Example Statistical Analysis Plan – Unification of Evidence
CDER Center for Clinical Trial Innovation (C3TI) Bayesian Statistical Analysis Demonstration Project

Overview

When a clinical trial has multiple endpoints that are not easily combined into a single ordinal scale, it is standard practice to analyze the endpoints separately, providing separate evidence for treatment effect on each endpoint. Then the results are informally combined to judge overall evidence for treatment benefit. The Bayesian approach has a potential advantage of allowing investigators and regulatory authorities to define the condition that would determine therapeutic success, then computing the Bayesian posterior probability that the condition is satisfied.

Study Design

Study XY-02 is a double-blind, parallel-group, two-treatment, randomized controlled trial of drug (group B) vs. placebo (group A). The primary analysis is intent-to-treat and the outcomes are survival time (with follow-up up to 3 years), infection (within 90 days), and patient performance status (within 30 days; PS). The disease setting is a high mortality one, and infections and PS are frequently not assessable as a result. So death is counted in both outcomes by making it the highest level of the ordinal outcome. The infection outcome thus has 3 levels (alive and infection-free for 90 days; alive and infection within 90 days; death), and PS has 6 levels. One could say that these outcomes are infection penalized for death and PS penalized for death.

A Bayesian approach provides a unique opportunity to state the condition that would change clinical practice, then to compute the probability the condition is satisfied. The condition can represent a compound assertion. For drug B, success is taken to be that the drug is superior on mortality, or non-inferior on mortality and superior on either infections or PS. The condition for success is then $A \cup (B \cap (C \cup D))$ using the definitions below, and letting HR be a hazard ratio and OR be an odds ratio:

- $\cup$ = union (or)
- $\cap$ = intersection (and)
- A = any reduction in mortality (HR < 1.0)
- B = a reduction or only a small increase in mortality (HR < 1.1)
- C = a reduction in infection (OR < 1)
- D = an improvement in PS (OR < 1)

In this statistical plan, if the C and D effects go in opposite directions but one of them is favorable, the treatment may be considered effective, if the desired combinations of A and B are satisfied. One may choose to require that neither C nor D show a large detrimental effect.

Statistical Analysis

A flexible Bayesian proportional hazards model will be used to analyze the drug effect on survival time. The other 2 endpoints will be analyzed with proportional odds models. The prior distributions for log (HR) and log (OR) will be normal with mean zero and variance chosen to reflect the unlikeliness of very

large benefits or very large harms, i.e., probabilities of only 0.025 that HR or OR exceed 4 or that they are less than $\frac{1}{4}$.

Within-patient correlations among the 3 outcomes will be modeled using either a Gaussian copula or by having separate subject-level random effects for each outcome, with these random effects having a multivariate normal distribution with correlations between-outcomes. The new method of Afonso PM, D Rizopoulos et al, Stat in Med 2025; 44:e70057, https://doi.org/10.1002/sim.70057 should also be considered.

The Bayesian posterior probability of success on the combination of 3 endpoints will be computed from the fraction of posterior draws for which the success criterion was met for the 3 efficacy parameters. Note that Bayesian power and required sample size can be computed by simulation.

Bayesian posterior probabilities of benefit (and non-inferiority for mortality) will also be computed, and 0.95 highest posterior density uncertainty intervals computed for the HR and 2 ORs.

The decision criterion for the primary efficacy analysis is based on the posterior probability that the combination of conditions listed above holds, Pr($A \cup \left(B \cap (C \cup D)\right)$), exceeding a context-specific threshold (typically between 0.9 and 0.99).

Software

The 3 connected outcome models and their dependencies will be coded in Stan. Example code will be added at a later date.

Bayesian Markov Chain Monte Carlo simulation will be run in 4 independent chains with 4000 iterations per chain. Diagnostics including trace plots and Rhat value will be used to check for convergence of posterior distributions.