Supplement 2

Statistical Analysis Plan (Original 19.2.2008)

Statistical Analysis Plan: amendment 10-year follow-up (all amendments shown in *italic*) 11.3.2019

Saija Hurme, Biostatistician, MSc

A Randomized Multicenter Study: Laparoscopic Gastric Bypass vs. laparoscopic sleeve gastrectomy in the treatment of morbid obesity

Aim of the study

The aim of the study is to compare two different operative techniques regarding weight loss, resolution of comorbidities and complications. The primary endpoint is weight loss evaluated by %EWL (excessive weight loss, %). The secondary endpoints include resolution of associated comorbidities, improvement of QOL, mortality, and morbidity of the procedures.

Amendment to study aims at 10-year follow-up

The additional aim of the 10-year long-term follow-up was to compare the prevalence / cumulative incidence of gastroesophageal reflux (GERD), esophagitis, and Barrett's esophagus assessed by an upper gastrointestinal endoscopy and histopathology, GERD symptoms using a validated questionnaire, and proton pump inhibitor use between LSG and LRYGB.

Hypothesis and meaning of the study

The results of this current RCT study enable comparison on the long-term data between the current golden standard of laparoscopic gastric bypass and laparsocopic sleeve gatrectomy. The hypothesis of the study is that as sleeve can be considered less traumatic and easier and faster to perform compared to RYGB, it could become the procedure of choice in treating morbid obesity provided that long-term results of SG were comparable with those of RYGB. The results can directly be applied to patient care.

Additional post hoc outcomes at 10 years

Endoscopic findings of esophagitis and Barrett's esophagus including histopathology, reflux symptoms using a validated questionnaire (GERD-Health Related Quality of Life Questionnaire, GERD-HRQL), and PPI use will be assessed at 10 years. If evaluated clinically necessary, patients may also undergo pH-monitoring and manometry studies. A validated questionnaire (AUDIT) assessing patient alcohol use will be used at long-term follow-up.

Sample size calculation

Sample size calculations are based on test of mean difference of bypass and sleeve operation in excessive weight loss (EWL) in one year. The mean of 60 and standard deviation of 20 in bypass group were assumed and α-level of 0.05 were used in calculations. By using an equivalence design with a margin of equivalence of 15% (-9 to 9), a sample size of 108 patients per group is needed for 90% power. When 10% drop out is taken account, 120 patients per group were recruited for the study.

Statistical analyses

Continuous variables will be characterized by treatment and time-point using means and standard deviations (SD) or medians and range of values, and in case of categorical variables frequencies and percentages will be used.

Primary endpoint of the study is weight loss evaluated by %EWL (excess weight loss, %) and groups will be compared using equivalence trial setting. *For comparing the groups, the 95% confidence intervals (95% CI) for the difference between the study groups will be calculated at every time-point and the equivalence will be evaluated using the pre-defined margins of equivalence -9 to 9. Original plan at study initiation of using the 90% confidence interval was revised to 95% confidence interval based on the discussion in the peer review process of 5-year results of the study.* When analyzing the data from the longer follow-up, repeated measurements ANOVA will be used *(PROC MIXED in SAS)* and confounding factors will be taken into account in the analyses, if needed. *Excess weight at baseline, operation type, time (within factor) and the interaction of operation and time will be included in the model as well as diabetes status and study site as between factors and other potential covariates, if needed.*

In analyses of secondary outcomes associations between categorical variables will be tested using Pearson's χ^2 -test and in case of small frequencies Fisher's exact test will be used. Differences between groups in normally distributed continuous variables will be tested using independent samples t-test and in case of non-normally distributed variables Mann-Whitney U test will be used. When the data from the longer follow-up will be analyzed, repeated measurements ANOVA techniques will be used *(PROC MIXED in SAS)* to analyze continuous variables and confounding factors will be taken into account in the analyses if needed. *All of the models will include operation, time (within factor), interaction of operation and time, and study site as between factors. Models will also include other covariates as between factors, if needed. Results will be quantified using model-based mean estimates with 95% CI. The results will be presented separately for both operations and time points only, if the interaction term in the model is statistically significant. Otherwise, the estimates of the main effects of operation and time will be used. Transformation of the variables will be used when needed to achieve the normality of the residuals. Assumptions for models will be checked with the evaluation*

of studentized residuals. Step-down Bonferroni method of Holm will be used to adjust the p-values in pairwise comparisons.

Subgroup of patients with type 2 diabetes (T2DM) at baseline will also be analyzed separately regarding the variables related to weight. The effect of preoperative duration of T2DM on weight related variables and remission of diabetes will be evaluated.

Additional sub-group analyses will be performed for the data, if needed.

Two-tailed p-values will be used and p-values less than 0.05 will be considered statistically significant. The main analyses will be based on the intention-to-treat principle, i.e. all patients will be analyzed in their initial randomized intervention group. *For the primary outcome of %EWL, a per-protocol analysis will also be performed by excluding from the analysis all the patients, who have undergone a conversion to another bariatric procedure.* Missing data will be excluded from the analyses *but sensitivity analyses will be made for primary endpoint using multiple imputation. Multivariate imputation by fully conditional specification method will be performed. The predictive mean matching method will be used to construct 10 imputed datasets and a linear mixed model for repeated measures will be fitted for each. Results will be combined for the inference and compared to the original analyses.* Statistical analyses will be performed using SAS System for Windows, Version 9.2 or later (SAS Institute Inc., Cary, NC).

SLEEVEPASS Statistical analysis plan:

Summary of clinical amendments (11.3.2019) at 10-year follow-up, please see above for the detailed statistical additions in *italic*

Amendment to study aims at 10-year follow-up

The additional aim of the 10-year long-term follow-up was to compare the prevalence / cumulative incidence of gastroesophageal reflux (GERD), esophagitis, and Barrett's esophagus assessed by an upper gastrointestinal endoscopy and histopathology, GERD symptoms using a validated questionnaire, and proton pump inhibitor use between LSG and LRYGB.

Additional post hoc outcomes at 10 years

Endoscopic findings of esophagitis and Barrett's esophagus including histopathology, reflux symptoms using a validated questionnaire (GERD-Health Related Quality of Life Questionnaire, GERD-HRQL), and PPI use will be assessed at 10 years. If evaluated clinically necessary, patients may also undergo pH-monitoring and manometry studies. A validated questionnaire (AUDIT) assessing patient alcohol use will be used at long-term follow-up.