Statistical Science
2024, Vol. 39, No. 4, 566–567
https://doi.org/10.1214/24-STS956
Main article: https://doi.org/10.1214/24-STS945

Rejoinder: Protocols for Observational Studies: Methods and Open Problems

Dylan S. Small

I would like to thank Matias Cattaneo, Rocío Titiunik and Ben Hansen for their insightful comments. I am honored to have these seminal contributors to observational studies research discuss my paper. Their astute comments advance best practices for observational study protocols.

1. PROTOCOLS FOR REGRESSION DISCONTINUITY DESIGNS

Cattaneo, Titiunik and their collaborators have made many pioneering advances in regression discontinuity designs (e.g., [2–4] and [7]) and brought into wider use these valuable designs by writing lucid reviews and textbooks (e.g., [5, 8] and [6]). Their valuable comment provides a comprehensive outline of what should go into a protocol for a regression discontinuity design and the considerations to think about. I hope protocols based on Cattaneo and Titiunik's outline become widely used.

Cattaneo and Titiunik make the good point that sample splitting of the entire data may not be as useful for regression discontinuity designs as for other observational studies because the "localization" of the analysis to scores near the discontinuity reduces the effective sample size. For observational studies based on ignorability, [1] propose pilot designs in which a carefully chosen part of the data that would not otherwise be used in the analysis is used as pilot data to plan the analysis. [1]'s idea might be adapted to regression discontinuity designs by selecting as pilot data a subset of the data that is away from the discontinuity so that its removal from the analysis sample would not contribute to much loss of effective sample size. For example, suppose the local randomization framework is used where the discontinuity in the running variable S is at $S = 6$ and the window in which effective randomization is thought to take place is $6 \leq S \leq 10$. Then all or a randomly chosen subset of the data outside of the window, $S < 6$ and $S > 10$, could be examined before writing the protocol. This pilot data could be useful for deciding which pre-treatment covariates should be adjusted for in the analysis to improve efficiency and which subgroups should be considered in heterogeneity analyses. The pilot data could also potentially be used in selecting outcomes by doing analyses that assume ignorability if one were to assume that even though these ignorability analyses may be biased, they give a correct ordering of how the treatment affects the outcomes.

2. CONVINCING RESEARCHERS TO PREPARE A PROTOCOL

Section 1 of Hansen's comment provides a lucid list of ten reasons for researchers to prepare a protocol. More researchers doing observational studies might be willing to try out protocols if they feel they do not have to develop how to approach the protocol from scratch. Some resources are the following. [11] provide a fillable Word template for an observational study protocol. Hansen and his collaborators have written clear, well thought out observational study protocols that are good models [9, 10]. Cattaneo and Titiunik's comment provide a good road map for a regression discontinuity design protocol.

In Section 4 of his comment, Hansen illuminates several aspects of multiple testing for primary, secondary and exploratory hypotheses. I agree with most of what Hansen says in this section but am less sure about his argument that

> Allocating allowances for type 1 errors among "primary" and "secondary" hypotheses, a process the full research team can contribute to, gives those designations tangible meaning.

This differs from the usual practice of having a separate budget for type 1 errors among primary and secondary hypotheses (or not controlling for multiplicity at all among secondary hypotheses which Hansen cogently advises against). While I see the value in Hansen's suggestion, it could be an excruciating choice for researchers to choose how much type 1 error to budget for the primary vs. secondary hypotheses. Particularly in the social sciences where the selection of primary vs. secondary hypotheses in a protocol is less common than in the medical sciences, it might convince more researchers to select primary vs. secondary hypotheses in a protocol if they are given a fixed 0.05 Type 1 error budget for each of the primary and secondary hypotheses rather than having to make a choice of how much to budget for each.

Dylan Small is Universal Furniture Professor, Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (e-mail: dsmall@wharton.upenn.edu).