CID Case Study: External and Synthetic Placebo-Controlled Study in Subjects with Alopecia Areata

Study Design:

This proposed trial in subjects with severe alopecia areata (AA) is a phase 3 clinical trial to evaluate the benefit and risk of a higher dose than previously evaluated in clinical trials (hereafter referred to as the lower dose). The trial will not include an internal randomized placebo group but will instead make comparisons with an external control arm constructed using placebo data from completed trials, and a synthetic control arm constructed using placebo data from completed trials and extrapolation from Bayesian longitudinal data modeling for a later timepoint that was not evaluated in the completed trials. The trial will follow subjects for 48 weeks. At baseline, patients will be randomly assigned to either the higher dose or the lower dose. At Week 24:

- Patients who were randomly assigned to the higher dose at baseline will remain on the higher dose through Week 48.
- Patients who were randomly assigned to the lower dose at baseline and found to be responders at Week 24 will remain on the lower dose through Week 48.
- Patients who were randomly assigned to the lower dose at baseline and found to be non-responders at Week 24 will be re-randomized to receive the higher or lower dose for the remainder of the trial.

The proposed comparators are as follows:

- External placebo control at Week 24 (based on placebo data through Week 24 from completed trials)
- Synthetic placebo control at Week 36 (based on placebo data through Week 24 from completed trials and Bayesian longitudinal data modeling to predict synthetic placebo data at Week 36)
- External lower-dose group at Week 48 (based on data from completed trials from patients who were non-responders on the lower dose at Week 24 and remained on the lower dose through Week 48) to augment the data from the lower-dose non-responders who were re-randomized to the lower dose for use in the comparative analysis versus the lower-dose non-responders who were re-randomized to the higher dose.

The primary endpoint is the proportion of patients who achieved a Severity of Alopecia Tool (SALT) score of 20 or lower (SALT ≤20) on the higher dose compared with the corresponding proportion in the external placebo control at Week 24. The secondary endpoint is the proportion of patients who achieved SALT ≤20 for the higher dose compared with the proportion in the synthetic placebo control at Week 36.

Additional endpoints compare the response based on SALT ≤20 at Week 48 and change in SALT score from Week 24 to Week 48 between the lower-dose non-responders who were re-randomized to the higher dose and the lower-dose non-responders who were re-randomized to the lower dose augmented with the data from the external control group of lower-dose non-responders who remained on the lower dose.

The proposed analyses are as follows:

- The Week 24 analysis will apply inverse probability of treatment weighting using propensity scores (IPTW-PS) to create weighted populations that are balanced on baseline characteristics and compare the treatment group to an external placebo control group using a Bayesian model.
- The Week 36 analysis will use a Bayesian longitudinal model for extrapolation to Week 36 and then follow a similar approach as the Week 24 analysis to compare the treatment group to a synthetic placebo control group.

- The Week 48 analysis will use a Bayesian approach with a robust mixture prior to compare the lower-dose non-responders who were re-randomized to the higher dose to the lower-dose non-responders who were re-randomized to the lower dose augmented with the data from the external control group of lower-dose non-responders who remained on the lower dose.
- The analysis will report means, standard deviations, and 95% credible intervals for each group, along with the mean differences.

Innovative Characteristics:

FDA considers the following trial design features innovative:

- Use of external and synthetic placebo data
- Bayesian longitudinal data modeling to generate synthetic placebo data
- Bayesian parametric modeling to analyze primary and secondary endpoints
- Use of propensity scores and IPTW for causal inference

Potential Benefits of Design:

- Use of external and synthetic placebo data allows all patients to receive active treatment, rather than randomly assigning additional patients to placebo. In previously completed AA trials, the placebo response rate has been consistently low and stable in the severe AA patient population.
- Without a concurrent placebo arm, the number of patients needed for the trial will be reduced.

Considerations for the Proposed Design:

- Will comparisons to the external placebo control group lead to valid inferences (i.e., is the patient population comparable and are potential sources of bias or confounding adequately addressed)?
- How reliable is model-based extrapolation and the resulting synthetic placebo data for statistical comparison?
- Will a comparison between the lower-dose non-responders who were re-randomized to the higher dose and the lower-dose non-responders who were re-randomized to the lower dose augmented with the data from the external control group of lower-dose non-responders who remained on the lower dose lead to valid inference at Week 48?
- Does this study design have adequate operating characteristics, such as the Type I error and statistical power, under the proposed modeling assumptions?
- Will the design be adequate to characterize the safety of the higher dose for the proposed patient population?

Simulations:

The sponsor conducted simulation studies to evaluate how well the proposed analysis methods would perform for the proposed primary and secondary endpoints (SALT ≤ 20 at Week 24 and Week 36, respectively). Simulations were conducted using three different assumptions about the true treatment effect, including scenarios representing “no treatment effect” and “promising treatment effect.”

After estimating propensity scores using a frequentist logistic regression model, the rest of the analysis (including applying the decision rule using the posterior distribution) was repeated 2,000 times to approximate the probability of the posterior decision rule producing the correct or incorrect outcome.

The simulations also tested different scenarios by varying: 1) baseline variables in the external placebo control group; 2) baseline SALT scores; and 3) the presence of concomitant variables (unmeasured confounders). These simulations assessed the robustness of the analysis method's statistical properties, such as sufficient statistical power and type I error control, when causal inference assumptions are violated.

Discussion:

The trial was designed to eliminate the need to assign patients to placebo while evaluating the safety and effectiveness of a higher dose of the drug. While FDA supports minimizing the exposure of subjects to ineffective treatments, the clinical trial must still generate reliable evidence to support regulatory decision-making. The key issue is whether the proposed design is suitable for characterizing the benefit and risk of the new dosing regimen.

The following points need to be considered:

1. Innovative designs that aim for efficiency gains and/or forgo internal control depend heavily on assumptions. These assumptions should be judiciously evaluated and verified. Assumptions that are inherently difficult (if not outright impossible) to verify should be justified based on an agreed-upon set of principles. While the proposed trial's reliance on data from previous placebo-controlled trials with consistently low placebo response rates mitigates some concern regarding the data's fitness for use, assumptions about population comparability still require thorough scrutiny. Although the data on the placebo response at timepoints later than Week 24 are very limited, a secondary endpoint using comparison with a synthetic control may provide an interesting test case for the utility of a synthetic placebo control arm.
2. Although data from historic trials demonstrated that the placebo response rate is consistent and low, and thus it may be reasonable to consider an external placebo control for Week 24 comparisons, there is insufficient data that an external control comprised of lower-dose non-responders who remain on the lower dose for the assessment at Week 48 would be appropriate or interpretable. Use of an internal randomized concurrent control for the dose-increase assessment would therefore be preferable, which is incorporated in the study design.
3. Because the trial's key operating characteristics cannot be easily calculated in closed form, extensive simulations are needed. However, simulations depend on the scenarios and assumptions chosen, and it is impossible to test every possible way the assumptions could fail. The value of the simulation study depends on both the time invested and the range of scenarios tested, as well as the resources available to conduct these simulations. Simulations should cover an appropriate range of scenarios.