Statistical Analysis Plan

Official Title: Semaglutide Improves Metabolic Abnormalities and Fertility in Obese Infertile Women With Polycystic Ovary Syndrome: a Prospective, Randomized, Open, Controlled Study

NCT number: Not assigned yet

Document Date: Jan 18th, 2023

Statistical Analysis Plan

A sample size of 45 patients (15 in each group) provided power of 95% for the coprimary and secondary endpoints. The primary endpoints are percentage change in bodyweight and achievement of weight reduction of at least 5% at 12 weeks, thus referring to the data of previous liraglutide, with setting α=0.05 and β= 0.1, we calculate about 15 patients in each group. Since the exfoliation rate of similar studies in the past is about 20%~30%, there should be 21 patients in each group. In this study, 25 subjects will be included in each group, with a total of 75.

Outcomes are assessed using intention-to-treat analysis (ie, the full set of all randomly assigned patients). Observation periods include the in-trial period for 12 weeks (ie, while in the trial, regardless of treatment discontinuation) and the follow-up period for 24 weeks. All results from statistical analyses on confirmatory endpoints are accompanied by two-sided 95% CIs and corresponding *p* values (superiority defined as $p<0{\cdot}05$). Exploratory secondary endpoint analyses are not controlled for multiple comparisons and should not be used to infer definitive treatment effects.

Two estimands (the treatment policy estimand and the trial product

estimand) are used to assess treatment efficacy, and accounted differently for intercurrent events and missing data. The treatment policy estimand, which quantified average treatment effect among all randomly assigned patients, regardless of adherence to treatment (patients in trial; intention to treat) is used to assess the superiority of semaglutide 1.0 mg versus metformin for the primary and secondary confirmatory endpoints in a predefined hierarchical order. Continuous endpoints are analysed using an analysis of covariance model with randomised treatment, stratification groups, and the interaction between stratification groups as factors and baseline endpoint value as covariate. Missing data are imputed 1000 times from retrieved patients of the same randomised treatment and the results are combined using Rubin's rules. Categorical endpoints are analysed by logistical regression using randomised treatment, stratification groups, and the interaction between stratification groups as factors, and the baseline endpoint value as a covariate. Analyses are done using PASS (version 15.0). The trial product estimand model the average treatment effect in all randomly assigned patients, assuming that patients have remained on treatment for the duration of the trial. Continuous endpoints are analysed using a mixed model for repeated measurements with same factors and covariates as the treatment policy estimand all nested within visit, and categorical endpoints are analysed using the predicted values from the mixed model for repeated measurements by

logistic regression with treatment and stratification groups as factors and baseline endpoint as covariate. There is no data monitoring committee.