Internship on sequential emulated trials

Iqraa Meah[1*] and François Petit[1*]

[1]Université Paris Cité and Université Sorbonne Paris Nord, Inserm, INRAE, Center for Research in Epidemiology and StatisticS (CRESS), Paris

[*]Correspondence at iqraa.meah@inserm.fr or francois.petit@inserm.fr

January 16, 2025

Internship Details

Location:
Faculté de Médecine - Site Villemin
Université Paris Cité
10 Av. de Verdun, 75010 Paris

Duration:
6 months

Supervisors:
Iqraa Meah, François Petit

1 Context

Randomized Controlled Trials (RCTs) are often presented as the gold standard for studying the causal relationship between a treatment and an outcome. As much as they are desired, they are also difficult or sometimes impossible to carry out due to financial, logistical, or ethical constraints. To compensate for these use case limitations, researchers can use observational data when they are available. This kind of data provides records of real-world patient experiences that can be exploited to emulate a target trial, a concept discussed extensively by Hernán and Robins (2016). As a prompt summary, emulated trials use the methodology of RCTs to clearly define the causal question that can be addressed using a set of observational data on the effects of an intervention. This emulation comes with limitations. Some assumptions required for causal inference in emulated trials are inherently unverifiable and must rely on expert judgment, introducing a degree of uncertainty.

With that in mind, one main statistical challenge is to come up with estimators of the treatment effect that incorporate correction for the confounding that occur when the treatment assignment is not random, as is the case when working with observational data. Addressing this issue is a central topic in causal inference, and extensive research continues in this field.

Another kind of bias can be due to the exploitation of the data rather than to its nature. This is the case for the so-called immortal-time bias which occurs when there is a time gap between treatment assignment and the treatment initiation. Not accounting for this gap can result in bias

that disproportionately favors one group over the other, often benefiting the treatment group. To address this bias, particularly in observational studies, a method known as cloning and censoring has gained popularity. This approach involves creating a clone of the patient in the control group during the period they are waiting to initiate treatment. Once the patient begins treatment, the control group clone is censored, effectively correcting for the misclassification caused by the delay.

Immortal time bias is particularly likely to occur when studying transplantation as a treatment using observational data. For example, if a person who eventually receives a transplant is classified in the treatment group from the start, even though their actual transplant date is later, this misclassification would artificially favor the treatment group. The survival time between being placed on the transplant waiting list and the actual transplant would be incorrectly attributed as survival time under treatment. In this context, along with the cloning and censoring practice, conducting a sequence of emulated trials with time-varying T_0 —representing the start time of each trial—can be an effective strategy to reduce the immortal-time bias (Doubeni, Corley, Quinn, et al. 2018; Suissa 2008).

Performing sequential emulated trials is a promising methodology that has gained increasing attention in the literature in recent years. For instance, Thomas et al. (2020) provides a comprehensive review, and Keogh et al. (2023) compares the performance of these trials with well-established epidemiological methods, such as marginal structural models. Nonetheless, statistical investigations about the properties of treatment effect estimators coming from pooled results of a sequence of emulated trials remain highly desirable.

2 Goals

We propose a six-month internship within the CRESS METHODS team to focus on sequential emulated trials. The primary objective is to develop an initial understanding of the convergence properties of an estimator pooled from a sequence of emulated trials. To achieve this, we expect the student to begin with a literature review, leveraging references such as those cited above, to build a solid foundation in causal inference. Subsequently, the student will conduct simulation studies to recreate various strata of sequential trials, aiming to empirically evaluate the convergence behavior under different trial scenarios. Finally, we also expect/encourage the student to contribute to other team activities such as a biweekly reading group on causal inference and the team seminar.

3 References

References

Hernán, Miguel A. and James M. Robins (2016). "Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available: Table 1". In: *American Journal of Epidemiology* 183.8, pp. 758–764.

Doubeni, Chyke A, Douglas A Corley, Victoria P Quinn, et al. (2018). "Effectiveness of screening colonoscopy to prevent colorectal cancer among Medicare beneficiaries aged 70 to 79 years". In: *Annals of Internal Medicine* 169.9, pp. 603–611.

Suissa, Samy (2008). "An unjustified benefit: immortal time bias in the analysis of time-dependent events". In: *Statistical Methods in Medical Research* 17.5, pp. 487–496.

Thomas, Laine E. et al. (2020). "Matching with time-dependent treatments: A review and look forward". In: *Statistics in Medicine* 39.17, pp. 2350–2370.

Keogh, Ruth H. et al. (2023). "Causal inference in survival analysis using longitudinal observational data: Sequential trials and marginal structural models". In: *Statistics in Medicine* 42.13, pp. 2191–2225.