**AUGMENTING RANDOMIZED TRIALS WITH REAL-WORLD DATA: A SIMULATION STUDY EVALUATING METHODS FOR HYBRID CONTROL ARM ANALYSES**

Authors: Benjamin Ackerman, Rafael Sauter, Martina Fontana, Ignazio Craparo, Brian Hennessy

Affiliation: Janssen Research & Development

**Longer Abstract (499 words):**

Randomized trials (RCTs) are considered the gold standard for estimating causal effects of new therapies and interventions, yet statistical challenges remain in detecting treatment effects among rare disease populations, particularly when outcomes also occur infrequently. In such cases, innovative approaches exist to supplement trials with evidence from real-world data (RWD) by constructing a hybrid control arm, retaining the benefits of randomization while increasing study sample size and power. Identifying suitable RWD for this use case is critical, and data appropriateness are highly dependent on the design of the candidate RCT, its study population, and its primary outcome measure. Even with high-quality RWD, differences in study populations may exist, and must be properly accounted for to ensure comparability. In this work, we present propensity score-type weighting methods to align study populations when conducting hybrid control arm analyses.

When augmenting RCT control arms with RWD in a hybrid control analysis, it is important to first identify any baseline patient covariate imbalances. In particular, when prognostic factors of the outcome are imbalanced between the two studies, then naively pooling the control arms together without any population adjustment would result in a biased treatment effect estimate. Such imbalances can be accounted for with propensity scores by modeling the probability of study membership (RCT vs. RWD) conditional of the observed prognostic factors and weighting the RWD patients by the odds of their propensity score. This ensures that the distribution of covariates in the weighted RWD sample is more similar to the demographic profile of the RCT. The outcome model is then fit on the combined data of RCT and weighted RWD patients. When all prognostic factors are accounted for, and the propensity score model is correctly specified, the augmented treatment effect is unbiased.

In our proposed analysis method, the hybrid control group consists of both unit-weighted RCT patients and propensity score-weighted RWD patients. The scale of the RWD weights is a function of the study sample sizes, and if one study is much larger than the other, inflated variability among the combined control arm could result in lower power and type-1 error. To address this, we assess methods to rescale the RWD weights, including a frequentist modified power-prior approach that accounts for residual biases between RWD and RCT control subjects when determining a down-weighting factor.

Operating characteristics of the proposed methods are established in a simulation study, where RCT and RWD samples are generated with baseline covariates that are increasingly prognostic of the outcome, and are also increasingly imbalanced between the studies. In doing so, we illustrate conditions where use of the proposed methods to augment trials with RWD yield unbiased estimates with greater precision than RCT-only analyses while maintaining the type-1 error. We provide guidance on how to adequately quantify covariate imbalance and how to accordingly justify the appropriateness of a propensity score approach. We highlight key criteria when selecting suitable RWD based on study population comparability, and highlight practical considerations and limitations when implementing the proposed methods for hybrid control arm analyses.

**Shorter Abstract (237 words):**

Randomized trials (RCTs) are considered the gold standard for estimating causal effects of new therapies and interventions, yet statistical challenges remain in detecting treatment effects among rare disease populations, particularly when outcomes also occur infrequently. In such cases, innovative approaches exist to supplement trials with evidence from real-world data (RWD) by constructing a hybrid control arm, retaining the benefits of randomization while increasing study sample size and power. Identifying suitable RWD for this use case is critical, and data appropriateness are highly dependent on the design of the candidate RCT, its study population, and its primary outcome measure. Even with high-quality RWD, differences in study populations must be properly accounted for and eligibility criteria must be aligned to ensure comparability. Furthermore, it is crucial that all imbalanced baseline characteristics that are prognostic of the outcome are observed in both the RWD and RCT in order to adequately align study populations. In this work, we present propensity score-type weighting and frequentist borrowing methods to align study populations when conducting hybrid control arm analyses. Using simulation techniques, we generate studies under varying degrees of covariate imbalance and illustrate conditions where use of these methods to augment trials with RWD yield greater precision than RCT-only analyses while maintaining the type-1 error. We present key criteria when selecting suitable RWD based on study population comparability, and highlight practical considerations and limitations when implementing the proposed methods for hybrid control arm analyses.