Causal Inference with Time-to-Event Outcomes: A Debiased Machine Learning Framework for Treatment Parameters

Author: Meimei Chen, PhD Candidate, Macquarie University
Supervisors: Yuanyuan Gu, Henry Cutler

ECR Pre-Congress Session iHEA 2025

Discussed by Nhung Nghiem, Australian National University

19-23 July 2025, Bali, Indonesia

Background

- The effectiveness of health interventions
- Advancements in data
- Integrate machine learning (ML) with causal inference

Objective

- Aims to adapt the double machine learning (DML) framework for use with survival data

→ critical in health economics research for understanding causal effects when the outcome variables are subject to censoring.

Significance

- Allows for the use of a broad array of state-of-the-art ML methods
- Ensure that the estimator remains unbiased and normally distributed asymptotically
- Allows for the derivation of valid confidence intervals and hypothesis testing.

Methodology

- This study develops a new framework for causal analysis with right-censored time-to-event data based on DML.
- Use the three fundamental assumptions of Rubin’s potential outcome framework (i.e., exchangeability, consistency, and positivity)
- **New development**: assume that right censoring is independent of the outcome given all covariates.

- Provide a Neyman orthogonal score function for the parameters of interest, nuisance parameters, and censoring indicator → To remove bias introduced by ML methods
- The Neyman orthogonality ensures that the estimator is locally insensitive to the values of nuisance parameters.

- Used simulated dataset to evaluate the accuracy and stability of the method.
- Showed that various state-of-the-art ML models, including random forests, gradient boosting machines, and deep neural networks, can be integrated to estimate the nuisance parameters without compromising the consistency of the parameters of interest.
- Compared the performance of the method with several conventional statistical methods on simulated datasets.
- Compared the method to alternatives based on absolute bias, mean squared error, and standard deviation.

→ The results showed that the method had relatively smaller bias and lower variance.

Conclusion

- The study stated that it successfully extended the econometric causal inference framework DML to incorporate time-to-event outcomes, showing robust performance on both simulated and real-world datasets (*results pending*).
- This method offers a powerful tool for health economists, enabling more accurate causal inference in studies involving time-to-event data.

Strengths and Limitations

- The research aim was clear and well-motivated
- The research design was appropriate
- The manuscript was well-written
- The method was well-supported by theoretical background, but of note, I am an applied researcher so providing comments from this perspective.
- The results suggested advantages of DML over other benchmark models

Objectives and Aims

- Clarify objectives vs aims (“Aims to adapt the DML framework for use with survival data”)
- Objectives, e.g., to design a novel estimator or to evaluate the method

Literature review

- Detailed & thorough, such as Wager and Athey (2018)
- But the specific literature that this work extends or builds upon could be clearer
- Such as this work extends the paper by Chernozhukov et al. (2018).
- Which study was used to develop the method “DML framework to censored time-to-event data”?
- Some repetitions between the introduction and the related literature section, especially about DML

Methods

- State what is the final model: such as the model include equations (1) and (2), with the extension in equation (3)
- Provide codes on Github or Appendix
- Number of covariates for the simulation seems to be very small, d=10 and 20, that raises a question does ML needed for this work?
- Should benchmark against Cox model or g-formula model (without ML components)

Methods

- New assumptions re right censoring: How this assumption was validated? It was not clear to me which study that this study was built upon regarding the censoring.
- Compare to real datasets, or previous papers with available data → more external validations to the model

- “Our approach is grounded in this framework, allowing for clear definitions of estimands and applicability to both randomised and observational studies.”?

3.4.2 Censoring Mechanism

In our assumptions, censoring can be affected by covariates W and treatment A. We can treat it as the "event" in survival analysis and define its conditional hazard function $m(t)$ as:

$$\begin{aligned} m(t \mid A, W) &= P(\tilde{T} = t, \Delta = 0 \mid \tilde{T} \geqslant t, A, W) \\ &= P[dM(t) = 1 \mid N(t-1) = 0, M(t-1) = 0, A, W] \end{aligned}$$

We can estimate the censoring mechanism using the same method as the conditional hazard function of the outcome variable.

J Causal Inference. Author manuscript; available in PMC: 2023 Aug 25.

Published in final edited form as: J Causal Inference. 2022 Dec 9;10(1):415–440. doi: 10.1515/jci-2022-0004

Doubly robust estimators for generalizing treatment effects on survival outcomes from randomized controlled trials to a target population

Dasom Lee [1], Shu Yang [2,*], Xiaofei Wang [3]

Author information Article notes Copyright and License information

PMCID: PMC10457100 NIHMSID: NIHMS1875530 PMID: 37637433

Abstract

In the presence of heterogeneity between the randomized controlled trial (RCT) participants and the target population, evaluating the treatment effect solely based on the RCT often leads to biased quantification of the real-world treatment effect. To address the problem of lack of generalizability for the treatment effect estimated by the RCT sample, we leverage observational studies with large samples that are representative of the target population. This article concerns evaluating treatment effects on survival outcomes for a target population and considers a broad class of estimands that are functionals of treatment-specific

Algorithm 2: Double Machine Learning for Survival Data

input : Data $\mathcal{O}$, Neyman orthogonal score ψ from (3.3)

output: $\hat{\theta}_0$

for $k = 1, \ldots, K$ **do**

Divide the sample into K-fold random partition $(I_k)_{k=1}^{K}$ such that the size of each fold I_k is $n = N/K$;

Define $I_k^c := \{W_1, \ldots, W_N\} \backslash I_k$;

Construct an ML estimator $\hat{\eta}_{0,k} = \hat{\eta}_0((W_i)_{i \in I_k^c}) = (\hat{n}(t), \hat{m}(t), \hat{g}(a, w))$ of η_0 on I_k^c;

Construct the estimator $\hat{\theta}_{0,k}$ on I_k as the solution of $E_{n,k}[\psi(W; \hat{\theta}_{0,k}, \hat{\eta}_{0,k})] = 0$;

if *achievement of exact 0 is not possible* **then**

Define the estimator $\hat{\theta}_{0,k}$ of θ_0 as an approximate ϵ_N-solution:

$\|E_{n,k}[\psi(W; \hat{\theta}_{0,k}, \hat{\eta}_{0,k})]\| \leqslant \inf_{\theta \in \Theta} \|E_{n,k}[\psi(W; \theta, \hat{\eta}_{0,k})]\| + \epsilon_N$,

where $\epsilon_N = o(\delta_N N^{-1/2})$ and $(\delta_N)_{N \geqslant 1}$ is some sequence of positive constants converging to zero.

end

end

Aggregate the estimators:

$\hat{\theta}_0 = \frac{1}{K} \sum_{k=1}^{K} \hat{\theta}_{0,k}$.

Simulated datasets:

- The goal of the simulation exercise
- The rationale for each parameter, equation
- Theory behind these equations
- Probably check: Wager and Athey (2018)

- “Estimate the nuisance parameters using state-of-the-art ML methods. ..The selected algorithms include Random Survival Forests (RSF), Gradient Boosted Models (GBM), and Deepsurv which leverages Deep Neural Networks to survival analysis.” (p15)

→Some clarifications needed.

→How test, validation and training datasets were created? How did the sample was partition to these datasets, such as 80% for training and 20% for test?

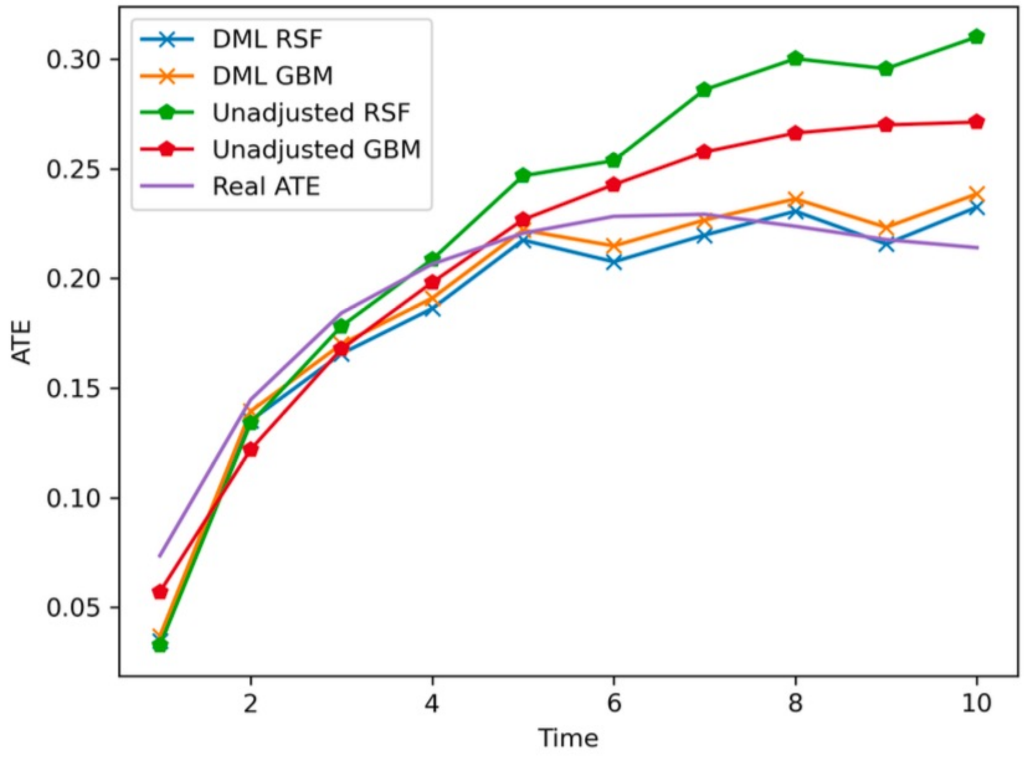
- “The hyperparameters are selected based on minimising the error or maximising the prediction accuracy on the inner test set. Once the optimal hyperparameters are determined, the model is evaluated on the outer test set.” (p16)

- Reference for hyperparameters optimisation
- Tool from ML literature
- Python package for tunning parameters & running the ML models

- Evaluation matrix: The C-index
- Other index: F1, AUC
- Matrix for continuous vs categorical outcomes

Results

- Max time =10
- Confidence interval for the k-fold cross-validation
- Deepsur model
- Cox L1 model
- Explain the ATE



- Does that make sense when DML RSF bias ~ 1/3 of AIPCW
- Standard errors seem to be very small, eg, IPCW (2000 iterations, ATE=0.62, SD=0.000)

The following table shows the results comparing with the two conventional statistical methods with $d = 10$:

N	**DML RSF**	**DML GBM**	**DML Cox-l1**	**IPCW**	**AIPCW**
200	**0.057**(0.002)	0.061(0.001)	0.061(0.002)	0.126(0.004)	0.088(0.002)
500	**0.037**(0.000)	0.049(0.001)	0.039(0.000)	0.093(0.002)	0.070(0.001)
1000	**0.032**(0.001)	0.034(0.001)	0.035(0.001)	0.076(0.001)	0.069(0.001)
1500	**0.030**(0.000)	0.031(0.000)	**0.030**(0.000)	0.072(0.001)	0.057(0.001)
2000	0.016(0.000)	**0.014**(0.000)	0.021(0.000)	0.062(0.000)	0.045(0.000)

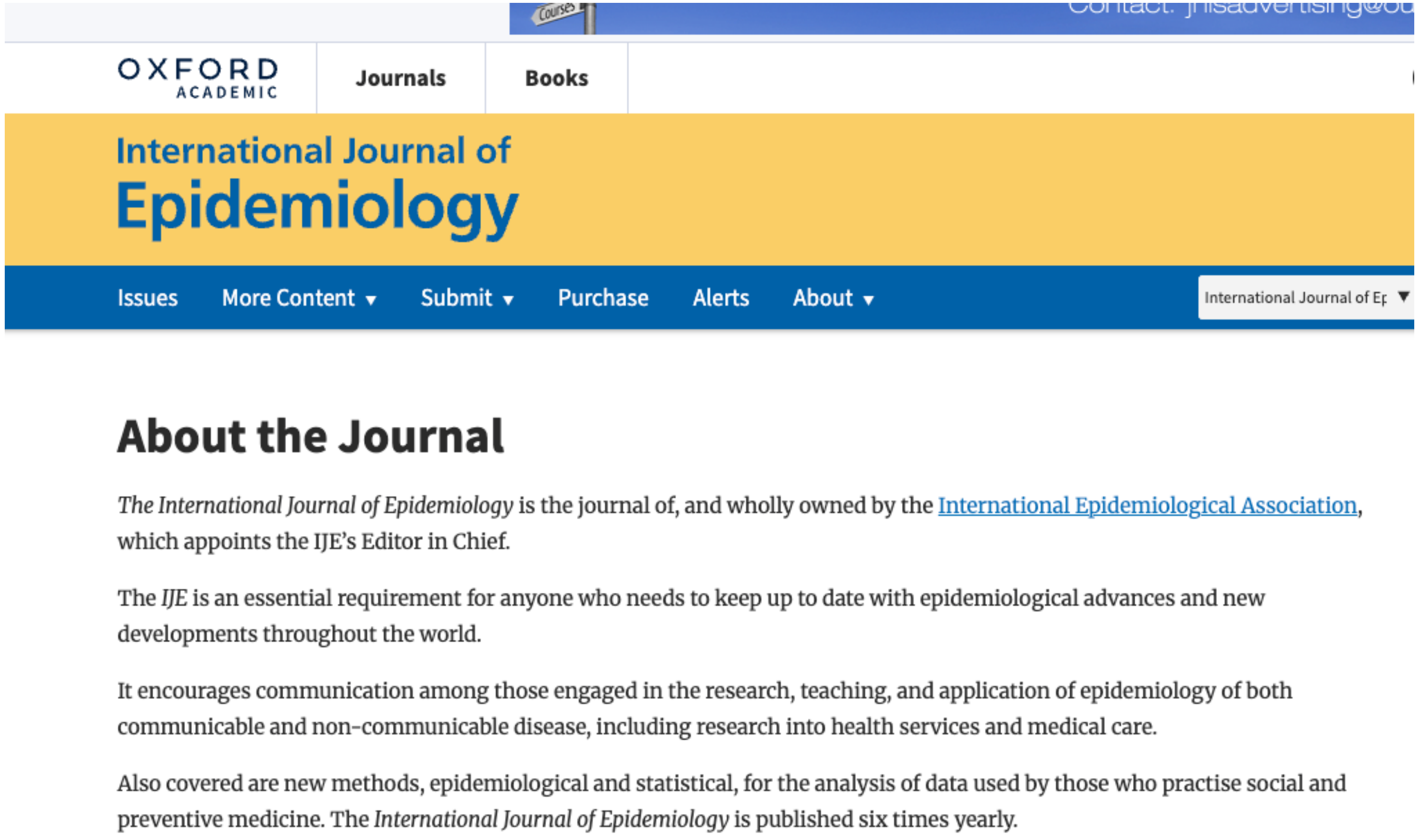
- Limit with the number of features
- Trade-off in computational time

Real-world data

- Discontinuation of medications
- Non-recurring events → implications for discontinuation of medications
- Traditional survival analysis example, such as vaccine effectiveness (Time-varying factor, competing risks)

Pass on to Meimei for clarifications!

Potential journals to consider



OXFORD ACADEMIC | Journals | Books

International Journal of Epidemiology

Issues | More Content | Submit | Purchase | Alerts | About

About the Journal

The International Journal of Epidemiology is the journal of, and wholly owned by the International Epidemiological Association, which appoints the IJE's Editor in Chief.

The *IJE* is an essential requirement for anyone who needs to keep up to date with epidemiological advances and new developments throughout the world.

It encourages communication among those engaged in the research, teaching, and application of epidemiology of both communicable and non-communicable disease, including research into health services and medical care.

Also covered are new methods, epidemiological and statistical, for the analysis of data used by those who practise social and preventive medicine. The *International Journal of Epidemiology* is published six times yearly.

Health Economics

HOME | ABOUT | AUTHORS | BROWSE

Overview

Members of the International Health Economics Association are eligible for a special subscription rate. For rates, and to subscribe, please follow this link and select 'Other'.

Aims and Scope

This Journal publishes articles on all aspects of health economics: theoretical contributions, empirical studies and analyses of health policy from the economic perspective. Its scope includes the determinants of health and its definition and valuation, as well as: the demand for and supply of health care; planning and market mechanisms; micro-economic evaluation of individual procedures and treatments; and evaluation of the performance of health care systems.

Contributions should be original and innovative. As a rule, the Journal does not include routine applications of cost-effectiveness analysis, discrete choice experiments and costing analyses. Editorials, which are regular features, should be concise and topical. Occasionally, commissioned reviews are published, and special issues bring together contributions on a single topic.

Short Research Articles, also known as Health Economics Letters, are concise articles of new research findings that make a significant contribution to knowledge. As a rule, Health Economics Letters only considers contributions that can attain the standard quality expected of this Journal within the concise format. Health Economics Letters does not routinely consider opinion or commentary pieces.

WILEY
Online Library

38 results for **""machine learning""** anywhere published in **"Health Economics"**

ARTICLES & CHAPTERS (38)

Filters

Refine Search

Relevance

Export Citation(s)

SHORT RESEARCH ARTICLE

Predicting vaccine hesitancy from area-level indicators: A machine learning approach

Health economics published many ML studies in the last two years

→ the literature may need to be updated

→ the novelty of this research might need to be revisited

RESEARCH ARTICLE Open Access

A Double Machine Learning Approach for the Evaluation of COVID-19 Vaccine Effectiveness Under the Test-Negative Design: Analysis of Québec Administrative Data

Cong Jiang, Denis Talbot, Sara Carazo, Mireille E. Schnitzer

Statistics in Medicine | Volume 44, Issue 5

First published: 21 February 2025

Abstract ⌄

RESEARCH ARTICLE

Evaluating the effectiveness of ground motion intensity measures through the lens of causal inference

Henry V. Burton, Jack W. Baker

Earthquake Engineering & Structural Dynamics | Volume 52, Issue 15

First published: 09 August 2023

Abstract ⌄

Thank you!