



OPEN

Development and validation of a scoring system for pre-surgical and early post-surgical prediction of bariatric surgery unsuccess at 2 years

Fabio Bioletto¹✉, Marianna Pellegrini¹, Chiara D'Eusebio¹, Stefano Boschetti², Farnaz Rahimi², Antonella De Francesco², Simone Arolfo³, Mauro Toppino³, Mario Morino³, Ezio Ghigo¹ & Simona Bo¹

Bariatric surgery (BS) is an effective treatment for morbid obesity. However, a simple and easy-to-use tool for the prediction of BS unsuccess is still lacking. Baseline and follow-up data from 300 consecutive patients who underwent BS were retrospectively collected. Supervised regression and machine-learning techniques were used for model development, in which BS unsuccess at 2 years was defined as a percentage of excess-weight-loss (%EWL) < 50%. Model performances were also assessed considering the percentage of total-weight-loss (%TWL) as the reference parameter. Two scoring systems (NAG-score and ENAG-score) were developed. NAG-score, comprising only pre-surgical data, was structured on a 4.5-point-scale (2 points for neck circumference ≥ 44 cm, 1.5 for age ≥ 50 years, and 1 for fasting glucose ≥ 118 mg/dL). ENAG-score, including also early post-operative data, was structured on a 7-point-scale (3 points for %EWL at 6 months $\leq 45\%$, 1.5 for neck circumference ≥ 44 cm, 1 for age ≥ 50 years, and 1.5 for fasting glucose ≥ 118 mg/dL). A 3-class-clustering was proposed for clinical application. In conclusion, our study proposed two scoring systems for pre-surgical and early post-surgical prediction of 2-year BS weight-loss, which may be useful to guide the pre-operative assessment, the appropriate balance of patients' expectations, and the post-operative care.

Bariatric surgery (BS) is a recognized and effective treatment for patients with morbid obesity, i.e., for those individuals with a body mass index (BMI) ≥ 40 kg/m² or a BMI ≥ 35 kg/m² with comorbidities, who are refractory to lifestyle change and drug therapy¹. Several surgical techniques are currently available, but the two most commonly performed are sleeve gastrectomy (SG) and Roux-en-Y gastric bypass (RYGB)^{1–4}.

Determining the predictors of BS weight outcomes is an area of active debate^{1,5,6}. A reliable prediction of long-term BS results would facilitate the pre-operative assessment, the appropriate balance of patients' expectations, and the post-operative care^{1,4}. Several baseline characteristics have been associated with poor weight loss, such as increased fasting glucose levels^{7,8} and increased waist circumference values⁹. Other predictors, such as sex, age, and psychiatric comorbidities (psychiatric diseases, personality disorders, eating disorders) were associated with weight loss in some studies^{10,11} but not in others^{12,13}. An interesting focus should be reserved to pre-surgical BMI, as its association with weight loss following BS has variably been reported to be direct^{10,14}, inverse^{15–17} or absent^{18–20}. Moreover, post-operative weight loss trajectories, usually measured up to 6 months, have been proved to be a useful tool for early prediction of weight outcomes in the longer term^{21–28}. However, post-surgical weight changes are remarkably variable in their trends, with significant differences among individuals^{29,30}; therefore, long-term BS success may still be achieved in patients showing a poor early weight loss, and vice versa^{21–28}.

In clinical practice, it is often difficult to take into account all these factors comprehensively during patient assessment. Furthermore, collinearities between explanatory variables are possible, and factors significantly

¹Department of Medical Sciences, University of Turin, Corso Dogliotti 14, 10126 Turin, Italy. ²Dietetic Unit, "Città della Salute e della Scienza" Hospital, Turin, Italy. ³Department of Surgical Sciences, University of Turin, Turin, Italy. ✉email: fabio.bioletto@unito.it

associated with the outcome at univariate analyses may lose their statistical significance when included in multivariate models. Another important issue in predictive models is to establish the relative strength of each considered variable in terms of outcome prediction; model calibration with proper tuning of predictors' relative weights is a crucial step for any subsequent practical application.

Several equations for post-operative weight loss prediction have been proposed in recent years^{21,31–36}; however, the retrieved evidence was often limited by a short follow-up time (up to 1 year after BS in most cases)^{21,31–33,35} or by a relatively small sample size^{21,33,35}. Moreover, post-operative weight loss and model predictors were considered as continuous measures in most cases^{31–36}; this represents a further limitation for a comfortable use of these models in the clinical practice, in which dichotomous evaluations, conclusions, or choices are most often needed.

To the best of our knowledge, a simple and easy-to-use estimation tool for pre-surgical and early post-surgical prediction of long-term BS weight loss is still lacking. The aim of the present study was to propose two scoring systems, the NAG-score and the ENAG-score, respectively for pre-BS and early post-BS prediction of weight loss after a 2-year follow-up.

Methods

Study design and patient management. In this retrospective observational study, we collected data from the clinical records of the first 300 patients who underwent BS at the General Surgery Department of the “Città della Salute e della Scienza” Hospital of Turin, University of Turin, from the 1st January 2016 and met the below-reported criteria.

Inclusion criteria were: (1) age from 18 to 65 years old; (2) BMI ≥ 35 kg/m² with comorbidities or BMI ≥ 40 kg/m², with numerous unsuccessful attempts to lose weight; (3) a minimum 2-year follow-up at our Obesity Unit. Exclusion criteria were: (1) secondary causes of obesity (e.g., hypothalamic diseases, endocrine diseases); (2) associated comorbidities or diseases impacting on weight loss (for example, patients who performed BS in prevision of transplantation).

Patients gave their informed consent to the processing of their data. The study was approved by the local Ethics Committee (Comitato Etico Interaziendale A.O.U. Città della Salute e della Scienza di Torino—A.O. Ordine Mauriziano—A.S.L. Città di Torino) and was in accordance with the principles of the Declaration of Helsinki.

Patient management and data collection. Pre-operatively, data relative to eating behaviours, previous attempts of weight loss, and presence of comorbidities were collected²⁸. Weight, height, waist circumference, and neck circumference were assessed in all patients as anthropometric measures. A fasting blood sample was drawn from all patients, and glucose and lipid values were centrally measured.

The surgical approach consisted of either SG or RYGB. The type of surgery was based on the following patient features: BMI, age, gender, body fat distribution, presence of type 2 diabetes mellitus, hiatal hernia, gastroesophageal reflux disease, patient's expectations/realistic goals, long-term treatment for a coexisting disease or condition for which absorption and pharmacokinetics of drugs are of major concern. Follow-up visits at our obesity unit took place at 1, 2, 6, 12, and 24 months after BS. Details about patient management and surgical techniques have been previously described²⁸.

To the scope of model creation, BS success was defined as the percentage of excess-weight-loss (%EWL), which was calculated as:

$$[(\text{pre-BS weight} - \text{weight at the time of visit})/(\text{pre-BS weight-ideal weight})] \times 100.$$

Weight corresponding to the BMI = 25 kg/m² was considered as the ideal body weight. An unsuccessful weight loss after BS was defined as %EWL < 50%, according to guidelines¹.

The use of %EWL for the evaluation of BS weight-outcome has been recently questioned, mostly due to its dependency on pre-BS weight excess^{37,38}, with the percentage of total-weight-loss (%TWL), calculated as:

$$[(\text{pre-BS weight} - \text{weight at the time of visit})/(\text{pre-BS weight})] \times 100.$$

being proposed as an alternative parameter^{37,38}. However, there is still no consensus in literature about which threshold to adopt for the definition of BS unsuccess, since %TWL < 20% was proposed by some authors^{37–39}, while %TWL < 25% was suggested by others⁴⁰. Therefore, to the scope of this study, we opted to use %EWL as the reference parameter for model development, due to the availability of an unanimously recognized cut-off for the definition of BS unsuccess; nevertheless, in order to further strengthen the consistency of our results, final model performances were also assessed considering %TWL as the outcome of choice.

Measurements. Weight was measured with the patient wearing light clothes and no shoes to the nearest 0.1 kg by a digital scale with a capacity of 300 kg (Wunder Sa.Bi.srl). Height was measured to the nearest 0.1 cm with a Stadiometer SECA 220 measuring rod (Hamburg, Germany). Waist and neck circumferences were assessed by a plastic tape meter at the umbilicus level and under the cricoid cartilage, respectively. Type 2 diabetes mellitus and arterial hypertension were diagnosed in accordance with international guidelines. Obstructive sleep apnea (OSA) was hypothesized in the presence of excessive daytime sleepiness, snoring, and choking or gasping during sleep, enlarged neck circumference, and an intermediate to high-risk score at the STOP-Bang questionnaire⁴¹. The diagnosis was confirmed by a sleep-expert neurologist by means of further exams, according to international guidelines^{42,43}.

Statistical analysis. Patient characteristics were summarized using mean and standard deviation for continuous variables and percent values for categorical data.

Relevant predictors of BS outcomes were found through univariate and multivariate logistic regressions, using a stepwise backward selection. Optimal cut-points for continuous variables were found through a supervised machine-learning algorithmic approach; cut-offs for class distinction were automatically derived by Class-Attribute Contingency Coefficient (CACC) discretization algorithm as those maximizing separation between classes⁴⁴. Multivariate logistic regression was re-applied on discretized variables; integer and half-integer point scores were assigned upon normalization and rounding of regression beta-coefficients. Model calibration was evaluated by the Hosmer–Lemeshow test. A ten-fold cross-validation algorithm was adopted for internal validation, in order to provide an estimate of model performance on unseen data^{45,46}: after a random split of the original sample into ten groups, the modelling process was repeated starting from stepwise variable selection in nine of them, and its performance was evaluated in the tenth; the process was then repeated ten times, rotating the validation group at each round; final model performance was obtained as the average performance over the ten iterations⁴⁶. According to the TRIPOD statement⁴⁶, this internal validation approach was preferred to the more classical sample-split approach due to its better reliability in reducing the bias and the variability of performance estimates. Iterative Dichotomiser 3 (ID3) algorithm^{47,48} was applied to cluster risk classes of clinical relevance.

Statistical analysis was performed using STATA 17 (StataCorp, College Station, Texas, USA) and R 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). For the supervised machine-learning approach used for score creation, ‘arulesCBA’, ‘partykit’ and ‘rpart’ packages were used.

Results

Univariate analysis and variable discretization. The study comprised 32 males (10.7%) and 268 females (89.3%). The mean age at the moment of BS was 44.6 ± 10.3 years. Two hundred and thirty-three patients (77.7%) underwent SG, while the remaining 67 (22.3%) underwent RYGB.

At 24 months, 56 patients (18.7%) showed a %EWL < 50%, while 46 (15.3%) showed a %TWL < 20%. The association between relevant predictors and %EWL < 50%, which was adopted as the reference parameter for model development, was first explored through univariate logistic regression analysis. Neck circumference, age, frequency of OSA, and frequency of male sex were higher, while %EWL at 6-months lower, in patients with %EWL < 50% at 24 months (Table 1).

For continuous variables, the possible presence of significant threshold values was explored through the application of CACC discretization algorithm; if present, meaningful cut-offs for class distinction were automatically derived as those maximizing separation between classes. Not surprisingly, a meaningful categorization could be found for neck circumference, age, and %EWL at 6 months, i.e., those variables who were already significant when considered as continuous. The optimal cut-points retrieved were ≥ 44 cm for neck circumference (OR = 5.21, 95% CI 2.39–11.36), ≥ 50 years for age (OR = 3.64, 95% CI 2.00–6.64), and $\leq 45\%$ for %EWL at 6 months (OR = 12.21, 95% CI 6.11–24.40). Moreover, a significant dichotomization emerged for two other predictors, i.e., fasting glucose and waist circumference. The optimal cut-points retrieved were ≥ 118 mg/dL for fasting glucose (OR = 2.57, 95% CI 1.33–4.96) and ≥ 142 cm for waist circumference (OR = 2.35, 95% CI 1.07–5.16) (Table 1). The clinical significance of the algorithmically retrieved threshold values was qualitatively substantiated by descriptive analyses, in which the non-linear dependence between the predictors and the outcome was readily evident (Supplementary Fig. 1).

Development and internal validation of pre-surgical score. All pre-surgical variables showing a significant correlation with the outcome at univariate analysis were included in a multivariate regression model. Given the intention of developing a predictive score, continuous variables were included in the model according to their retrieved dichotomizations. After a stepwise backward selection, the variables retaining statistical significance were neck circumference ≥ 44 cm (OR = 4.21, 95% CI 1.85–9.55), age ≥ 50 years (OR = 3.03, 95% CI 1.62–5.67), and fasting glucose ≥ 118 mg/dL (OR = 2.06, 95% CI 1.01–4.18) (Table 2). All other variables (i.e., male sex, OSA, waist circumference ≥ 142 cm) were excluded as non-significant at multivariate analysis.

This model showed a moderate accuracy in the prediction of the outcome (AUC = 0.713). The Hosmer–Lemeshow test did not reveal any significant miscalibration ($p = 0.58$). Internal validation of the model was performed through ten-fold cross-validation; the final estimation of the model performance on unseen data, obtained as the average AUC over the ten iterations, was equal to 0.695, thus reassuring about a substantially null overfitting effect.

In order to simplify its clinical application, the three variables selected by the model were used to develop a discrete-point prediction score; integer and half-integer point scores were assigned upon normalization and rounding of regression beta-coefficients (Table 2). Due to the considered variable, this score was named “NAG-score” (Neck circumference, Age, Glucose), and was defined by the sum of all three components, on a 4.5-point-scale. Notably, this mild simplification did not lead to a significant reduction in the predictive power of the model since the AUC remained equal to 0.713.

A descriptive graph of the risk of unsuccessful weight loss after BS according to NAG-score is presented in Fig. 1; the performance of the model in predicting BS unsuccess was evaluated both in terms of %EWL and %TWL, with similar results. In addition, in order to simplify the interpretation of NAG-score and to identify the most clinically relevant risk classes, an algorithmic classification clustering was proposed. ID3 algorithm clustered the patients in three distinct risk classes (0–1 points, 1.5–2 points, 2.5–4.5 points), with no differences whether considering %EWL < 50% or %TWL < 20% as the reference outcome (Fig. 2); given their clinical correlates, we referred to them as “low-risk”, “intermediate-risk”, and “high-risk” classes, respectively. As reported in Table 3, when considering %EWL, the low-risk class comprised 184 patients, with a 10.3% risk of unsuccessful weight loss at 2-years from BS. The intermediate-risk class comprised 75 patients, with a 20.0% risk of unsuccessful post-surgical weight loss. The high-risk class comprised 41 patients, with a 54.7% risk of unsuccessful post-surgical

Variable	Successful weight loss (n = 244)	Unsuccessful weight loss (n = 56)	OR	95% CI	p-value
Age (years)	43.4 ± 10.1	49.9 ± 9.7	1.07	1.04–1.11	< 0.001
Age category (%) §					
< 50 years	71.7	41.1	1.00		
≥ 50 years	28.3	58.9	3.64	2.00–6.64	< 0.001
Male gender (%)	8.2	21.4	3.05	1.56–5.86	0.005
Smoking habits (%)					
Current smoker	57.4	55.4	1.00		
Past smoker	20.1	21.4	1.02	0.91–1.14	0.79
Never smoker	22.5	23.2	1.01	0.91–1.13	0.86
Job (%)					
Unemployed/housewife/ retired	38.1	39.3	1.00		
Employee	40.6	39.3	0.99	0.90–1.09	0.85
Manual worker	19.7	16.1	0.97	0.85–1.09	0.59
Other	1.6	5.4	1.27	0.94–1.71	0.12
Type of surgery (%)					
Gastric bypass	22.1	23.2	1.00		
Sleeve gastrectomy	77.9	76.8	0.94	0.54–1.71	0.86
Pre-surgical weight (kg)	118.3 ± 22.4	117.8 ± 25.4	1.00	0.99–1.01	0.89
Height (cm)	162.9 ± 8.7	163.9 ± 9.5	1.01	0.99–1.04	0.44
Pre-surgical BMI (kg/m ²)	44.6 ± 8.1	43.6 ± 7.7	0.98	0.95–1.02	0.41
Waist circumference (cm)	123.0 ± 14.6	125.3 ± 18.6	1.01	0.99–1.03	0.33
Waist circumference category (%) §					
< 142 cm	90.6	80.4	1.00		
≥ 142 cm	9.4	19.6	2.35	1.07–5.16	0.033
Neck circumference (cm)	38.5 ± 3.3	40.4 ± 4.4	1.15	1.07–1.23	< 0.001
Neck circumference category (%) §					
< 44 cm	93.4	73.2	1.00		
≥ 44 cm	6.6	26.8	5.21	2.39–11.36	< 0.001
OSA (%)	18.9	33.9	2.21	1.28–3.76	0.015
Arterial hypertension (%)	34.4	35.7	1.06	0.63–1.75	0.85
Fasting glucose (mg/dL)	106.7 ± 28.8	112.2 ± 31.2	1.01	1.00–1.02	0.20
Fasting glucose category §					
< 118 mg/dL	84.4	67.9	1.00		
≥ 118 mg/dL	15.6	32.1	2.57	1.33–4.96	0.005
Type 2 diabetes mellitus (%)	22.1	25.0	1.17	0.65–2.04	0.64
Total cholesterol (mg/dL)	198.3 ± 37.7	188.3 ± 37.4	0.99	0.98–1.00	0.07
HDL cholesterol (mg/dL)	47.4 ± 13.1	46.5 ± 12.3	0.99	0.97–1.01	0.63
Triglycerides (mg/dL)	146.4 ± 65.8	145.9 ± 50.3	1.00	0.99–1.01	0.96
%EWL at 6 months	59.9 ± 25.9	32.5 ± 48.2	0.94	0.93–0.96	< 0.001
Category of %EWL at 6 months (%) §					
%EWL > 45%	78.7	23.2	1.00		
%EWL ≤ 45%	21.3	76.8	12.21	6.11–24.40	< 0.001

Table 1. Association of pre-surgical and early post-surgical data with BS failure (defined as %EWL < 50% at 2 years), analyzed through univariate logistic regressions. §Cut-off derived through CACC discretization algorithm. *BMI* body mass index, *BS* bariatric surgery, *CACC* Class-Attribute Contingency Coefficient, *CI* confidence interval, *EWL* excess weight loss, *HDL* high-density lipoprotein, *OR* odds-ratio, *OSA* obstructive sleep apnea.

weight loss. The stratification performance of the model across different risk classes was overall preserved when adopting %TWL as the reference parameter; in this case, the risk of BS unsuccess was equal to 8.7%, 18.7%, and 39.0% in the low-risk, intermediate-risk, and high-risk class, respectively.

Development and internal validation of early post-surgical score. In order to develop an early post-surgical predictive score, %EWL at 6 months was added to pre-surgical variables in a further multivariate regression model. Again, given the intention of developing a predictive score, continuous variables were included in the model according to their retrieved dichotomizations. After a stepwise backward selection, the

Variable	OR	95% CI	p-value	β -coefficient	Normalized coefficient	Points for NAG-score
Neck circumference ≥ 44 cm	4.21	1.85–9.55	0.001	+ 1.437	1.990	+ 2
Age ≥ 50 years	3.03	1.62–5.67	0.001	+ 1.109	1.536	+ 1.5
Fasting glucose ≥ 118 mg/dL	2.06	1.01–4.18	0.046	+ 0.722	1.000	+ 1

Table 2. Prediction of BS failure, defined as %EWL < 50% at 2 years, by multivariate logistic regression after stepwise backward selection of pre-surgical data; NAG-score point assignment according to multivariate regression coefficients. BS bariatric surgery, CI confidence interval, EWL excess weight loss, OR odds-ratio.

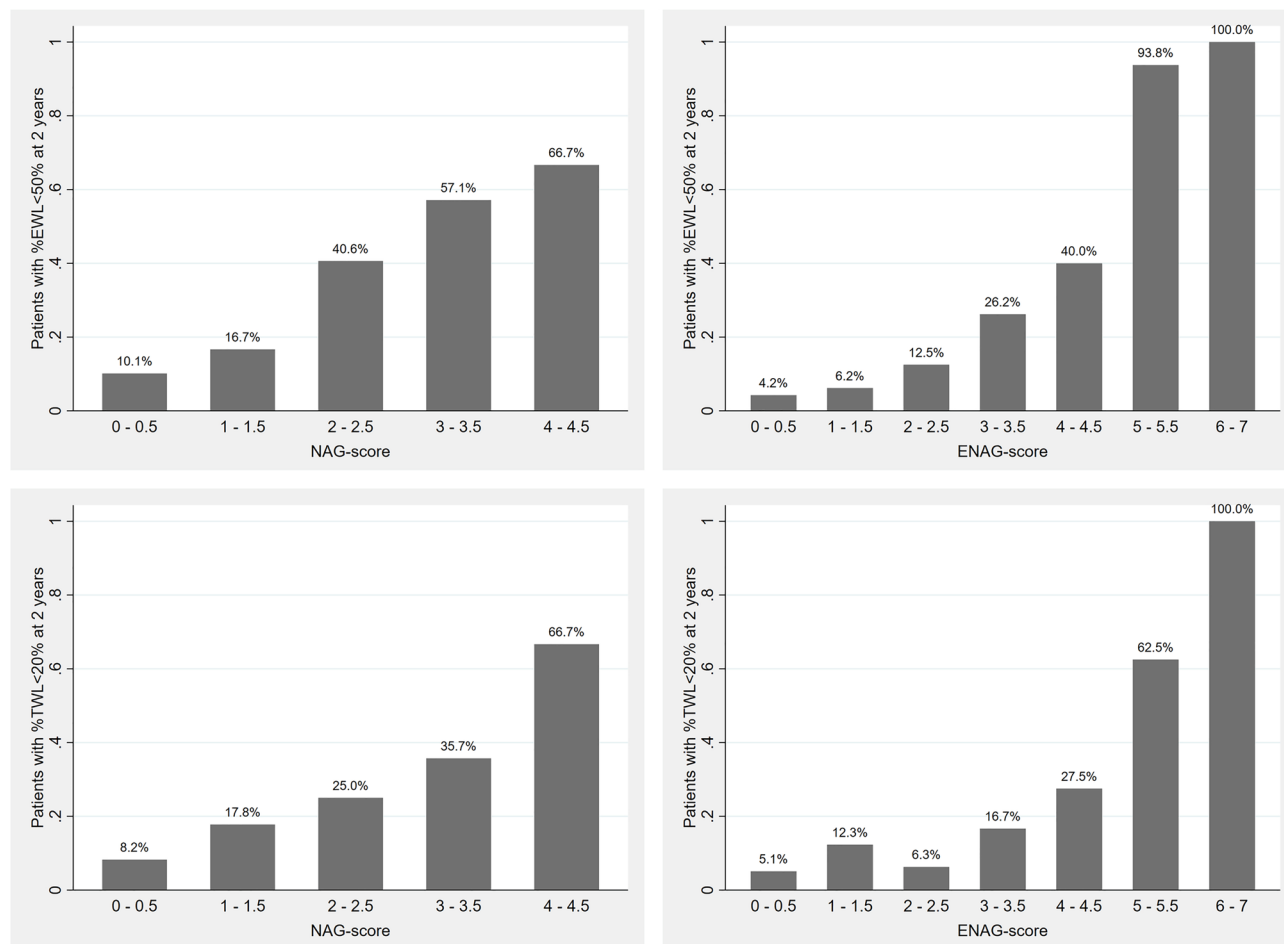


Figure 1. Descriptive graphs of BS unsuccessful risk, defined either as %EWL < 50% (upper row) or %TWL < 20% (lower row) at 2 years, according to NAG-score (left column) and ENAG-score (right column). BS bariatric surgery, EWL excess weight loss, TWL total weight loss.

variables retaining statistical significance were %EWL at 6 months $\leq 45\%$ (OR = 13.34, 95%CI 6.14–28.98), neck circumference ≥ 44 cm (OR = 3.81, 95% CI 1.52–9.57), age ≥ 50 years (OR = 2.40, 95% CI 1.18–4.88), and fasting glucose ≥ 118 mg/dL (OR = 3.40, 95% CI 1.43–8.07) (Table 4). All other variables (i.e., male sex, OSA, waist circumference ≥ 142 cm) were excluded as non-significant at multivariate analysis.

A significant increase in model accuracy could be noted (AUC = 0.846). The Hosmer–Lemeshow test did not reveal any significant miscalibration ($p = 0.68$). Internal validation of the model was performed through ten-fold cross-validation; the final estimation of the model performance on unseen data, obtained as the average AUC over the ten iterations, was equal to 0.817, thus reassuring about a modest overfitting effect.

The four variables selected by the model were used to develop a discrete-point prediction score (Table 4), which was named “ENAG-score” (Early loss, Neck circumference, Age, Glucose), and was defined by the sum of all four components, on a 7-point-scale. This mild simplification did not lead to a significant reduction in the predictive power of the model, since the AUC only slightly declined from 0.846 to 0.845.

A descriptive graph of the risk of unsuccessful weight loss after BS according to ENAG-score is presented in Fig. 1; the performance of the model in predicting BS unsuccessful was evaluated both in terms of %EWL and %TWL, with similar results. ID3 algorithm clustered the observations in three distinct and clinically relevant

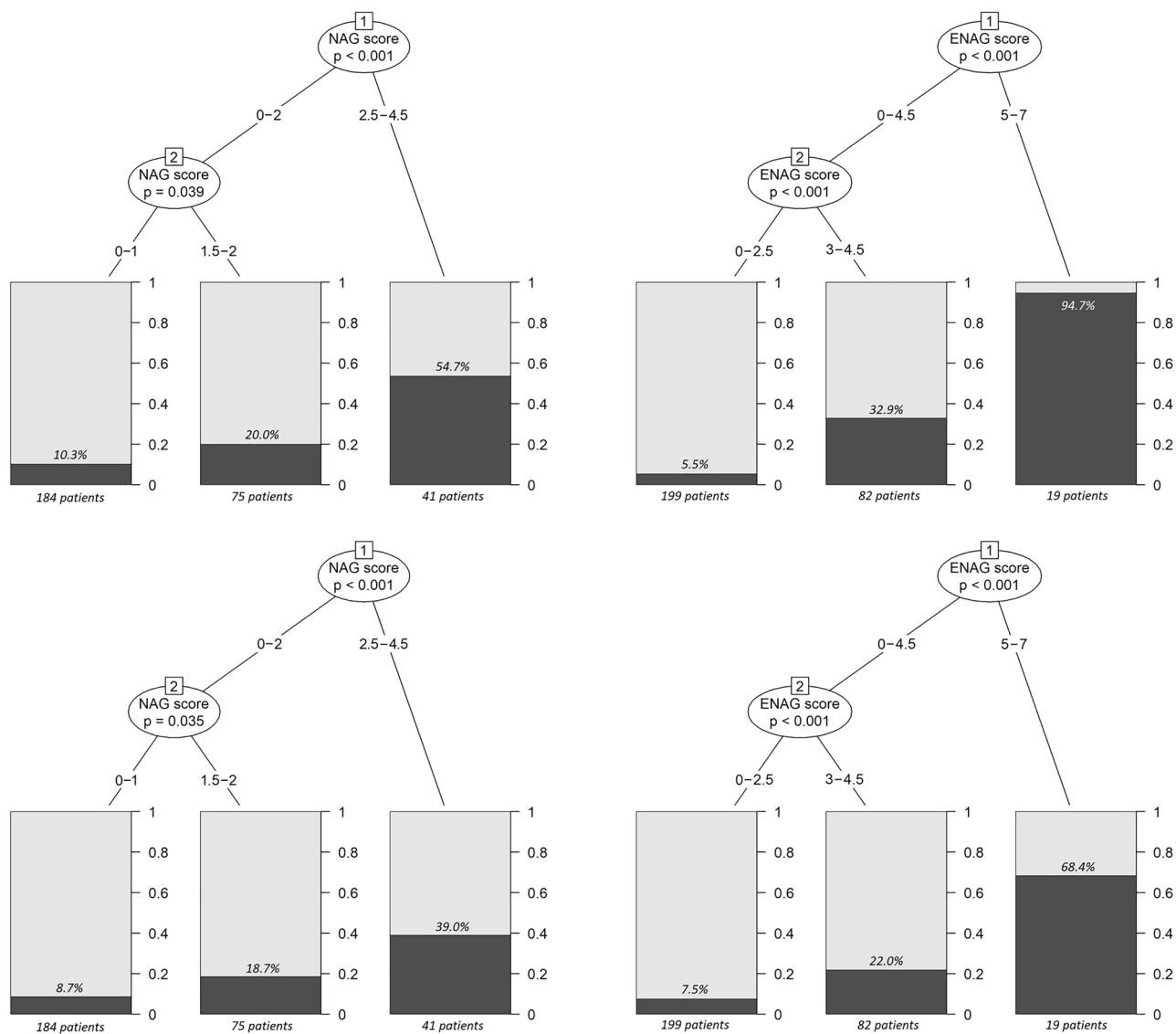


Figure 2. Stratification of BS unsuccessful risk, defined either as %EWL < 50% (upper row) or %TWL < 20% (lower row) at 2 years, based on NAG-score (left column) and ENAG-score (right column). Relevant cut-offs for class distinction and patients’ clustering were automatically retrieved through ID3 algorithm. BS bariatric surgery, EWL excess weight loss, ID3 Iterative Dichotomiser 3, TWL total weight loss.

Risk class as identified by ID3 algorithm	NAG-score	N° of patients	2-year BS unsuccessful risk (defined as %EWL < 50%)	2-year BS unsuccessful risk (defined as %TWL < 20%)
Low risk	0–1	184	10.3%	8.7%
Intermediate risk	1.5–2	75	20.0%	18.7%
High risk	2.5–4.5	41	54.7%	39.0%

Table 3. Pre-surgical estimation of unsuccessful weight loss at 2-years from BS according to NAG-score risk classes. BS bariatric surgery, EWL excess weight loss, ID3 Iterative Dichotomiser 3, TWL total weight loss.

risk classes (0–2.5 points, 3–4.5 points, 5–7 points), with no differences whether considering %EWL < 50% or %TWL < 20% as the reference outcome (Fig. 2); given their clinical correlates, these classes were labelled again as “low-risk”, “intermediate-risk”, and “high-risk”, respectively. As reported in Table 5, when considering %EWL, the low-risk class comprised 199 patients, with a 5.5% risk of unsuccessful weight loss at 2-years from BS. The intermediate-risk class comprised 82 patients, with a 32.9% risk of unsuccessful post-surgical weight loss. The high-risk class comprised 19 patients, with a 94.7% risk of unsuccessful post-surgical weight loss. The stratification performance of the model across different risk classes was overall preserved when adopting %TWL as the

Variable	OR	95% CI	p-value	β -coefficient	Normalized coefficient	Points for ENAG-score
Neck circumference ≥ 44 cm	3.81	1.52–9.57	0.004	+1.338	1.531	+1.5
Age ≥ 50 years	2.40	1.18–4.88	0.016	+0.874	1.000	+1
Fasting glucose ≥ 118 mg/dL	3.40	1.43–8.07	0.006	+1.223	1.399	+1.5
EWL at 6 months $\leq 45\%$	13.34	6.14–28.98	<0.001	+2.591	2.965	+3

Table 4. Prediction of BS failure, defined as %EWL < 50% at 2 years, by multivariate logistic regression after stepwise backward selection of pre-surgical and early post-surgical data; ENAG-score point assignment according to multivariate regression coefficients. BS bariatric surgery, CI confidence interval, EWL excess weight loss, OR odds-ratio.

Risk class as identified by ID3 algorithm	ENAG-score	N° of patients	2-year BS unsuccess risk (defined as %EWL < 50%)	2-year BS unsuccess risk (defined as %TWL < 20%)
Low risk	0–2.5	199	5.5%	7.5%
Intermediate risk	3–4.5	82	32.9%	22.0%
High risk	5–7	19	94.7%	68.4%

Table 5. Early post-surgical estimation of unsuccessful weight loss at 2-years from BS according to ENAG-score risk classes. BS bariatric surgery, EWL excess weight loss, ID3 Iterative Dichotomiser 3, TWL total weight loss.

reference parameter; in this case, the risk of BS unsuccess was equal to 7.5%, 22.0%, and 68.4% in the low-risk, intermediate-risk, and high-risk class, respectively.

Discussion

Two multivariate models for pre-surgical and for early post-surgical prediction of BS weight loss after a 2-year follow-up were developed and internally validated. To facilitate their clinical use, two simplified scoring systems (NAG-score and ENAG-score) were derived by assigning integer or half-integer points to each of the included predictors.

The first model, based only on pre-surgical data, showed a moderate overall accuracy for the prediction of the outcome of interest, with an AUC of 0.713. The retrieved score, i.e., the NAG-score, can be a simple and useful tool to stratify different pre-surgical risks for BS failure and may be helpful, during pre-operative assessment, for an appropriate balance of patient's expectations and for a personalization of the intensity of the follow-up. In particular, a higher-intensity follow-up might be considered in patients with intermediate-to-high risk of unsuccessful post-surgical weight-loss, since, possibly, these patients are those who may benefit the most from a closer dietary and medical counselling^{1,4,49,50}.

The second model, which included also early post-operative weight-loss data at 6 months, showed good overall accuracy for the prediction of the outcome of interest, with an AUC of 0.846. The retrieved score, i.e., the ENAG-score, can be of value for early prediction of long-term outcomes of the surgical procedure. This can be useful as a further guide for the refinement of the follow-up continuation and for the personalization of the therapeutic approach; in particular, it may be helpful as a guide for the avoidance of therapeutic inertia in patients with an intermediate-to-high probability of unsuccessful post-surgical weight loss. In fact, these patients are those who may benefit the most from stricter counselling and more intensive lifestyle management, and that may be potential candidates for an early start of adjunctive pharmacological treatments^{1,4,49,50}.

It is interesting to note that the algorithm of stepwise backward selection led to the inclusion of exactly the same set of pre-surgical factors in both models; in particular, the parameters that were retained as statistically significant were larger neck circumference (≥ 44 cm), older age (≥ 50 years), and higher fasting glucose levels (≥ 118 mg/dL). These results were coherent with previous findings by other authors^{7,8,10,11,28}. Even more interestingly, their predictive capacity remained significant after adding to the model a very robust parameter such as early post-operative weight loss; this further supported their relevance as independent predictors of BS outcomes in the longer term. On the other hand, male sex, OSA, and larger waist circumference (≥ 142 cm), though significant at univariate analysis, lost their significance in the multivariate model. This is not surprising, and it is likely a consequence of the multiple collinearities between these variables and those retained in the scores. Nevertheless, the fact that neck circumference performed better than waist circumference in the prediction of BS outcomes is noteworthy. Neck circumference, indeed, relates to oropharyngeal fat infiltration, which narrows the upper respiratory tract and is a more stable index because it is not affected by eating or body position or respiratory rates as the waist circumference measurement does⁵¹. Furthermore, the neck fat depot has been reported to be strongly associated with cardiometabolic and atherosclerotic diseases independent of visceral and whole-body obesity^{51–55}. Ectopic fat deposition is dysfunctional and associated with chronic sub-clinic inflammation, oxidative stress, and endothelial dysfunction, and upper-body subcutaneous fat delivers more free acids than visceral fat in the systemic circulation, thus contributing to increased insulin resistance and other dysmetabolic disorders^{53,55,56}. In particular, neck circumference appears to be a unique pathologic fat depot with

a unique genetic basis, independent of BMI⁵⁷. Finally, increased neck circumference is also a risk factor for OSA, which, in turn, is associated with increased cardiometabolic risk⁵³.

The main strength of our study was the simplicity of the proposed scoring systems, which was achieved upon the categorization of predictive variables through a robust supervised algorithmic approach. Moreover, the internal validation of our model, together with the assessment of its good calibration, conferred higher consistency to the obtained results, which were further strengthened by the reproducibility of risk-class stratification over two different BS-success defining parameters (%EWL and %TWL).

Our study had also some limitations. First, its retrospective design limited the possibility to take into account some other potential predictors, such as psychosocial data and physical activity levels; their inclusion might have led to a more complete and better-performing scoring system and could have allowed the assessment of causal effects through Mendelian randomization analysis procedures based on genomic background and environmental exposure data^{58–62}, which might be the subject for future research. Second, stronger evidence would have been achieved by considering a longer follow-up time; however, the time-point which has been examined (i.e., 2 years after BS) was longer than in other studies proposing equations or scoring systems for the prediction of BS outcomes^{21,31–36}, and was considered as a time-point of weight stabilization in most patients^{29,30,63,64}. Third, the sample size was not sufficient to develop different scoring systems for SG and RYGB; however, the type of intervention was taken into account during model development, and no significant differences emerged between the two procedures in terms of outcomes. Fourth, our study comprised only Caucasian patients; as there is extensive evidence of weight loss variability among ethnical groups^{65–68}, the generalizability of our data to other populations is uncertain.

In conclusion, our study proposed two simple scoring systems (the NAG-score and the ENAG-score) for pre-surgical and early post-surgical prediction of 2-year BS weight loss. The presented data supported their consistency as easy-to-use estimation tools. Further studies are needed to confirm their external validity on different patient cohorts; if so, their application in clinical practice might provide a simple and objective instrument for the evaluation of BS failure risk, which may be useful to guide pre-operative patient's assessment, to appropriately balance patient's expectations, and to manage more effectively the post-operative care.

Data availability

The dataset analyzed during the current study is available from the corresponding author on reasonable request.

Received: 3 June 2021; Accepted: 11 October 2021

Published online: 26 October 2021

References

- Fried, M. *et al.* Interdisciplinary European guidelines on metabolic and bariatric surgery. *Obes. Facts* **6**, 449–468 (2013).
- Pucci, A. & Batterham, R. L. Mechanisms underlying the weight loss effects of RYGB and SG: Similar, yet different. *J. Endocrinol. Invest.* **42**, 117–128 (2019).
- Osland, E., Yunus, R. M., Khan, S., Memon, B. & Memon, M. A. Weight loss outcomes in Laparoscopic Vertical Sleeve Gastrectomy (LVSG) versus laparoscopic Roux-en-Y Gastric Bypass (LRYGB) procedures: A meta-analysis and systematic review of randomized controlled trials. *Surg. Laparosc. Endosc. Percutaneous Tech.* **27**, 8–18 (2017).
- Di Lorenzo, N. *et al.* Clinical practice guidelines of the European Association for Endoscopic Surgery (EAES) on bariatric surgery: Update 2020 endorsed by IFSO-EC EASO and ESPCOP. *Surg. Endosc.* **34**, 2332–2358 (2020).
- Karpińska, I. A. *et al.* Is it possible to predict weight loss after bariatric surgery?—external validation of predictive models. *Obes. Surg.* <https://doi.org/10.1007/s11695-021-05341-w> (2021).
- Al-Khyatt, W., Ryall, R., Leeder, P., Ahmed, J. & Awad, S. Predictors of inadequate weight loss after laparoscopic gastric bypass for morbid obesity. *Obes. Surg.* **27**, 1446–1452 (2017).
- Sans, A. *et al.* Baseline anthropometric and metabolic parameters correlate with weight loss in women 1-year after laparoscopic Roux-en-Y gastric bypass. *Obes. Surg.* **27**, 2940–2949 (2017).
- Faria, G. *et al.* Fasting glycemia: A good predictor of weight loss after RYGB. *Surg. Obes. Relat. Dis.* **10**, 419–424 (2014).
- Barhouch, A. S. *et al.* Predictors of excess weight loss in obese patients after gastric bypass: A 60-month follow-up. *Obes. Surg.* **26**, 1178–1185 (2016).
- Masrur, M. *et al.* Factors associated with weight loss after metabolic surgery in a multiethnic sample of 1012 patients. *Obes. Surg.* **30**, 975–981 (2020).
- Colles, S. L., Dixon, J. B. & O'Brien, P. E. Hunger control and regular physical activity facilitate weight loss after laparoscopic adjustable gastric banding. *Obes. Surg.* **18**, 833–840 (2008).
- Ansar, H., Zamaninour, N., Pazouki, A. & Kabir, A. Weight loss after one anastomosis gastric bypass-mini gastric bypass (OAGB-MGB): Patient-related perioperative predictive factors. *Obes. Surg.* **30**, 1316–1323 (2020).
- Black, D. W., Goldstein, R. B. & Mason, E. E. Psychiatric diagnosis and weight loss following gastric surgery for obesity. *Obes. Surg.* **13**, 746–751 (2003).
- Alvarado, R. *et al.* The impact of preoperative weight loss in patients undergoing laparoscopic Roux-en-Y gastric bypass. *Obes. Surg.* **15**, 1282–1286 (2005).
- Diab, A., Rahman, F., Abdurasul, E. M. & Diab, F. H. The effect of age, gender, and baseline BMI on weight loss outcomes in obese patients undergoing intragastric balloon therapy. *Obes. Surg.* **29**, 3542–3546 (2019).
- Chen, E. Y. *et al.* Body mass index as a predictor of 1-year outcome in gastric bypass surgery. *Obes. Surg.* **19**, 1240–1242 (2009).
- Riess, K. P., Baker, M. T., Lambert, P. J., Mathiason, M. A. & Kothari, S. N. Effect of preoperative weight loss on laparoscopic gastric bypass outcomes. *Surg. Obes. Relat. Dis.* **4**, 704–708 (2008).
- Carlin, A. M., O'Connor, E. A., Genaw, J. A. & Kavar, S. Preoperative weight loss is not a predictor of postoperative weight loss after laparoscopic Roux-en-Y gastric bypass. *Surg. Obes. Relat. Dis.* **4**, 481–485 (2008).
- Sarela, A. I., Dexter, S. P. L., O'Kane, M., Menon, A. & McMahon, M. J. Long-term follow-up after laparoscopic sleeve gastrectomy: 8–9-year results. *Surg. Obes. Relat. Dis.* **8**, 679–684 (2012).
- Krimpuri, R. D. *et al.* Qualifying for bariatric surgery: Is preoperative weight loss a reliable predictor of postoperative weight loss? *Surg. Obes. Relat. Dis.* **14**, 60–64 (2018).
- Chen, C. Y. *et al.* Check point to get adequate weight loss within 6-months after laparoscopic sleeve gastrectomy for morbid obesity in Asian population. *Sci. Rep.* **10**, 2 (2020).

22. Hindle, A., de la Piedad Garcia, X. & Brennan, L. Early post-operative psychosocial and weight predictors of later outcome in bariatric surgery: A systematic literature review. *Obes. Rev.* **18**, 317–334 (2017).
23. Mor, A., Sharp, L., Portenier, D., Sudan, R. & Torquati, A. Weight loss at first postoperative visit predicts long-term outcome of Roux-en-Y gastric bypass using Duke weight loss surgery chart. *Surg. Obes. Relat. Dis.* **8**, 556–560 (2012).
24. Puzziferri, N. *et al.* Variations of weight loss following gastric bypass and gastric band. *Ann. Surg.* **248**, 233–242 (2008).
25. Ritz, P. *et al.* Early prediction of failure to lose weight after obesity surgery. *Surg. Obes. Relat. Dis.* **9**, 118–121 (2013).
26. Obeidat, F. & Shanti, H. Early weight loss as a predictor of 2-year weight loss and resolution of comorbidities after sleeve gastrectomy. *Obes. Surg.* **26**, 1173–1177 (2016).
27. Vuolo, G. *et al.* Is sleeve gastrectomy a therapeutic procedure for all obese patients?. *Int. J. Surg.* **30**, 48–55 (2016).
28. D'Eusebio, C. *et al.* What predicts the unsuccess of bariatric surgery? An observational retrospective study. *J. Endocrinol. Invest.* <https://doi.org/10.1007/s40618-020-01398-z> (2020).
29. Maciejewski, M. L. *et al.* Bariatric surgery and long-term durability of weight loss. *JAMA Surg.* **151**, 1046–1055 (2016).
30. Kindel, T., Lomelin, D., McBride, C., Kothari, V. & Thompson, J. The time to weight-loss steady state after gastric bypass predicts weight-loss success. *Obes. Surg.* **26**, 327–331 (2016).
31. Cottam, S. *et al.* The use of predictive markers for the development of a model to predict weight loss following vertical sleeve gastrectomy. *Obes. Surg.* **28**, 3769–3774 (2018).
32. Wise, E. S., Hocking, K. M. & Kavac, S. M. Prediction of excess weight loss after laparoscopic Roux-en-Y gastric bypass: Data from an artificial neural network. *Surg. Endosc.* **30**, 480–488 (2016).
33. Goulart, A. *et al.* Doctor, how much weight will i lose?—a new individualized predictive model for weight loss. *Obes. Surg.* **26**, 1357–1359 (2016).
34. Seyssel, K. *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.* **28**, 3393–3399 (2018).
35. Janik, M. R. *et al.* Setting realistic expectations for weight loss after laparoscopic sleeve gastrectomy. *Wideochirurgia I Inne Tech. Maloinwazyjne* **14**, 415–419 (2019).
36. Velázquez-Fernández, D. *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.* **29**, 2878–2885 (2019).
37. Grover, B. T. *et al.* Defining weight loss after bariatric surgery: A call for standardization. *Obes. Surg.* **29**, 3493–3499 (2019).
38. van Rijswijk, A. S., van Olt, N., Schats, W., van der Peet, D. L. & van de Laar, A. W. What is weight loss after bariatric surgery expressed in percentage total weight loss (%TWL)? A systematic review. *Obes. Surg.* **31**, 3833–3847 (2021).
39. Brissman, M., Beamish, A. J., Olbers, T. & Marcus, C. Prevalence of insufficient weight loss 5 years after Roux-en-Y gastric bypass: Metabolic consequences and prediction estimates: A prospective registry study. *BMJ Open* **11**, 1–8 (2021).
40. Tu, Y. *et al.* A total weight loss of 25% shows better predictivity in evaluating the efficiency of bariatric surgery. *Int. J. Obes.* **45**, 396–403 (2021).
41. Chung, F., Yang, Y. & Liao, P. Predictive performance of the stop-bang score for identifying obstructive sleep apnea in obese patients. *Obes. Surg.* **23**, 2050–2057 (2013).
42. Qaseem, A. *et al.* Diagnosis of obstructive sleep apnea in adults: A clinical practice guideline from the American College of Physicians. *Ann. Intern. Med.* **161**, 210–220 (2014).
43. Kapur, V. K. *et al.* Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: An American academy of sleep medicine clinical practice guideline. *J. Clin. Sleep Med.* **13**, 479–504 (2017).
44. Tsai, C.-J., Lee, C.-I. & Yang, W.-P. A discretization algorithm based on class-attribute contingency coefficient. *Inf. Sci. (Nij)* **178**, 714–731 (2008).
45. Steyerberg, E. W. *et al.* Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J. Clin. Epidemiol.* **54**, 774–781 (2001).
46. Moons, K. G. M. *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann. Intern. Med.* **162**, W1–W73 (2015).
47. Quinlan, J. R. Induction of decision trees. *Mach. Learn.* **1**, 81–106 (1986).
48. Ogheneovo, E. E. & Nlerum, P. A. Iterative dichotomizer 3 (ID3) decision tree: A machine learning algorithm for data classification and predictive analysis. *Int. J. Adv. Eng. Res. Sci.* **7**, 514–521 (2020).
49. Rudolph, A. & Hilbert, A. Post-operative behavioural management in bariatric surgery: A systematic review and meta-analysis of randomized controlled trials. *Obes. Rev.* **14**, 292–302 (2013).
50. Stewart, F. & Avenell, A. Behavioural interventions for severe obesity before and/or after bariatric surgery: A systematic review and meta-analysis. *Obes. Surg.* **26**, 1203–1214 (2016).
51. Namazi, N., Larjani, B., Surkan, P. & Azadbakht, L. The association of neck circumference with risk of metabolic syndrome and its components in adults: A systematic review and meta-analysis. *Nutr. Metab. Cardiovasc. Dis.* **28**, 657–674 (2018).
52. Aoi, S. *et al.* Association of changes in neck circumference with cardiometabolic risk in postmenopausal healthy women. *J. Atheroscler. Thromb.* **23**, 728–736 (2016).
53. Silva, A. *et al.* Neck circumference and 10-year cardiovascular risk at the baseline of the ELSA-Brasil study: Difference by sex. *Arg. Bras. Cardiol.* **115**, 840–848 (2020).
54. Shokri-Mashhadi, N., Moradi, S., Mohammadi, H., Ghavami, A. & Rouhani, M. Association between neck circumference and lipid profile: A systematic review and meta-analysis of observational studies. *Eur. J. Cardiovasc. Nurs.* <https://doi.org/10.1093/EURJCN/ZVAA018> (2021).
55. Shi, J. *et al.* Neck circumference as an independent predictor for NAFLD among postmenopausal women with normal body mass index. *Nutr. Metab.* **18**, 2 (2021).
56. Nielsen, S., Guo, Z., Johnson, C., Hensrud, D. & Jensen, M. Splanchnic lipolysis in human obesity. *J. Clin. Invest.* **113**, 1582–1588 (2004).
57. Liu, Y. *et al.* Genome-wide association study of neck circumference identifies sex-specific loci independent of generalized adiposity. *Int. J. Obes.* **45**, 1532–1541 (2021).
58. Wang, X. *et al.* Genetic support of a causal relationship between iron status and type 2 diabetes: A Mendelian randomization study. *J. Clin. Endocrinol. Metab.* <https://doi.org/10.1210/CLINEM/DGAB454> (2021).
59. Zhang, F. *et al.* Causal influences of neuroticism on mental health and cardiovascular disease. *Hum. Genet.* **140**, 1267–1281 (2021).
60. Zhang, F. *et al.* Genetic evidence suggests posttraumatic stress disorder as a subtype of major depressive disorder. *J. Clin. Invest.* <https://doi.org/10.1172/JCI145942> (2021).
61. Hou, L. *et al.* Exploring the causal pathway from ischemic stroke to atrial fibrillation: a network Mendelian randomization study. *Mol. Med.* **26**, 2 (2020).
62. Wu, Y. *et al.* Multi-trait analysis for genome-wide association study of five psychiatric disorders. *Transl. Psychiatry* **10**, 2 (2020).
63. Buchwald, H. *et al.* Bariatric surgery: A systematic review and meta-analysis. *J. Am. Med. Assoc.* **292**, 1724–1737 (2004).
64. Ballantyne, G. H. Measuring outcomes following bariatric surgery: Weight loss parameters, improvement in co-morbid conditions, change in quality of life and patient satisfaction. *Obes. Surg.* **13**, 954–964 (2003).
65. Anderson, W. A., Greene, G. W., Forse, R. A., Apovian, C. M. & Istfan, N. W. Weight loss and health outcomes in African Americans and whites after gastric bypass surgery. *Obesity* **15**, 1455–1463 (2007).

66. Harvin, G., Delege, M. & Garrow, D. A. The impact of race on weight loss after roux-en-Y gastric bypass surgery. *Obes. Surg.* **18**, 39–42 (2008).
67. Buffington, C. K. & Marema, R. T. Ethnic differences in obesity and surgical weight loss between African-American and Caucasian females. *Obes. Surg.* **16**, 159–165 (2006).
68. Guajardo-Salinas, G. E., Hilmy, A. & Martinez-Ugarte, M. L. Predictors of weight loss and effectiveness of Roux-en-Y gastric bypass in the morbidly obese Hispano-American population. *Obes. Surg.* **18**, 1369–1375 (2008).

Author contributions

F.B. conceptualized the work, performed the data analysis and contributed to the manuscript writing. M.P., C.D.E., S.Bos. and F.R. contributed to the data collection and manuscript revision. S.A., M.T. and M.M. contributed to data interpretation and manuscript writing. A.D.F. and E.G. supervised the manuscript drafting. S.B. contributed to work conceptualization, data analysis, manuscript writing and final draft supervision. All authors approved the manuscript in its final form.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-00475-4>.

Correspondence and requests for materials should be addressed to F.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021