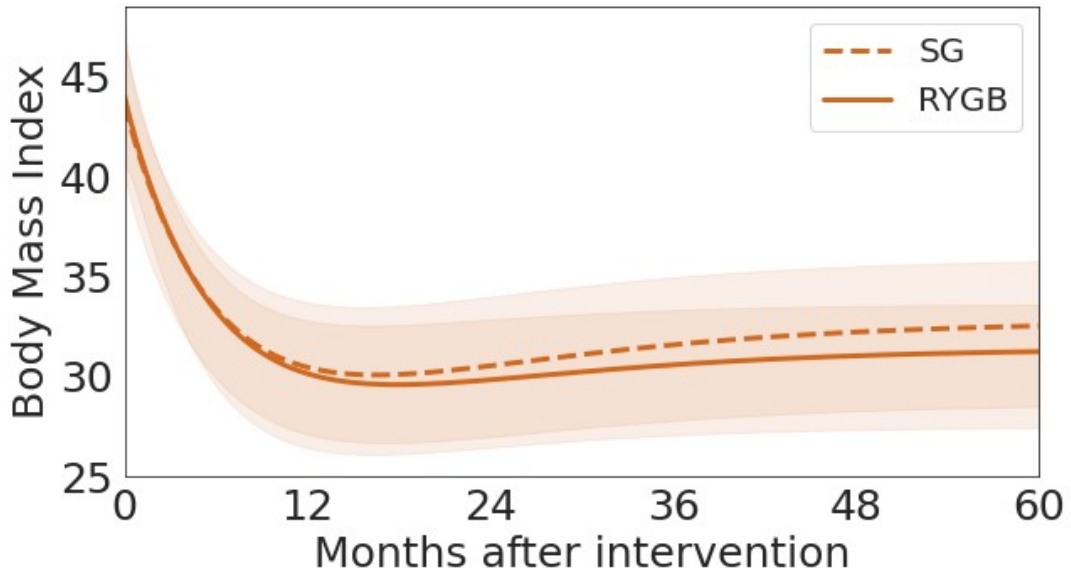
I. COD, São Paulo, Roux-en-Y gastric bypass, patients with type 2 diabetes only.



J. Mexico, Roux-en-Y gastric bypass.



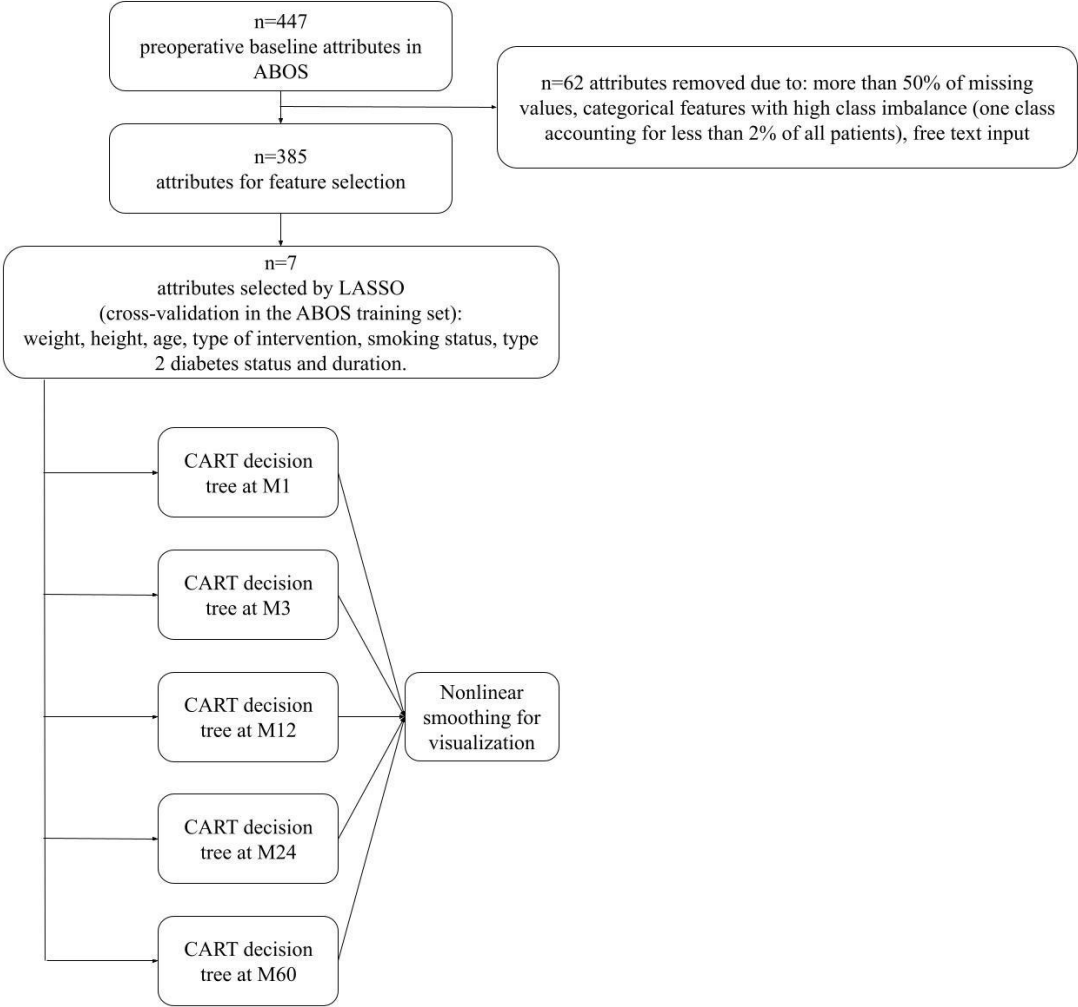
K. SleevePass (RCT), Roux-en-Y gastric bypass and sleeve gastrectomy.



L. SM-BOSS (RCT), Roux-en-Y gastric bypass and sleeve gastrectomy.

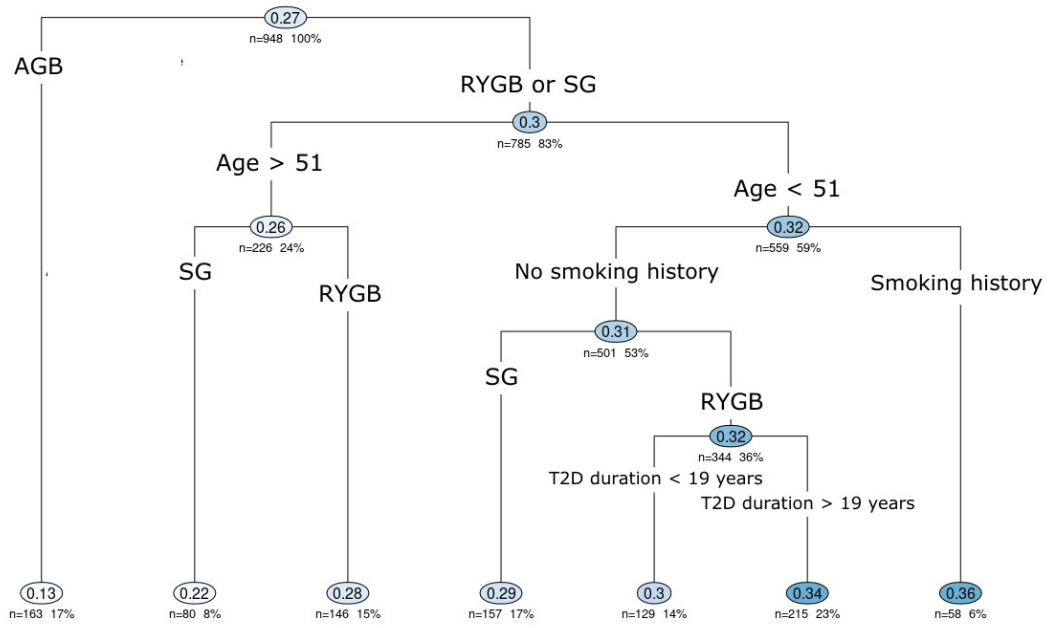
Appendix Figure 3: Machine learning pipeline diagram in A), and all decision trees for prediction of %TWL at B) month 12, C) month 24 and D) month 60

A)



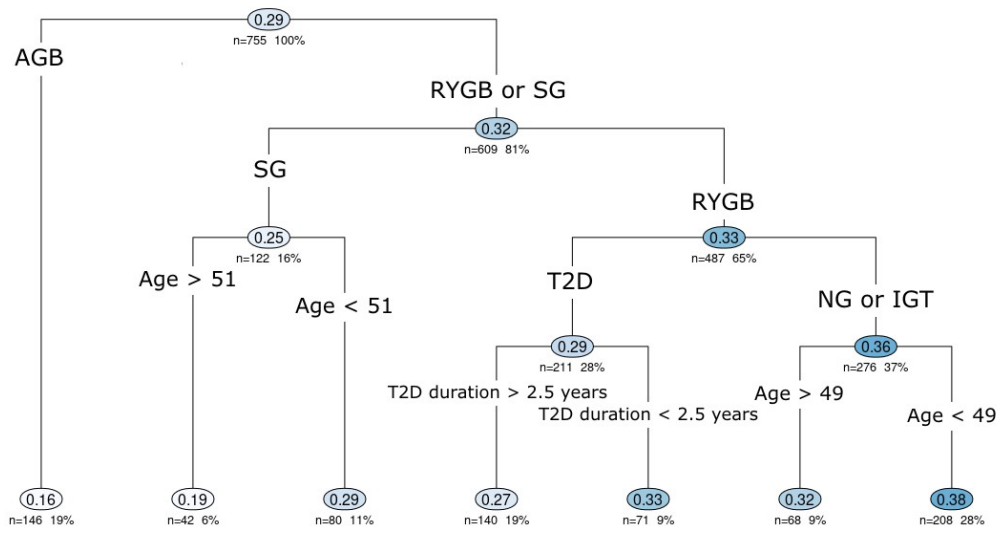
B)

TWL Prediction at M12



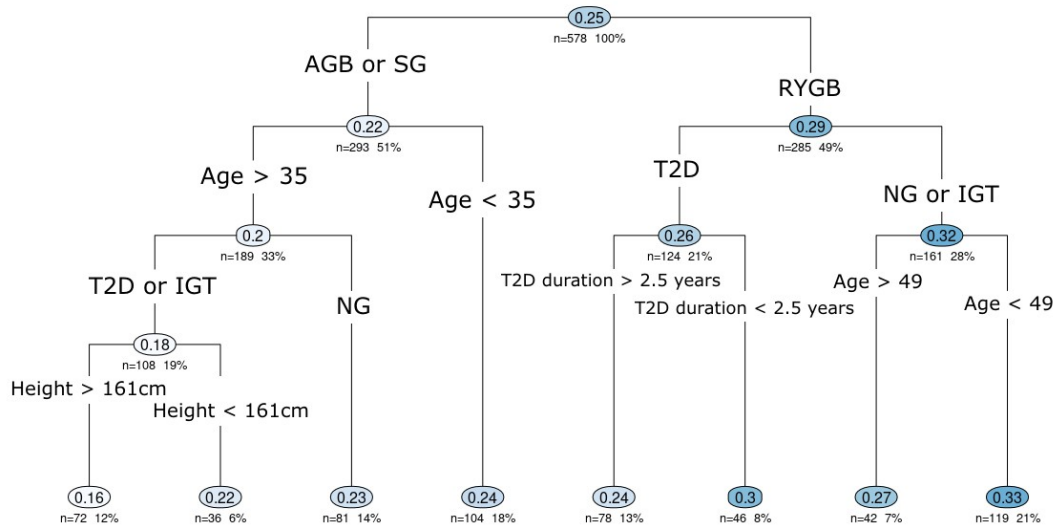
C)

TWL Prediction at M24



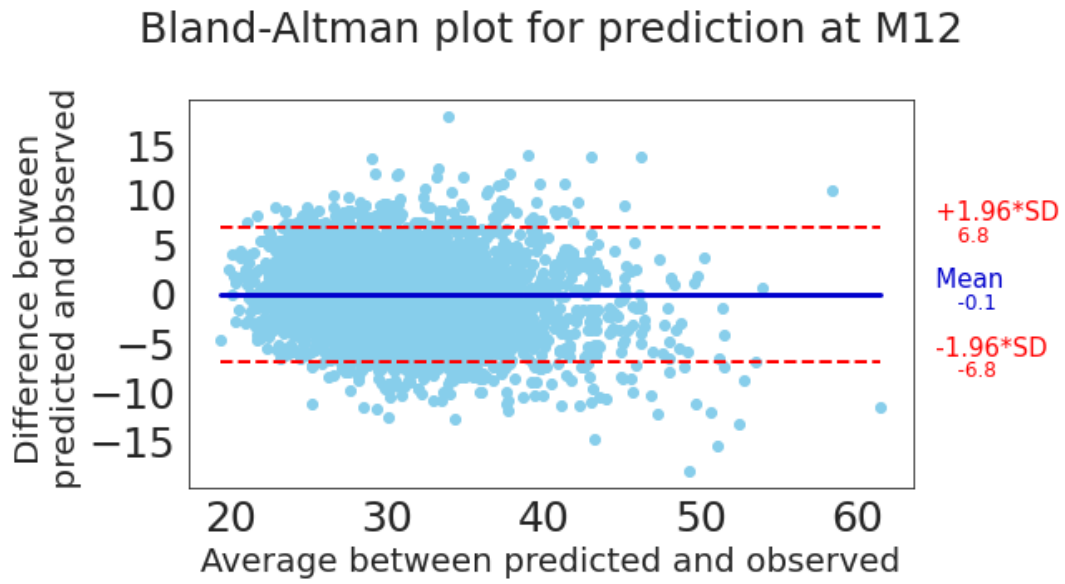
D)

TWL Prediction at M60



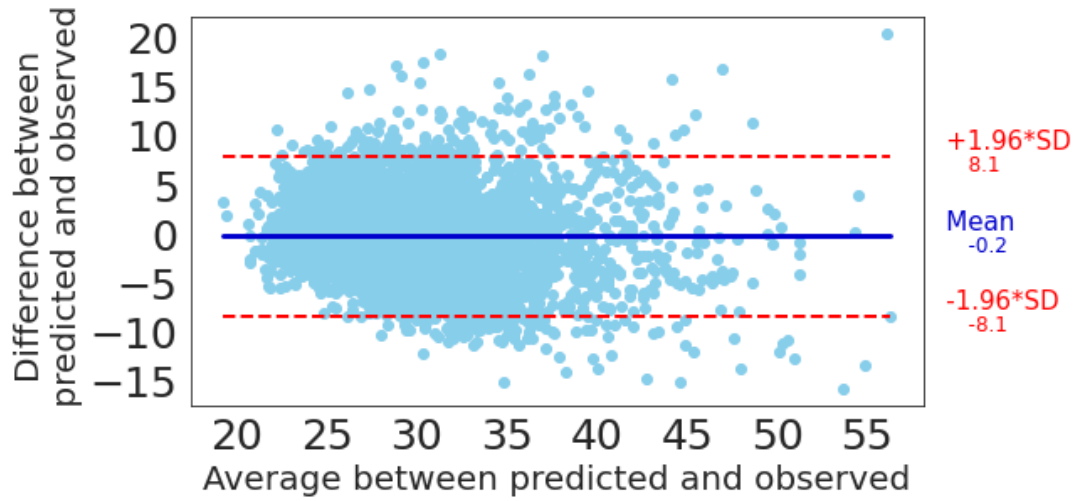
Appendix Figure 4: Bland-Altman plots at A) M12, B) M24 and C) M60 for all interventions

A) At M12



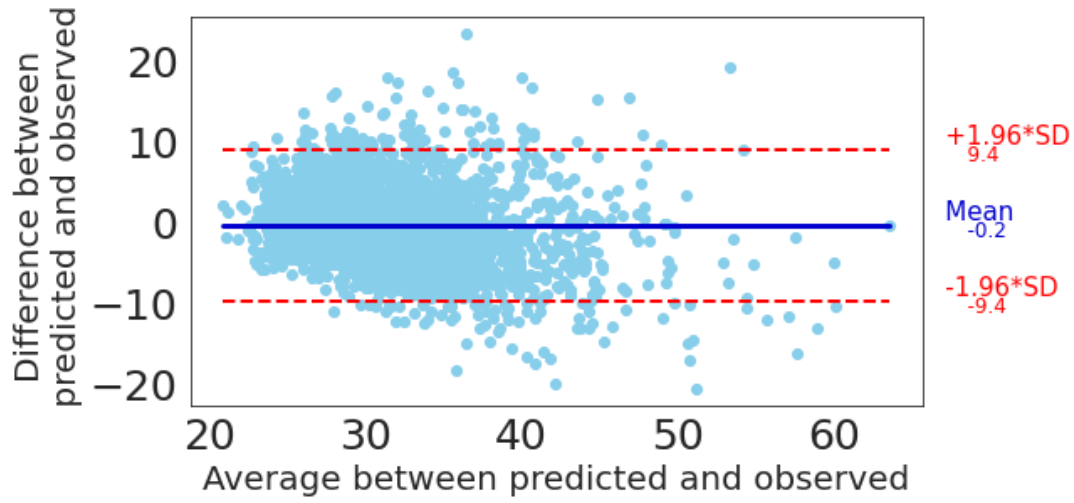
B) At M24

Bland-Altman plot for prediction at M24

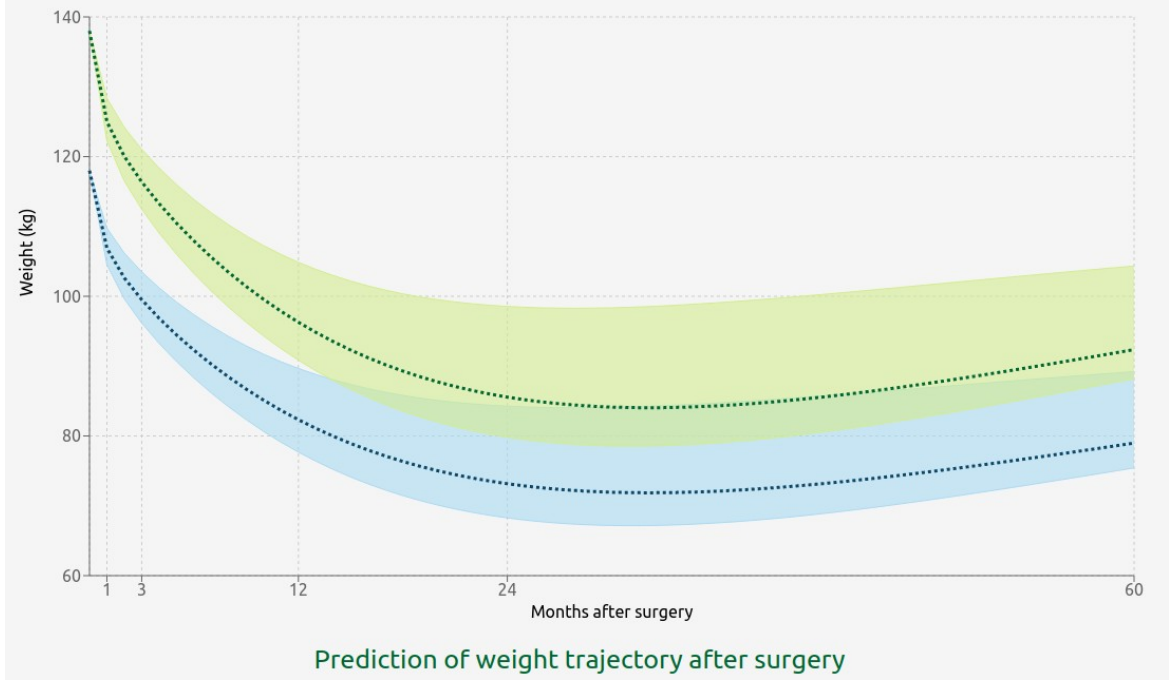


C) At M60

Bland-Altman plot for prediction at M60

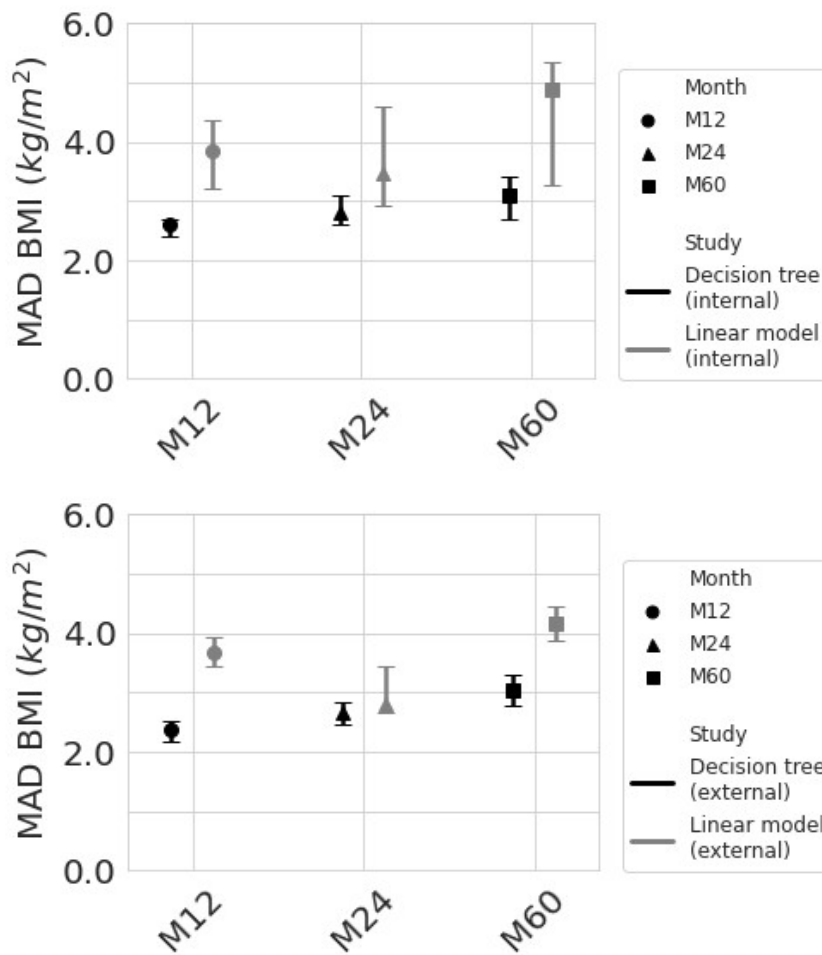


Appendix Figure 5: Predicted trajectory and IQR of prediction error, for a 40 years old patient with height 1.70m, non-smoker, with prediabetes, undergoing RYGB with preoperative weight 118kg (blue) and 138kg (green), corresponding to the 1st and 2nd tercile of weights in the ABOS cohort

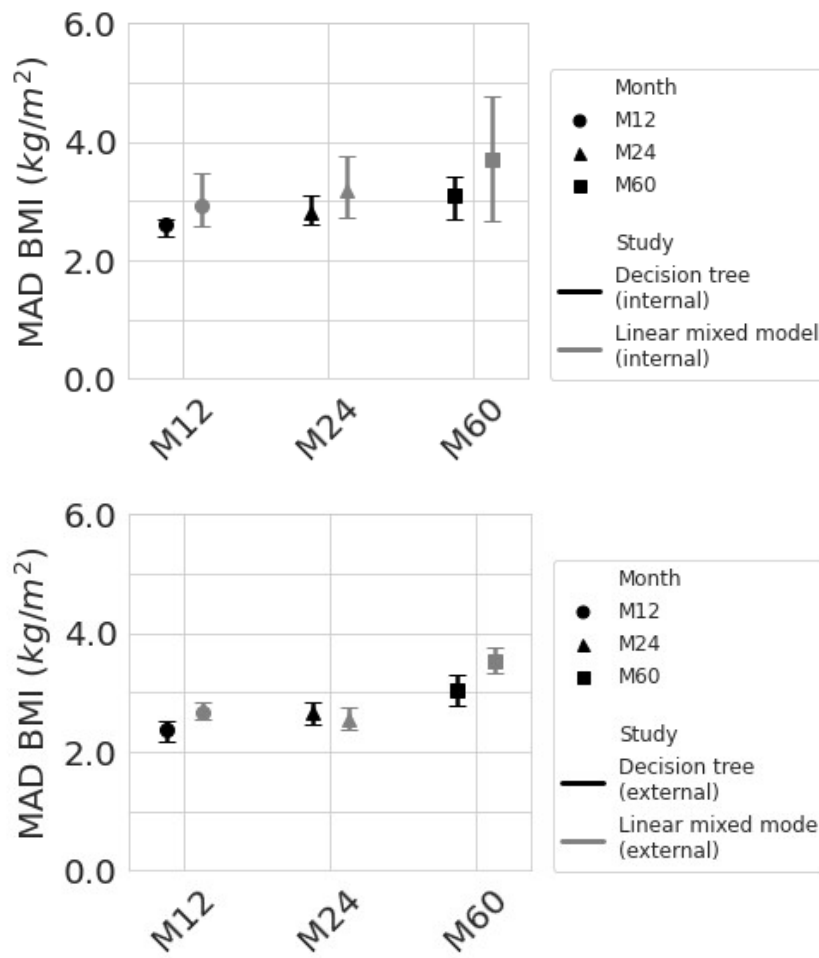


Appendix Figure 6: Accuracy of BMI prediction, in terms of MAD error, on internal test data as well as (available) external test data, comparing decision trees with A) linear regression, B) longitudinal model (mixed effect model), C) random forest, and D) CART decision tree with pruning (without LASSO)

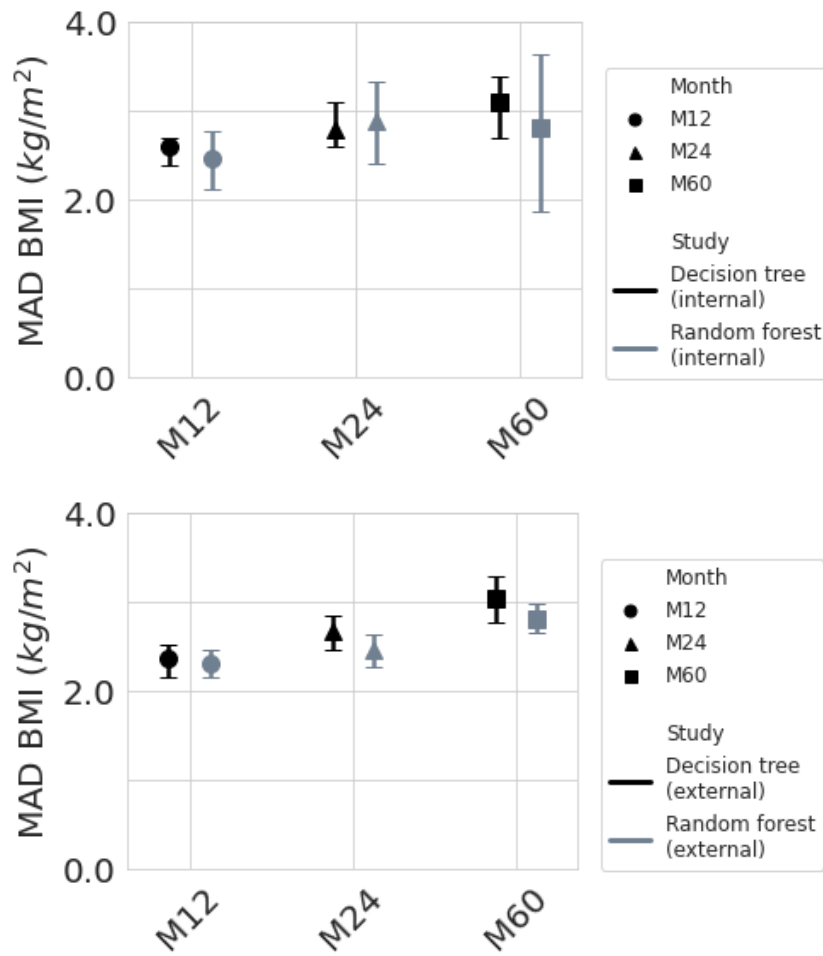
A) linear regressions on internal (top) and external (bottom) test data



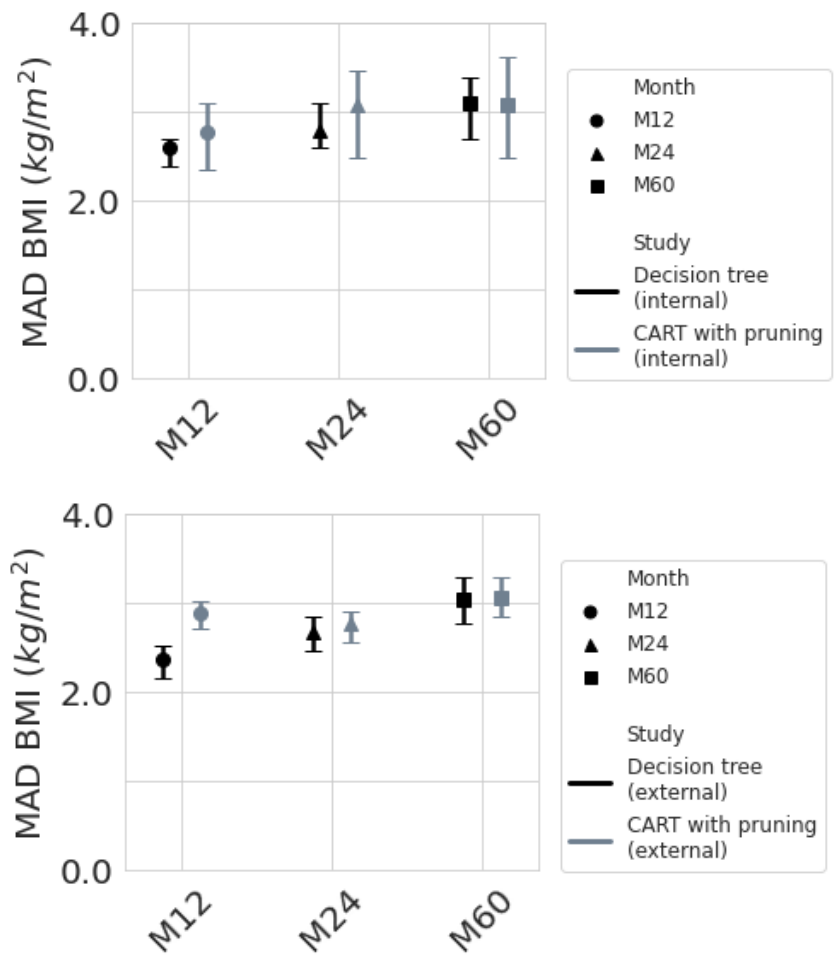
B) longitudinal model (mixed effect model) on internal (top) and external (bottom) test data



C) random forest on internal (top) and external (bottom) test data



D) CART decision tree with pruning (without LASSO) on internal (top) and external (bottom) test data



Appendix Table 1: Validation of existing prediction models on the ABOS cohort for corresponding type of intervention and date of prediction

Model	Type of operation	Month of prediction	ABOS sample size	Mean observed BMI (SD)	Mean predicted BMI (SD)	RMSE in kg/m ² (95% CI)	Normalized RMSE in % (95% CI)
Cottam ⁶	Sleeve gastrectomy	12	193	35.2 (7.9)	33.2 (5.9)	4.7 (4.3; 5.2)	13.3 (11.6; 14.9)
Goulard ⁷	Sleeve gastrectomy	12	193	35.2 (7.9)	31.5 (5.6)	5.6 (5.2; 6.2)	16.0 (14.3; 18.0)
Janik ⁸	Sleeve gastrectomy	12	193	35.2 (7.9)	31.0 (7.7)	6.1 (5.6; 6.6)	17.4 (15.7; 19.0)
Velazquez ⁹	RYGB	12	652	32.5 (6.1)	31.9 (4.5)	3.8 (3.6; 4.0)	11.8 (11.0; 12.6)
Wise ¹⁰	RYGB	12	652	31.4 (5.9)	30.2 (3.9)	3.9 (3.5; 4.3)	12.4 (11.3; 13.6)
Seysssel ¹¹	RYGB	12	652	32.5 (6.1)	31.5 (4.3)	3.8 (3.5; 4.0)	11.7 (10.8; 12.5)
Baltasar1 ¹²	Sleeve gastrectomy	24	162	35.1 (7.9)	31.1 (3.8)	7.0 (6.3; 7.6)	20.0 (17.7; 22.2)
Baltasar2 ¹²	Any	24	959	33.5 (7.3)	30.5 (2.9)	6.8 (6.4; 7.0)	20.2 (19.0; 21.3)
Baltasar3 ¹²	RYGB	24	606	31.8 (6.5)	30.7 (3.0)	5.2 (4.8; 5.5)	16.3 (15.1; 17.5)
Velazquez2 ⁹	RYGB	24	606	31.8 (6.5)	31.3 (3.9)	4.9 (4.6; 5.2)	15.5 (14.5; 16.5)
Seysssel2 ¹¹	RYGB	24	606	31.8 (6.5)	31.0 (3.6)	4.9 (4.7; 5.2)	15.5 (14.3; 16.5)
Seysssel3 ¹¹	RYGB	60	348	34.1 (6.9)	33.3 (3.7)	5.4 (5.0; 5.7)	15.8 (14.3; 17.3)

RMSE: root mean squared errors (measures prediction bias and standard deviation in BMI points; the lower, the more accurate).

Appendix Table 2: Comparison of literature models versus our model on the ABOS testing subset

Model	Type of operation	Month of prediction	ABOS (test) sample size	RMSE in kg/m ² of literature model (95% CI)	RMSE in kg/m ² of our model (95% CI)	p-value (test of forecast accuracy)
Cottam ⁶	Sleeve gastrectomy	12	41	5.3 (3.8; 6.7)	5.1 (4.1; 6.1)	0.22
Goulard ⁷	Sleeve gastrectomy	12	41	6.3 (4.8; 7.8)	5.1 (4.0; 6.0)	0.011
Janik ⁸	Sleeve gastrectomy	12	41	6.3 (5.6; 7.2)	5.1 (4.1; 6.1)	0.011
Velazquez ⁹	RYGB	12	124	4.0 (3.5; 4.5)	3.8 (3.3; 4.2)	0.13
Wise ¹⁰	RYGB	12	60	3.5 (2.9; 3.9)	3.6 (3.1; 4.1)	0.64
Seyssel1 ¹¹	RYGB	12	124	4.1 (3.5; 4.8)	3.8 (3.3; 4.2)	0.11
Baltasar1 ¹²	Sleeve gastrectomy	24	33	6.6 (5.6; 7.6)	5.5 (4.3; 6.7)	0.099
Baltasar2 ¹²	Any	24	187	6.8 (6.0; 7.4)	4.7 (4.0; 5.2)	<0.0001
Baltasar3 ¹²	RYGB	24	111	5.1 (4.3; 5.9)	4.3 (3.4; 5.1)	0.009
Velazquez2 ⁹	RYGB	24	111	4.8 (4.1; 5.6)	4.2 (3.7; 5.0)	0.002
Seyssel2 ¹¹	RYGB	24	111	4.9 (4.1; 5.7)	4.3 (3.6; 4.9)	0.007
Seyssel3 ¹¹	RYGB	60	55	4.3 (3.7; 4.8)	4.0 (3.0; 4.7)	0.12

RMSE: root mean squared errors (measures prediction bias and standard deviation in BMI points; the lower, the more accurate). P-values correspond to the Diebold-Mariano test.¹³

Appendix Table 3: Model-based means for primary and secondary outcomes after sleeve gastrectomy or Roux-en-Y gastric bypass at 1, 3 and 5 years (similar format as Table 2 from PMID: 33640917)

	Model-based mean in operations	M12	M36	M60	p
Predicted %EBMIL					
Sleeve Gastrectomy	58.0 (53.3; 62.8)	66.4 (59.9; 72.9)	56.4 (51.5; 61.2)	46.3 (38.5; 54.1)	
Gastric bypass	72.7 (67.8; 77.5)	76.3 (69.8; 82.7)	72.0 (67.2; 76.9)	67.8 (60.0; 75.6)	
Difference	-14.7 (-15.7; -13.7)	-9.8 (-11.2; -8.5)	-15.7 (-16.7; -14.7)	-21.5 (-23.2; -19.9)	< 0.0001**
P-value of difference		< 0.0001	< 0.0001	< 0.0001	
Predicted TWL					
Sleeve Gastrectomy	25.2 (23.4; 26.9)	28.7 (26.4; 31.2)	24.4 (22.6; 26.2)	20.1 (17.2; 23.0)	
Gastric bypass	31.8 (30.0; 33.5)	33.3 (30.9; 35.7)	31.4 (29.7; 33.2)	29.6 (26.7; 32.5)	
Difference	-6.6 (-6.9; -6.2)	-4.5 (-5.0; -4.0)	-7.0 (-7.4; -6.7)	-9.5 (-10.1; -8.9)	< 0.0001**
P-value of difference		< 0.0001	< 0.0001	< 0.0001	

**main effect of operation.

Model-based difference in predicted % EBMIL (respectively, in predicted TWL) at 5 years between SG and RYGB is statistically significant: -21.5%, $p < 0.0001$ (respectively -9.5%, $p < 0.0001$).

By comparison, using the same model on observed data, the model-based difference in observed % EBMIL (respectively, in observed TWL) at 5 years between SG and RYGB is statistically significant: -8.1%, $p < 0.0001$ (respectively -3.4%, $p < 0.0001$).

Methods: We used mixed effect models to estimate model-based marginal means, and marginal contrasts (e.g. difference), between SG and RYGB, on the joint SleevePass and SM-Boss populations, based on predicted %EBMIL and predicted TWL. Models included operation, time and interaction between operation and time, age, baseline BMI and baseline

diabetes status as fixed factors, and trial as random factor. Sex and centers were not available, contrary to initial mixed effect model from ¹⁴.

Appendix Table 4: Baseline attributes by categories, before feature selection

I	Patient description
	Inclusion date 1 DD/MM/YYYY
	ABOS phenotyping groups : - 1: type 2 diabetes - 2: glucose intolerant - 3: normoglycemic with BMI \geq 35 - 4: control, normoglycemic and BMI \leq 27 2 - 5: control, normoglycemic and $27 < \text{BMI} < 35$
	3 Year of birth
	Gender - M: male 4 - F: female
	5 Geographical origin of the patient
	6 Geographical origin of the patient's father
	7 Geographical origin of paternal grandfather
	8 Geographical origin of paternal grandmother
	9 Geographical origin of the patient's mother
	10 Geographical origin of maternal grandfather
	11 Geographical origin of maternal grandmother
	Socio-Professional Category: - 1: farmers - 2: artisans - traders and entrepreneurs - 3: liberal professions - civil service executives - professors, scientific professions - information, arts and entertainment professions - administrative and commercial company executives - engineers and technical company executives - 4: school teacher - intermediate health and social work professions - clergy, religious - administrative intermediary professions of the public service - administrative and commercial intermediary profession of companies - technician - foremen, supervisors - 5: civilian employees - public service agents - police - military - administrative employees - 6: workers or drivers - 7: retirees 12 - 8: no professional activity
	13 Height (cm)
	14 Ideal weight (calculated automatically)
	15 Current age
	Year of birth 16 DD/MM/YYYY
	17 Job
	18 Postal code
	19 City
II	Patient history
II.1	<i>Tobacco</i>

	Tobacco consumption: - smoker - no smoking - former smoker
20	
21	If smoker or former smoker: number of years of tobacco consumption
22	If smoker: number of cigarettes per day
23	If former smoker: number of cigarettes per day
24	Smoking cessation age
25	Number of packets per year
<i>II.2</i>	<i>Alcohol</i>
	Consumption during meals - no - yes
26	
	Consumption between meals - no - yes
27	
28	Number of glasses per month
	Type 1 alcohol: the type of alcohol the patient consumes the most - beer - wine - aperitif 16 degrees - aperitif 40 degrees - did not answer (DNA) - does not drink alcohol
29	
	Type 2 alcohol: other type of alcohol consumed - beer - wine - aperitif 16 degrees - aperitif 40 degrees - did not answer (DNA) - does not drink alcohol
30	
<i>II.3</i>	<i>Food</i>
	Binge eating episodes - yes - no - did not answer (DNA)
31	
32	If binge eating: the number of episodes
	Frequency: - day - week - month - year
33	
<i>II.4</i>	<i>Women</i>
34	Number of pregnancies
	Period disorders: - yes - no - did not answer (DNA)
35	
	Infertility: - yes - no
36	

	- did not answer (DNA)
	Menopause: - yes - no
37	- did not answer (DNA)
38	Menopause age
	Hysterectomy: - yes - no
39	- did not answer (DNA)
40	Age of hysterectomy
	Hormone treatment - no - contraception
41	- menopause
42	Date of last period
	Current hormone treatments for contraception - yes - no
43	- did not answer (DNA)
	Current hormone treatments for menopause - yes - no
44	- did not answer (DNA)
II.5	<i>Family</i>
	Obese father: - yes - no
45	- did not answer (DNA)
	Obese mother: - yes - no
46	- did not answer (DNA)
	Obese paternal grandfather: - yes - no
47	- did not answer (DNA)
	Obese maternal grandfather: - yes - no
48	- did not answer (DNA)
	Obese paternal grandmother: - yes - no
49	- did not answer (DNA)
	Obese maternal grandmother: - yes - no
50	- did not answer (DNA)
	Diabetic father: - yes - no
51	- did not answer (DNA)
52	Diabetic mother:

	- yes - no - did not answer (DNA)
53	Diabetic paternal grandfather: - yes - no - did not answer (DNA)
54	Diabetic maternal grandfather: - yes - no - did not answer (DNA)
55	Diabetic paternal grandmother: - yes - no - did not answer (DNA)
56	Diabetic maternal grandmother: - yes - no - did not answer (DNA)
57	Number of brother
58	Number of sister
59	Number of children
60	Number of obese brother
61	Number of obese sister
62	Number of obese children
63	Number of diabetic brother
64	Number of diabetic sister
65	Number of diabetic children
II.6	<i>History of obesity</i>
66	Birth weight (in grams)
67	Birth weight (categorical): - <2.5kg - >4kg - 2.5-4kg - did not answer (DNA)
68	Birth size (in centimeters)
69	Term: number of weeks pregnant
70	Initial age of weight gain: - before 7 years - Before puberty - puberty - adult
71	Weight at 20 years in kg
72	Maximum weight in kg
73	Maximum BMI
74	Previous diets: - no - from 1 to 3 diets - more than 3 diets
75	Age of first diet

	76	Maximum weight loss in a previous diet (kg)
	77	Age of onset of obesity
	78	<p>Circumstances of weight gain:</p> <ul style="list-style-type: none"> - stopping professional activity - stop sport - stop smoking - others - professional change - corticosteroid therapy - leaving the family home - depression - pregnancy - hysterectomy - immobilization
	79	Other circumstances of weight gain
	80	<p>Weight evolution the 6 months before surgery:</p> <ul style="list-style-type: none"> - stable - gained more than 4kg - lost more than 4kg - don't know
III		Medical history
<i>III.1</i>		<i>General</i>
	81	<p>History of respiratory failure:</p> <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
	82	<p>History of respiratory failure: informations</p> <ul style="list-style-type: none"> - restrictive ventilatory failure - nocturnal oxygenation and ventilation
	83	<p>History of respiratory failure: dyspnea:</p> <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
	84	<p>History of respiratory failure: if dyspnea: everyday activities:</p> <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
	85	<p>History of respiratory failure: if dyspnea: climbing stairs:</p> <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
	86	<p>History of respiratory failure: if dyspnea: normal walking:</p> <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
	87	<p>History of respiratory failure: if dyspnea: fast walking:</p> <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
	88	Breathing:

	<ul style="list-style-type: none"> - Asthma - Sleep apnea syndrome - Alveolar hypoventilation - Other
89	History of sleep apnea syndrome: <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
90	History of sleep apnea syndrome: informations
91	History of sleep apnea syndrome: with machinery: <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
92	History of sleep apnea syndrome: evocative signs: <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
93	Respiratory history: <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
94	Respiratory history: informations
95	High blood pressure history: <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
96	High blood pressure history: informations
97	High blood pressure: yes/no (binary)
98	Hyperlipidemia history: <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
99	Hyperlipidemia history: informations
100	Hyperlipidemia: yes/no (binary)
101	Cardiac history: <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
102	Cardiac history: informations
103	Coronaropathy informations
104	Osteoarticular history: <ul style="list-style-type: none"> - yes - no - did not answer (DNA)
105	Osteoarticular history: informations
106	Rheumatology: <ul style="list-style-type: none"> - low back pain - gonalgia - coxalgia
107	Rheumatology: functional impotence informations: <ul style="list-style-type: none"> - minimal - Average - Major

	- rheumatoid arthritis - Total
108	Genitourinary history: - yes - no - did not answer (DNA)
109	Genitourinary history: informations
110	Gynecology-fertility history: - yes - no - did not answer (DNA)
111	Gynecology-fertility history: informations
112	Liver history: - yes - no - did not answer (DNA)
113	Liver history: informations
114	Cirrhosis - yes - no - did not answer (DNA)
115	Steatosis - yes - no - did not answer (DNA)
116	NASH - yes - no - did not answer (DNA)
117	Fibrosis - yes - no - did not answer (DNA)
118	Digestive history - yes - no
119	Digestive history: informations
120	Digestive history: Gastroesophageal reflux disease - yes - no
121	Cancers history - yes - no
122	Cancers history: informations
123	Psychological history - yes - no
124	Psychological history: informations
125	Breakdown: informations
126	Vascular diseases history - yes - no

	127	Vascular diseases history: informations
	128	Stroke: yes/no (binary)
	129	Phlebitis: yes/no (binary)
	130	Allergies history - yes - no
	131	Allergies history: informations
	132	Surgery for obesity in the past: - yes - no - did not answer (DNA)
	133	Surgery for obesity in the past: informations
	134	Other medical history - yes - no
	135	Other medical history: informations
III.2		<i>Diabetes</i>
	136	Diabetes history: - yes - no - did not answer (DNA)
	137	Diabetes history: informations
	138	If diabetes: age of discovery of diabetes
	139	Remission
	140	Type 2 diabetes: yes/no (binary)
III.3		<i>Surgery</i>
	141	Surgery history: - Appendectomy - Cholecystectomy - Caesarean - Other
IV		Initial visit interview
IV.1		<i>Tobacco</i>
	142	Smoking satus (as of preop visit): - yes - no - did not answer (DNA)
	143	If the patient smokes: number of cigarettes per day
IV.2		<i>Psychological</i>
	144	Depression - present - past - no
	145	Night eating syndrome - present - past - no
	146	Binge eating syndrome - present - past - no

<i>IV.3</i>	<i>Physical activity</i>
147	Activity score
148	Activity status: - inactive - active - very active
<i>IV.4</i>	<i>Eating behavior</i>
149	Hunger VAS
150	Satiety VAS
151	Pleasure VAS
152	Hyperphagia - yes - no - did not answer (DNA)
153	Snacking - yes - no - did not answer (DNA)
154	compulsive and binge eating - yes - no - did not answer (DNA)
155	Anorexia / Food disgust - yes - no - did not answer (DNA)
156	Cognitive restriction - yes - no - did not answer (DNA)
157	Tachyphagia - yes - no - did not answer (DNA)
158	Eating disorders informations
<i>IV.5</i>	<i>Alcohol</i>
159	Alcohol consumption: - never - opportunistic - regular - daily - weekly - monthly
160	If alcohol consumption: number of glasses per day
161	If alcohol consumption: number of glasses per week
162	If alcohol consumption: quantity in g / week
163	Alcohol_status: (Hepatology. 2018; 67: 2141-2149) 0 if g/week =0 1 if 0 < g/week < 140 for women or 0 < g/week < 210 for men 2 if g/week ≥ 140 for men or ≥ 210 for men

<i>IV.6</i>	<i>Quality of life</i>
164	Beck (21 questions)
165	Coop score (58 questions)
166	DEBQ (Dutch Eating Behaviour Questionnaire)
167	DEBQ score (33 questions)
168	QEWPR score
169	HAD score
170	BDI2 score
171	SF36 score
172	EQVOD score
173	BES score
174	IPAQ score
175	BAECKE score
176	RICCIGAGNON score
177	EPICES score
V	Clinical evaluation
<i>V.1</i>	<i>General</i>
178	Dyslipidemia 0: no 1: yes
179	Dyslipidemia: number of treatments - 0: no treatment - 1: only one treatment - 2: two or more treatments
180	Diabetes: number of treatments - 0: no treatment - 1: only one treatment - 2: two or more treatments
181	Diabetic (ADA criteria) (Diabetes Care 2017;40(Suppl. 1):S11–S24 DOI: 10.2337/dc17-S005): 0 for normoglycaemic 1 for glucose intolerant 2 diabetic
182	Systolic blood pressure (mmHg)
183	Dystolic diastolic blood pressure (mmHg)
184	Hypertension: number of treatments - 0: no treatment - 1: only one treatment - 2: 2 treatments - 3: 3 or more treatments
185	Hypertension - no - yes
186	Absolute coronary heart disease risk
187	Relative coronary heart disease risk
188	Weight at surgery
189	BMI at surgery (kg/m ²)
190	Waist circumference in cm

	191	Hip circumference (cm)
	192	% weight loss ((before surgery)
	193	% excess weight loss ((before surgery)
	194	Grip Test - dominant arm
	195	Grip Test - right arm
	196	Grip Test - left arm
V.2		<i>Bioelectrical impedance analysis</i>
	197	Weight in kg
	198	% fat mass
	199	Adipose tissue in kg
	200	Lean body mass in kg
	201	% CUN-BAE (Clínica Universidad de Navarra-Body Adiposity Estimator)
	202	% BAI (Body Adiposity Index)
V.3		<i>Symptoms</i>
	203	Feces (number per day)
		Nocturnal feces
		- no
	204	- yes
		Vomiting
		- yes
		- no
	205	- did not answer (DNA)
	206	Vomiting (number per week)
		Excessive volume (vomiting)
		- yes
		- no
	207	- did not answer (DNA)
		Drank while eating (vomiting)
		- yes
		- no
	208	- did not answer (DNA)
		Ate too fast (vomiting)
		- yes
		- no
	209	- did not answer (DNA)
		Other (vomiting):
		- no
	210	- did not answer (DNA)
		Asthenia
		- yes
		- no
	211	- did not answer (DNA)
		Hair loss
		- yes
		- no
	212	- did not answer (DNA)
		Amyotrophy
		- yes
		- no
	213	- did not answer (DNA)

	Paresthesia - yes - no 214 - did not answer (DNA)
	Dumping syndrome - yes - no 215 - did not answer (DNA)
	Hypoglycemia - yes - no 216 - did not answer (DNA)
	Flatulence - no 217 - yes
	Pyrosis - no 218 - yes
	Abdominal pain: - no 219 - yes
	220 Other symptoms
	Gastrointestinal wall: - Eventration - Suppuration 221 - Other
	Sleep apnea syndrome: - no 222 - yes
	Sleep apnea syndrome: with machinery: - no 223 - yes
	Cirrhosis - no 224 - yes
V.4	<i>Paraclinical</i>
	Abdominal echography done - no 225 - yes
	Abdominal echography anomaly - no 226 - yes
	Gallstone - no 227 - yes
	Kidney stone disease - no 228 - yes
	229 Other (Abdominal echography)
	Upper gastrointestinal series done - no 230 - yes
	231 Upper gastrointestinal series anomaly

	- no - yes
232	Hiatal hernia - no - yes
233	Reflux - no - yes
234	Dilatation - no - yes
235	Esophageal contraction - no - yes
236	Real inflation recovered
237	Inflation dont
238	Other (Upper gastrointestinal series)
239	Stenosis: - no - yes
240	Leak: - no - yes
241	Anastomosis: - standard - permeable - moderate shrinkage - dilated stenosis
242	Gastric stump: - standard - not dilated
243	Fibroscopy done: - no - yes
244	Anomaly - no - yes
245	Barrett's esophagus - no - yes
246	Hiatal hernia - no - yes
247	Duodenal ulcer - no - yes
248	Gastritis 0: standard 1: gastric 2: ulcerated
249	Intestinal transit done - no - yes

250	Intestinal transit anomaly - no - yes
251	History of surgical loop induced reflux - no - yes
252	Helicobacter pylori kind of test: - biopsie: biopsy - breath-test: breath-test - serology
253	Helicobacter pylori - no - yes
254	Pulmonary function testing done - no - yes
255	Pulmonary function testing anomaly - no - yes
256	Alveolar hypoventilation - no - yes
257	Other
258	Liver biopsy done - no - yes
259	Liver biopsy not interpretable - 0: interpretable - 1: uninterpretable
260	Blade ID
261	Kind of biopsy - Surgical Biopsy - Punch Biopsy
262	Size (mm)
263	Biopsy: number of hepatic portal tracts
264	Steatosis percentage
265	Liver biopsy: type of steatosis - macrovacuolar - mediovacuolar - microvacuolar - mixed - mixed with predominantly macrovacuolar - mixed with predominantly microvacuolar
266	Liver biopsy: topography of steatosis - non-systematic - panlobular - zone 1 (periportal) - zone 3 (centrolobular)
267	Liver biopsy: steatosis score - 0 - 1 - 2 - 3

	Liver biopsy: lobular inflammation score (kleiner method) - 0 - 1 - 2 - 3
268	
	Liver biopsy: ballooning score (kleiner method) - Not present - A few - Many
269	
	Liver biopsy: NAS score - 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8
270	
	Liver biopsy: fibrosis score (metavir method) - 0 - 1 - 2 - 3 - 4
271	
	Liver biopsy: brunt score - 0 - 1 - 2 - 3
272	
	Liver biopsy: iron overload - Absent - Slight - Moderate
273	
	Liver biopsy: iron overload: location - Mesenchymal - Parenchymal - Mixed with predominantly mesenchymal - Mixed with predominantly parenchymal
274	
	liver biopsy: cholestasis - no - yes
275	
VI	Biological evaluation
VI.1	<i>Metabolic</i>
	276 Hba1C (%)
	277 Leptin (ng/mL)
VI.2	<i>Liver</i>
	278 Total bilirubin (mg/l)
	279 Direct bilirubin (mg/l)
	280 Uric acid (mg/l)
	281 carbon dioxide (mMol/l)
	282 lactate dehydrogenase (U/l)

	283	Amylase (U/l)
	284	Aspartate transaminase (U/l)
	285	Alanine aminotransferase (U/l)
	286	Alkaline pH (new)
	287	Gamma GT (U/l)
	288	Alkaline pH (old)
VI.3		<i>Pancreas</i>
	289	Amylase (U/l)
	290	Lipase (U/l)
VI.4		<i>Specific</i>
	291	D-Xylose (g/L)
	292	Oxaluria (mmol/24h)
VI.5		<i>Ionogram</i>
	293	Fasting blood glucose (mmol/l)
	294	Urea (g/l)
	295	Creatinine (mg/l)
	296	Na (sodium) (mmol/l)
	297	K (potassium) (mmol/l)
	298	Cl (chlorine) (mmol/l)
	299	Protein (g/l)
	300	Albumin (g/l)
	301	Calcemia (mg/l)
	302	Phosphorus (mg/l)
	303	Average calculated between the fasting blood glucose of the HPO and that of the ionogram (mmol/l)
VI.6		<i>Lipidic</i>
	304	Total cholesterol (mMol/l)
	305	HDL cholesterol (mMol/l)
	306	LDL cholesterol (mMol/l)
	307	Triglycerides (mMol/l)
	308	Apolipoprotein A1 (g/l)
VI.7		<i>Blood</i>
	309	leukocytes (10.9/l)
	310	Red blood cells (10.12/l)
	311	Hemoglobin (g/dl)
	312	Hematocrit (%)
	313	Platelets (10.9/l)
	314	Neutrophils (10.9/l)
	315	Eosinophils (10.9/l)
	316	Basophils (10.9/l)
	317	lymphocytes (10.9/l)
	318	monocytes (10.9/l)
	319	APTT (activated partial thromboplastin time) for cases (sec.)

	320	APTT (activated partial thromboplastin time) for controls (sec.)
	321	PR (Prothrombin Ratio) (%)
	322	Fibrinogen (g/l)
VI.8		OGTT
	323	Fasting blood glucose 10 min before
	324	Fasting blood glucose 5 min before
	325	Fasting blood glucose (mMol/l)
	326	blood glucose 15 min after (Oral glucose tolerance test) (mMol/l)
	327	blood glucose 30 min after (Oral glucose tolerance test) (mMol/l)
	328	blood glucose 60 min after (Oral glucose tolerance test) (mMol/l)
	329	blood glucose 90 min after (Oral glucose tolerance test) (mMol/l)
	330	blood glucose 120 min after (Oral glucose tolerance test) (mMol/l)
	331	blood glucose 180 min after (Oral glucose tolerance test) (mMol/l)
	332	Fasting C-peptide 10 min before (ng/ml)
	333	Fasting C-peptide 5 min before (ng/ml)
	334	Fasting C-peptide (ng/ml)
	335	C-peptide 15 min after (Oral glucose tolerance test) (ng/ml)
	336	C-peptide 30 min after (Oral glucose tolerance test) (ng/ml)
	337	C-peptide 60 min after (Oral glucose tolerance test) (ng/ml)
	338	C-peptide 90 min after (Oral glucose tolerance test) (ng/ml)
	339	C-peptide 120 min after (Oral glucose tolerance test) (ng/ml)
	340	C-peptide 180 min after (Oral glucose tolerance test) (ng/ml)
	341	Fasting plasma insulin 10 min before (IRMA) (mUI/l)
	342	Fasting plasma insulin 5 min before (IRMA) (mUI/l)
	343	Fasting plasma insulin (IRMA) (mUI/l)
	344	plasma insulin after 15 min (Oral glucose tolerance test) (mUI/l)
	345	plasma insulin after 30 min (Oral glucose tolerance test) (mUI/l)
	346	plasma insulin after 60 min (Oral glucose tolerance test) (mUI/l)
	347	plasma insulin after 90 min (Oral glucose tolerance test) (mUI/l)
	348	plasma insulin after 120 min (Oral glucose tolerance test) (mUI/l)
	349	plasma insulin after 180 min (Oral glucose tolerance

	test) (mUI/l)
350	Fasting proinsulin 10 min before (pmol/l)
351	Fasting proinsulin 5 min before (pmol/l)
352	Fasting proinsulin (pmol/l)
353	proinsulin after 30 min (Oral glucose tolerance test) (pmol/l)
354	proinsulin after 120 min (Oral glucose tolerance test) (pmol/l)
355	xylose (g/l)
356	xylose after 15 min (g/l)
357	xylose after 30 min (g/l)
358	xylose after 60 min (g/l)
359	xylose after 90 min (g/l)
360	xylose after 120 min (g/l)
361	xylose after 180 min (g/l)
362	Homa2IR: insulin resistance Diabetes Care, vol. 36, no 8, 2013, p. 2324-30 (PMID 23564921, DOI 10.2337/dc12-0607)
363	Homa2S: insulin sensitivity Diabetes Care, vol. 36, no 8, 2013, p. 2324-30 (PMID 23564921, DOI 10.2337/dc12-0607)
364	Homa2IR: beta-cell function Diabetes Care, vol. 36, no 8, 2013, p. 2324-30 (PMID 23564921, DOI 10.2337/dc12-0607)
VI.9	Anemia
365	TSH (thyroid-stimulating hormone) (microUI/ml)
366	Ferritin (ng/ml)
367	Folate (vitamin B9) (ng/ml)
368	Vitamin B12 (ng/ml)
369	Transthyretin (prealbumin)
370	CRP (mg/l)
371	Haptoglobin (g/l)
372	Alpha2 macroglobulin (g/l)
373	Vitamin B1 (nmol/l)
374	Vitamin B6 (nmol/l)
375	Vitamin A (µmol/l)
376	Vitamin E (µmol/l)
377	Vitamin C (µmol/l)
378	Vitamin D (ng/ml)
379	PTH (ng/ml)
380	Erythrocyte magnesium (mg/l)
381	Serum zinc (µg/100ml)
382	Zn glob (µg/100ml)
383	Citraturie
384	Fer serique (µg/100ml)
385	transferrine (g/l)

	386	CTF (µg/dl)
	387	CSS (%)
VI.10		<i>Urinary ionogram</i>
	388	Diuresis (L)
	389	Na (mmol/l)
	390	K (mmol/l)
	391	Cl (mmol/l)
	392	Glucose (g/l)
	393	Urea (g/l)
	394	Creatinine (mg/l)
	395	Proteinuria (g/24H)
	396	Microalbuminuria (mg/24H)
	397	creatinine clearance (Salazar Corcoran method) (mL/min)
	398	creatinine clearance (MDRD method) (ml/min)
	399	D-xylose absorption test (g/l)
VI.11		<i>Hormonal analysis</i>
	400	Testosterone (ng/mL)
	401	Testosterone bioavailable (nmol/L)
	402	SBP (nmol/L)
	403	Inhibin B (pg/mL)
	404	FSH (UI/L)
	405	LH (UI/L)
	406	Estradiol (pg/mL)
	407	Prolactin (ng/mL)
VII		Treatments
	408	Diabetic (ADA criteria) (Diabetes Care 2017;40(Suppl. 1):S11–S24 DOI: 10.2337/dc17-S005): - 0: normoglycaemic - 1: glucose intolerant - 2: diabetic
	409	Antidiabetic drugs - yes - no
	410	Oral anti-diabetic medication - no - yes
	411	Insulin - 0: no - 1: yes
	412	Number of anti-diabetic treatments (number of molecules): - 0: for no treatment - 1: for a single treatment (a single molecule) - 2: for 2 or more treatments (molecules)
	413	Names of antidiabetic drugs
	414	Dyslipidemia:

	- no - yes
415	Anti-dyslipidemic medication - no - yes
416	Number of lipid-lowering treatments (number of molecules): - 0: for no treatment - 1: for a single treatment (a single molecule) - 2: for 2 or more treatments (molecules)
417	Names of lipid-lowering drugs
418	Hypertensive: - no - yes
419	Antihypertensives medication - no - yes
420	Number of antihypertensive treatments (number of molecules): - 0: for no treatment - 1: for a single treatment (a single molecule) - 2: for 2 or more treatments (molecules)
421	Names of antihypertensives drugs
422	Oral contraception - no - yes
423	Hormonal treatments at the time of surgery - for contraception - yes - no - did not answer (DNA)
424	Hormonal treatments at the time of surgery - for menopause - yes - no - did not answer (DNA)
425	Treatment for vitamin A - no - yes
426	Treatment for vitamin B1 - no - yes
427	Treatment for vitamin B6 - no - yes
428	Treatment for vitamin B9 - no - yes
429	Treatment for vitamin B12 - no - yes
430	Treatment for vitamin D - no - yes

	Treatment for iron - no 431 - yes
	Treatment for calcium - no 432 - yes
	Treatment for PPI (Proton Pump Inhibitors) - no 433 - yes
	Thyroid treatment - no 434 - yes
	435 Names of thyroid hormones
	Antibiotics medication - no 436 - yes
	Oral anticoagulants - no 437 - yes
	Antidepressants - no 438 - yes
	439 Names of antidepressants
	Hypnotic drugs - no 440 - yes
	441 Names of hypnotics
	Analgesics - no 442 - yes
	443 Names of analgesics
	Others treatments - yes 444 - no
	445 Names of others treatments (molecules)
VIII	Type of intervention
	Type of intervention - gastric band - gastric bypass 446 - sleeve gastrectomy
	447 Type of intervention: informations

Appendix table 5: TRIPOD checklist for model development and validation

Item		Section/Topic		Checklist Item	Page
Title and abstract					
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.		1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.		2
Introduction					
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.		4
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.		4
Methods					
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.		5
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.		App. 3
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.		5,8,9
	5b	D;V	Describe eligibility criteria for participants.		5, App. Figure 1
	5c	D;V	Give details of treatments received, if relevant.		NA
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.		5
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.		NA
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.		9, App. Table 4
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.		NA
Sample size	8	D;V	Explain how the study size was arrived at.		5,8,9
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.		7,8

Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	6
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	6,7
	10c	V	For validation, describe how the predictions were calculated.	App. Figure 3
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	7,8
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	NA
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	NA
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	Table 1
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	App. Figure 1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	Table 1
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	Table 1, Fig 1
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	Table 10
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	NA
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	App. Figure 3
	15b	D	Explain how to use the prediction model.	App. Figure 3, Web app
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	Table 2, Table 3, Figure 2
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	NA
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	12
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	App Table 1 and 2,

	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	11,12
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	12
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	Web app
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	13

Appendix Table 6: Baseline characteristics, among 633 expected at 5 years, comparing the complete cases with those with missing data at one or more follow-up visits

	Complete cases	≥1 missing visit	p*
n	412	221	
Age (mean (SD))	42.5 (11.5)	39.4 (12.0)	0.014
Height (cm) (mean (SD))	167.2 (9.1)	168.2 (9.5)	1.0
Type of operation (%)			0.96
AGB	123 (29.9)	58 (26.2)	
RYGB	262 (63.6)	139 (62.9)	
SG	27 (6.6)	24 (10.9)	
BMI at intervention (mean (SD))	47.2 (6.9)	48.5 (8.3)	0.25
Type 2 diabetes (at BL) (%)	159 (38.7)	65 (29.4)	0.18
Years with T2D (at BL) (mean (SD))	29.0 (21.2)	31.5 (19.9)	1.0
Smoker (%)	59 (14.3)	37 (16.7)	1.0

* Multiple testing correction (Bonferroni correction).

BL = baseline data

A logistic regression to predict “missing data at one or more follow-up visits” versus “complete cases”, in a multivariable analysis with the above variables, among the patients expected at 5 years. As a result, only the age was significantly associated with missing follow-up data : OR 0.98 (95% CI 0.96-1.00).

Appendix Table 7: List of authors in charge of the access and verification of each cohort

The datasets generated, analyzed or both during the current study are not publicly available since they are subject to national data protection laws and restrictions imposed by the ethics committee to ensure data privacy of the study participants. However, they can be applied for through an individual project agreement with the principal investigator of the University Hospital of Lille, France (francois.pattou@univ-lille.fr).

Dataset	Authors in charge of access and verification
ABOS	Patrick Saux, Pierre Bauvin, Tomy Soumphonphakdy, Philippe Preux, Violeta Raverdy, H�el�ene Verkindt, Robert Caiazzo, Fran�ois Pattou.
BAREVAL	Patrick Saux, Pierre Bauvin, Tomy Soumphonphakdy, Florence Galtier, David Nocca.
NOK	Anne Jacobs, Daan Jacobs, Valerie Montpellier.
SGH	Patrick Saux, Pierre Bauvin, Tomy Soumphonphakdy, Lee Phong Ching, Lim Chin Hong.
SOS	Johanna C. Andersson-Assarsson, Lena Carlsson, Per-Arne Svensson.
PRECOS	Patrick Saux, Pierre Bauvin, Tomy Soumphonphakdy, Guelareh Dezfoulian, Mihaela Moldovanu, Severine Andrieux, Julien Couster, Marie Lepage.
Roma	Patrick Saux, Pierre Bauvin, Tomy Soumphonphakdy, Erminia Lembo, Ornella Verrastro, Geltrude Mingrone.
Lyon	Patrick Saux, Pierre Bauvin, Tomy Soumphonphakdy, Maud Robert.
COD	Patrick Saux, Pierre Bauvin, Tomy Soumphonphakdy, Riccardo Cohen.
Mexico	Patrick Saux, Pierre Bauvin, Tomy Soumphonphakdy, Carlos Zerrweck.
SleevePass	Patrick Saux, Pierre Bauvin, Tomy Soumphonphakdy, Paulina Salminen.
SM-BOSS	Patrick Saux, Pierre Bauvin, Tomy Soumphonphakdy, Ralph Peterli.

References

1. Expert Committee on the D, Classification of Diabetes M. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 2003; **26 Suppl 1**: S5-20.
2. Sjostrom L, Larsson B, Backman L, et al. Swedish obese subjects (SOS). Recruitment for an intervention study and a selected description of the obese state. *Int J Obes Relat Metab Disord* 1992; **16**(6): 465-79.
3. Cohen RV, Pereira TV, Aboud CM, et al. Effect of Gastric Bypass vs Best Medical Treatment on Early-Stage Chronic Kidney Disease in Patients With Type 2 Diabetes and Obesity: A Randomized Clinical Trial. *JAMA Surg* 2020; **155**(8): e200420.
4. Cohen RV, Pereira TV, Aboud CM, et al. Gastric bypass versus best medical treatment for diabetic kidney disease: 5 years follow up of a single-centre open label randomised controlled trial. *EClinicalMedicine* 2022; **53**: 101725.
5. Zerrweck C, Herrera A, Sepulveda EM, Rodriguez FM, Guilbert L. Long versus short biliopancreatic limb in Roux-en-Y gastric bypass: short-term results of a randomized clinical trial. *Surg Obes Relat Dis* 2021; **17**(8): 1425-30.
6. Cottam S, Cottam D, Cottam A, Zaveri H, Surve A, Richards C. The Use of Predictive Markers for the Development of a Model to Predict Weight Loss Following Vertical Sleeve Gastrectomy. *Obes Surg* 2018; **28**(12): 3769-74.
7. Goulart A, Leao P, Costa P, et al. Doctor, How Much Weight Will I Lose?-a New Individualized Predictive Model for Weight Loss. *Obes Surg* 2016; **26**(6): 1357-9.
8. Janik MR, Rogula TG, Mustafa RR, Saleh AA, Abbas M, Khaitan L. Setting realistic expectations for weight loss after laparoscopic sleeve gastrectomy. *Wideochir Inne Tech Maloinwazyjne* 2019; **14**(3): 415-9.
9. Velazquez-Fernandez D, Sanchez H, Monraz F, et al. Development of an Interactive Outcome Estimation Tool for Laparoscopic Roux-en-Y Gastric Bypass in Mexico Based on a Cohort of 1002 Patients. *Obes Surg* 2019; **29**(9): 2878-85.
10. Wise ES, Hocking KM, Kavic SM. Prediction of excess weight loss after laparoscopic Roux-en-Y gastric bypass: data from an artificial neural network. *Surg Endosc* 2016; **30**(2): 480-8.
11. Seyssel K, Suter M, Pattou F, et al. A Predictive Model of Weight Loss After Roux-en-Y Gastric Bypass up to 5 Years After Surgery: a Useful Tool to Select and Manage Candidates to Bariatric Surgery. *Obes Surg* 2018; **28**(11): 3393-9.
12. Baltasar A, Perez N, Serra C, Bou R, Bengochea M, Borrás F. Weight loss reporting: predicted body mass index after bariatric surgery. *Obes Surg* 2011; **21**(3): 367-72.
13. Harvey D, Leybourne S, Newbold P. Testing the equality of prediction mean squared errors. *Int J Forecasting* 1997; **13**(2): 281-91.
14. Wolnerhanssen BK, Peterli R, Hurme S, et al. Laparoscopic Roux-en-Y gastric bypass versus laparoscopic sleeve gastrectomy: 5-year outcomes of merged data from two randomized clinical trials (SLEEVEPASS and SM-BOSS). *Br J Surg* 2021; **108**(1): 49-57.