

Peer Review

Review of: "Impact of Telemedicine on Post-Bariatric Surgery Outcomes and Patient Satisfaction during the COVID-19 Pandemic: A Retrospective Study"

Evgenii Arsentev¹

1. Independent researcher

1. What the work claims

The authors report a single-centre series from the Hospital do Espírito Santo de Évora: patients operated on in 2020 with Roux-en-Y gastric bypass whose postoperative follow-up, because of the COVID-19 pandemic, ran through telemedicine (in practice, as the Instruments subsection describes it, telephone consultation). Of a population of 108 patients, 80 agreed to take part and form a convenience sample (Materials and methods, "Characterization of the sample"). Clinical data came from the electronic record, and a satisfaction questionnaire was administered by telephone. The paper reports weight and comorbidity change from baseline to three years (Table 3), surgical safety events (Results, paragraph after the comorbidity counts), and the share of patients satisfied with telemedicine follow-up (Table 4: 71.3% satisfied, 28.7% not satisfied).

The headline conclusion, in the Abstract and again in the Discussion, is that follow-up by telemedicine "did not modify the results of bariatric surgery", and that telemedicine is therefore a reasonable option for routine post-bariatric monitoring. The Discussion goes further and states that the absence of differences in weight loss "clearly shows that monitoring via telemedicine during the COVID-19 pandemic did not affect the results." The paper also attributes the comorbidity improvement to the bypass itself.

2. Whether the conclusion holds on the data

Some things here are useful. The cohort is homogeneous in procedure (all RYGB, Results, first paragraph), the three-year horizon is longer than most pandemic-era telemedicine reports, and the comorbidity

counts in the text agree with the percentages in Table 2 and Table 3. I checked them: $67/80 = 83.75\%$ (Table 2, row Hypertension, 83.8%); $63/80 = 78.75\%$ (row Diabetes, 78.8%); $66/80 = 82.5\%$ (row Dyslipidemia, 82.5%); $14/80 = 17.5\%$ (row OSAS, 17.5%). At three years, $4/80 = 5\%$, $9/80 = 11.25\%$, $5/80 = 6.25\%$, and $2/80 = 2.5\%$, which match Table 3, column "3 years", rows HTA, Diabetes, Dyslipidemia, and OSAS. The sex-weighted means in Table 1 also reconcile with the totals: $(68 \times 83.6 + 12 \times 74.2) / 80 = 6575.2 / 80 = 82.19$, matching Table 1, row %EWL, column Total, 82.2. So the arithmetic is internally consistent. That is worth saying because the problems lie elsewhere.

The central problem is design. The conclusion is comparative (telemedicine did not change outcomes), but the study contains no comparison group. Every included patient received telemedicine follow-up; this is an inclusion criterion. Without a contemporaneous or historical cohort followed face to face, with the same procedure, the same outcome definitions, and the same time horizon, the paper can describe what happened to telemedicine-followed patients. It cannot estimate the effect of telemedicine. The Discussion leans on agreement with "data previously published in this center", but I could not find those data cited with a reference number, and no numbers from them appear anywhere in the text. A statement of equivalence also needs a pre-specified margin and a test; neither is present. Absence of evidence of a difference, in a study that never measured a difference, is not evidence of no difference.

The significance tests that are reported answer a different question. Table 3 tests baseline against three years within the same patients (Student's *t* for weight and BMI, McNemar for the four comorbidities). These show that patients lost weight and that medication-treated comorbidity fell after gastric bypass. Nobody doubts that. The *p*-values say nothing about the telemedicine component, and the title ("Impact of Telemedicine on Post-Bariatric Surgery Outcomes") and the Abstract conclusion should not rest on them.

Confounding and selection are not addressed at all. Twenty-eight of 108 patients (26%) are missing, and the manuscript does not say whether they refused, could not be reached, had died, had been lost to follow-up, or had had complications. A patient who regained weight or had a poor experience is plausibly less likely to answer a telephone call from the surgical team. That would inflate both the weight results and the satisfaction figure. A comparison of the 80 responders with the 28 non-responders on the record-based variables (age, sex, baseline BMI, 30-day events) is feasible from the hospital database and would take one table.

The exposure itself is undefined. I could not find how many telemedicine contacts each patient had, over what period, by which professionals (surgeon, nurse, nutritionist, psychologist are all mentioned in the

Introduction), whether any in-person visits were mixed in, or when the service returned to face-to-face follow-up. The Conclusions describe telemedicine as "videoconferencing, messaging, and other digital means," while the Instruments subsection says the questionnaire concerned "the telephone consultation." These are different interventions. Without a dose or a description, even a future comparative study could not replicate the exposure.

Outcome definitions are thin. The comorbidity paragraph in the Results defines dyslipidemia by patients "taking medication to control the disease," and it is not stated whether the same medication-based definition applies to diabetes and hypertension, or whether laboratory values (HbA1c, fasting glucose, blood pressure, lipid profile) were used. Remission criteria for type 2 diabetes after metabolic surgery are well established and would make the 78.8% to 11.3% change interpretable. The baseline prevalences themselves are high for a bariatric cohort (Table 2: diabetes 78.8%, hypertension 83.8%), which makes me wonder whether prediabetes or insulin resistance were counted as diabetes. The authors should say. OSAS at 17.5% (Table 2) is, by contrast, low for a population with a mean BMI of 42.8 kg/m², which suggests under-diagnosis at baseline rather than true prevalence, and the three-year figure then means little.

The weight results also need their definitions. %EWL depends on the reference weight (BMI 25, ideal body weight by Metropolitan tables, or other); none is given. For orientation only, using the group means in Table 3 and a BMI-25 reference, $(42.8 - 28.4) / (42.8 - 25) = 14.4 / 17.8 = 80.9\%$, which is close to the reported 82.2%, so a BMI-25 convention is plausible, but it should be stated rather than inferred. The source of the three-year weight is not stated either: measured in clinic, taken from the record, or self-reported by telephone. Self-reported weight in people after bariatric surgery is systematically biased downward, and this matters for a paper whose main quantitative claim is about weight.

The Abstract does not match the Results in two places. It says there was "a 75% loss of excess weight"; Table 1 (row %EWL, column Total) and Table 2 (row EWL (%)) both give 82.2. It also says that "50% of the subjects" maintained "excellent levels" of comorbidity control; I could not find any 50% figure or any definition of "excellent levels" in the Results or tables. And "surgical outcomes were not significant" is ambiguous: I think the intended meaning is that adverse events were rare, but no test was performed on them, so "not significant" is the wrong word.

The surgical safety paragraph (Results, after the comorbidity counts) reports no 30-day complications and, "three years after surgery," one readmission, two emergency visits, no reoperation, and no death. It is unclear whether these are cumulative over three years or events in the third year, and whether they

were counted among the 80 or among all 108. With 80 patients, zero events still leave an upper 95% bound of about 3.75% by the rule of three ($3/80 = 0.0375$). That is not the same as safety equivalence.

The satisfaction result is the most directly telemedicine-related finding and the weakest reported. Table 4 gives one dichotomy, satisfied versus not satisfied, from a questionnaire that the Instruments subsection describes as eight-dimensional with 16 Likert items. The threshold for "satisfied" is not given. The dimension scores, item distributions, and Cronbach's alpha (which the Statistical analysis subsection says was computed) are not reported. Satisfaction was measured by telephone, by the treating service, roughly three years after the follow-up it asks about (surgery in 2020, ethics approval 31/01/2023 per the Institutional Review Board Statement). Recall bias and social desirability bias both push in the same direction. The Discussion also interprets the 28.7% dissatisfied as reflecting unclear patient understanding of telemedicine; that may be true, but no data in the paper test it. A small numerical point: $57/80 = 71.25\%$ and $23/80 = 28.75\%$, so the reported 71.3% and 28.7% fit $n = 57$ and $n = 23$. Please give the counts.

Finally, the Statistical analysis subsection lists chi-square, Student's t , and ANOVA for "various hypotheses", but the hypotheses are never stated, and I could find no chi-square or ANOVA result in the Results. The text reports that %EWL was higher in women (Table 1: 83.6 versus 74.2) without a test, confidence interval, or comment on the small male group ($n = 12$). The weight and BMI comparisons in Table 3 are within-subject, so a paired t -test is required; the note says only "Student's t test". No confidence intervals or effect sizes appear anywhere.

In short: the descriptive data are plausible and internally consistent, but the causal and comparative claims in the title, Abstract, and Discussion go well beyond what a single-arm, 74%-response, telephone-surveyed series can support.

3. Reproducibility

Data. The Data Availability statement says data are available on reasonable request and are not public for privacy reasons. For patient data, that is acceptable. Still, an aggregated, de-identified table (per-patient baseline and three-year weight, BMI, and binary comorbidity status) would carry little re-identification risk at $n = 80$ and would let readers check Table 3.

Analysis. SPSS version 28 is named. That is the only analytic detail. The hypotheses, the test used for each variable, the handling of non-normal variables after the Shapiro–Wilk step, missing-data handling, and the significance level are not given. I could not reconstruct the analysis from the text.

Instrument. The satisfaction questionnaire is attributed to reference [17] (Abecasis et al., 2021). The items, the eight dimensions, the scoring, and the cut-off for "satisfied" should be reproduced in an appendix; the manuscript mentions Appendix A for the consent form, but as rendered on Qeios, I could not see any appendix content. The Ethical considerations subsection also states that permission to use the instruments was not requested from their authors, on a free-use argument under Portuguese copyright law. I am not in a position to judge the legal point, but most journals will ask for explicit permission from the instrument developers, and it is simple to obtain.

Design description. The paper calls itself retrospective (title, Abstract), "observational, cross-sectional, quantitative" (Study design), and then says data were collected "throughout the postoperative period" and "supplemented by data collected afterwards". The Results mention two assessments and then list three time points ("baseline (before surgery), three years after surgery, and three years after the COVID-19 pandemic"). A reader cannot tell which data were abstracted retrospectively from records and which were collected prospectively by interview after the 2023 ethics approval. For an observational study, following the STROBE checklist would resolve most of this: a participant flow diagram, stated eligibility and sources, variable definitions, efforts to address bias, handling of missing data, and numbers at each stage.

Setting and dates. The operation year (2020) and the ethics approval date are given; recruitment and interview dates are not. The timing of the three-year assessment relative to the telemedicine period is not given.

4. What specifically to fix

1. Add a comparator or change the claim. The best option is a historical cohort from the same centre (for example, RYGB patients operated on in 2018–2019 with face-to-face follow-up), with the same outcome definitions at three years, analysed with adjustment for baseline BMI, age, sex, and diabetes. If that is impossible, retitling the paper as a descriptive series and removing "Impact of" from the title, "did not modify the results" from the Abstract, and "clearly shows that monitoring via telemedicine ... did not affect the results" from the Discussion.
2. If the claim is "no worse than in-person follow-up", pre-specify a non-inferiority margin for %EWL or %TWL and report the confidence interval of the difference.
3. Methods, Characterization of the sample: add a flow diagram, 108 → 80, with reasons for each exclusion or non-response, and a table comparing responders with non-responders on record-based variables.

4. Describe the exposure: modality (telephone, video, messaging), number and timing of contacts, professionals involved, whether any in-person visits occurred, and when face-to-face follow-up resumed. Make the Conclusions consistent with it.
5. Abstract, Results: correct the 75% excess weight loss figure to match Table 1 and Table 2 (82.2), and either define and support the "50% ... excellent levels" statement or delete it.
6. Replace "surgical outcomes were not significant" with the event counts.
7. Table 3: state that the weight and BMI tests are paired; give the mean change with a 95% CI; give n at each time point.
8. Define comorbidity presence and remission (medication, laboratory, or clinical criteria) for each of diabetes, hypertension, dyslipidemia, and OSAS, and cite the remission criteria used. Explain the baseline diabetes prevalence of 78.8% (Table 2).
9. State the %EWL reference weight and the source of the three-year weight (measured versus self-reported).
10. Surgical safety paragraph: say whether the one readmission and two ED visits are cumulative over three years, what caused them, and which denominator applies.
11. Table 4 needs much more. Report counts, the scoring rule, the satisfied/not-satisfied threshold, mean scores per dimension, and Cronbach's alpha per dimension. The text refers to this table as "Table 1"; fix the cross-reference. Use one decimal separator throughout (Tables 1–4 use commas, the text uses points; Table 4 mixes both).
12. Statistical analysis subsection: list the hypotheses and the test for each. Either report the chi-square and ANOVA results or remove them from the methods.
13. Table 1: if the sex difference in %EWL is to be discussed, test it and note the imbalance (68 women versus 12 men).
14. Study design subsection: settle on one design label and describe clearly which data are record-based and which come from the later telephone interview. Correct "two assessments" versus the three time points listed.
15. Check references. ACS-NSQIP is cited to [16], which is André & Ribeiro, "E-health: ICT as a mechanism for health evolution"; that looks like a numbering error. Reference [28] (Pujol-Rafols et al., on conversion of adjustable gastric banding to RYGB) is cited for comorbidity resolution after primary RYGB, which it does not appear to address. The claim that telemedicine-followed patients "had better weight loss results than those who received face-to-face follow-up" is attributed to [22]; please confirm that study made such a comparison.

16. Use the standard abbreviation RYGB rather than RYGP.
17. Limitations: expand beyond cross-sectional design and telephone difficulty. At minimum, name the absence of a control group, non-response, recall and social desirability bias in the satisfaction measure, and possible self-report of weight.
18. Obtain written permission from the developers of the satisfaction questionnaire and put the instrument in an appendix.

5. Recommendation

Major revisions. The data describe a real and relevant clinical situation, and the numbers I could check are internally consistent. But the paper's main conclusion is comparative and causal, and the design has no comparison group, an unexplained 26% non-response, an undefined exposure, and an under-reported satisfaction measure. This is version 2 on Qeios. I reviewed it as it stands and could not see from the text what changed since version 1. With a historical control cohort, a STROBE-conformant account of selection and variables, and claims matched to the design, this could become a useful single-centre report. Without a comparator, it should be reframed as a descriptive series, and the conclusions softened accordingly.

Declarations

Potential competing interests: No potential competing interests to declare.