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

Min Lu

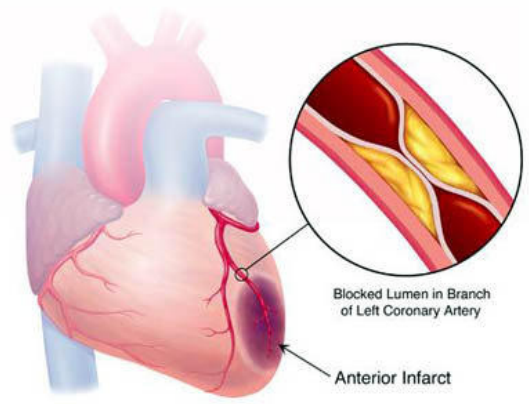
Division of Biostatistics, University of Miami

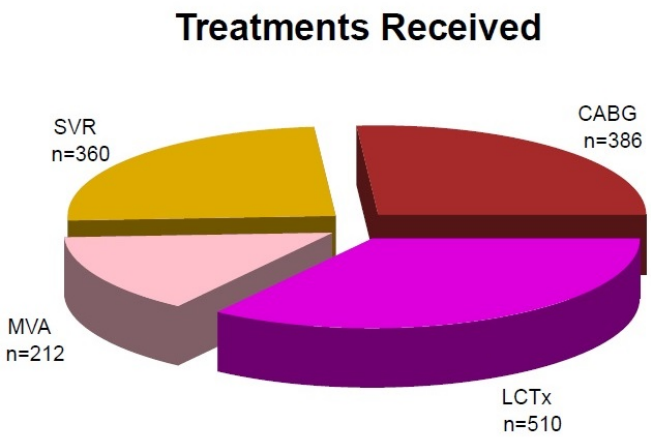
Joint work with Eugene H. Blackstone and Hemant Ishwaran

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)**





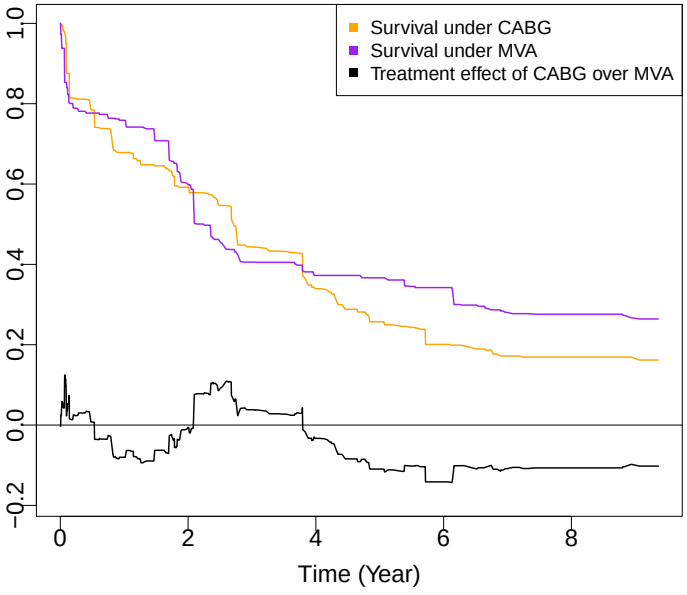
- What is the average treatment efffect (ATE)?
- What is the individual treatment efffect (ITE)?
- Have patients received optimal treatments?

Treatment effect on survival outcome

Let $\{(\mathbf{X}_1, Z_1, T_1, \delta_1), \ldots, (\mathbf{X}_n, Z_n, T_n, \delta_n)\}$ 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: $T_i^o \perp C_i^o | (\mathbf{X}_i, Z_i)$. Let $T^o(j)$ denote the potential outcome (event time) under treatment $Z = j$

Units	Observed			Potential Outcomes		Treatment Effects
	Outcome		Treatment	Therapy 1 ...	Therapy M	for j over k
1	T_1	δ_1	Z_1	$T_1^o(1)$	$T_1^o(M)$	$S_j(t\|\mathbf{x}_1) - S_k(t\|\mathbf{x}_1)$
⋮	⋮			⋮	⋮	⋮
i	T_i	δ_i	Z_i	$T_i^o(1)$	$T_i^o(M)$	$S_j(t\|\mathbf{x}_i) - S_k(t\|\mathbf{x}_i)$

The individual treatment effect (ITE) $\tau_{j,k}(t, \mathbf{x})$



Weak Unconfoundedness Assumption

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

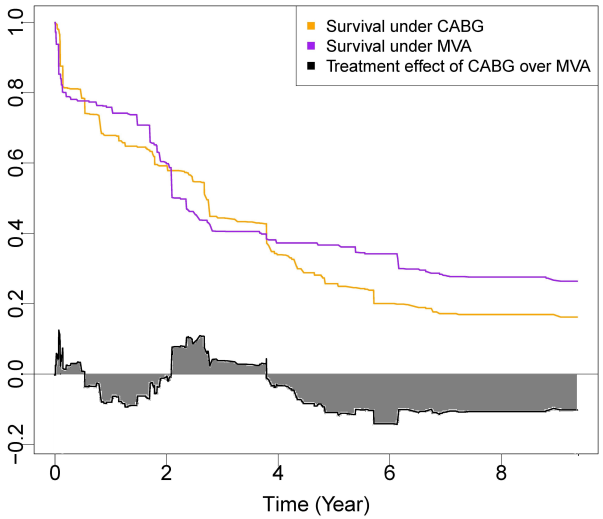
$$\mathbf{1}_{\{Z=j\}} \perp T^o(j) \mid \mathbf{X}$$

Definition

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

$$\begin{aligned}
\tau_{j,k}(t, \mathbf{x}) &= \mathbb{P}\{T^o(j) > t|\mathbf{X} = \mathbf{x}\} - \mathbb{P}\{T^o(k) > t|\mathbf{X} = \mathbf{x}\} \\
&= \mathbb{P}\{T^o > t|\mathbf{X} = \mathbf{x}, Z = j\} - \mathbb{P}\{T^o > t|\mathbf{X} = \mathbf{x}, Z = k\} \\
&= S(t|\mathbf{x}, Z = j) - S(t|\mathbf{x}, Z = k)
\end{aligned}$$

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

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

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

$$\tau_{j,k}(t) = E_{\mathbf{X}}\Big[\tau_{j,k}(t, \mathbf{X}) \Big| \mathbb{P}\{Z = j|\mathbf{x}\} > 0, \mathbb{P}\{Z = k|\mathbf{x}\} > 0\Big].$$

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

Treatment eligibility

A unique feature of our study was the availability of expert knowledge for defining treatment eligibility

Table: Expert knowledge used for determining treatment eligibility

Treatment	Expert Knowledge Eligibility Criteria
CABG	(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
SVR*	Anterior wall akinesia/dyskinesia; left ventricular end-diastolic diameter$>$6 cm
MVA	3+/4+ mitral regurgitation (MR) present
LCTx*	Age$<$70 years; NYHA functional class III/IV; creatinine level$<$1.7 mg$\cdot dL^{-1}$

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

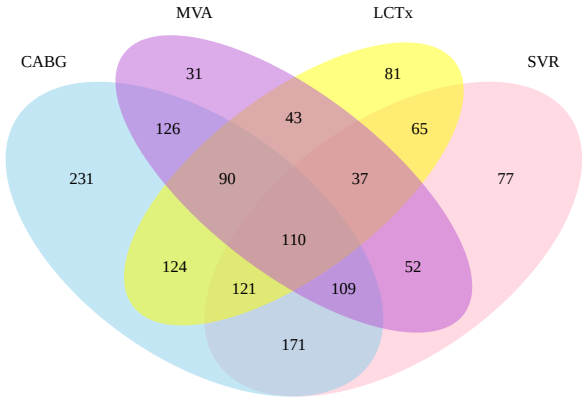


Fig: Venn diagram summarizing eligiblity status defined using expert knowledge

- Let $\mathbf{E}_{n\times M} = \{E_{ij}\}$ 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 $\hat{\mathbb{P}}\{Z = j|\mathbf{X}_i\} \leq C$ or $\hat{\mathbb{P}}\{Z = k|\mathbf{X}_i\} \leq C$

Estimating treatment eligibility $\mathbb{P}\{Z = j|\mathbf{x}\}$ using random forest

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

Approach II: Random forest distance approach

Let d_i^A be the count of the edges from i to the closest common ancestor of i and i'. Similarly, let $d_{i'}^A$ count the edges from i' to the closest (i, i') common ancestor. Define $D_{i,i'}^A = d_i^A + d_{i'}^A$. Let d_i^R and $d_{i'}^R$ be the count of the edges from i and i' to the root node and define $D_{i,i'}^R = d_i^R + d_{i'}^R$. The distance is defined as

$$d_{i,i'} = \frac{D_{i,i'}^A}{D_{i,i'}^R}.$$

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,

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

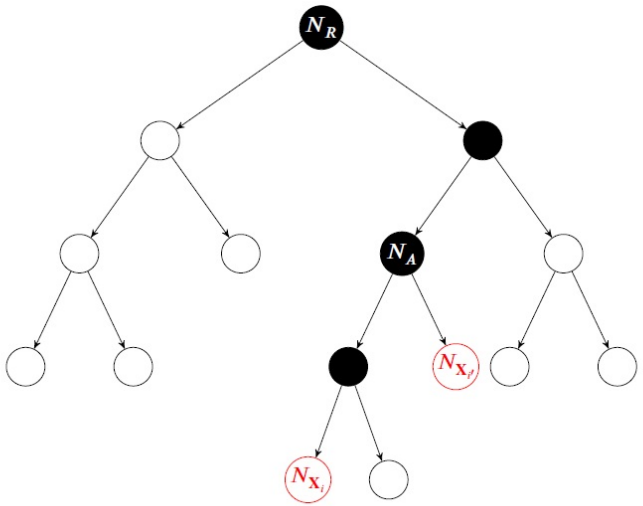


FIG : *Example illustrating random forest distance between* $\mathbf{X}_i$ *and* $\mathbf{X}_{i'}$.

The distance between $\mathbf{X}_i$ and $\mathbf{X}_{i'}$ 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 $d_{i,i'} = (2+1)/(4+3) = 3/7$

Cutoff criteria and validation

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

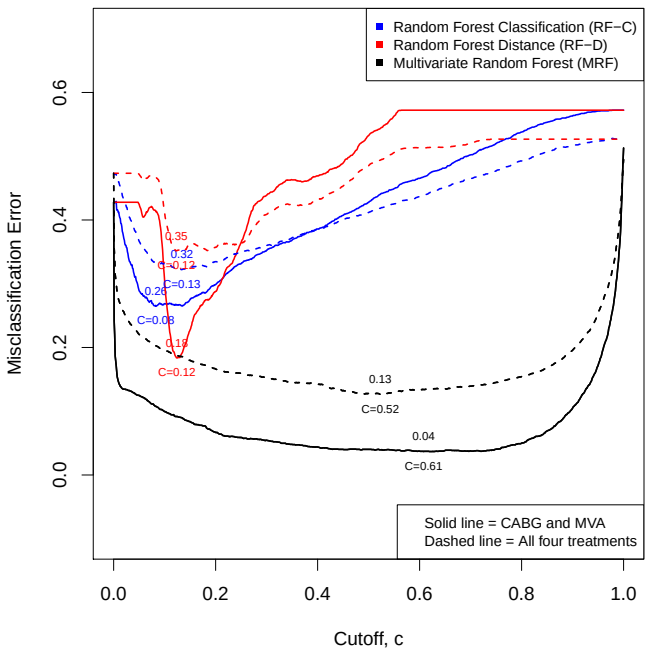


Fig: Misclassification error as a function of the cutoff value c

- 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:

$$C^* = \underset{0<c<1}{\arg\min} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \sum_{j' \in M'} \mathbf{I}\left(E_{ij'} \neq \mathbf{1}_{\{\hat{\mathbb{P}}\{Z=j'|\mathbf{X}_i\}>c\}}\right) \right\}$$

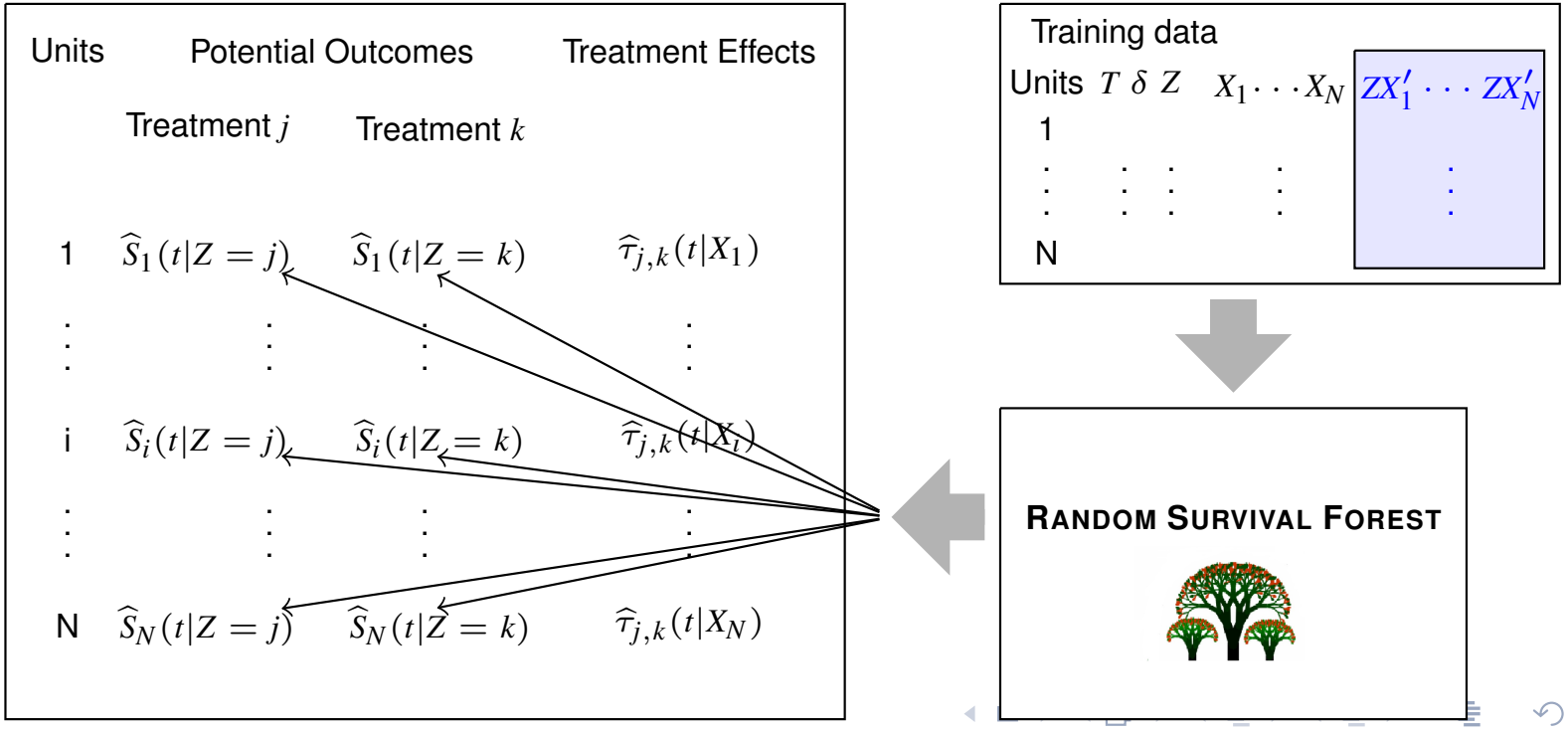
Table: Cutoff values for estimating treatment eligibility

Method	Cutoff Value	Misclassification Error	
		CABG MVA	All four treatments
RF-C	0.08	0.26	0.32
RF-D	0.12	0.18	0.35
MRF	0.61	0.04	0.13

Counterfactual survival analysis using random survival forests

We estimate the survival function $S(t|\mathbf{x}, Z)$ 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 $\mathbf{X}$ to the design matrix to grow random survival forest The counterfactual ITE estimate is defined as

$$\hat{\tau}_{j,k}(t, \mathbf{X}_i) = \hat{S}(t|\mathbf{X}_i, Z_i = j) - \hat{S}(t|\mathbf{X}_i, Z = k)$$



Average treatment effect on the treated (ATT)

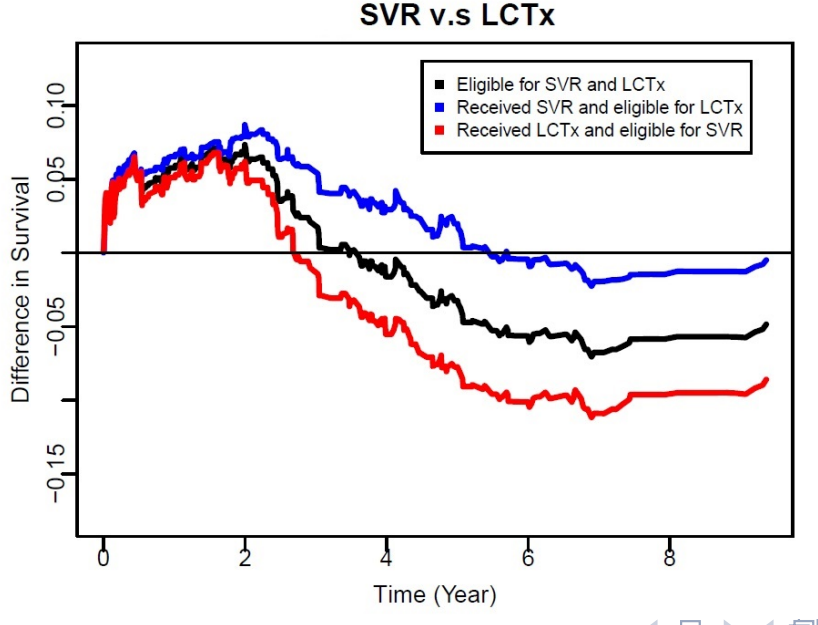
Definition

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

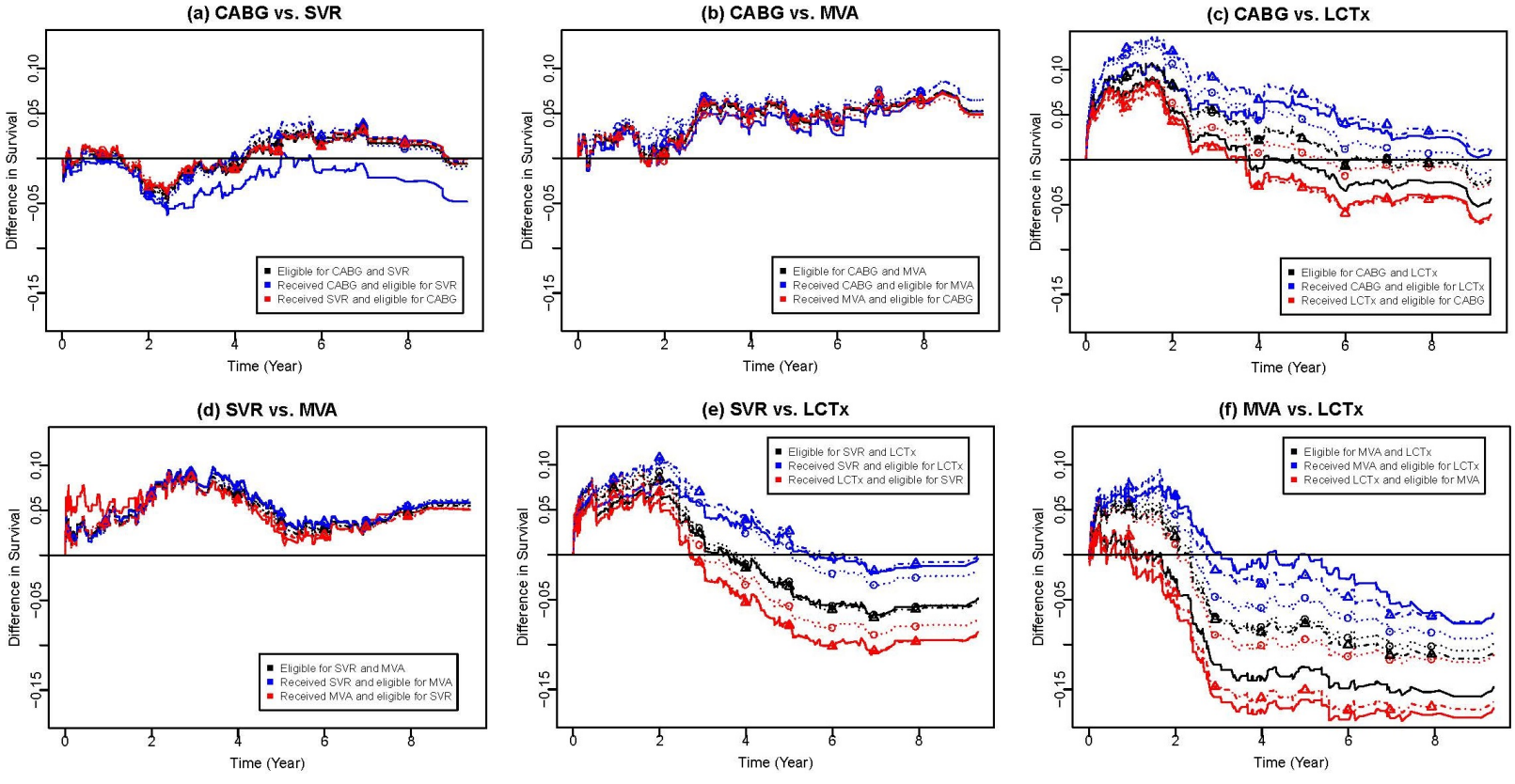
$$\tau_{j\textcircled{k}}(t) = E_{\mathbf{X}}\Big[\tau_{j,k}(t,\mathbf{X})\big|Z=j,\ \mathbb{P}\{Z=j|\mathbf{x}\}>0, \mathbb{P}\{Z=k|\mathbf{x}\}>0\Big]$$

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

$$\tau_{j\textcircled{k}}(t) = E_{\mathbf{X}}\Big[\tau_{j,k}(t,\mathbf{X})\big|Z=k,\ \mathbb{P}\{Z=j|\mathbf{x}\}>0, \mathbb{P}\{Z=k|\mathbf{x}\}>0\Big]$$



Results



Multivariate Random Forest (MRF)
Random Forest Classification (RF−C)
Random Forest Distance (RF−D)

Results

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

$$\text{ATE}^o_{jk} = \tau_{j,k}([0, t_0]) \quad \text{ATE before } t_0 \text{ (black line)}$$

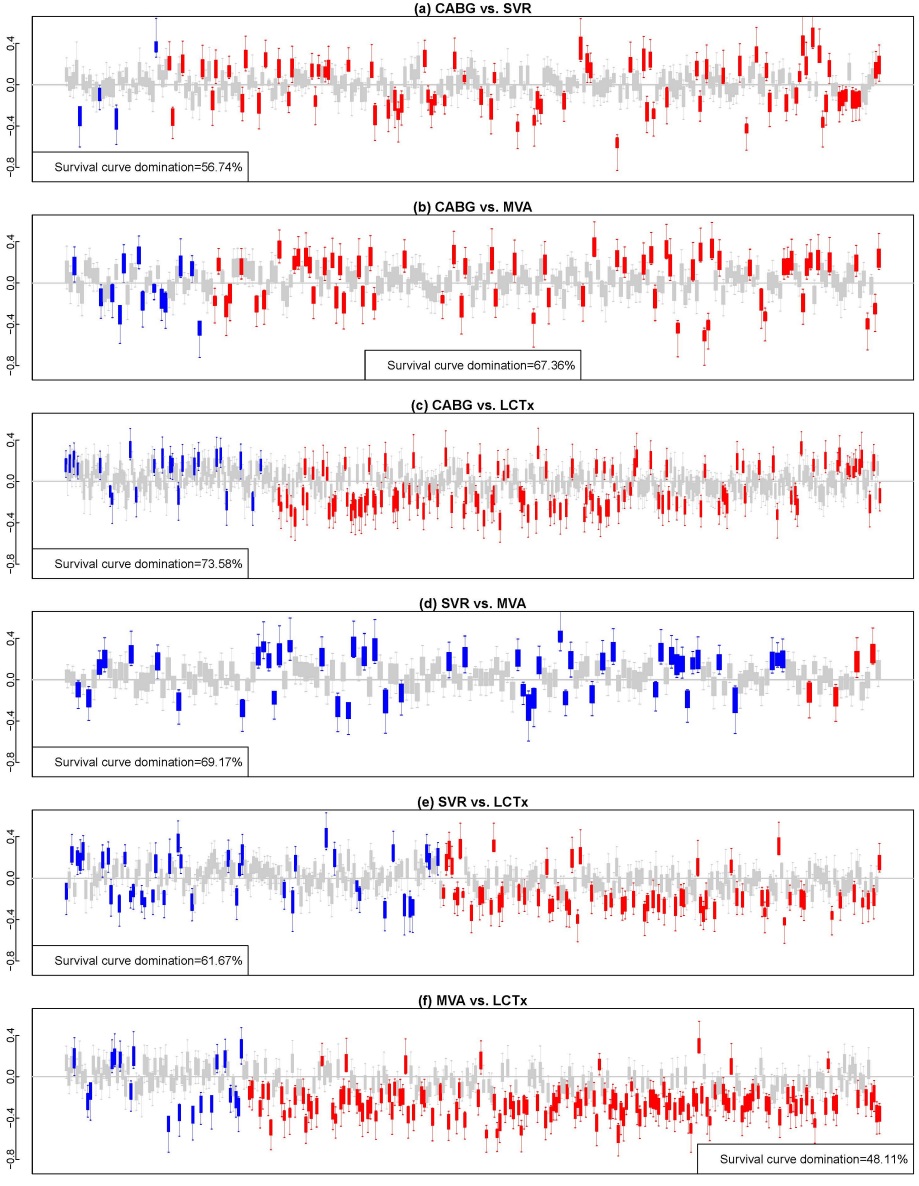
$$\text{ATT}^o_{jk} = \tau_{\textcircled{j}k}([0, t_0]) \quad \text{ATT before } t_0 \text{ where } j \text{ is the treated (blue line)}$$

$$\text{ATT}^o_{kj} = \tau_{j\textcircled{k}}([0, t_0]) \quad \text{ATT before } t_0 \text{ where } k \text{ is the treated (red line)}$$

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

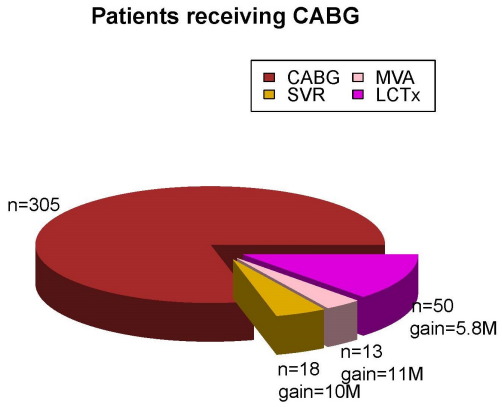
Treatment j vs. k	ATE^o_{jk}			ATT^o_{jk}		ATT^o_{kj}	
	MRF	RF-C	RF-D	Mean	SE	Mean	SE
(a) CABG vs. SVR	0.31	0.29	0.60	-2.67	3.74	0.70	0.93
(b) CABG vs. MVA	4.88	5.06	5.21	4.20	2.89	5.02	1.55
(c) CABG vs. LCTx	0.85	3.67	3.50	5.85	2.26	-0.74	1.11
(d) SVR vs. MVA	5.95	5.49	5.47	5.97	1.41	5.70	5.61
(e) SVR vs. LCTx	-1.40	-0.55	-1.08	2.57	1.52	-4.81	1.53
(f) MVA vs. LCTx	-11.80	-6.08	-6.81	-0.84	2.62	-14.97	1.36

Results

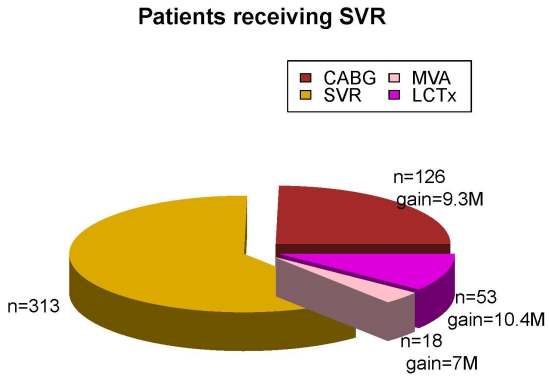


Confidence intervals for individual treatment effects $\hat{\tau}_{j,k}(t, \mathbf{x})$ at $t = 5$ years. Each subfigure indicates a pairwise comparison for treatment j versus k. Red and blue indicate patients with significant treatment effect (p-value $< .05$), where blue are from treatment j group 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

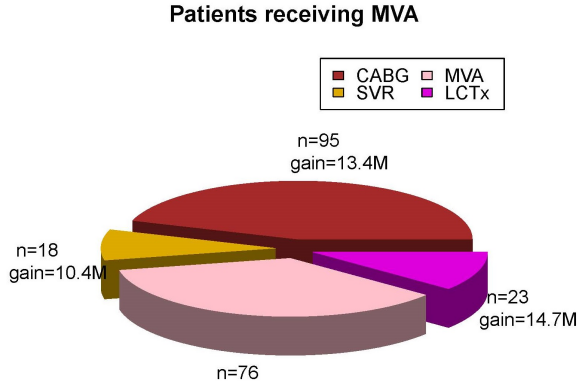
Results



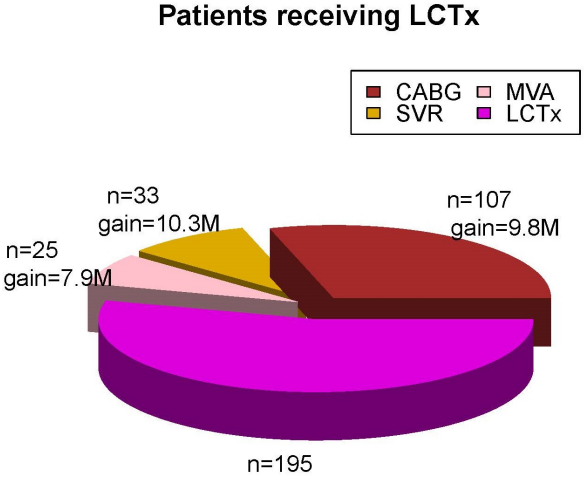
Overall, 21% potentially gain 9.2 months (M=months)



Overall, 39% potentially gain 9.2 months (M=months)



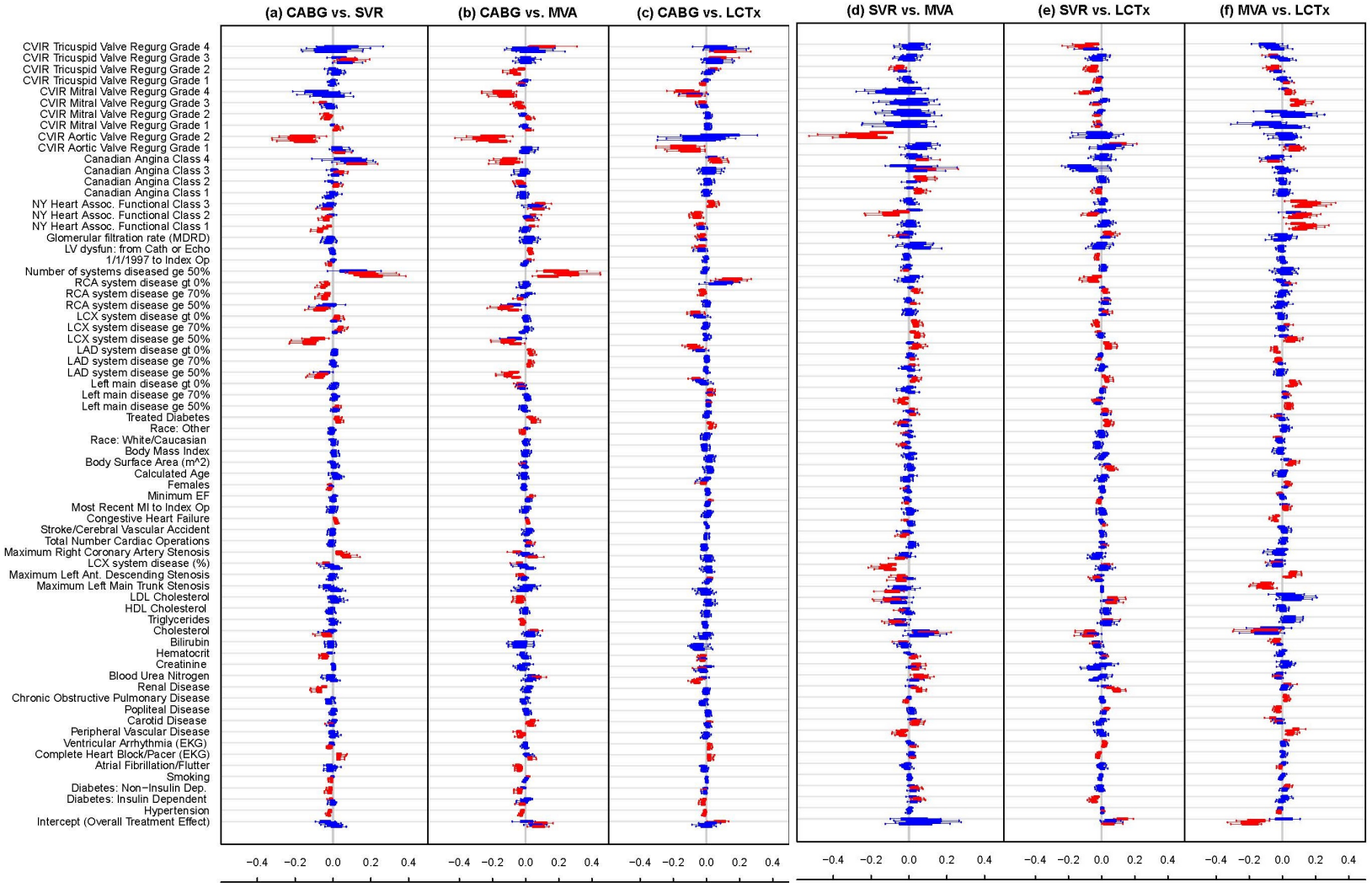
Overall, 64% potentially gain 13.1 months (M=months)



Overall, 46% potentially gain 9.8 months (M=months)

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



Subgroup analysis

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. CATE^{o}_{jk} equals the conditional ATE before t_0, conditioned on subgroup criteria

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

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

Treatment decisions

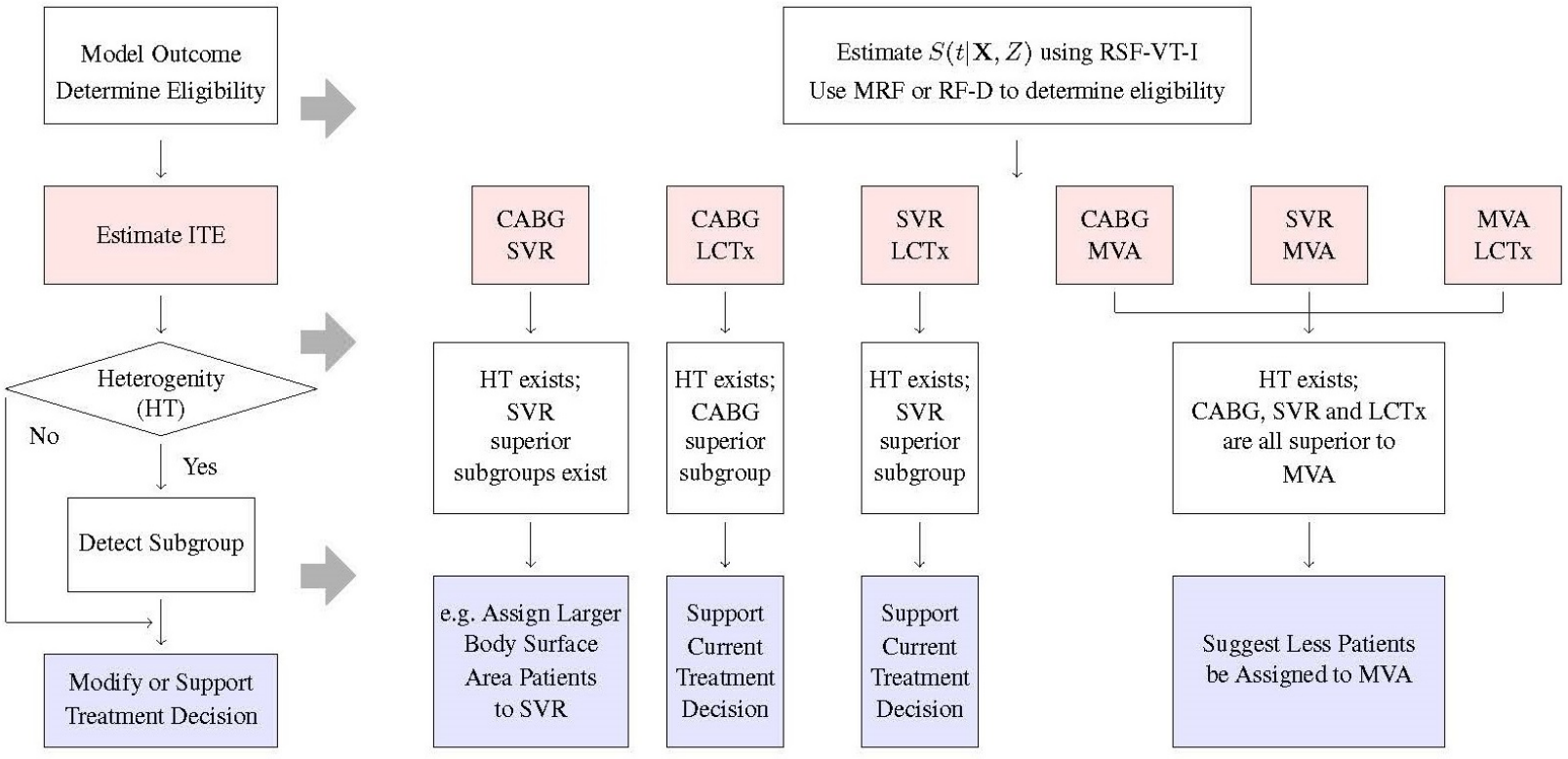


Fig: Paradigm for Individual Causal Inference and Treatment Decision Making for Ischemic Cardiomyopathy.

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

Reference

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Foster, J. C., Taylor, J. M., and Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in Medicine*, 30(24):2867–2880.

Friedman, J.H. and Fisher, N.I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing* **9** 123–143.

Hill, J. and Su, Y.S. (2013). Assessing lack of common support in causal inference using Bayesian nonparametrics: Implications for evaluating the effect of breastfeeding on children's cognitive outcomes. *Ann. Appl. Stat.* **7** 1386–1420.

Ishwaran H., Kogalur U.B., Blackstone E.H., and Lauer M.S. (2008). Random survival forests. *Ann. Appl. Stat.* **2** 841–860.

Lu M., Sadiq S., Feaster D.J. and Ishwaran H. (2017). Estimating individual treatment effect in observational data using random forest methods. *arXiv preprint arXiv:1701.05306*. (Accepted by Journal of Computational and Graphical Statistics)

Lu M., Blackstone E.H. and Ishwaran H. (2017). Treatment decision in ischemic cardiomyopathy: causal inference using random survival forests. *(Submitted in August 2017)*.

Parast, L., and Griffin, B. A. (2017), "Landmark estimation of survival and treatment effects in observational studies," Lifetime Data Analysis, 23(2), 161–182.

Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling: Springer Series in Statistics*. Berlin: Springer.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Yoon, D. Y., Smedira, N.G., Nowicki, E.R. et al. (2010). Decision support in surgical management of ischemic cardiomyopathy. *J. Thoracic and Cardiovascular Surgery* **139** 283—293.

Thank you all very much!

International Conference on Health Policy Statistics

January 10–12, 2018
Charleston, South Carolina