IFSO

ORIGINAL CONTRIBUTIONS

CrossMark

Sensitivity and Specificity of 50% Excess Weight Loss (50%EWL) and Twelve Other Bariatric Criteria for Weight Loss Success

Arnold W. van de Laar[1] · A. S. van Rijswijk[1] · H. Kakar[2] · S. C. Bruin[1]

Abstract

Background Criteria for bariatric weight loss success are numerous. Most of them are arbitrary. None of them is evidence-based. Our objective was to determine their sensitivity and specificity.

Methods Thirteen common bariatric weight loss criteria were compared to a benchmark reflecting the gold standard in bariatric surgery. We used an elaborate baseline BMI-independent weight loss percentile chart, based on retrospective data after laparoscopic Roux-en-Y gastric bypass (LRYGB), performed between 2007 and 2017. Percentile curves p31.6 (patients' expectation), p25 (interquartile range), p15.9 (1 standard deviation (SD) below median), and p10.9 (surgeons' goal) were used as possible cutoff for success to determine true or false positive and negative results beyond 1 year.

Results We operated 4497 primary LRYGB patients, with mean follow-up 22 (± 1 SD 19; range 0–109) months, 3031 patients with last result ≥ 1 year, 518 ≥ 5 years. For all four cutoff percentile curves for success, specificities were low (2–72%) for criteria < 35 body mass index (BMI), ≥ 25 percentage excess BMI loss (%EBMIL), ≥ 50%EBMIL, ≥ 15 percentage total weight loss (%TWL), ≥ 20%TWL, ≥ 25 percentage excess weight loss (%EWL), and high (83–96%) for < 30 BMI. No criterion had > 80% specificity and sensitivity for a cutoff above p15.9. For p15.9, they were both > 80% for criteria ≥ 10 BMI reduction and ≥ 50%EWL, both > 90% for ≥ 25%TWL and ≥ 35 percentage alterable weight loss (%AWL). All criteria had high sensitivities for all cutoff percentile curves (87–100%), except < 30 BMI (65–78%).

Conclusions For the first time, common bariatric criteria for weight loss success were systematically validated. Most criteria recognized success very well (high sensitivities), but ≥ 15%TWL, ≥ 20%TWL, < 35BMI, ≥ 25%EWL, ≥ 25%EBMIL, and ≥ 50%EBMIL left too many poor responders unnoticed (low specificities). Bariatric weight loss success is best assessed by comparing results to percentile curve 1 SD below median (p15.9) in a bariatric baseline BMI-independent weight loss percentile chart. Criteria ≥ 35%AWL and ≥ 25%TWL came close to that curve, both with > 90% sensitivity and specificity. Among others, criterion ≥ 50%EBMIL did not.

Keywords Specificity · Sensitivity · Weight loss · Success · Responder · Criteria · Bariatric · LRYGB · 50%EWL · 50%EBMIL · 20%TWL · 35%AWL · Alterable weight loss · Weight regain · Chart · Percentile · %AWL · %TWL · %WL · %EWL · %EBMIL · Gold standard

Introduction

Success of a bariatric procedure should be measured, both primarily and in the long term, by postoperative weight loss, improvement of (metabolic) comorbidities, impact on quality-of-life, and complications caused by the operation. For three of these effects, success can readily be defined, with standardized criteria for comorbidities, validated questionnaires for

✉ Arnold W. van de Laar
arnoldvandelaar@usa.net

A. S. van Rijswijk
anne-sophie.vanrijswijk@slz.nl

H. Kakar
hala_kakar@hotmail.nl

S. C. Bruin
sjoerd.bruin@slz.nl

1   MC Slotervaart Hospital, Louwesweg 6,
     1066EC Amsterdam, The Netherlands

2   University of Amsterdam, Spui 21,
     1012WX Amsterdam, The Netherlands

🌐 Springer

quality-of-life, and clear classifications for postoperative complications [1–4]. For bariatric weight loss however, defining success remains troublesome.

Fundamentally, successful results would lay around or above the average outcome that can be expected with a specific therapy. A criterion for success should serve as a clear cutoff between those results below average that are still acceptable (successful because close enough below average) and those that are not (unsuccessful because too far below average: poor responders).

This simple premise becomes problematic when applied to bariatric weight loss. First of all, there is no average weight loss outcome in bariatric surgery. Each type of procedure yields different results [5]. Common criteria for bariatric weight loss success do not differentiate for that. To validate them, a choice has to be made on a benchmark procedure to compare them with. One that generally is accepted as gold standard in bariatric surgery. The laparoscopic Roux-en-Y gastric bypass (LRYGB) could be considered as such. Until very recently, it was the most performed bariatric procedure worldwide, with average outcome between that of the increasingly popular sleeve gastrectomy and malabsorptive procedures like duodenal switch [5, 6].

Second, bariatric weight loss is dynamic over time [7]. Early weight loss success would therefore look different from success after 1, 5, or 10 years. To validate weight loss criteria, they should be compared to a benchmark that is dynamic as well. A weight loss percentile chart, similar to growth charts in children's healthcare, would fill that need, picturing the course and spread of weight loss over time in one graph.

A third major problem is commonly overlooked. With common outcome metrics, postoperative weight loss results are influenced by differences in body mass index (BMI) at baseline. There is strong evidence that this distorting effect can be significant in bariatric surgery, where patients typically show a wide range of baseline BMI, from 35 kg/m^2 to more than double [8]. To validate bariatric criteria, they should be compared to a benchmark that is influenced by this effect as least as possible. Different researchers proved that outcome metric percentage total weight loss (%TWL) is affected less than percentage excess BMI loss (%EBMIL) [9–13]. In 2011, Baltasar et al. described a metric that seemed to be not affected at all [14]. It was found with results of 7410 Spanish bariatric patients and validated with outcome of 265 different subjects [15]. In 2013, van de Laar described a less elaborate baseline BMI-independent metric called percentage alterable weight loss (%AWL), based on a similar algorithm found with results of 8945 American subjects and validated with outcome of 500 Dutch LRYGB patients [16, 17]. It was confirmed for sleeve gastrectomy and lower BMI's and proven superior to other weight loss metrics when compared to metabolic outcome after LRYGB [18, 19]. An overview of this relatively new weight loss metric is presented in Table 1.

These findings paved the way for constructing a new kind of bariatric benchmark; a baseline BMI-independent, time-dependent percentile chart [20]. Such benchmark (provided that it is based on outcome of a large and representative cohort of patients after the gold standard bariatric operation) would be ideal to test criteria for bariatric weight loss success.

With charts based on American and Dutch LRYGB data, it was already demonstrated that popular criteria ≥ 50%EBMIL and < BMI35 are weak in recognizing insufficient weight loss [20, 23]. At time, all further evidence on the validity of common bariatric criteria is expert opinion only [24]. In this study, we updated our elaborated LRYGB percentile chart and used this benchmark as a representation of the bariatric gold standard to systematically establish the sensitivity and specificity of commonly used criteria for bariatric weight loss success.

Materials and Methods

Benchmark

To validate common criteria for bariatric weight loss success, one needs a benchmark to compare them to preferably based on outcome of the gold standard bariatric procedure. We chose LRYGB outcome for this purpose, expressed in a %AWL-based percentile chart, being both independent of common differences in baseline BMI among patients and taking into account the dynamic weight changes over time. Such gold standard benchmark should be based on a cohort that is both sufficiently large and representative. We therefore used our own large LRYGB cohort and compared our results with outcome of landmark studies in bariatric research. For that purpose, we searched bariatric literature for studies that reported LRYGB outcome of at least 250 patients, with more than 1-year follow-up and expressed with %AWL or %TWL, two metrics least influenced by differences in baseline BMI.

We retrospectively analyzed data from all consecutive primary LRYGB patients operated in our bariatric center of excellence from the beginning of our bariatric program in 2007, until January 2017. Revision procedures were excluded, because results could have been influenced by preceding bariatric surgery. Patients were urged to lose weight preoperatively to reduce the operative risk. All procedures were standardized, with 4-cm × 8-cm gastric pouch, 50-cm biliary limb, and 150-cm alimentary limb. Patients were weighed with the same type of scale at all pre- and postoperative visits. All data were collected prospectively in a central electronic database. We used the measurement at first preoperative visit for baseline weight [21]. We specified age, gender, and type 2 diabetes (defined with the American Diabetes Association criteria) as potential confounders for weight loss, to illustrate the representability of the cohort [1, 25]. We only used the last measured weight for each patient postoperatively. Thus, all

Table 1 Background information on the percentage alterable weight loss (%AWL) metric

The percentage alterable weight loss (%AWL) metric in a nutshell			
What?	The %AWL formula is body mass index (BMI) based and resembles the formulas of common metrics % excess BMI loss (%EBMIL) and % total weight loss (%TWL).		
	%EBMIL=	100%×	BMI loss / (BMI − 25)
	%AWL=	100%×	BMI loss / (BMI − 13)
	%TWL=	100%×	BMI loss / (BMI − 0)
How?	The %AWL algorithm was found in 2013, by analyzing nearly 9000 gastric bypass patients from the American Bariatric Outcomes Longitudinal Database (BOLD) and validated with outcome of over 3000 bariatric patients [16–18, 20].		
Why?	With %EBMIL, heavier bariatric patients appear to have (on average) lower weight loss results than lighter patients, while the opposite occurs (to a lesser extent) with %TWL [9–13]. These little-known features lead to biased conclusions when comparing unequally heavy subjects, favoring lighter patients with %EBMIL and heavier with %TWL [8, 21, 22]. With %AWL, heavier, and lighter patients turn out to have comparable results.		
Pros	%AWL allows comparing results of heavier (groups of) patients to those of lighter (groups of) patients without preoperative BMI bias.		
	%AWL allows expressing bariatric outcome in percentile charts. Percentile charts are extremely intuitive and conceivable [20, 23].		
	The metabolic effect on type 2 diabetes mellitus after bariatric surgery matches the bariatric effect (weight loss) better with %AWL, than with %EBMIL or %TWL [19].		
Cons	Its rationale is not clear. The denominator (BMI − 13) kg/m² might reflect the portion of an adult human body that is affected by weight loss (the "alterable" part: fat, muscles, water), with 13 kg/m² being the "inert" part (bone, connective tissues).		
	%AWL is not an intuitive metric, whereas %EBMIL and %TWL are more conceivable.		
	%AWL is not (yet) a well-known outcome measure.		

patients contributed to the chart equally, even if lost in follow-up, with only one result. In this way, the number of postoperative visits (and therefore the compliance to follow-up) could not introduce a bias. Weight loss was calculated with five different metrics:

1. BMI reduction = (baseline weight − last weight) / (body length)^2.
2. %TWL = 100% × BMI reduction / baseline BMI.
3. %AWL = 100% × BMI reduction / (baseline BMI − 13), as described by van de Laar et al. [16, 17].
4. %EBMIL = 100% × BMI reduction / (baseline BMI − 25), as suggested by Deitel et al. [26].
5. Percentage excess weight loss (%EWL) = 100% × (baseline weight − last weight) / (baseline weight − ideal weight), based on the 1983 gender-specific Metropolitan Life insurance tables for ideal weight, using "medium frame" and subtracting 2.54 cm (the tables were based on people wearing "shoes with 1-in. heels") [27].

Out of these five metrics, only %AWL was proven to render baseline BMI-independent outcome and therefore was used for the percentile chart [17, 22]. We created percentile curves using Excel (Microsoft Office 2016) by determining percentiles for results grouped by consecutive intervals of 100 days and smoothened them using polynomial trendlines. We used the trendline formulas to judge the adequacy of each curve.

Criteria

From a 2015 review of bariatric literature on the subject by Mann et al., we chose 11 popular criteria for weight loss success: ≥ 50%EWL, ≥ 25%EWL, ≥ 50%EBMIL, ≥ 25%EBMIL, ≥ 20%TWL, ≥ 15%TWL, ≥ 10-kg/m² BMI reduction, thresholds < 30 BMI and < 35 BMI, < 10% weight regain relative to the absolute nadir weight (%WRnadir), and < 25% regain of the initial maximum weight loss (%WRmax) [28]. We added ≥ 25%TWL and ≥ 35%AWL.

Bariatric weight loss criteria are not very useful during the period of rapid initial weight loss. They are rather meant to assess the lowest (nadir) weight that results from it and the mid- and long-term weight thereafter. More than 95% of LRYGB patients reach their nadir weight only after the first year [20, 23]. We therefore used the benchmark chart only from 1 year upward to test the 11 weight loss criteria. For the two weight regain criteria < 10%WRnadir and < 25%WRmax, we compared the nadir results of the percentile curves with the corresponding results at 5 years, as suggested by Aminian et al. [29].

Cutoff

We looked at four %AWL percentile curves as possible cutoff between success (true positive) and failure (true negative). First, we asked 200 of the primary LRYGB patients preoperatively from which weight up they would expect to be disappointed in the long run. They were subjects taking part in a

different study on expectations and preferences with respect to bariatric surgery. With those answers and with their weight and height at baseline, we calculated their mean "expected disappointing %AWL" and applied that to our ≥ 5-year outcome to select the patients' expectation cutoff percentile. Second, we used the 25th percentile (p25), being the lower border of the interquartile range (IQR) as cutoff, as suggested by Mor et al. [30]. Third, we chose percentile p15.9 as cutoff, representing 1 standard deviation (SD) below median. Finally, we applied ≥ 50%EBMIL, the criterion most commonly used by surgeons, to our outcome and used the percentage of results after the first year that were < 50%EBMIL to indicate the surgeons' goal cutoff percentile [24].

Sensitivity and Specificity

We expressed sensitivity and specificity of the 11 criteria for successful weight loss as percentages and calculated them with the number of false positive results (successful according to the criterion, but below the cutoff percentile curve) and false negative results (unsuccessful according to the criterion, but above the cutoff percentile curve), with respect to each of the four cutoff percentiles. We considered a criterion "inadequate" for a specific cutoff if either sensitivity or specificity was below 60%. We considered a criterion "useful" for a specific cutoff if both sensitivity and specificity were above 80%.

Results

Benchmark

Until January 2017, we performed 4497 primary LRYGB procedures. These patients had mean age 43 (range 18–66) years, mean baseline BMI 43.4 (range 34.5–71.5) kg/m^2, and 16.7% were male, 10.3% were super-obese (BMI ≥ 50 kg/m^2), and 22.6% had type 2 diabetes. There was no missing data for these variables. Time between first preoperative visit and operation was median 5.4 months. Preoperatively, they lost mean 6.2 (± 1 SD 4.1) kg. The longest possible follow-up per patient was median 24.3 months, mean 31.3 (range 0.0–108.9) months, while the longest actual follow-up was mean 22.0 (± 1 SD 19.3) months. One patient died within 30 days from bowel ischemia (0.02% mortality). Of the 3535 patients that were ≥ 1-year postoperative, 3031 (85.7%) had ≥ 1-year follow-up. Of the 518 patients that were ≥ 5-year postoperative, 304 (58.7%) had ≥ 5-year follow-up. All 4497 last weight loss results are presented in Fig. 1, with percentile curves p50 (median), p + 1 SD, and p − 1 SD.

Our data came close to outcome reported in literature. We found seven studies reporting %TWL of more than 250 patients, more than 1 year after LRYGB [7, 31–36]. These

results are presented together with weight loss of our patients with last results at 0.5, 1, 2, 3, 5, and 7 years in Table 2.

Our data confirmed that %AWL was the only metric that is baseline BMI-independent. For %TWL, %EWL, %EBMIL, BMI, and BMI-reduction, we found significant differences in mean outcome between the heaviest 1000 patients (mean baseline BMI 50.0 kg/m^2) and the lightest 1000 (mean baseline BMI 38.0 kg/m^2): + 3.4%TWL, − 11.8%EWL, − 19.2%EBMIL, + 6.9-kg/m^2 BMI, and + 5.1-kg/m^2 BMI reduction (all $p < 0.0001$; Mann-Whitney U test); but not for %AWL: − 0.1%AWL ($p = 0.896$).

Cutoff

In total, 3031 patients (67.4%) had their last measurement after the first year. Of these results, 10.9% were < 50%EBMIL. We therefore used p10.9 as the surgeons' goal cutoff percentile curve. Characteristics of the subgroup of 200 patients that were interviewed about their expectations after bariatric surgery were mean age 45 (range 18–66) years, mean baseline BMI 42.2 (range 35.0–58.8) kg/m^2, 22.0% male, 8.0% super-obese, and 12.0% type 2 diabetes. The distribution of their long-term expected disappointing weight loss results is presented in Fig. 2. The mean value was 34.6%AWL (or 23.9 (range 3.2–44.1) %TWL), which corresponded with percentile p31.6 of the 304 results at ≥ 5 years. We therefore used p31.6 as the patients' expectation cutoff percentile curve. Percentile curves p50 (median), p31.6 (patients' expectation cutoff), p25 (IQR cutoff), p15.9 (1 SD below median cutoff), and p10.9 (surgeons' goal cutoff) from 1 year up are presented in Fig. 3, together with all last results beyond 1 year that correspond with ≥ 50%EBMIL. The nadir results of these four cutoff percentile curves were 42.2%AWL at 17 months, 39.8%AWL at 16.5 months, 36.4%AWL at 15.5 months, and 33.8% at 15 months, respectively.

Sensitivity and Specificity

Sensitivity and specificity of the 11 criteria for weight loss success based on the 3031 results beyond 1 year and with reference to each of the four cutoff percentile curves are presented in Fig. 4. Weight loss criteria ≥ 25%EWL, ≥ 25%EBMIL, ≥ 15%TWL, and < 35BMI could be considered inadequate for all four cutoffs, as their specificities were (far) below 60%. Only four criteria ≥ 10 kg/m^2, ≥ 50%EWL, ≥ 25%TWL, and ≥ 35%AWL could be considered useful, as their sensitivities and specificities were all above 80%, but only for cutoffs p15.9 and p10.9.

All percentile curves showed weight regain. Compared to the < 10%WRnadir criterion, at 5 years, the four cutoff percentile curves gained 7.8, 8.5, 9.2, and 9.8% from their absolute nadir result for p31.6, p25, p15.9, and p10.9, respectively. Compared to the < 25%WRmax, at 5 years, the cutoff

Fig. 1 All 4497 last weight loss results after laparoscopic Roux-en-Y gastric bypass expressed as percentage alterable weight loss (%AWL) with percentile curves p50 (median), and ± 1 standard deviation (SD)
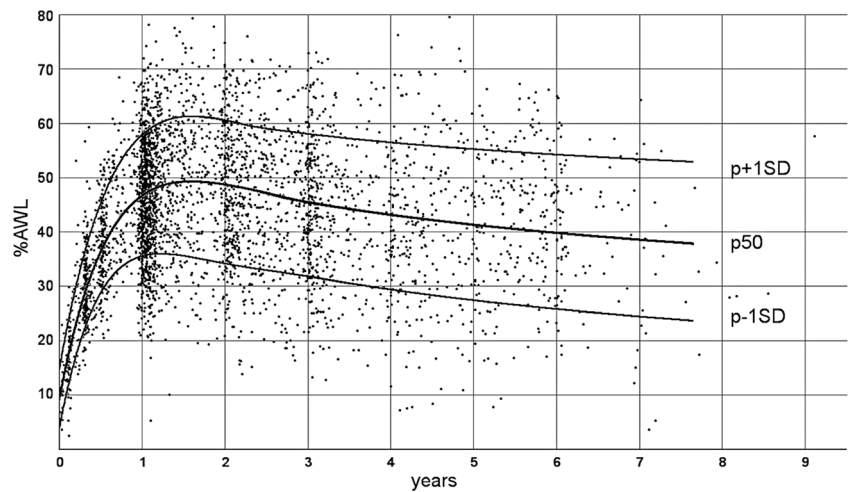
percentile curves p31.6, p25, p15.9, and p10.9 lost 18.6, 22.1, 27.0, and 31.4% from their initial maximum weight loss.

Discussion

For the first time, an attempt was made to systematically validate common bariatric weight loss criteria for success. We compared them to mid-term results (1–9 years), of a large cohort ($n = 3031$), of a wide variety of patients (BMI 35–72 kg/m^2; 18–66 years; 1/6 male; 1/5 type 2 diabetes), after the gold standard bariatric procedure (LRYGB). Results were (very) disappointing for criteria $\geq 50\%$EBMIL, $\geq 25\%$EBMIL, $\geq 25\%$EWL, $\geq 15\%$TWL, $\geq 20\%$TWL, and < 35BMI. They all had low specificities, leaving too many poor responders unnoticed. Popular criterion $\geq 50\%$EBMIL differentiated successful results with ease (sensitivities > 96%), but missed 28–66% of the truly unsuccessful results. This

Table 2 Mean weight loss results and standard deviation (SD) after laparoscopic Roux-en-Y gastric bypass (LRYGB) at 0.5, 1, 2, 3, 5, and 7 years (± 10%) and compared to results from seven landmark studies on percentage total weight loss (%TWL) after LRYGB. Body mass index (BMI), percentage excess weight loss (%EWL), percentage excess BMI loss (%EBMIL), percentage alterable weight loss (%AWL)

Weight loss mean (± 1 SD)	0.5 year 164–201 days	1 year 329–402 days	2 years 657–804 days	3 years 986–1205 days	5 years 1644–2009 days	7 years 2301–2812 days
$n=$	147	735	387	425	191	45
BMI-reduction (kg/m^2)	11.1 (± 2.3)	13.9 (± 3.7)	14.4 (± 4.5)	13.5 (± 4.4)	12.8 (± 4.8)	12.0 (± 5.3)
BMI (absolute) (kg/m^2)	31.7 (± 4.2)	29.3 (± 4.5)	28.9 (± 4.7)	29.8 (± 4.6)	31.3 (± 5.3)	32.8 (± 5.9)
%EWL (%)	56.8 (± 13.1)	69.0 (± 16.5)	71.1 (± 18.7)	66.6 (± 18.8)	60.3 (± 21.0)	55.3 (± 21.8)
%EBMIL (%)	65.0 (± 16.3)	78.8 (± 19.8)	81.3 (± 22.4)	75.8 (± 22.1)	68.7 (± 24.9)	62.2 (± 24.5)
%AWL (%)	37.6 (± 7.3)	46.1 (± 10.3)	47.6 (± 12.0)	44.6 (± 12.1)	41.1 (± 14.0)	37.7 (± 14.5)
%TWL (%)	26.0 (± 4.8)	32.0 (± 7.2)	33.1 (± 8.5)	31.1 (± 8.6)	28.8 (± 9.9)	26.6 (± 10.4)
Landmark LRYGB study						
Sjöström et al. 2007 [7] (mean %TWL; $n = 265$)	27%TWL	32%TWL	31%TWL	29%TWL	27%TWL	26%TWL
Kolotkin et al. 2009 [31] (mean %TWL; $n = 308$)			34.2%TWL			
Nelson et al. 2012 [32] (mean %TWL; $n = 77,406$)	26%TWL	35%TWL	36%TWL			
Courcoulas et al. 2013 [33] (median %TWL; $n = 1738$)				31.5%TWL		
Thereaux et al. 2015 [34] (mean %TWL; $n = 352$)					27.4–29.7%TWL	
Wood et al. 2016 [35] (mean %TWL; $n = 726$)						22.5%[a]
Ames et al. 2017 [36] (median %TWL; $n = 305$)		31.5%TWL	31.2%TWL			

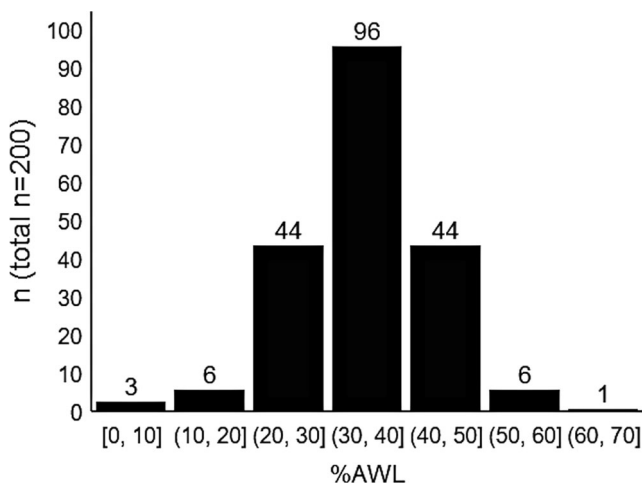[a] Follow-up 7–12 (median 9.3) years

Fig. 2 The distribution of the expected disappointing long-term weight loss results expressed as percentage alterable weight loss (%AWL) of 200 bariatric patients interviewed preoperatively about their postoperative expectations

inadequate specificity is clearly visible in Fig. 3, where many of the dots representing ≥ 50%EBMIL results lay below all four cutoff percentile curves (false positive results). Remarkably, the old ≥ 50%EWL criterion (based on life insurance tables for ideal weight) had better specificities, very likely because the %EBMIL metric is influenced more by differences in baseline BMI than %EWL [10, 22]. This study again proves how important this distorting effect can be in assessing bariatric weight loss results [8, 12, 16, 21, 23].

Criteria for weight loss success express treatment goals. They are used as simple tools to conveniently judge real-life results. The benchmark to test them should therefore be as close to real-life outcome as possible. As explained before, there is strong evidence that, although bariatric patients
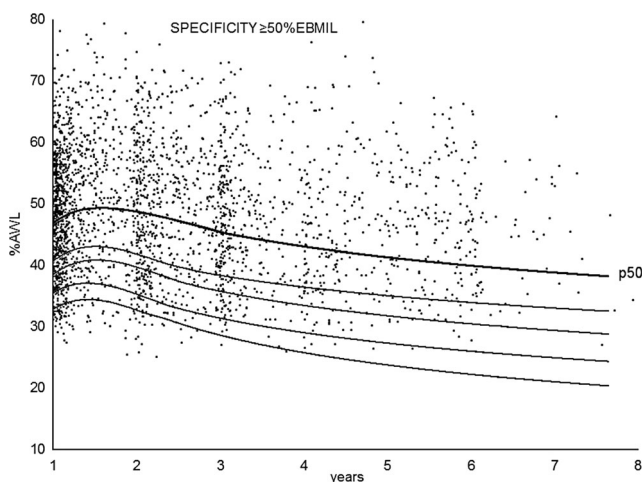

Fig. 3 Percentile curves (from top to bottom) p50 (median), p31.6 (patients' expectation cutoff), p25 (inter quartile range cutoff), p15.9 (1 standard deviation below median cutoff), and p10.9 (surgeons' goal cutoff), from 1 year up, expressed as percentage alterable weight loss (%AWL). The dots are all > 1-year %AWL results that correspond with ≥ 50 percentage excess body mass index loss (≥ 50%EBMIL)

typically show a wide variety of BMI at baseline, their weight loss essentially is baseline BMI-independent. We therefore constructed a benchmark that was baseline BMI-independent as well. It might explain why the most adequate criteria found in this study (≥ 35%AWL and ≥ 25%TWL) were based on metrics least influenced by baseline BMI. Bariatric weight loss is also essentially dynamic over time. We therefore constructed a benchmark that expressed our outcome against time, using percentile curves. This might explain why none of the criteria were fully sensitive and specific, as they are all static over time. Still, some turned out to be quite useful among the dynamic real-life results, like ≥ 50%EWL, > 10-kg/m^2 BMI reduction, ≥ 35%AWL, and ≥ 25%TWL.

A limitation of our study is that we only used LRYGB outcome for our benchmark, while the criteria are not meant to be procedure specific. As bariatric outcome varies for different procedures, we needed a gold standard to compare them to and considered the LRYGB procedure as such. Criteria commonly used in bariatric surgery are also not specified for gender, age, ethnicity, comorbidities, or other confounders on bariatric outcome. It was therefore important that our bariatric cohort was both large enough and sufficiently representative to test them. The external validation with landmark studies on LRYGB outcome showed that this was the case. It is important that other researchers confirm our findings, with other cohorts, that are large and representative as well; and with outcome after other procedures, like gastric banding, sleeve gastrectomy, and duodenal switch.

Another limitation of this study was the disbalance between short and longer-term results. Fortunately, this was caused more by the lower numbers of procedures performed in the earlier days of our bariatric program than by a lack of follow-up.

As the criteria tested are widely used by bariatric professionals and researchers, they represent a consensus. Then, they could help identify the most appropriate cutoff in a percentile-based benchmark. Although criteria ≥ 25%EBMIL, ≥ 25%EWL, ≥ 15%TWL, and < 35 BMI seemed to be inadequate for all four cutoff percentiles, the other weight loss criteria clearly returned more acceptable specificities for p15.9 and p10.9 than for p25 and p31.6. Furthermore, the weight regain criterion < 25%WRmax could be situated between p25 (22%) and p15.9 (27%). This all suggests that the %AWL p15.9 (1 SD below median) is the most expedient cutoff percentile curve for defining success. It would mean that at any time after the operation, 84.1% of all bariatric patients can be categorized successful, whatever BMI they had before the operation. In our LRYGB chart, only the ≥ 35%AWL and ≥ 25%TWL criteria came close to that curve, both with > 90% sensitivity and specificity.

Specificities of the ≥ 25%EBMIL, ≥ 25%EWL, ≥ 15%TWL, ≥ 20%TWL, and < 35BMI criteria were so low that they should be abandoned all together. The ≥ 10-kg/m^2

Cut off	Criterion	≥25 %EBMIL	≥25 %EWL	≥15 %TWL	<35 BMI	≥20 %TWL	≥50 %EBMIL	<30 BMI	≥10 Kg/m²	≥50 %EWL	≥35 %AWL	≥25 %TWL
p31.6	Sensitivity	100	100	100	97	100	100	78	98	99	100	99
	Specificity	2	3	9	34	27	34	83	57	55	61	69
p25	Sensitivity	100	100	100	96	100	100	74	96	98	98	96
	Specificity	3	4	11	39	35	43	88	66	66	74	79
p15.9	Sensitivity	100	100	100	93	99	98	68	92	94	94	91
	Specificity	4	7	18	47	52	61	94	80	83	95	95
p10.9	Sensitivity	100	100	100	92	98	97	65	89	91	90	87
	Specificity	6	10	25	52	65	72	96	88	92	100	99

Fig. 4 Sensitivity and specificity (expressed as percentages) of 11 criteria for weight loss success with reference to four cutoff percentile curves p31.6 (patients' expectation cutoff), p25 (inter quartile range cutoff), p15.9 (1 standard deviation below median cutoff), and p10.9 (surgeons' goal cutoff). Body mass index (BMI); percentage excess BMI loss (%EBMIL); percentage excess weight loss (%EWL); percentage total weight loss (%TWL); percentage alterable weight loss (%AWL). Red: inadequate criteria with sensitivity or specificity < 60%. Green: useful criteria with sensitivity and specificity ≥ 80%

criterion has merits, with better specificities than ≥ 50%EBMIL and sensitivities comparable to ≥ 35%AWL. The < BMI30 threshold had low sensitivities (< 80%), but high specificities (> 80%) for all four cutoffs, including the patients' expectation cutoff p31.6. It therefore should be useful in preoperatively informing patients on what (not) to expect: that not reaching below 30BMI is not exceptional. With only one postoperative measurement per patient, we could not calculate sensitivities and specificities for the weight regain criteria. The < 25%WRmax criterion did fit the benchmark between nadir and 5 years; the < 10%WRnadir criterion did not.

Popular criteria ≥ 50%EBMIL, ≥ 25%EBMIL, ≥ 20%TWL, and < 35BMI all had low specificities, leaving too many unsuccessful results unnoticed. This was never demonstrated before and calls for a radical change in how we define bariatric weight loss success. Looking for yet another static criterion might not be the right way to go. Our second best static criterion ≥ 25%TWL had 91% sensitivity and 95% specificity, yet Wood et al. and Sjöström et al. showed that mean LRYGB outcome can drop below 25%TWL beyond 7 years [7, 35]. This suggests that static criteria suited for mid-term outcome might not be appropriate for long-term results.

A better way to define success would be with elaborate baseline BMI-independent percentile charts, based on large, representative cohorts of bariatric patients. They present success in a dynamic way, with a cutoff between good and poor responders (percentile curve 1 SD below median) that evolves over time. When similar charts are available for other types of bariatric procedures, it will become possible to compare success of different patients, with different preoperative BMIs, after different bariatric procedures, at different postoperative intervals, in a clear and simple way. Cumbersome criteria for weight regain would become redundant. All patients have to do to stay successful, even if some weight regain occurs, is to keep above the cutoff curve. Given the relative complexity of the %AWL algorithm, this all can only be possible digitally, with applications designed to convert individual postoperative weight results that are hard to compare, in to percentile results that can easily be compared, plotted in a chart.

Compliance with Ethical Standards

Informed Consent Informed consent was obtained from all individual participants included in the study. The Institutional Review Board approved this study. For this type of study (retrospective studies), formal consent is not required.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Marathe PH, Gao HX, Close KL. American Diabetes Association standards of medical care in diabetes 2017. J Diabetes. 2017;9(4): 320–4.
2. Marek RJ, Heinberg LJ, Lavery M, et al. A review of psychological assessment instruments for use in bariatric surgery evaluations. Psychol Assess. 2016;28(9):1142–57.
3. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann Surg. 2004;240(2): 205–13.
4. Ballantyne GH. Measuring outcomes following bariatric surgery: weight loss parameters, improvement in co-morbid conditions, change in quality of life and patient satisfaction. Obes Surg. 2003;13(6):954–64.
5. Buchwald H, Estok R, Fahrbach K, et al. Weight and type 2 diabetes after bariatric surgery: systematic review and meta-analysis. Am J Med. 2009;122(3):248–56.
6. Welbourn R, Pournaras DJ, Dixon J, et al. Bariatric surgery worldwide: baseline demographic description and one-year outcomes from the Second IFSO Global Registry Report 2013–2015. Obes Surg. 2017; https://doi.org/10.1007/s11695-017-2845-9. (Epub ahead of print)

7. Sjöström L, Narbro K, Sjöström CD, et al. Effects of bariatric surgery on mortality in Swedish obese subjects. N Engl J Med. 2007;357:741–52.

8. van de Laar AW. The %EBMIL/%EWL double booby-trap. A comment on studies that compare the effect of bariatric surgery between heavier and lighter patients. Obes Surg. 2015;26:612–3.

9. Dallal RM, Quebbemann BB, Hunt LH, et al. Analysis of weight loss after bariatric surgery using mixed-effects linear modeling. Obes Surg. 2009;19:732–7.

10. van de Laar A, de Caluwé L, Dillemans B. Relative outcome measures for bariatric surgery. Evidence against excess weight loss and excess bodymass index loss from a series of laparoscopic Roux-en-Y gastric bypass patients. Obes Surg. 2011;21(6):763–7.

11. Hatoum IJ, Kaplan LM. Advantages of percent weight-loss as a method of reporting weight-loss after Roux-en-Y gastric bypass. Obesity. 2013;21(8):1519–25.

12. Sczepaniak JP, Owens ML, Shukla H, et al. Comparability of weight loss reporting after gastric bypass and sleeve gastrectomy using BOLD data 2008–2011. Obes Surg. 2015;25(5):788–95.

13. Corcelles R, Boules M, Froylich D, et al. Total weight loss as the outcome measure of choice after Roux-en-Y gastric bypass. Obes Surg. 2016;26(8):1794–8.

14. Baltasar A, Perez N, Serra C, et al. Weight loss reporting: predicted body mass index after bariatric surgery. Obes Surg. 2011;21(3): 367–72.

15. Molina López A, Sabench Pereferrer F, Vives Espelta M, et al. Usefulness of Baltasar's expected body mass index as an indicator of bariatric weight loss surgery. Obes Surg. 2016;26(11):2712–7.

16. van de Laar AW. Algorithm for weight-loss after gastric bypass surgery considering body mass index, gender, and age from the Bariatric Outcome Longitudinal Database (BOLD). Surg Obes Relat Dis. 2014;10(1):55–61.

17. van de Laar AW, Dollé MH, de Brauw LM, et al. Validating the alterable weight loss (AWL) metric with 2-year weight loss outcome of 500 patients after gastric bypass. Obes Surg. 2014;24(7): 1085–9.

18. Park JY, Kim YJ. Reply to the letter to editor entitled "The %EBMIL/%EWL double-booby trap. A comment on studies that compare the effect of bariatric surgery between heavier and lighter patients". Obes Surg. 2016;26(3):614–6.

19. van de Laar AW, de Brauw LM, Meesters EW. Relationships between type-2 diabetes remission after gastric bypass and different weight loss metrics: arguments against excess weight loss in metabolic surgery. Surg Obes Relat Dis. 2016;12(2):274–82.

20. van de Laar AW, de Brauw M, Bruin SC, et al. Weight-independent percentile chart of 2880 gastric bypass patients: a new look at bariatric weight loss results. Obes Surg. 2016;26(12):2891–8.

21. van de Laar AW, Dollé MH, de Brauw LM, et al. Which baseline weight should be preferred as reference for weight loss results? Insights in bariatric weight loss mechanisms by comparing primary and revision gastric bypass patients. Obes Surg. 2015;25(4):687–93.

22. van de Laar A. Bariatric Outcomes Longitudinal Database (BOLD) suggests excess weight loss and excess BMI loss to be inappropriate outcome measures, demonstrating better alternatives. Obes Surg. 2012;22(12):1843–7.

23. van de Laar AW, Acherman YI. Weight loss percentile charts of large representative series: a benchmark defining sufficient weight loss challenging current criteria for success of bariatric surgery. Obes Surg. 2014;24(5):727–34.

24. Brethauer SA, Kim J, El Chaar M, et al. ASMBS Clinical Issues Committee. Standardized outcomes reporting in metabolic and bariatric surgery. Obes Surg. 2015;25(4):587–606.

25. Livhits M, Mercado C, Yermilov I, et al. Preoperative predictors of weight loss following bariatric surgery: systematic review. Obes Surg. 2012;22(1):70–89.

26. Deitel M, Gawdat K, Melissas J. Reporting weight loss 2007. Obes Surg. 2007;14:565–8.

27. Weigley ES. Average? Ideal? Desirable? A brief overview of height-weight tables in the United States. J Am Diet Assoc. 1984;84(4):417–23.

28. Mann JP, Jakes AD, Hayden JD, et al. Systematic review of definitions of failure in revisional bariatric surgery. Obes Surg. 2015;25(3):571–4.

29. Aminian A, Jamal M, Augustin T, et al. Failed surgical weight loss does not necessarily mean failed metabolic effects. Diabetes Technol Ther. 2015;17(10):682–4.

30. Mor A, Sharp L, Portenier D, et al. 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. 2012;8(5): 556–60.

31. Kolotkin RL, Crosby RD, Gress RE, et al. Two-year changes in health-related quality of life in gastric bypass patients compared with severely obese controls. Surg Obes Relat Dis. 2009 Mar-Apr;5(2):250–6.

32. Nelson DW, Blair KS, Martin MJ. Analysis of obesity-related outcomes and bariatric failure rates with the duodenal switch vs gastric bypass for morbid obesity. Arch Surg. 2012;147(9):847–54.

33. Courcoulas AP, Christian NJ, Belle SH, et al. Weight change and health outcomes at 3 years after bariatric surgery among individuals with severe obesity. JAMA. 2013;310(22):2416–25.

34. Thereaux J, Czernichow S, Corigliano N, et al. Five-year outcomes of gastric bypass for super-super-obesity (BMI ≥ 60 kg/m^2): a case matched study. Surg Obes Relat Dis. 2015;11(1):32–7.

35. Wood GC, Benotti PN, Lee CJ, et al. Evaluation of the association between preoperative clinical factors and long-term weight loss after Roux-en-Y gastric bypass. JAMA Surg. 2016;151(11):1056–62.

36. Ames GE, Heckman MG, Diehl NN, et al. Guiding patients toward the appropriate surgical treatment for obesity: should presurgery psychological correlates influence choice between Roux-en-Y gastric bypass and vertical sleeve gastrectomy? Obes Surg. 2017;27(10):2759–67.