Precision of the Predicted Individual Treatment Effect

Gerd K. Rosenkranz

Center for Medical Statistics, Informatics and Intelligent Systems (CeMSIIS)
Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria
gerd.rosenkranz@meduniwien.ac.at
gk.rosenkranz@gmx.de

EFSPI Meeting on Recent Developments in Biomarkers and Subgroup Analyses in Drug Development
Gothenburg, Sweden, March 20, 2019

MEDIZINISCHE UNIVERSITÄT WIEN

Overview

The problem

"This leads directly to a related criticism of the present controlled trial—that it does not tell the doctor what he wants to know. It may be so constituted as to show without any doubt that ***treatment A is on the average better than treatment B****. On the other hand, that result does not answer the practicing doctor's question—****what is the most likely outcome when this drug is given to a particular patient?****"*

SIR AUSTIN BRADFORD HILL (1897-1991) [1]

Prediction

From each of $n = 1, \ldots, N$ subjects in the trial we observe a copy $(y_n, \mathbf{x}_n, t_n)$ of $(Y, \mathbf{X}, T)$ with distribution

$$P(Y, \mathbf{X}, T) = P(Y|\mathbf{X}, T)P(\mathbf{X})P(T)$$

i.e., $\mathbf{X}$ and T are assumed to be independent

A *predictor* $\eta(\mathbf{x}, t, \mathcal{L})$ derived from $\mathcal{L} = \{(y_n, \mathbf{x}_n, t_n), n = 1, \ldots, N\}$ assigns an outcome y to every input $(\mathbf{x}, t)$.

The *prediction error* of a future observation $(Y, \mathbf{X}, T) \sim P$ independent of $\mathcal{L}$ is given by

$$\text{PE}(\mathcal{L}) = E_P\left[\left(Y - \eta(\mathbf{X}, T, \mathcal{L})\right)^2\right]$$

Prediction error decomposition

For $Y = f(\mathbf{x}, t) + e$ with e independent of $(\mathbf{X}, T)$, $E[e] = 0$ and $E[e^2] = \sigma^2$, one gets

$$\begin{aligned} \mathrm{PE}(\mathcal{L}) &= E_P\left[\left(Y - \eta(\mathbf{X}, T, \mathcal{L})\right)^2\right] \\ &= \sigma^2 + E_P\left[\left(f(\mathbf{X}, T) - \eta(\mathbf{X}, T, \mathcal{L})\right)^2\right] \end{aligned}$$

or

”Prediction error = random error + model error”

Estimation of the prediction error

The naive estimator ("apparent error") of the prediction error

$$\widehat{\mathrm{PE}}(\mathcal{L}) = \frac{1}{N}\sum_{n=1}^{N}\left(y_n - \eta(\mathbf{x}_n, t_n, \mathcal{L})\right)^2$$

is *overoptimistic*, since the predictor η has been determined from the same data that are used to estimate the precision error, in particular if η was determined by minimizing the expression above.

A correction can be obtained from cross-validation or bootstrap [2]

Predicted individual treatment effect

Let $Y(t)$ be the *potential outcome* under treatment $T = t \in \{-1, 1\}$. The *individual treatment effect* is then given by $Y(1) - Y(-1)$.

The *predicted individual treatment effect* (PITE) in an individual with biomarkers $\mathbf{X} = \mathbf{x}$ is defined as

$$\Delta(\mathbf{x}) = E[Y(1) - Y(-1)|\mathbf{X} = \mathbf{x}]$$

The difference $Y(1) - Y(-1)$ is generally *not observable* since a subject is receiving either treatment 1 or treatment -1 (e.g., in parallel group studies), but its expectation is estimable under certain conditions.

We assume that $\Delta(\mathbf{x}) < 0$ reflects an advantage of treatment 1 over -1.

Prediction error of the PITE

$$\begin{aligned}\text{PEP}(\mathcal{L}) &= E_P\left[\left(Y(1) - Y(-1) - (\eta(\mathbf{X}, 1, \mathcal{L}) - \eta(\mathbf{X}, -1, \mathcal{L}))\right)^2\right] \\ &= E_P\left[\left(Y(1) - \eta(\mathbf{X}, 1, \mathcal{L})\right)^2\right] + E\left[\left(Y(-1) - \eta(\mathbf{X}, -1, \mathcal{L})\right)^2\right] \\ &\quad - 2E_P\left[\left(Y(1) - \eta(\mathbf{X}, 1, \mathcal{L})\right)\left(Y(-1) - \eta(\mathbf{X}, -1, \mathcal{L})\right)\right]\end{aligned}$$

Since T is independent of $\mathbf{X}$, the first and the second term are estimable from subjects receiving treatment 1 or -1 respectively: the missing measurements are missing completely at random.

The third term is *not estimable* because only one treatment is assigned to a subject. An upper bound can be obtained by applying Schwartz's inequality:

$$2\sqrt{E_P\left[\left(Y(1) - \eta(\mathbf{X}, 1, \mathcal{L})\right)^2\right] E\left[\left(Y(-1) - \eta(\mathbf{X}, -1, \mathcal{L})\right)^2\right]}$$

An alternative approach [3]

Note that if $P[T = 1] = 1/2$, the following identity holds:

$$\begin{aligned} E_P[2TY|\mathbf{X} = \mathbf{x}] &= 2E_P[Y|\mathbf{X} = \mathbf{x}, T = 1]P[T = 1] \\ &\quad - 2E_P[Y|\mathbf{X} = \mathbf{x}, T = 0]P[T = -1] \\ &= E_P[Y(1) - Y(-1)|\mathbf{X} = \mathbf{x}] \\ &= \Delta(\mathbf{x}) \end{aligned}$$

This suggests to estimate $\Delta(\mathbf{x})$ with $D(\mathbf{x})$ obtained from minimizing

$$\sum_{n=1}^{N} \big(2t_n y_n - D(\mathbf{x}_n)\big)^2,$$

the *modified outcome model*.

A special class of predictors

Now consider the class of predictors

$$\eta(\mathbf{x}, t) = \gamma' W(\mathbf{x}) + t\delta' W(\mathbf{x})/2$$

with $W : \mathcal{R}^K \to \mathcal{R}^K$ containing an intercept[1]. Then

$$D(\mathbf{x}) = \eta(\mathbf{x}, 1) - \eta(\mathbf{x}, -1) = \delta' W(\mathbf{x})$$

is a predictor of $\Delta(\mathbf{x})$. One can estimate δ directly from minimizing

$$\sum_{n=1}^{N} \left(2t_n y_n - \delta' W(\mathbf{x}_n)\right)^2$$

[1] $W = (1, \mathbf{x})'$ yields a linear predictor.

Estimates of the prediction error of the PITE

By using $Y = \frac{1}{2}\big((1+T)Y(1) + (1-T)Y(-1)\big)$, $T^2 = 1$, and the independence of T and $Y(t)$ one can show

$$\begin{aligned}
\mathrm{PEP}(\mathcal{L}) &= E_P\big[\big(Y(1) - Y(-1) - D(\mathbf{X})\big)^2\big] \\
&= E\big[\big(2YT - D(\mathbf{X})\big)^2\big] - E_P\big[\big(Y(1) + Y(-1)\big)^2\big] \\
&\leq E\big[\big(2YT - D(\mathbf{X})\big)^2\big] - E_P\big[Y(1) + Y(-1)\big]^2
\end{aligned}$$

A naive estimator of the first term is

$$\frac{1}{N}\sum_{n=1}^{N}\big(2t_n y_n - \hat{\delta}' W(\mathbf{x}_n)\big)^2$$

This estimator is *too optimistic* since δ has been estimated to minimize it. It can be corrected by *cross-validation* or *bootstrap* [2].

Correcting the optimism by regular bootstrap

1. Draw B bootstrap samples $\mathcal{L}_b^*$, $b = 1, \ldots, B$, from $\mathcal{L}$.
2. Estimate $\hat{\delta}_b^*$ from each sample $\mathcal{L}_b^*$ by minimizing

$$\sum_{n=1}^{N} \left(2t_{bn}^* y_{bn}^* - \delta' W(\mathbf{x}_{bn}^*)\right)^2$$

3. Estimate the optimism by

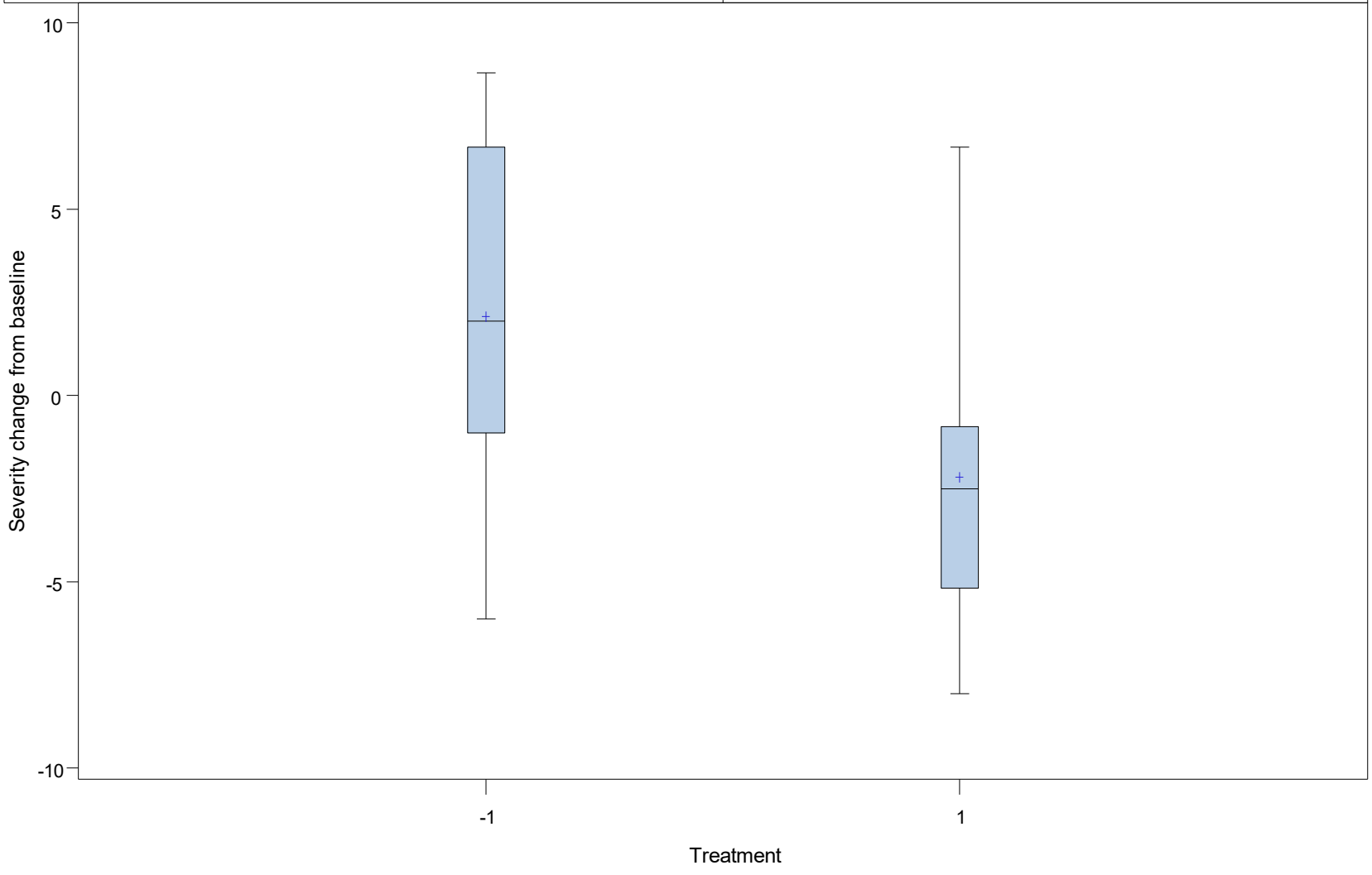
$$\widehat{\mathrm{OP}}(\mathcal{L}) = \frac{1}{BN} \sum_{b=1}^{B} \sum_{n=1}^{N} \left(2t_n y_n - \hat{\delta}_b^{*\prime} W(\mathbf{x}_n)\right)^2 - \frac{1}{BN} \sum_{b=1}^{B} \sum_{n=1}^{N} \left(2t_{bn}^* y_{bn}^* - \hat{\delta}_b^{*\prime} W(\mathbf{x}_{bn}^*)\right)^2$$

An Alzheimer dataset

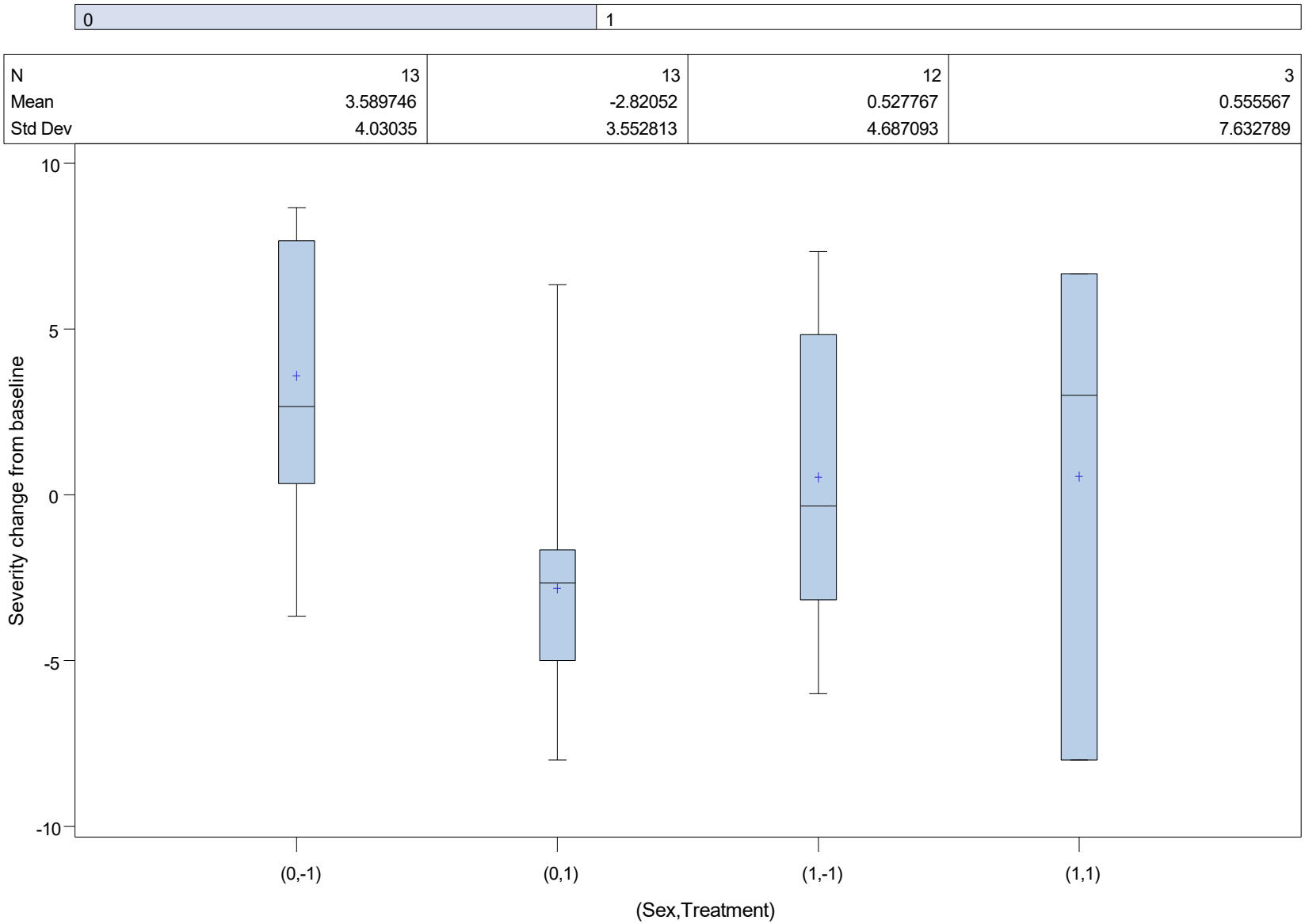
- This dataset referenced and provided in [4] is taken from the clinical development of an Alzheimer's drug undertaken by AbbVie. Unfortunately, references to clinical trial databases or a publication are missing.
- There were 41 subjects treated, 25 receiving placebo ($t = -1$) and 16 receiving test treatment ($t = 1$).
- Four *candidate biomarkers* were recorded at baseline: Disease severity (with high values indicating severe cognitive impairment), age, sex and presence or absence of a genetic marker.
- The response of interest is *change in disease severity* from baseline to end of study where a negative change refers to improvement.

Boxplots and summary statistics

N	25	16
Mean	2.119996	-2.1875
Std Dev	4.541872	4.440504



Boxplots and summary statistics by sex



Predictor for the Alzheimer data

We investigate a *linear predictor* of severity change based on the values of the biomarkers mentioned above, namely

$$\eta(\mathbf{x}, t) = \gamma_0 + \sum_{k=1}^{4} \gamma_k x_k + t\left(\delta_0 + \sum_{k=1}^{4} \delta_k x_k\right)$$

We estimate the relevant parameters δ_k by minimizing

$$\frac{1}{41} \sum_{n=1}^{41} \left\{ 2t_n y_n - \delta_0 - \sum_{k=1}^{4} \delta_k x_{nk} \right\}^2$$

to obtain the predictor

$$D(\mathbf{x}) = \hat{\delta}_0 + \sum_{k=1}^{4} \hat{\delta}_k x_k$$

Parameter estimation

Table 1: Parameter estimates of linear predictor for Alzheimer data

Parameter	Marker	Estimate	StdErr
δ_0	—	14.518442	12.42435
δ_1	Severity	0.016837	0.161010
δ_2	Age	-0.286140	0.153540
δ_3	Carrier(Y)	1.015439	2.672602
δ_4	Sex(M)	5.513353	2.754234

Results

- The naive estimate of the prediction error equals 62.93, the estimator of the upper bound from Schwartz's inequality is 104.19
- To compensate for the optimism we run 1000 bootstrap samples and obtained a correction of 17.32, resulting in a corrected prediction error estimate of 80.25, which is not great given the actual range of responses
- Considering the standardized estimates, age and sex seem to be the most influential markers
- Female patients are predicted an extra gain of 5 units on the severity scale under test treatment as compared to male patients while one year of age increases efficacy of test treatment by about 0.3 units
- If one conducts a forward variable selection that enters variables as long as the Bayesian Information Criterion (BIC) decreases, age and sex are retained in the model

Variable selection

Table 2: Prediction error estimate of the PITE for predictors with 1 to 4 variables for the Alzheimer data as determined by forward selection. The optimism is calculated from 1000 bootstrap samples.

# Variables	Naive estimate	Optimism	Corrected estimate
1	69.39	13.15	82.54
2	63.20	15.02	**78.22**
3	62.95	16.72	79.67
4	62.93	17.32	80.25

Noteworthy is the trade-off between the naive estimate, which decreases, and optimism, which increases with the number of variables in the model.

Simulations

Following [3], data were generated from

$$Y = \left(\gamma_0 + \sum_{k=1}^{9} \gamma_k X_k\right)^m + \left(\delta_0 + \sum_{k=1}^{9} \delta_k X_k + \delta_{12} X_1 X_2\right) + e$$

with

- $X_k \sim N(0, 1)$
- $e \sim N(0, 2)$
- $\delta_0 = 0.4$, $\delta_2 = 0.8$, $\delta_3 = -0.8$, $\delta_k = 0$ for $k > 3$
- $\gamma_0 = 1/\sqrt{6}$, $\gamma_1 = \gamma_2 = 2/\sqrt{6}$, $\gamma_k = 0$ for $k > 3$
- $m = 1, 2$
- $\delta_{12} = 0, 0.8$

Simulation results

m	δ_{12}	$N = 100$	$N = 1000$
1	0	9.62, 8.99	8.12, 8.39
1	0	9.62, 8.59	8.12, 8.03
2	0	9.67, 8.87	8.17, 8.32
1	0.8	12.46, 11.59	10.67, 10.91
2	0.8	12.48, 11.46	10.72, 10.85

Table 3: Upper bounds of prediction error. Left number: approximation by Schwartz inequality after fitting the full prediction model. Right number: approximation using the modified outcome fit. The second simulation was run with $\gamma_k = 0$ for all k.

Conclusions

- We investigated two upper bounds of the prediction error of the individual treatment effect for continuous data, equally allocated treatments and a special but common type of predictors.
- None of these upper bounds is uniformly superior to the other as shown by simulations.
- Results should be extendable to other data types (i.e., binary, survival).
- There is still room for improvement.

References

[1] A Bradford Hill.

Reflections on the controlled trial.

Annals of Rheumatoid Diseases, 25:107–113, 1966.

[2] B Efron and R Tibshirani.

An Introduction to the Bootstrap.

Chapman and Hall/CRC, 1993.

[3] L Tian, AA Alizadeh, AJ Gentles, and R Tibshirani.

A simple method for estimating interactions between a treatment and a large number of covariates.

Journal of the American Statistical Association, 109:1517–1532, 2014.

[4] P M Schnell, Q Tang, W W Offen, and B P Carlin.

A Bayesian credible subgroup approach to identifying patient subgroups with positive treatment effect.

Biometrics, 72:1026–1036, 2016.