A Bayesian Machine Learning Approach for Estimating Heterogeneous Survivor Causal Effects

Xinyuan Chen[1], **Michael O. Harhay**[2,3], **Guangyu Tong**[4], and **Fan Li**[4,5]

[1]Department of Mathematics and Statistics, Mississippi State University
[2]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania
[3]Clinical Trials Methods and Outcomes Lab, Palliative and Advanced Illness Research Center, Perelman School of Medicine, University of Pennsylvania
[4]Department of Biostatistics, Yale University School of Public Health
[5]Center for Methods in Implementation and Prevention Science, Yale University School of Public Health

Introduction

Motivated by the Acute Respiratory Distress Syndrome Network (ARDSNetwork) ARDS respiratory management (ARMA) trial [1], we developed a flexible machine learning approach using the Bayesian additive regression trees (BART) [2] under the principal stratification framework to estimate the average causal effect and heterogeneous causal effects among the always-survivors stratum when clinical outcomes are subject to truncation. In the analysis, we found that the low tidal volume treatment had an overall benefit for patients sustaining acute lung injuries on the outcome of time to returning home, but substantial heterogeneity in treatment effects among the always-survivors, driven most strongly by sex and the alveolar-arterial oxygen gradient at baseline ($AaDO_2$, a physiologic measure of lung function and pulmonary pathology).

Causal Assumptions and Setup

Consider the ARMA trial (two-arm randomized) with $N = 861$ participants. T_i: binary treatment for participant i (1: treatment, mechanical ventilation treatment with a volume of 6 mL/kg of predicted body weight; 0: control, mechanical ventilation treatment with a volume of 12 mL/kg of predicted body weight). $Y_i(\boldsymbol{t})$: non-mortality outcome of days to returning home (DTRH) that would be observed under treatment assignment $\boldsymbol{t}$. $D_i(\boldsymbol{t})$: survival status of participant i at the time that the measurement of the non-mortality outcome (1: alive; 0: death) was taken. $\boldsymbol{X}_i$: baseline characteristics of participant i, including (i) demographic information, (ii) respiratory measures, and (iii) physiological measures. We make the following assumptions:

Assumption 1. *(Stable Unit Treatment Value Assumption). Let $\boldsymbol{t}$ and $\boldsymbol{t}'$ be any two possible treatment assignments. If $t_i = t_i'$, then $D_i(\boldsymbol{t}) = D_i(\boldsymbol{t}')$ and $Y_i(\boldsymbol{t}) = Y_i(\boldsymbol{t}')$.*

Assumption 2. *(Randomization). The assignment variable T_i is independent of all potential outcomes $\{D_i(1), D_i(0), Y_i(1), Y_i(0)\}$, given baseline characteristics $\boldsymbol{X}_i$.*

Assumption 3. *(Monotonicity). $\mathbb{P}(D_i(1) \geq D_i(0)|\boldsymbol{X}_i = \boldsymbol{x}) = 1, \forall\, \boldsymbol{x} \in \mathcal{X}$, where $\mathcal{X}$ is the support of $\boldsymbol{X}$.*

Using the principal stratification framework [3], each participant can be classified into distinct principal strata (S_i), always-survivors ($S_i = 11$): participants who would survive under either treatment status; protected ($S_i = 10$): participants who would survive under treatment but would die under control; harmed ($S_i = 01$): participants who would die under treatment but would survive under control; never-survivors ($S_i = 00$): participants who would die under either treatment status. For always-survivors, a causal estimand, the Survivor Average Causal Effect (SACE), is defined as

$$\Delta_{SACE} = \mathbb{E}[Y_i(1) - Y_i(0)|S_i = 11].$$

Further, the individualized treatment effect for individual i, the Conditional Survivor Average Causal Effect (CSACE), is defined as

$$\Delta_{CSACE}(\boldsymbol{x}) = \mathbb{E}[Y_i(1) - Y_i(0)|\boldsymbol{X}_i = \boldsymbol{x}, S_i = 11].$$

Variations in $\Delta_{CSACE}(\boldsymbol{x})$ measure the degree of treatment effect heterogeneity among the always-survivors. and may provide useful evidence for tailoring treatment rules for future participants.

Methodology

We consider the Bayesian principal stratification framework [4], where two sets of models are to be specified: the distribution of $Y(0)$ and $Y(1)$ given S and $\boldsymbol{X}$ (the Y-model), and the distribution of S given $\boldsymbol{X}$ (the S-model). According to values of T_i and D_i, each participant can be reclassified into: $\mathcal{O}(1,1)$, assigned to the treatment and survived; $\mathcal{O}(1,0)$, assigned to the treatment and died; $\mathcal{O}(0,1)$, assigned to the control and survived; $\mathcal{O}(0,0)$, assigned to the control and died. For participants in $\mathcal{O}(1,0)$ and $\mathcal{O}(0,1)$, S_i are fully inferred Assumption 3. For participants in $\mathcal{O}(1,1)$ and $\mathcal{O}(0,0)$, S_i require imputation. Denote $\pi_{i,s} = \mathbb{P}(S_i = s|\boldsymbol{X}_i, \boldsymbol{\theta})$ and $f_{i,st} = \mathbb{P}(Y_i(t)|S_i = s, \boldsymbol{X}_i, \boldsymbol{\theta})$, for $s = 00, 10, 11$ and $t = 0, 1$. The posterior distribution of model parameters, $\boldsymbol{\theta}$, can be generically written as

$$\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{S}^{obs}, \boldsymbol{T}, \boldsymbol{X}) \propto \mathbb{P}(\boldsymbol{\theta}) \times \prod_{i\in\mathcal{O}(1,1)} \left(\pi_{i,11}f_{i,111} + \pi_{i,10}f_{i,101}\right) \times \prod_{i\in\mathcal{O}(1,0)} \pi_{i,00} \times \prod_{i\in\mathcal{O}(0,1)} \pi_{i,11}f_{i,110} \times \prod_{i\in\mathcal{O}(0,0)} \left(\pi_{i,10} + \pi_{i,00}\right),$$

where $\mathbb{P}(\boldsymbol{\theta})$ is the prior distribution. Posterior inference on $\boldsymbol{\theta}$ involving imputing missing S_i is performed via a nested Probit modeling approach. Two additional latent variables Z and W to be augmented for each participant, where

$$\{Z_i|m_Z(\bullet), \boldsymbol{X}_i\} \sim \mathcal{N}\left(m_Z(\boldsymbol{X}_i), 1\right), \text{ and } \begin{cases} S_i = 00, & \text{if } Z_i > 0 \\ S_i = 10 \text{ or } 11, & \text{if } Z_i \leq 0 \end{cases}$$

$$\{W_i|m_W(\bullet), \boldsymbol{X}_i\} \sim \mathcal{N}\left(m_W(\boldsymbol{X}_i), 1\right), \text{ and } \begin{cases} S_i = 10, & \text{if } W_i > 0 \\ S_i = 11, & \text{if } W_i \leq 0 \end{cases}$$

Here, $m_Z(\bullet)$ and $m_W(\bullet)$ are conditional mean functions to be estimated for the S-model. Further, $\mathbb{P}(S_i = 00|\bullet) = \mathbb{P}(Z_i > 0|m_Z(\boldsymbol{X}_i))$, $\mathbb{P}(S_i = 10|\bullet) = \mathbb{P}(Z_i \leq 0|m_Z(\boldsymbol{X}_i))\mathbb{P}(W_i > 0|m_W(\boldsymbol{X}_i))$, and $\mathbb{P}(S_i = 11|\bullet) = \mathbb{P}(Z_i \leq 0|m_Z(\boldsymbol{X}_i))\mathbb{P}(W_i \leq 0|m_W(\boldsymbol{X}_i))$ where $\Phi(\bullet)$ is the standard normal cumulative distribution function. For the Y-models,

$$\{Y_i(t)|S_i = s, m_{st}(\bullet), \boldsymbol{X}_i\} \sim \mathcal{N}\left(m_{st}(\boldsymbol{X}_i), \sigma_{st}^2\right),$$

where $t = 0, 1$ for $s = 11$, and $t = 1$ for $s = 10$; $m_{st}(\bullet)$ are conditional mean functions to be estimated for $Y_i(t)$, and σ_{st}^2 is the variance parameter. A BART estimator is proposed to estimate the mean functions nonparametrically to reduce bias:

$$m(\boldsymbol{X}) = \sum_{j=1}^{J} h(\boldsymbol{X}; \mathcal{T}_j, \mathcal{M}_j),$$

where $\mathcal{T}_j$ is a binary tree and $\mathcal{M}_j$ are terminal node parameters of $\mathcal{T}_j$. A Gibbs sampling procedure (Metropolis-within-Gibbs) was proposed for posterior inference. Posteriors of the SACE and CSACE are also obtained.

Results

In the ARMA trial, 30.3% of participants can be directly identified as always-survivors, and another 30.6% can only be identified by the model (total likely always-survivors: 60.9%). The posterior mean of SACE is -23.87 days and confirms that the low tidal volume treatment leads to on average 24 days in reductions on DTRH among the always-survivors, indicating an overall benefit.

Results (cont'd)

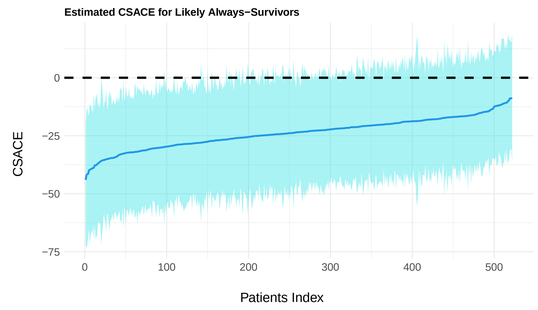

Fig. 1: Posterior means of $\Delta_{CSACE}(\boldsymbol{X})$ (darker blue) with corresponding 95% credible intervals (lighter blue) likely always-survivors.

Figure 1 shows the posterior mean and 95% credible intervals of $\Delta_{CSACE}(\boldsymbol{X})$ likely always-survivors. The CSACEs range from -46.94 to -8.27 days, suggesting heterogeneity in response to the low tidal volume treatment. We further explore the potential relationship between CSACEs and covariates using a classification and regression tree (CART) model with CSACEs as the response variable. Results showed that $AaDO_2$ and sex have the largest effects on CSACE.

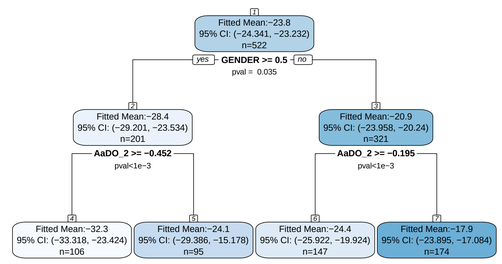

Fig. 2: Final CART model fit to the posterior mean DTRHs (in days) between the low tidal volume treatment and the traditional tidal volume treatment. Values in each node correspond to the posterior mean and 95% credible intervals for the average CSACE for the subgroup of individuals represented in that node.

Figure 2 illustrates the results of the final tree estimates based on the top 2 covariates that are the main drivers of the heterogeneity in CSACE, where the final R^2 between the tree fit, and the posterior mean CSACE of low tidal volume treatment versus traditional tidal volume treatment was 78.9%. Male participants or participants with higher $AaDO_2$ tend to experience greater treatment benefit from the low tidal volume treatment.

Future Directions

Acknowledgements

References

[1] Roy G Brower et al. "Acute Respiratory Distress Syndrome Network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome". In: *New England Journal of Medicine* 342.18 (2000), pp. 1301–1308.

[2] Hugh A. Chipman, Edward I. George, and Robert E. McCulloch. "BART: Bayesian additive regression trees". In: *The Annals of Applied Statistics* 4.1 (2010), pp. 266–298.

[3] Constantine E. Frangakis and Donald B. Rubin. "Principal stratification in causal inference". In: *Biometrics* 58.1 (2002), pp. 21–29.

[4] Constantine E. Frangakis, Donald B. Rubin, and Xiao-Hua Zhou. "Clustered encouragement designs with individual noncompliance: Bayesian inference with randomization, and application to advance directive forms". In: *Biostatistics* 3.2 (June 2002), pp. 147–164. ISSN: 1465-4644.