A Causally Formulated Hazard Ratio Estimation through Backdoor Adjustment on Structural Causal Model

Riddhiman Adib [1], Paul Griffin [2], Sheikh Iqbal Ahamed [1], Mohammad Adibuzzaman [2]
[1] Department of Computer Science, Marquette University, USA
[2] Regenstrief Center for Healthcare Engineering, Purdue University, USA

Published in: Machine Learning for Healthcare, 2020

Introduction

- **Survival Analysis:** A set of statistical approaches used to investigate the time it takes for an event of interest to occur
  - **Survival Curve:** graphically reports the hazard in a population by plotting the fraction of the population that survived in the treatment and the control group over time
  - **Hazard Ratio:** quantitatively reports comparative hazard rates between two levels of treatment, estimated through Cox Proportional Hazards Model

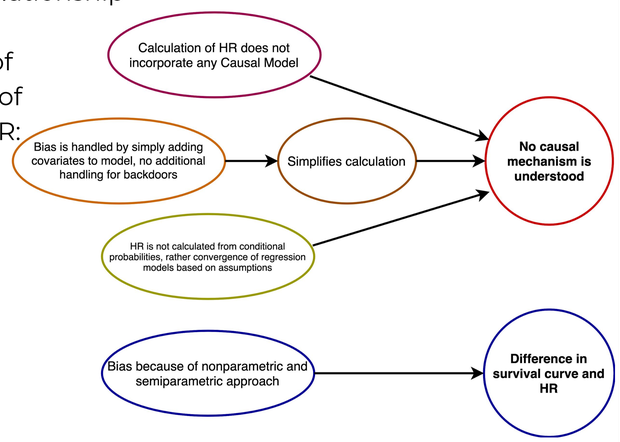
- **Issues with HR in Observational Studies:**
  - Lack of causal interpretation from HR
    - Does not address whether the treatment is ”causing” the hazard of the outcomes
    - Difference in Randomised studies (e.g. RCTs) and. observational studies

- **Causally Formulated Hazard Ratio:**
  - SCM for “simulating” RCTs from observational data
  - *Hypothesis:* SCM can aid in generation of causally interpretable HR.

Background and Related Works

- Overview of similar previous works:
  - Survival curve through IPW with no adjusted survival time, no HR, no SCM [1]
  - Kaplan Meier estimator through IPW without involving SCM [2]
  - No predominant existing approach to explain HR for causal relationship

- Summary of limitations of standard HR:



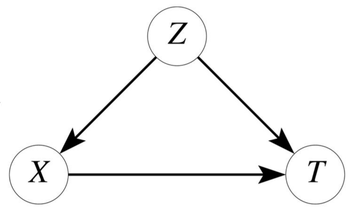
Problem Statement

How can we compute compute the HR for an observational study, which clearly conveys causal effect of treatment on outcome?

- *Our approach:* We plan to leverage the SCM with explicitly declaring our assumptions and adjusting for the right confounders
- Assumptions:
  - the observational data is available
  - no hidden confounders (causal sufficiency)
  - the SCM is fully specified

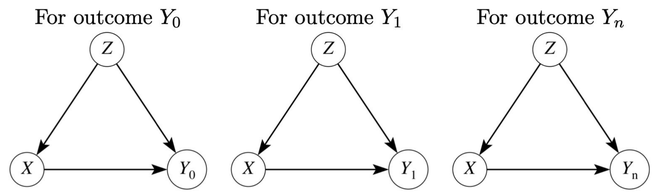
Methods

- A simple observational study:
  - treatment X (dichotomous)
  - outcome in survival time, aka, time-to-event T (continuous)
  - a single confounder Z (dichotomous)

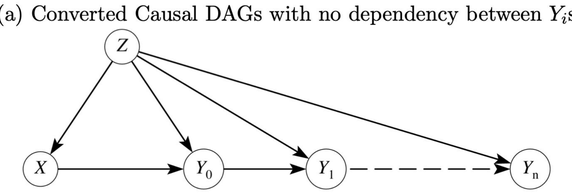


Transformation of single study to multiple studies

- From survival time, we calculate the ***i***-th day survival,***i*** being number of days, generating new Causal DAGs
- For each of the newly generated Causal DAGs *(fig (a) or (b) below)*, we adjust for the confounder using the backdoor adjustment formula
- We get adjusted probabilities ***P_adj***, adjusted counts ***C_adj***, and adjusted survival time ***T_adj***
- ***P_adj*** --→ survival curve through Kaplan Meier fitter.
- ***T_adj*** --→ Cox PH Model --→ HR *(We do not need confounder Z anymore since outcome is adjusted accordingly)*



(a) Converted Causal DAGs with no dependency between Y_is



(b) Converted single Causal DAG with dependencies between Y_is

Experiments

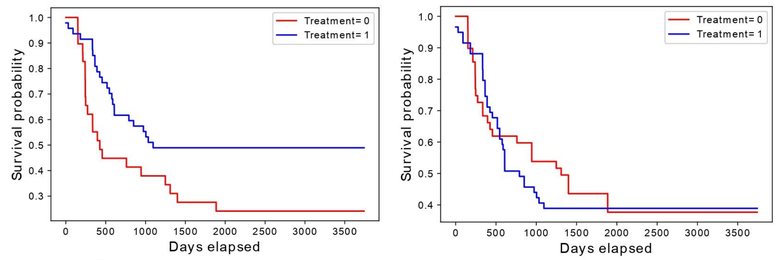


Figure: Unadjusted versus adjusted survival curve for Ewing dataset

	Original model		Adjusted model
	Without confounding	With confounding	After adjustment
Hazard Ratio	0.53 (0.30-0.96)	1.12 (0.59-2.11)	1.04 (0.57-1.87)

Figure: Hazard Ratio calculated in different ways for Ewing dataset

- In both simulated and ewing dataset, traditional methods generate a difference in survival curves and hazard ratios calculated
- Our proposed methodology adjusts properly and generates an appropriate and synchronous hazard ratio and survival curve

Discussion

- Our proposed approach alters the original SCM into multiple SCMs with different endpoints
  - Doing so enables us to calculate conditional probabilities and thus backdoor adjustment on SCM
  - Since our method only uses treatment and adjusted outcome, the HR calculated reflects true and direct causal relationship of treatment and outcome
- Through the backdoor adjustment, we get rid of biases from confounders and look at causal effect of treatment on outcome
- *Limitation*: Knowledge of true Causal DAG
  - Active research has been going on to develop statistical and computational algorithms of causal structure learning

References

1. Makuch, R. W. (1982). Adjusted survival curve estimation using covariates. Journal of chronic diseases, 35(6), 437-443.
2. Satten, G. A., & Datta, S. (2001). The Kaplan–Meier estimator as an inverse-probability-of-censoring weighted average. The American Statistician, 55(3), 207-210.

For more information about our project, please email:
Riddhiman Adib (riddhiman.adib@marquette.edu)