Universiteit Utrecht

Methodology & Statistics for the Behavioural, Biomedical and Social Sciences

Thesis proposal

Causal Inference with Observational Data: Confidence Intervals in Emulated Target Trials

Author:

Florian Metwaly

Student-Nr.: 0778265

Supervisor:

Oisín Ryan

Wouter van Amsterdam

Journal: *BMC Medical Research Methodology*

October 10, 2024

Wordcount: 762

Proposal

Randomised Control Trials (RCT) are the gold standard of causal inference in all fields of science. However, due to practical and ethical constraints, it is not always feasible to conduct an RCT (M. A. Hernán & Robins, 2016; Sanson-Fisher, Bonevski, Green, & D'Este, 2007). In health sciences, large-scale observational datasets such as electronic health records databases, are increasingly used for causal analysis when RCTs are not feasible (Bakker, Goossens, O'Kane, Uyl-de Groot, & Redekop, 2021). Target Trial Emulation (TTE; M. A. Hernán and Robins 2016) is a methodological framework which enables robust causal analysis of observational data: By explicitly mimicking an idealized RCT, TTE allows researchers to adjust for immortal time bias (M. A. Hernán, Sauer, Hernández-Díaz, Platt, & Shrier, 2016) previously common in such settings (M. A. Hernán & Robins, 2016). These features of TTE lead to TTE quickly becoming the standard methodological framework for causal analysis in epidemiology and biomedical sciences, particularly in assessing medical treatments' safety and efficacy (Bakker et al., 2021).

One of the core concepts in Target Trial Emulation is that of "time-zero"; the moment at which eligibility to enroll in the hypothetical trial is assessed and follow-up time for treated and control units begins (M. A. Hernán & Robins, 2016). Correct alignment of time-zero allows researchers to avoid the introduction of selection bias (M. A. Hernán, 2018). In an RCT, time-zero is typically aligned by design, but this is not the case in observational data. For instance, if studying the effectiveness of taking against not taking a COVID-19 vaccine on time-to-covid-infection, individuals who are never vaccinated may be eligible to start follow-up at many different time points, while others may be unvaccinated for a period, before later becoming vaccinated. Sequential TTE (M. A. Hernán & Robins, 2016) has been developed to deal with such cases: The researcher emulates a series of trials, each starting on a different calendar date, ensuring that an individual's data is used possibly many times in each relevant emulated trial.

While sequential TTE is methodologically sound (Fu, 2023), it poses several practical challenges for applied researchers. First, sequential TTE typically involves copying each individuals data multiple times to construct the different emulated trial datasets. When TTE

is used in the context of population-scale electronic health databases (Dickerman, García-Albéniz, Logan, Denaxas, & Hernán, 2023; Xie, Bowe, & Al-Aly, 2023), this can involve copying millions of records potentially hundreds of times, posing a considerable computational burden in most standard statistical software packages, such as R (Wickham, 2020). Second, since the analytic dataset now contains the same individuals entered multiple times, the observations are no longer independent, and therefore standard approaches to inference are not possible. In the TTE literature, it is typically recommended to use non-parametric bootstrap approaches to obtain confidence intervals and perform inference (Maringe et al., 2020). However, that too is a considerable practical challenge to researchers, and may be computationally expensive, particularly when the dataset is large.

In this research project we will investigate computationally efficient and statistically valid methods of performing statistical inference in the context of sequential Target Trial Emulation. The project consists of two parts: First, we will implement sequential Target Trial Emulation in the software package Julia. Although R packages for sequential TTE have been developed (e.g. *TrialEmulation*; Su, Rezvani, Seaman, Starr, and Gravestock 2024), we believe that Julia, as it is developed to run fast and handle memory more efficiently (Stefan Karpinski, 2012), may yield considerable computational advantages to R, allowing sequential TTE to run faster and with less memory allocation than existing R implementations. Benchmark tests of our implementation with that of the *TrialEmulation* package will establish whether our expectations are met. Second, we will use a simulation study to investigate the relative performance of the non-parametric bootstrap with naïve and sandwich-type estimators (Danaei, Rodríguez, Cantero, Logan, & Hernán, 2013; M. Hernán, Brumback, & Robins, 2000) for the variance of our estimates. Our simulation study will vary the size of the observational datasets, and the degree to which information from different individuals is used multiple times. We will assess the coverage rates, width of the confidence intervals, power, and Type-I error rates between different methods. This leads to the following research questions: 1) Is our Julia implementation of sequential TTE more computationally efficient than the R implementation? 2) Does non-parametric bootstrap outperform naïve and sandwich-type estimators consistently across different settings? Our expectation is that the more computationally intensive bootstrap standard errors will yield robust infer-

ence across conditions, while the naïve method will consistently underestimate the degree of uncertainty. We expect the bootstrap confidence intervals to generally perform better than the sandwich estimators on the statistical measures. A main interest is, if the sandwich estimator may outperform the non-parametric bootstrap under certain conditions, and if so, whether a consistent pattern can be found there. However, due to the computational cost of bootstrapping, which requires resampling, it is important to evaluate whether the improvements in coverage, Type-I error rate, and statistical power justify the longer computational runtime. This trade-off between statistical performance and computational efficiency will be a key factor in determining whether the bootstrapped confidence intervals are worth using in practice, particularly for large-scale trials.

Bibliography

Bakker, L. J., Goossens, L. M. A., O'Kane, M. J., Uyl-de Groot, C. A., & Redekop, W. K. (2021, September). Analysing electronic health records: The benefits of target trial emulation. *Health Policy and Technology*, *10*(3), 100545. doi: 10.1016/j.hlpt.2021.100545

Danaei, G., Rodríguez, L. A. G., Cantero, O. F., Logan, R., & Hernán, M. A. (2013, February). Observational data for comparative effectiveness research: An emulation of randomised trials of statins and primary prevention of coronary heart disease. *Statistical Methods in Medical Research*, *22*(1), 70–96. doi: 10.1177/0962280211403603

Dickerman, B. A., García-Albéniz, X., Logan, R. W., Denaxas, S., & Hernán, M. A. (2023, September). Evaluating Metformin Strategies for Cancer Prevention: A Target Trial Emulation Using Electronic Health Records. *Epidemiology*, *34*(5), 690. doi: 10.1097/EDE.0000000000001626

Fu, E. L. (2023, August). Target Trial Emulation to Improve Causal Inference from Observational Data: What, Why, and How? *Journal of the American Society of Nephrology*, *34*(8), 1305. doi: 10.1681/ASN.0000000000000152

Hernán, M. A. (2018, February). How to estimate the effect of treatment duration on survival outcomes using observational data. *BMJ*, k182. doi: 10.1136/bmj.k182

Hernán, M. A., & Robins, J. M. (2016, April). Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available: Table 1. *American Journal of Epidemiology*, *183*(8), 758–764. doi: 10.1093/aje/kwv254

Hernán, M. A., Sauer, B. C., Hernández-Díaz, S., Platt, R., & Shrier, I. (2016, November). Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *Journal of Clinical Epidemiology*, *79*, 70–75. doi: 10.1016/j.jclinepi.2016.04.014

Hernán, M. , Brumback, B., & Robins, J. M. (2000, September). Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men. *Epidemiology*, *11*(5), 561.

Maringe, C., Benitez Majano, S., Exarchakou, A., Smith, M., Rachet, B., Belot, A., &

Leyrat, C. (2020, October). Reflection on modern methods: trial emulation in the presence of immortal-time bias. Assessing the benefit of major surgery for elderly lung cancer patients using observational data. *International Journal of Epidemiology*, *49*(5), 1719–1729. doi: 10.1093/ije/dyaa057

Sanson-Fisher, R. W., Bonevski, B., Green, L. W., & D'Este, C. (2007, August). Limitations of the Randomized Controlled Trial in Evaluating Population-Based Health Interventions. *American Journal of Preventive Medicine*, *33*(2), 155–161. doi: 10.1016/j.amepre.2007.04.007

Stefan Karpinski, A. E. J. B., Viral Shah. (2012). *Why We Created Julia.* Retrieved 2024-10-10, from `https://julialang.org/blog/2012/02/why-we-created-julia/`

Su, L., Rezvani, R., Seaman, S. R., Starr, C., & Gravestock, I. (2024, February). *TrialEmulation: An R Package to Emulate Target Trials for Causal Analysis of Observational Time-to-event Data.* arXiv. (arXiv:2402.12083 [stat])

Wickham, H. (2020). *Advanced R, Second Edition* (2nd ed.). New York: Chapman and Hall/CRC. doi: 10.1201/9781351201315

Xie, Y., Bowe, B., & Al-Aly, Z. (2023, March). Molnupiravir and risk of hospital admission or death in adults with covid-19: emulation of a randomized target trial using electronic health records. *BMJ*, *380*, e072705. (Publisher: British Medical Journal Publishing Group Section: Research) doi: 10.1136/bmj-2022-072705

Additional Sources

- Austin, P. C. (2016). Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. *Statistics in Medicine, 35*(30), 5642–5655. https://doi.org/10.1002/sim.7084

- Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine, 34*(28), 3661–3679. https://doi.org/10.1002/sim.6607

- Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine, 19*(9), 1141–1164. https://doi.org/10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F

- Danaei, G., Rodríguez, L. A. G., Cantero, O. F., Logan, R., & Hernán, M. A. (2013). Observational data for comparative effectiveness research: An emulation of randomised trials of statins and primary prevention of coronary heart disease. *Statistical Methods in Medical Research, 22*(1), 70–96. https://doi.org/10.1177/0962280211403603

- DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science, 11*(3), 189–228. https://doi.org/10.1214/ss/1032280214

- Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429246593

- Hall, P. (1988). Theoretical Comparison of Bootstrap Confidence Intervals. *The Annals of Statistics, 16*(3), 927–953.

- Hernan, M. A., & Robins, J. M. (2024). *Causal Inference: What If*.

- Keogh, R. H., Gran, J. M., Seaman, S. R., Davies, G., & Vansteelandt, S. (2023). Causal inference in survival analysis using longitudinal observational data: Sequen-

tial trials and marginal structural models. *Statistics in Medicine, 42*(13), 2191–2225. https://doi.org/10.1002/sim.9718

- Mandel, M. (2013). Simulation-Based Confidence Intervals for Functions With Complicated Derivatives. *The American Statistician, 67*(2), 76–81. https://doi.org/10.1080/00031305.2013.783880

- Mestdagh, M., Verdonck, S., Duisters, K., & Tuerlinckx, F. (2015). Fingerprint resampling: A generic method for efficient resampling. *Scientific Reports, 5*(1), 16970. https://doi.org/10.1038/srep16970

- Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine, 38*(11), 2074–2102. https://doi.org/10.1002/sim.8086

- Naimi, A. I., Cole, S. R., & Kennedy, E. H. (2017). An introduction to g methods. *International Journal of Epidemiology, 46*(2), 756–762. https://doi.org/10.1093/ije/dyw323

- Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research, 22*(3), 278–295. https://doi.org/10.1177/0962280210395740