Individualized Treatment Effects Using a Non-parametric Bayesian Approach

Ravi Varadhan
Nicholas C. Henderson

Division of Biostatistics & Bioinformatics
Department of Oncology
Johns Hopkins University

July 27, 2017

Acknowledgements

- Drs. Tom Louis and Gary Rosner
- Patient-Centered Outcomes Research Institute (PCORI) Award (ME-1303-5896)
- Dr. Dan Scharfstein and the CTSA Supplement Award

Why is HTE Important?

The paradox of the clinical trial is that it is the best way to assess whether an intervention works, but is arguably the worst way to assess who benefits from it. (Mant, 1999)

Heterogeneity of Treatment Effect (HTE)

- Heterogeneity of Treatment Effect (HTE) refers to variability in treatment response that is attributable to observable differences in patient characteristics.
- Consistency of treatment effect across key patient subgroups - subgroup analysis - most prevalent form of HTE assessment
- Tests for treatment-covariate interactions - less common.

HTE Goals/Questions

- Our view: HTE assessment encompasses a broad range of related goals and questions.
- This goes beyond conventional subgroup analysis and treatment-covariate interactions.

Key questions of interest include:

- Quantifying overall heterogeneity in treatment response.
- Estimating the proportion of patients that benefit from treatment
- Detection of cross-over (qualitative) interactions.
- Estimation and prediction of individualized treatment effects.
- Optimal allocation of treatments to individuals

Modeling HTE

- In contrast to subgroup analysis, many important HTE questions could be directly addressed if a sufficiently rich model describing patient outcomes were available.
- Bayesian nonparametric methods are well-suited to provide this individual-level view of HTE.
- Bayesian nonparametrics allow construction of flexible models for patient outcomes coupled with probability modeling of all unknown quantities
- **Motivation of this work:** Develop a flexible, non-parametric approach that can address many of the previously highlighted HTE goals.

Why Bayes?

- The phrase 'heterogeneity of treatment effects' implies an underlying distribution of treatment effects
- Thus a Bayesian framework appears natural
- Emphasizes estimation of treatment effect heterogeneity rather than hypothesis testing.
- Well-suited to estimation with many parameters and small subgroups. Tends to prevent "over-fitting".

Why Bayes?

- Direct probability statements for questions of interest:
  - e.g., what is the probability that a given individual will benefit from the treatment?
- Customized treatment recommendations - can utilize the posterior for each individual, can directly weigh efficacy versus safety.

Time-to-Event Data and Notation

- Our focus here is on cases where patient outcomes are recorded as time-to-events.

- For the i^{th} patient, we observe $\quad Y_i = \min\{T_i, C_i\}$

  T_i - failure time
  C_i - censoring time
  $\delta_i = 1$ if $T_i \leq C_i$, and $\delta_i = 0$ if $T_i > C_i$
  A_i - treatment assignment, $A_i = 1$ or $A_i = 0$
  $\mathbf{x}_i$ - a collection of baseline covariates

Accelerated Failure Time (AFT) Models and Individualized Treatment Effects

- We assume patients are randomly assigned to one of two treatments $A = 1$ or $A = 0$.
- Consider the AFT model for log-failure time T_i

$$\log T_i = \underbrace{m(A_i, \boldsymbol{x}_i)}_{\text{Regression Function}} + \underbrace{W_i}_{\text{Error Term}} \; ; \qquad E(W_i) = 0$$

- The "Individualized Treatment Effect" $\theta(\mathbf{x}_i)$ for the i^{th} patient is the difference between expected log-failure under treatment $A = 1$ and expected log-failure time under treatment $A = 0$

$$\begin{aligned} \theta(\boldsymbol{x}_i) &= E\big[\log T_i | A_i = 1, \boldsymbol{x}_i\big] - E\big[\log T_i | A_i = 0, \boldsymbol{x}_i\big] \\ &= m(1, \boldsymbol{x}_i) - m(0, \boldsymbol{x}_i). \end{aligned}$$

- The ratio of expected failure times offers a more interpretable measure of treatment effect

$$\xi(\mathbf{x}_i) = \frac{E\big[T_i | A_i = 1, \boldsymbol{x}_i\big]}{E\big[T_i | A_i = 0, \boldsymbol{x}_i\big]} = \exp\{\theta(\mathbf{x}_i)\}$$

Accelerated Failure Time (AFT) Models

- In contrast to Cox-proportional hazards model, AFT models provide a direct relationship between survival times and patient covariates.
- AFT models have a nice interpretation as a regression with log-time as the response.
- They provide interpretable measures of treatment effect: i.e., differences in expected log-survival time or ratios in expected survival,

Modeling the Regression Function

- Our flexible approach to modeling the regression function $m(A_i, \boldsymbol{x}_i)$ of the AFT model is to use Bayesian additive regression trees (BART)

Advantages of BART:

- Good at handling interactions and non-linearities
- Very effective as an "off-the-shelf" method - works quite well without any hyperparameter tuning.
- Seamlessly incorporates both discrete and continuous predictors.
- Automatically provides measures of uncertainty despite the complex nature of the model.

Using the Nonparametric AFT model to assess HTE

The posterior distribution of all unknowns in the AFT model can be used to assess a variety of questions.

For example,

- Point estimates of covariate-specific treatment effects
- The "distribution" of treatment effects
- Proportion of patient expected to benefit from treatment
- Qualitative interactions.
- The posterior can potentially be utilized in an individualized decision analysis

Application: The SOLVD trial

- A placebo-controlled trial studying the efficacy of the drug Enalapril in chronic heart failure patients
- $2,569$ enrolled in the treatment trial and $4,228$ enrolled in the prevention trial
- We utilized 18 patient covariates common to both trials (e.g., age, gender, ejection fraction)

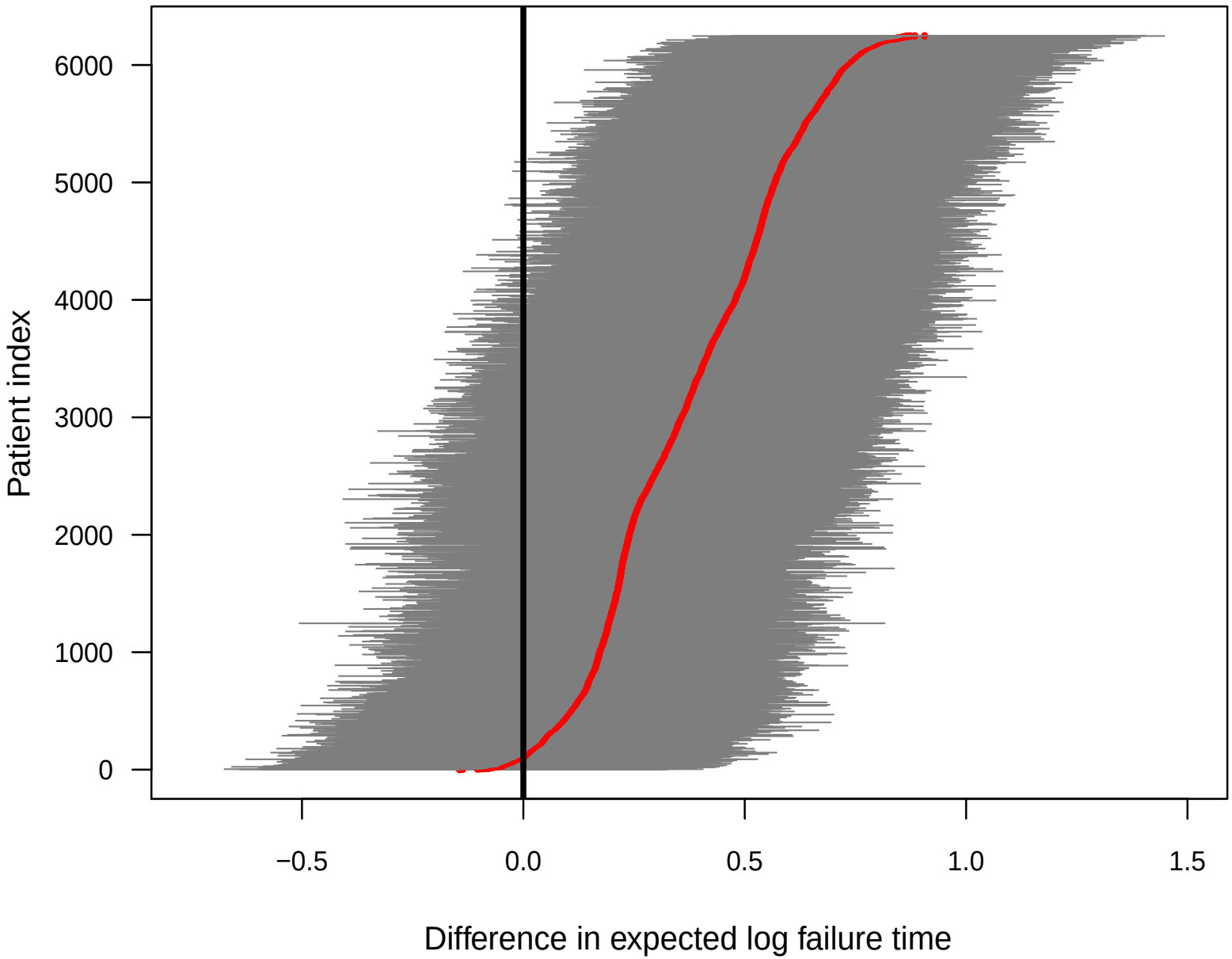
The AFTrees package

The methods discussed here are implemented using the R package: **AFTrees**.

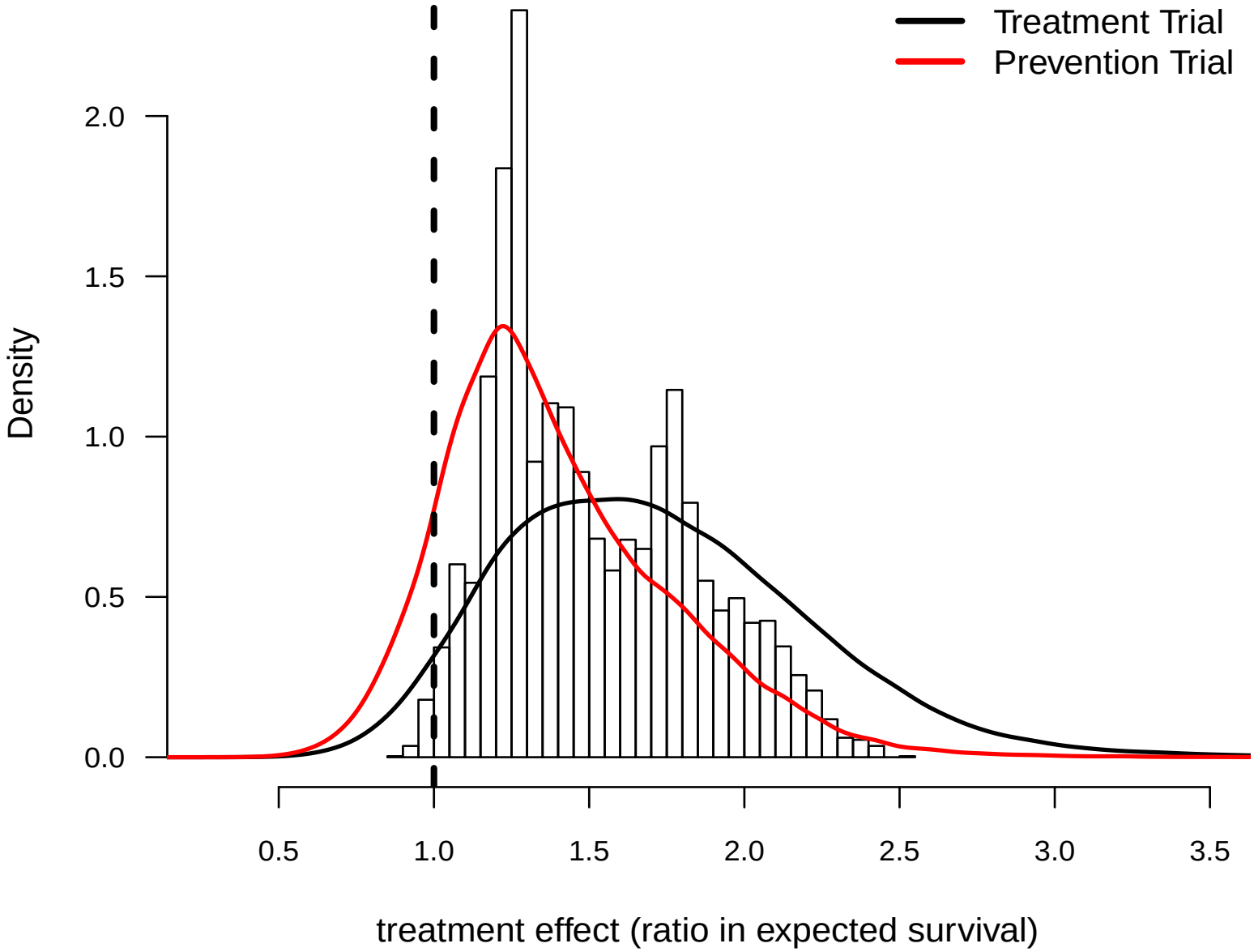
```
## An example:

library(AFTrees)
solvd.fit <- AFTrees(X, y, status, ndpost = 2000)

## X - design matrix
## y - follow-up time
## status - event indicator (1 if event, 0 otherwise)
## ndpost - number of posterior draws
```

The AFTrees package is available for download at
`http://www.hteguru.com/software`

Individualized treatment effect estimates for all patients in the SOLVD treatment and prevention trials.



Distribution of treatment effects in the SOLVD trials.



Proportion Benefiting

- The proportion of patients benefiting is the proportion of individuals with a positive treatment effect (i.e., $\theta(\mathbf{x}_i) > 0$)

$$Q = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\theta(\mathbf{x}_i) > 0\} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\xi(\mathbf{x}_i) > 1\}$$

- Alternatively, one could define the proportion benefiting relative to a clinically relevant threshold $\varepsilon > 0$, i.e., $Q_\epsilon = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\theta(\mathbf{x}_i) > \varepsilon\}$.
- An estimate of Q is obtained from taking the area under the curve to the right of 1 in the graph of treatment effect distribution (shown on the previous slide).
- The estimated proportions of patients benefiting were 95.6% and 89.1% in the SOLVD-T and SOLVD-P trials respectively.

Evidence of Benefit

	Treatment Trial	Prevention Trial
$P\{\xi(\mathbf{x}_i) > 1 \mid \text{data}\} \in (0.99, 1]$	51.38	20.47
$P\{\xi(\mathbf{x}_i) > 1 \mid \text{data}\} \in (0.95, 0.99]$	24.69	23.71
$P\{\xi(\mathbf{x}_i) > 1 \mid \text{data}\} \in (0.75, 0.95]$	20.08	41.98
$P\{\xi(\mathbf{x}_i) > 1 \mid \text{data}\} \in (0.25, 0.75]$	3.85	13.84
$P\{\xi(\mathbf{x}_i) > 1 \mid \text{data}\} \in [0, 0.25]$	0.00	0.00

Table: For each trial, the percentage of patients whose estimated posterior probability of treatment benefit lies within each of the intervals $(0.99, 1]$, $(0.95, 0.99]$, $(0.75, 0.95]$, $(0.25, 0.75]$, and $[0, 0.25]$.

Evidence of Differential Treatment Effect

- For patient i, the posterior probability of a greater than average treatment effect may be defined as

$$D_i = P\{\theta(\mathbf{x}_i) \geq \bar{\theta} | \text{data}\}, \qquad \bar{\theta} = \frac{1}{n}\sum_{i=1}^{n} \theta(\mathbf{x}_i)$$

  and the posterior probability of a "differential" treatment effect is

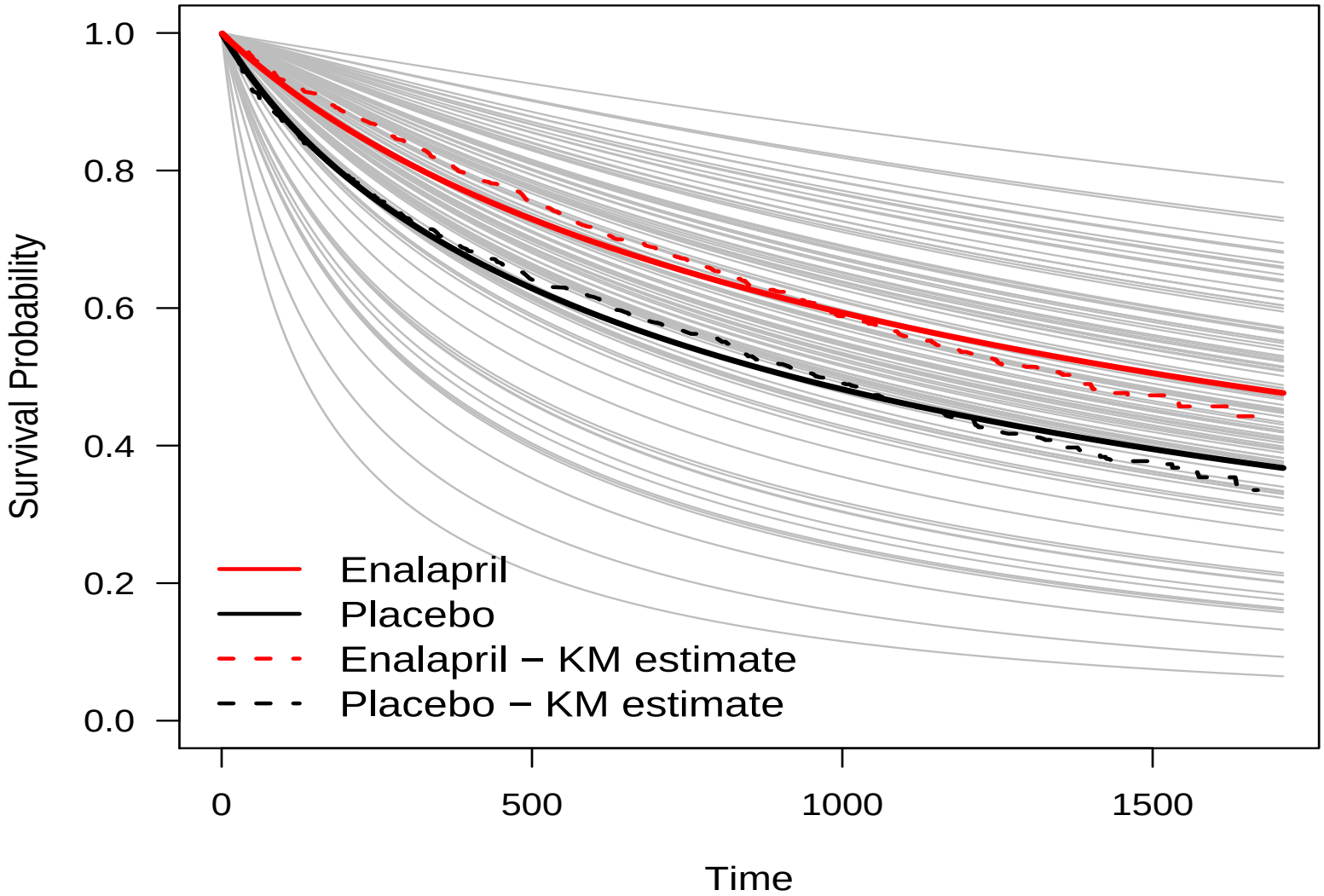
$$D_i^* = \max\{1 - 2D_i, 2D_i - 1\}.$$

- Note that D_i^* will be close to 1 whenever D_i is either close to 1 or close to 0.

	Treatment Trial	Prevention Trial
$D_i^* > 0.95$	19.36	7.30
$D_i^* > 0.80$	41.93	31.58

Table: For each trial, the percentage of patients that show "strong" (i.e., $D_i^* > .95$) and "mild" (i.e., $D_i^* > 0.80$ but $D_i^* \leq 0.95$) evidence of differential treatment effect.

Individual-Specific Posterior Survival Curves in SOLVD treatment trial



Examining Qualitative Interactions

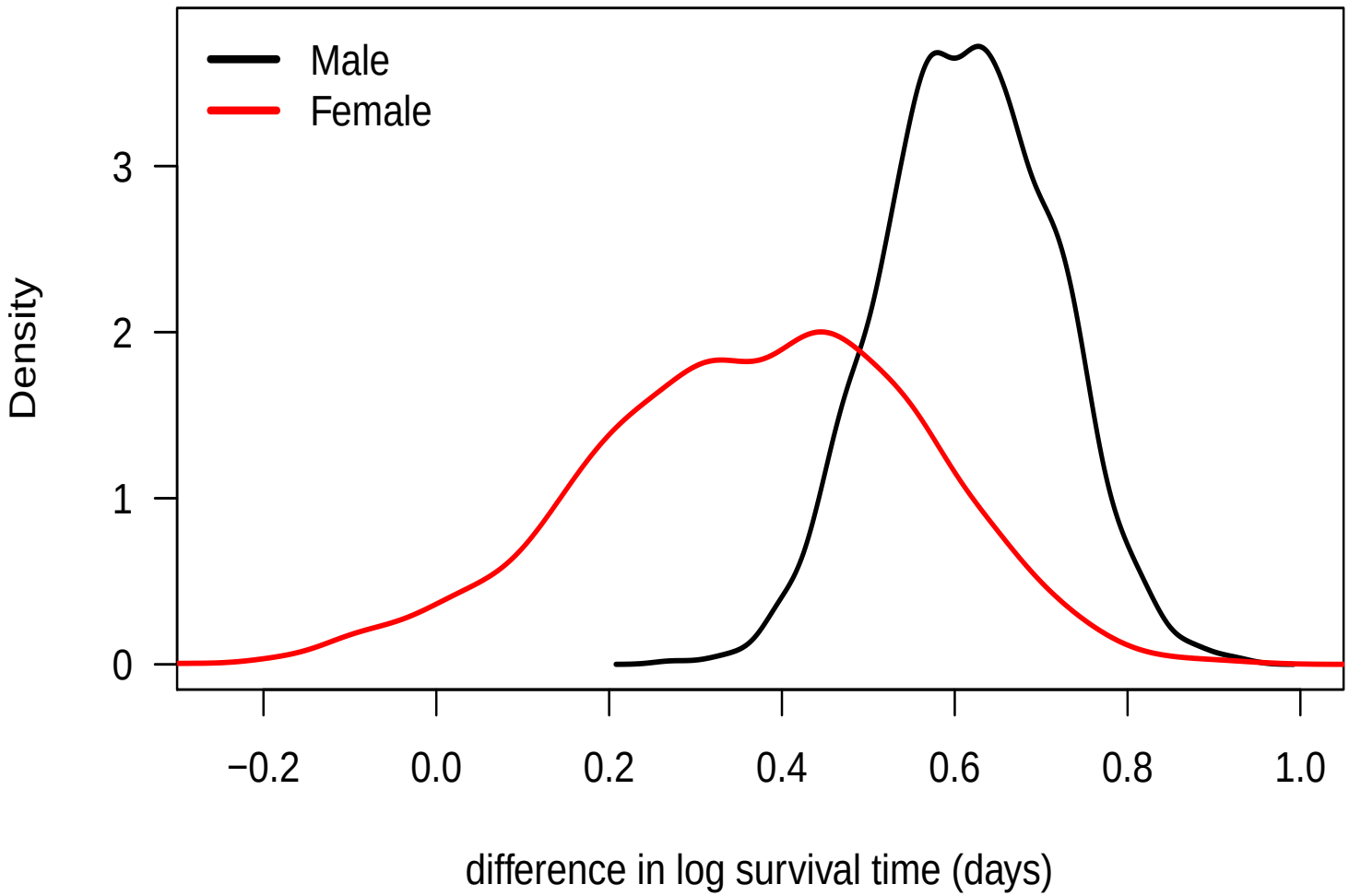
- Beyond quantitative heterogeneity, examination of qualitative interactions is often of key interest.
- **Qualitative Interaction:** occurs when the treatment effect in one subgroup has a different sign than in another subgroup.
- The presence of qualitative interactions can be examined by looking at the posterior histogram.
- For pre-specified subgroups of interest such as male vs. female, we can look at the posteriors of the subgroup-level treatment effects

$$\theta_{male} = \frac{1}{N_{male}} \sum_{i \in male} \theta(\mathbf{x}_i)$$

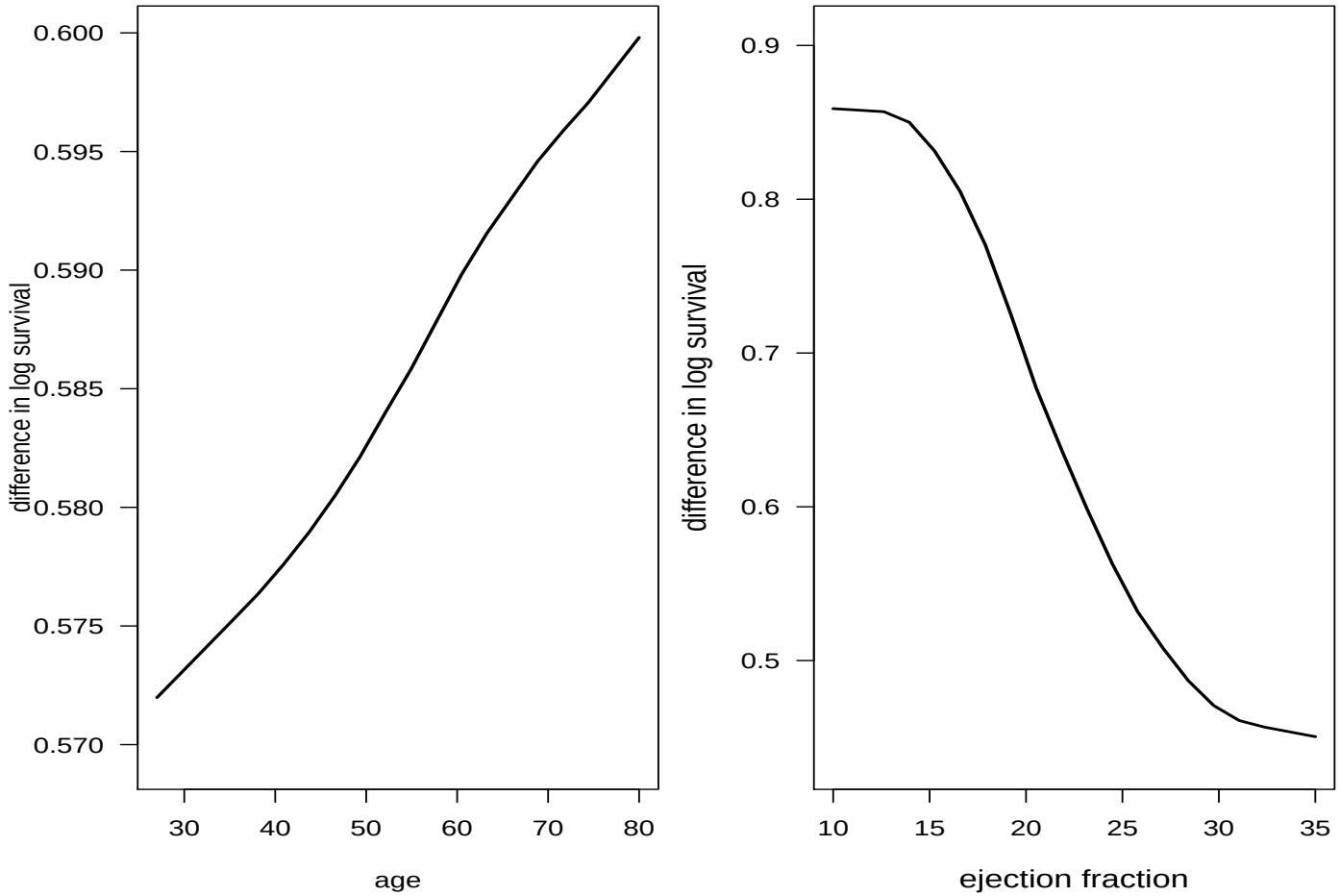
$$\theta_{female} = \frac{1}{N_{female}} \sum_{i \in female} \theta(\mathbf{x}_i)$$

SOLVD data: posterior of θ_{male} and θ_{female}

$$P\{\text{sign}(\theta_{male}) \neq \text{sign}(\theta_{female})|data\} = 0.035$$



Variable Importance: Partial Dependence Plots



Summary

- A flexible Bayesian approach for HTE assessment
- Fully non-parametric
- Default hyper-parameter settings, hence minimal user input
- Easy to use R package: **AFTrees**
  `http://www.hteguru.com/software`
- Give it a try and send us feedback!