Treatment Decision in Ischemic Cardiomyopathy: Causal Inference Using Random Survival Forests

Min Lu

Division of Biostatistics, University of Miami

Joint work with Eugene H. Blackstone and Hemant Ishwaran
Research Questions

We base our analysis on data from 1468 patients who were treated for ischemic cardiomyopathy at Cleveland Clinic from 1997 to 2007. Treatments include:

- Coronary artery bypass grafting alone (CABG)
- CABG plus mitral valve anuloplasty (MVA)
- CABG plus surgical ventricular reconstruction (SVR)
- Listing for cardiac transplantation (LCTx)

- What is the average treatment effect (ATE)?
- What is the individual treatment effect (ITE)?
- Have patients received optimal treatments?
Let \( \{(X_1, Z_1, T_1, \delta_1), \ldots, (X_n, Z_n, T_n, \delta_n)\} \) denote the data. The observed survival time \( T_i = \min(T_i^o, C_i^o) \), where \( T_i^o \) is the true event time and \( C_i^o \) is the true censoring time. We assume \( C_i^o: T_i^o \perp C_i^o | (X_i, Z_i) \). Let \( T^o(j) \) denote the potential outcome (event time) under treatment \( Z = j \).
The individual treatment effect (ITE) $\tau_{j,k}(t, x)$

The individual treatment effect (ITE) at time $t$ for covariate $x$ for treatment $j$ over treatment $k$ is defined as $\tau_{j,k}(t, x) = S_j(t|x) - S_k(t|x)$, where $S_j(t|x) = \mathbb{P}\{T^o(j) > t | X = x\}$ is the survival function. Under weak unconfoundedness,

$$\tau_{j,k}(t, x) = \mathbb{P}\{T^o(j) > t | X = x\} - \mathbb{P}\{T^o(k) > t | X = x\}$$

$$= \mathbb{P}\{T^o > t | X = x, Z = j\} - \mathbb{P}\{T^o > t | X = x, Z = k\}$$

$$= S(t|x, Z = j) - S(t|x, Z = k)$$

Weak Unconfoundedness Assumption

We say that weak unconfoundedness holds, if for all $j \in \{1, \ldots, M\}$,

$$\mathbf{1}_{\{Z=j\}} \perp T^o(j) | X$$
The average treatment effect (ATE) $\tau_{j,k}(t)$

Integrating over $t \in [0, t_0]$, we define the ITE before time $t_0$ as

$$\tau_{j,k}([0, t_0], x) = \int_0^{t_0} \tau_{j,k}(t, x) \, dt$$

which can be interpreted as the difference in the number of years alive before time $t_0$ for treatment $j$ over $k$. Typically, $t_0$ is chosen to equal the maximum observed follow-up time.

**Definition**

Define the average treatment effect (ATE) at time $t$ for treatment $j$ over treatment $k$, as

$$\tau_{j,k}(t) = \mathbb{E}_X \left[ \tau_{j,k}(t, X) \mathbb{P}\{Z = j|X\} > 0, \mathbb{P}\{Z = k|X\} > 0 \right].$$

We define the ATE before time $t_0$ as $\tau_{j,k}([0, t_0]) = \int_0^{t_0} \tau_{j,k}(t) \, dt$. 

---

**Random Survival Forest Causal Inference**

Min Lu

1. Introduction
2. Definitions & notations
3. Treatment eligibility
4. Treatment effect estimation
5. Results
6. Discussion
7. Reference
A unique feature of our study was the availability of expert knowledge for defining treatment eligibility.

### Table: Expert knowledge used for determining treatment eligibility

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Expert Knowledge Eligibility Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>CABG</td>
<td>(a) Ischemic symptoms (angina); viable myocardium with diseased but by-passable coronary arteries. If (a) was not available, eligibility was determined using: (b) ACC/AHA guidelines for CABG based on angina and coronary artery disease</td>
</tr>
<tr>
<td>SVR*</td>
<td>Anterior wall akinesia/dyskinesia; left ventricular end-diastolic diameter &gt; 6 cm</td>
</tr>
<tr>
<td>MVA</td>
<td>3+/4+ mitral regurgitation (MR) present</td>
</tr>
<tr>
<td>LCTx*</td>
<td>Age &lt; 70 years; NYHA functional class III/IV; creatinine level &lt; 1.7 mg·dL⁻¹</td>
</tr>
</tbody>
</table>

* Treatments where expert knowledge is considered less accurate for determining eligibility

### Let \( E_{n \times M} = \{E_{ij}\} \) denote the expert eligibility data from our \( n = 1468 \) patients for the \( M = 4 \) treatments, where \( E_{ij} \in \{0, 1\} \)

### Typically overlap is determined in practice by using a cutoff value \( 0 < C < 1 \). Patients are excluded from ITE and ATE calculations if \( \hat{P}\{Z = j|X_i\} \leq C \) or \( \hat{P}\{Z = k|X_i\} \leq C \)
Estimating treatment eligibility $\mathbb{P}\{Z = j|x\}$ using random forest

- **Approach I: Random forest classification (RF-C) approach.** Our first approach uses the treatment received $Z_i$ as the outcome and $X_i$ as features and fits a random forest classification (RF-C) model to estimate $\mathbb{P}\{Z = j|x\}$

- **Approach II: Random forest Distance (RF-D) approach.** The general idea is to assign patient $i$’s eligibility for treatment $j$ by using the “random forest distance” of $i$ to treatment $j$ patients

- **Approach III: Multivariate random forest (MRF).** We directly model expert knowledge by using the expert data $\{E_{i,j}\}$ as multivariate outcomes in a $M$-multivariate classification analysis
Approach II: Random forest distance approach

Let $d^A_i$ be the count of the edges from $i$ to the closest common ancestor of $i$ and $i'$. Similarly, let $d^A_{i'}$ count the edges from $i'$ to the closest ($i, i'$) common ancestor. Define $D^A_{i,i'} = d^A_i + d^A_{i'}$. Let $d^R_i$ and $d^R_{i'}$ be the count of the edges from $i$ and $i'$ to the root node and define $D^R_{i,i'} = d^R_i + d^R_{i'}$. The distance is defined as

$$d_{i,i'} = \frac{D^A_{i,i'}}{D^R_{i,i'}}.$$ 

The forest distance is defined as the forest averaged distance, which we denote by $\bar{d}_{i,i'}$. We define the probability of assigning $i$ to treatment $j$ by the closeness of $i$ to treatment $j$ patients,

$$\hat{P}\{Z = j | X_i\} = \frac{\sum_{i': Z_{i'} = j} (1 - \bar{d}_{i,i'})}{\sum_{i'} (1 - \bar{d}_{i,i'})}.$$ 

The distance between $X_i$ and $X_{i'}$ is the ratio of the number of edges connecting the red nodes to the ancestor, $N_A$, to the number of edges connecting the red nodes to the root node, $N_R$. Thus

$$d_{i,i'} = \frac{(2 + 1)}{(4 + 3)} = 3/7.$$
Cutoff criteria and validation

- We use a constant $0 < C < 1$ and say that patient $X_i$ is eligible for treatment $j$ if $\hat{P}\{Z = j|X_i\} > C$

- Let $M' = \{j_1, j_2\}$ denote the subset of treatment groups corresponding to CABG and MVA. We define the CABG and MVA cutoff as follows:

$$C^* = \arg\min_{0 < c < 1} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \sum_{j' \in M'} I(E_{ij'} \neq 1 \{\hat{P}\{Z=j'|X_i\} > c\}) \right\}$$

### Table: Cutoff values for estimating treatment eligibility

<table>
<thead>
<tr>
<th>Method</th>
<th>Cutoff Value</th>
<th>Misclassification Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>CABG</td>
<td>0.08</td>
<td>0.26</td>
</tr>
<tr>
<td>MVA</td>
<td>0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>All four treatments</td>
<td>0.61</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Fig: Misclassification error as a function of the cutoff value $c$
Counterfactual survival analysis using random survival forests

We estimate the survival function $S(t|x, Z)$ using virtual twin random survival forest interactions, denoted as RSF-VT-I where we add all possible interactions between the treatment variable $Z$ and covariates $X$ to the design matrix to grow random survival forest. The counterfactual ITE estimate is defined as

$$\hat{\tau}_{j,k}(t, X_i) = \hat{S}(t|X_i, Z_i = j) - \hat{S}(t|X_i, Z_i = k)$$
Average treatment effect on the treated (ATT)

**Definition**

Define the average treatment effect on the treated (ATT) at time $t$ for the treated $j$, for treatment $j$ over treatment $k$, as

$$
\tau_{j,k}(t) = \mathbb{E}_X[\tau_{j,k}(t,X)|Z = j, \mathbb{P}\{Z = j|x\} > 0, \mathbb{P}\{Z = k|x\} > 0]
$$

Likewise, the ATT for the treated $k$, for treatment $j$ over $k$, is

$$
\tau_{j,k}(t) = \mathbb{E}_X[\tau_{j,k}(t,X)|Z = k, \mathbb{P}\{Z = j|x\} > 0, \mathbb{P}\{Z = k|x\} > 0]
$$

---

**SVR vs LCTx**

- Eligible for SVR and LCTx
- Received SVR and eligible for LCTx
- Received LCTx and eligible for SVR
Results

(a) CABG vs. SVR
(b) CABG vs. MVA
(c) CABG vs. LCTx
(d) SVR vs. MVA
(e) SVR vs. LCTx
(f) MVA vs. LCTx

- **Multivariate Random Forest (MRF)**
- **Random Forest Classification (RF–C)**
- **Random Forest Distance (RF–D)**
The areas under the black, blue, and red lines of previous figure equal the ATE and ATT before \( t_0 \) (the maximum observed follow-up time), and thus represent the difference in number of years alive before \( t_0 \)

\[
\begin{align*}
\text{ATE}^{o}_{jk} &= \tau_{j,k}([0, t_0]) \quad \text{ATE before } t_0 \text{ (black line)} \\
\text{ATT}^{o}_{jk} &= \tau_{k}([0, t_0]) \quad \text{ATT before } t_0 \text{ where } j \text{ is the treated (blue line)} \\
\text{ATT}^{o}_{kj} &= \tau_{j}([0, t_0]) \quad \text{ATT before } t_0 \text{ where } k \text{ is the treated (red line)}
\end{align*}
\]

Table: Difference in number of months alive before maximum follow-up time, \( t_0 = 9.36 \) years

<table>
<thead>
<tr>
<th>Treatment j vs. k</th>
<th>( \text{ATE}^{o}_{jk} )</th>
<th>( \text{ATT}^{o}_{jk} )</th>
<th>( \text{ATT}^{o}_{kj} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MRF</td>
<td>RF-C</td>
<td>RF-D</td>
</tr>
<tr>
<td>(a) CABG vs. SVR</td>
<td>0.31</td>
<td>0.29</td>
<td>0.60</td>
</tr>
<tr>
<td>(b) CABG vs. MVA</td>
<td>4.88</td>
<td>5.06</td>
<td>5.21</td>
</tr>
<tr>
<td>(c) CABG vs. LCTx</td>
<td>0.85</td>
<td>3.67</td>
<td>3.50</td>
</tr>
<tr>
<td>(d) SVR vs. MVA</td>
<td>5.95</td>
<td>5.49</td>
<td>5.47</td>
</tr>
<tr>
<td>(e) SVR vs. LCTx</td>
<td>-1.40</td>
<td>-0.55</td>
<td>-1.08</td>
</tr>
<tr>
<td>(f) MVA vs. LCTx</td>
<td>-11.80</td>
<td>-6.08</td>
<td>-6.81</td>
</tr>
</tbody>
</table>
Confidence intervals for individual treatment effects $\hat{\tau}_{j,k}(t, x)$ at $t = 5$ years. Each subfigure indicates a pairwise comparison for treatment $j$ versus $k$. Red and blue indicate patients with significant treatment effect (p-value $< .05$), where blue are from treatment $j$ group and red are from treatment group $k$. Thus, blue and red boxes correspond to some of the patients from blue and red lines in previous figure.
Fig. Identifying patients who received optimal treatment and those who did not. Optimal therapy is defined as eligible treatment maximizing restricted mean survival time (RMST). Pie charts display gain in months for alternative optimized therapies and their respective sample sizes. If optimized treatment is the assigned treatment, gain is defined as zero.
Treatment effect heterogeneity test

(a) CABG vs. SVR
(b) CABG vs. MVA
(c) CABG vs. LCTx
(d) SVR vs. MVA
(e) SVR vs. LCTx
(f) MVA vs. LCTx

- CABG: Coronary Artery Bypass Graft
- SVR: Saphenous Vein Bypass
- MVA: Mitral Valve Annuloplasty
- LCTx: Left Coronary Artery Transplantation

Variables:
- Random Survival Forest
- Causal Inference

[Tables and graphs with various medical data points such as age, gender, blood pressure, and other cardiovascular metrics are shown, indicating differences and trends between the treatment groups.]
We fit a bump hunting model (Friedman and Fisher, 1999; Duong, 2015) for subgroup analysis. To improve efficiency of the algorithm, we only used variables found important by using random forest variable selection. The estimated ITE was used for the outcome and all pre-treatment covariates as independent variables.

Table: Subgroup detection using bump hunting after variable selection. $\text{CATE}_{jk}^o$ equals the conditional ATE before $t_0$, conditioned on subgroup criteria

<table>
<thead>
<tr>
<th>Treatment $j$ vs. $k$</th>
<th>Subgroup</th>
<th>$\text{CATE}<em>{jk}^o/\text{ATE}</em>{jk}^o$</th>
<th>Size/Total</th>
<th>% in $j$</th>
<th>% in $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CABG vs. SVR</td>
<td>BSA $&gt; 2.23$</td>
<td>-4.08/0.31</td>
<td>44/246</td>
<td>28.57</td>
<td>16.51</td>
</tr>
<tr>
<td></td>
<td>Regurgitation Grade $&gt; 0$</td>
<td>-7.26/0.31</td>
<td>31/246</td>
<td>10.71</td>
<td>12.84</td>
</tr>
<tr>
<td></td>
<td>Blood Urea Nitrogen $&lt; 30$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Creatinine $&lt; 1.8$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BMI $&gt; 27.04$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GFR $&gt; 44.75$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CABG vs. LCTx</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BSA $&gt; 1.83$</td>
<td>5.31/0.85</td>
<td>125/406</td>
<td>59.18</td>
<td>21.75</td>
</tr>
<tr>
<td></td>
<td>BMI $&gt; 27.77$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blood Urea Nitrogen $&lt; 25$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LDL $&lt; 133.31$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVR vs. LCTx</td>
<td>BSA $&gt; 1.83$</td>
<td>7.66/-1.40</td>
<td>60/292</td>
<td>30.37</td>
<td>12.10</td>
</tr>
<tr>
<td></td>
<td>BMI $&gt; 27.77$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>55.29 $&lt; \text{GFR} &lt; 120.80$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BSA=body surface area (m$^2$); BMI=body mass index; GFR=glomerular filtration rate; LDL=low-density lipoprotein cholesterol
Treatment decisions

Fig: Paradigm for Individual Causal Inference and Treatment Decision Making for Ischemic Cardiomyopathy.
Concluding remarks

- One contribution of this paper is to offer estimation methods for eligibility to treatments under the scenario that some treatments may have either gold standard expert knowledge, or controversial knowledge for judging eligibility.

- For personalized treatment decision and dynamic causal procedure of treatment effect, we develop a virtual twin random survival forest, extended to include interactions between treatment variables and all pre-treatment covariates.

- A key insight of this paper is to judge current treatment decisions using pairwise ATT comparisons.


Thank you all very much!