Evaluating Treatment Benefit Predictors using Observational Data: Contending with Identification and Confounding Bias

Yuan Xia *, Mohsen Sadatsafavi† and Paul Gustafson*

† Faculty of Pharmaceutical Sciences, *Department of Statistics, University of British Columbia, Vancouver, BC CANADA

Problem Setting

Treatment benefit is defined as the conditional risk reduction for treated individuals given their characteristics, compared to the conditional risk they would face under identical conditions without the treatment (CATE), denoted as

$$\mathrm{E}[B \mid X = x] = \mathrm{E}[Y^{(1)} \mid X = x] - \mathrm{E}[Y^{(0)} \mid X = x],$$

where $Y^{(a)}$ is counterfactual outcome under treatment $A = a$. A treatment benefit predictor (TBP) is a function of X, denoted as $h(x)$, which predicts $\mathrm{E}[B \mid X = x]$.

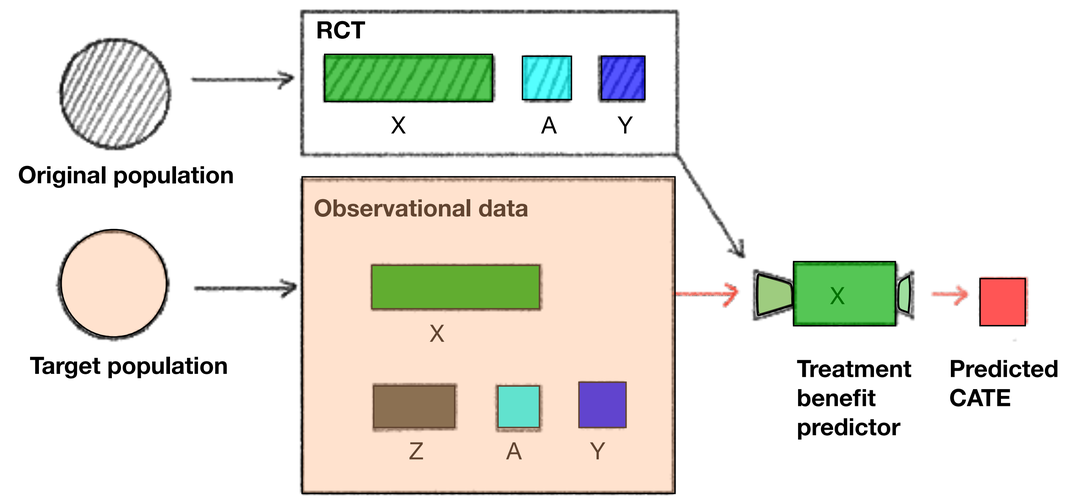

Figure 1: A pre-specified TBP was developed from unknown original population and will be evaluate on the target population using observational data.

How to evaluate the predictive performance of a pre-specified $h(x)$ using observational data (Y, A, X, Z) on the target population?

Assumptions

Key assumptions to evaluate $h(x)$ using observational data are as follows: (1) No interference; (2) Consistency; and (3) Conditional exchangeability given the set of variables $X \cup Z$.

Predictive Performance Metrics

Discrimination

The Concentration of the Benefit Index (C_b) [1] is defined as

$$C_b = 1 - \frac{\mathrm{E}[B]}{\mathrm{E}[B\eta(H)]},$$

where $H := h(X)$, $\eta(H) = 2F_H(H) - f_H(H)$, $F_H(\cdot)$ denotes cumulative distribution function of H, and $f_H(\cdot)$ is probability mass function of H.

Calibration

A $h(x)$ can be considered moderately calibrated [2] if for any h,

$$\mathrm{E}[B \mid H = h] = h,$$

where $\mathrm{E}[B \mid H = h]$ is the moderate calibration function.

Unlike the observed outcome Y, the individual treatment benefit B is unknown. How can we identify predictive performance measures using observational data (Y, A, X, Z)?

Calculate $\mathrm{E}[B \mid X = x, Z = z] = \mathrm{E}[Y \mid A = 1, X = x, Z = z] - \mathrm{E}[Y \mid A = 0, X = x, Z = z]$, which hold due to the assumptions.

TBPs Evaluation Results

Confounding bias might occur when X alone is insufficient to control for confounding and denote the bias as a function of X, which is

$$\text{bias}(x) = \mathrm{E}[Y \mid A = 1, X = x] - \mathrm{E}[Y \mid A = 0, X = x] - \mathrm{E}[B \mid X = x].$$

What is the bias of predictive performance metrics due to not fully control for confounding?

Target population 1

(Binary variables Y, A, X_1, X_2, and Z)

	$h_1(x_1, x_2)$	$h_2(x_1, x_2)$	$h_3(x_1, x_2)$	$h_4(x_1, x_2)$
$\tilde{C}_b$ **(No)**	**0.331**	**0.331**	**0.345**	**0.345**
C_b **(Yes)**	**0.607**	**0.607**	**0.63**	**0.63**

Table 1: Note that h_1 is the mean of covariates, h_2 is designed to be moderately calibrated, h_3 is designed to be strongly calibrated, and h_4 is design to be incorrect $\mathrm{E}[B \mid X = x]$.

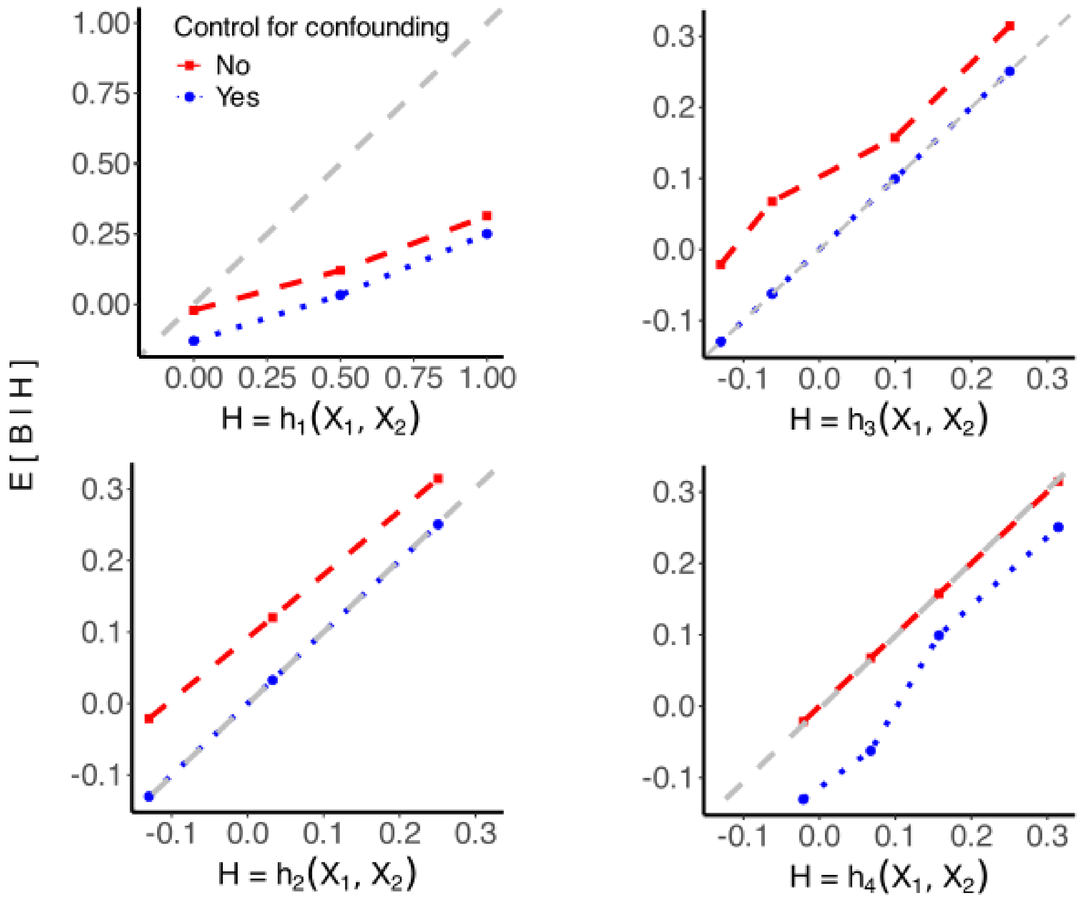

Figure 2: The moderate calibration plots for the four TBPs, where the blue curves represent $\mathrm{E}[B \mid H]$ and red curves represent $\tilde{\mathrm{E}}[B \mid H]$.

How confounding bias and the bias of performance metrics are influenced by the strength of confounding?

Target population 2

(Binary variables Y, A, X, and Z, with α and β control for strength of confounding)

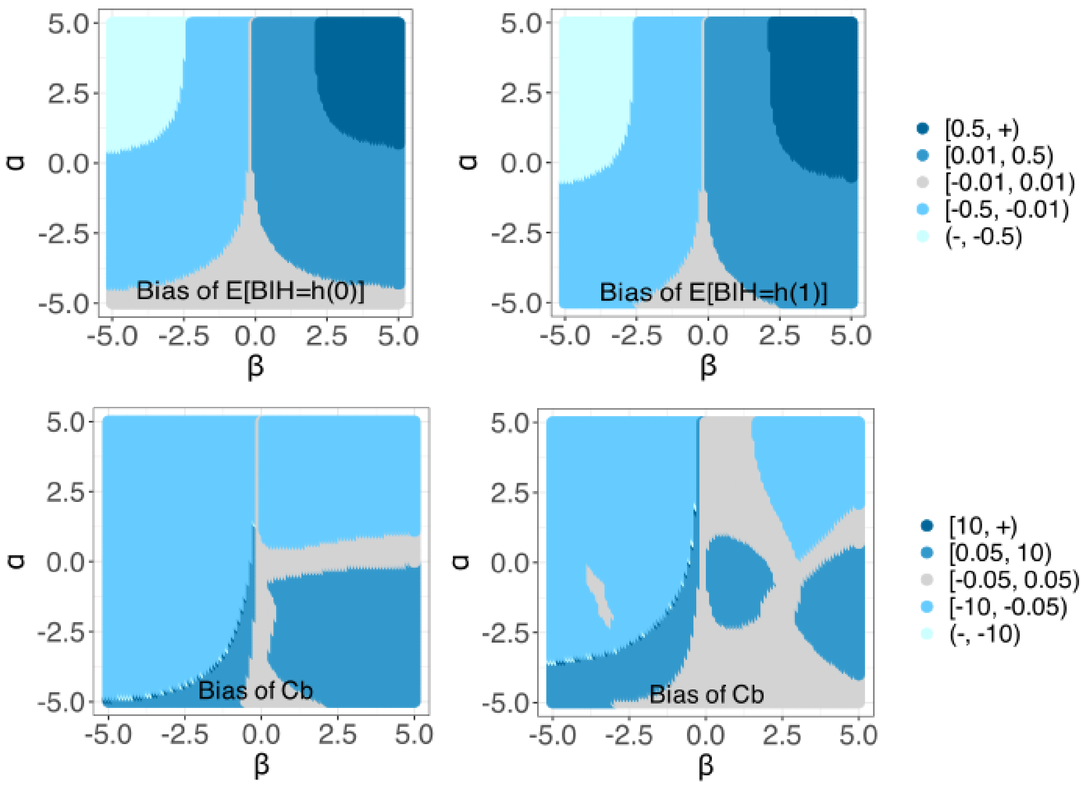

Figure 3: Heatmap of performance metrics for $h(x) := \mathrm{E}[B \mid X = x]$. Plots on row 1 show the bias of moderate calibration curve, when joint distribution of X and Z is $p = (0.3, 0.1, 0.4, 0.2)$. Plots on row 2 show the bias of C_b when $p = (0.3, 0.1, 0.4, 0.2)$ (left) and when $p = (0.300, 0.008, 0.194, 0.498)$ (right).

Main Take-home Message

- This study demonstrated how to evaluate pre-specified TBPs using observational data.
- The absence of full confounding control leads to bias in identifying C_b and $\mathrm{E}[B \mid H]$.
- The behaviour of biases in predictive performance metrics is more complex than when targeting commonly encountered estimands such as the average treatment effect.

References

[1] Sadatsafavi, M.; Mansournia, M. A.; Gustafson, P. *Statistics in Medicine* **2020**, *39*, 1362–1373.

[2] Van Calster, B.; Nieboer, D.; Vergouwe, Y.; De Cock, B.; Pencina, M. J.; Steyerberg, E. W. *Journal of Clinical Epidemiology* **2016**, *74*, 167–176.