Title

Causal healthcare analytics for Real-World Evidence with Targeted Learning: A cross-disciplinary, cross-sector approach

PROJECT DESCRIPTION

Abstract

The National Institute for Health and Care Excellence (NICE)) provides guidance on best practices in health and social care, including public health, to the NHS in England and Wales which may be based on assessment of complex real-world evidence (RWE) for new health technologies and treatments. The use of real-world data (RWD), in addition to randomised controlled trials (RCTs), has become increasingly popular for biomedical researchers, regulatory bodies and pharmaceutical industry. Evaluating reliable RWE from RWD is challenging due to lack of treatment randomisation, intercurrent events, and informative loss to follow-up. Targeted Learning (TL) integrates mathematical statistics, machine learning, and causal inference and offers an ideal step-by-step framework to address these challenges.

TL has been widely and successfully applied with USA industry and the USA Food and Drug Administration (FDA). Building on our growing implementation of TL methods at UoE (supervisors Dr Beentjes and Dr Khamseh), we bring together external partners and end-users from NICE (Dr Stephen Duffield, Dr Michael Merchant), pharmaceutical industry (Dr Di Zhang, Teva Pharmaceuticals, former mathematical statistician at FDA) and academia (Prof Mark van der Laan, UC Berkeley's School of Public Health). The aim of this project is to develop and apply TL estimators to real-world healthcare problems for the benefit of patients.

Introduction

RWE is relevant for a variety of scenarios such as pragmatic trials, commonly used for label changes on already approved products, including drug repurposing, to reduce burden on patients and healthcare professionals, and for designing new healthcare policies.

Current NICE RWE best practice frameworks emphasise the importance of good causal inference methods including references to methodologies such as multi-variate regression and propensity score estimation. However, these traditional methods can suffer from model-misspecification, implying the inferred estimates may be false negatives, false positives, or even have the incorrect sign, relative to the unknown ground truth. Where biases are introduced, this could lead to lack of power for detection of a beneficial intervention, or lead to over- or underestimation of the efficacy of treatment intervention potentially with serious consequences for downstream healthcare policy.

The Targeted Learning (TL) Roadmap offers a step-by-step guide to estimating causal effects from RWD with minimal bias [1]. TL has been successfully applied in high-impact real-world settings in the USA to generate RWE in collaboration with FDA [1], with the US government CoVPN/Biostatistics Team [2], with industry partners [3] (for

quantifying antibody levels predicting COVID-19 vaccine efficacy), and for research [4] (HIV Testing and Treatment in Rural Africa). TL offers a unified framework that equally applies to RWD and to RCTs affected by loss-to-follow-up, dropouts, or non-compliance. Additionally, TL can generate valid and reliable RWE from RCTs that use RWD such as pragmatic trials, commonly used for label changes on already approved products, including drug repurposing, to reduce burden on patients and healthcare professionals.

Research challenge

NICE provides guidance on best practices in health and social care, including public health, to the NHS in England and Wales, which may rely on assessment of complex RWE for new health technologies and treatments. Evidence generated from randomised controlled trials (RCTs) remains the gold standard. However, RCTs may not be sufficient to address the research question of interest, due to e.g., a small number of eligible patients, ethical considerations, and differences in populations and care pathways which may impact reliable transferability to the target population in the NHS. The NICE RWE framework (06/2022) aims "to use real-world data to resolve gaps in knowledge and drive forward access to innovations for patients." Regulators, pharmaceutical companies and academia have a pressing need to access mathematically rigorous techniques to analyse RWE in a robust, transparent, and reliable manner.

The opportunity of developing and applying novel Targeted Learning estimators benefits all sides of healthcare evidence submission: (i) NICE in their assessment of RWE submissions, (ii) pharmaceutical companies in the preparation of RWE submission to NICE, (iii) those collecting/analysing data in the NHS, and (iv) patients.

Data & Methodology

We will develop and apply a series of novel semi-parametric estimators in the framework of TL, establish their theoretical properties, and establish their performance through realistic simulations and applications to real-world data. Examples of such estimators include

- The causal effect of treatment on outcome in the presence of a high-dimensional set of potential confounders. This applies to questions in observational healthcare data (DataLoch), and genomic medicine (UK Biobank, All-of-Us, to identify DNA variants causal of trait or disease).
- The causal effect of treatment on composite outcomes applied for quantification of multiple endpoints in clinical trials, or for increasing power in detecting for genomic variants associated with multiple related diseases or traits.
- Estimators of causal effects in the presence of non-ignorable missingness, this is important for extracting RWE from routinely collected NHS data (DataLoch).
- Assessment of the robustness of and estimator of interest (e.g., with respect to model misspecification or low sample size) through realistic simulations and the development of a theory of diagnostic estimators.

The performance of these estimators will be evaluated on realistic CPRD (Clinical Practice Research Datalink) high-fidelity synthetic data sets with known ground truth.

Responsible AI/ethical considerations

The TL framework allows for the construction of reliable and robust estimators with theoretical guarantees which rely on minimal assumptions. These estimators are designed to minimise bias due to model misspecification and are transparent by construction.

Produced code and software will be made publicly available on GitHub, as part of our TMLE.jl software package and environment [6], and will include unit testing, continuous integration, and version control, for robustness and reliability.

Expected outcomes & Impact

The development of novel semi-parametric estimators in the framework of TL, and the establishment of their theoretical and numerical properties. Step-by-step explainable methodology and software for application to real-world routine healthcare data, as well as genomic data, to a standard usable by regulatory bodies and pharmaceutical companies

Impact: By (i) informing NICE RWE best practice recommendations around up-to-date state-of-the-art causal inference techniques, (ii) explore the possibility of providing NICE free access to mathematically rigorous, safe, reliable, and robust software tools to help NICE assess relevant submissions supported by RWE. Academics and clinical researchers at the NHS can generate new RWE supporting their RCTs.

References

[1] Gruber, S. *et al.* (2023). Targeted Learning: Toward a Future Informed by Real-World Evidence. Statistics in Biopharmaceutical Research, 16(1), 11–25. https://doi.org/10.1080/19466315.2023.2182356

[2] Fong, Y. *et al.* Immune correlates analysis of the ENSEMBLE single Ad26.COV2.S dose vaccine efficacy clinical trial. *Nat Microbiol* **7**, 1996–2010 (2022). https://doi.org/10.1038/s41564-022-01262-1

[3] Gilbert, P.B. *et al.*, Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. *Science* **375**,43-50(2022).DOI:10.1126/science.abm3425

[4] Havlir, D.V., Balzer, L.B. *et al.* HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa. *N Engl J Med* **381(3)**, 219-229 (2019). DOI: 10.1056/NEJMoa1809866

[5] **Software**: TMLE.R (Gruber S, van der Laan M (2012). "tmle: An R Package for Targeted Maximum Likelihood Estimation." Journal of Statistical Software, 51(13), 1–35. doi:10.18637/jss.v051.i13)

[6] **Software:** TMLE.jl (Labayle, O., Beentjes, S., Khamseh, A., & Ponting, C. (2024), https://github.com/olivierlabayle/TMLE.jl)