CID Case Study: External Control in Diffuse B-Cell Lymphoma

Study Design:

The proposed trial is a randomized, open-label, multicenter trial in patients with first-line diffuse large B-cell lymphoma. Patients are to be randomized 2:1 to treatment vs. control. The primary endpoint of the study is Investigator-assessed progression-free survival (PFS), defined as the time from randomization to the first occurrence of progression or relapse, using the 2014 Lugano classification for Malignant Lymphoma (Cheson et al. 2014), or death from any cause, whichever occurs first.

The key secondary endpoint is overall survival (OS). The analysis population for OS will be augmented by patients from an external control arm so that approximately half of the patients in the resulting control group are comprised of patients from the external control. The external control arm will be partially concurrent with the planned trial. The planned analysis of OS utilizes a Bayesian commensurate prior with a Weibull model (Lewis et al. 2019) to dynamically borrow information from the external control arm. Furthermore, propensity score matching will be conducted to select external control patients for inclusion in the analysis. Inference will be based on the posterior mean and 95% credible interval of the posterior distribution of the hazard ratio.

Innovative Characteristics:

FDA considers the following trial design features to be innovative, making it appropriate to review the design under the Complex Innovative Trial Design (CID) pilot meeting program:

- Use of external control data
- Use of a commensurate prior for borrowing data
- Use of a Bayesian parametric model as the primary analysis of a secondary endpoint

Potential Benefits of Design:

- If the model assumptions are met, borrowing patients' data from an external control arm reduces the number of patients necessary to randomize to the control arm of the proposed trial to achieve a specified power.
- The dynamic borrowing approach may mitigate the risk of borrowing patient data that is not compatible with that observed in the proposed trial.

Considerations for the Proposed Design:

- Is the difference in trial initiation dates between the two trials likely to result in populations that are incompatible due to availability of new therapies, the advent of COVID-19, or other predictable factors?
- Is the proposed parametric model compatible with previous data in this disease area?
- Is the proposed approach for borrowing appropriate and interpretable?
- Is the proposed use of the propensity score appropriate and interpretable?
- Is the proposed design robust to deviations from the model assumptions?

- What is the simulated Type I error of this design under various plausible deviations from these model assumptions?

Simulations:

The Sponsor conducted simulations to assess the operating characteristics of the proposed model under a variety of scenarios. Initial simulations focused on power and Type I error under various differences in OS curve behavior between the two control arms. FDA requested further simulations which assessed the operating characteristics of the model under deviations from the various model assumptions, including that these simulations implement all parts of the model including the proposed propensity score matching. The Sponsor conducted new simulations which characterized the model performance under deviations from the following model assumptions: similarity in patient populations, assumed Weibull distribution, unmeasured confounding, and the linear form of the propensity score model.

Discussion:

Innovative designs, such as those proposed under the CID program, often require stronger assumptions than designs commonly considered for regulatory decision-making. In addition, key operating characteristics such as power and Type I error may not have closed-form analytical expressions. Consequently, simulations are necessary to understand the operating characteristics of these designs. In this case, the Sponsor provided simulations to understand these operating characteristics in the case of violations from the various model assumptions, namely the proposed Weibull distribution, the linear form of the propensity score model, the assumption of no unmeasured confounding, and the assumed similarity in patient populations. These simulations facilitated discussion between the Sponsor and FDA on modeling choices and practical considerations for assessing the results.

In general, FDA prefers trial designs and analyses which require minimal assumptions and which result in straightforward interpretation of the treatment effect in the associated population. In this case, a consideration was whether the propensity score could be used as a covariate in the Weibull model for overall survival. In this case, FDA believed that use of the propensity score as a covariate would make results difficult to interpret and communicate. Consequently, the Sponsor specified propensity score matching as the method for adjusting for baseline differences in populations.

While simulations are important for assessing operating characteristics intractable to analytical assessment, many model assumptions are more tenable if supported by expert clinical input, historical data, or thoughtful plans for trial implementation. For instance, the chosen covariates for the propensity score model required clinical rationale based on expert clinical opinion and literature. In addition, the assumption of the Weibull distribution was supported by results from trials in this disease area which appeared to be reasonably fit with the Weibull distribution. The assumption of similarity in patient populations was bolstered by the Sponsor's plan to prioritize enrolling patients in the same sites for both the randomized arms and external control arm when possible. While rationale and simulations provide crucial support in designs with strong assumptions, ultimately many of these assumptions are unverifiable. Careful review of the final results will be necessary to further understand the strengths and limitations of this design.

References:

Cheson BD, Fisher RI, Barrington SF, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: The Lugano classification. J Clin Oncol 2014; 32: 3059; 68.

Lewis, CJ, Sarkar, S, Zhu, J, Carlin, BP et al. Borrowing from historical control data in cancer drug development: A cautionary tale and practical guidelines. Statistics in Biopharmaceutical Research, 11(1): 67–78, 2019