TECHNION
Israel Institute
of Technology

TEL AVIV UNIVERSITY

Heterogeneous Treatment Effect in Time-to-Event Outcomes: Harnessing Censored Data with Recursively Imputed Trees

Tomer Meir, Uri Shalit, and Malka Gorfine

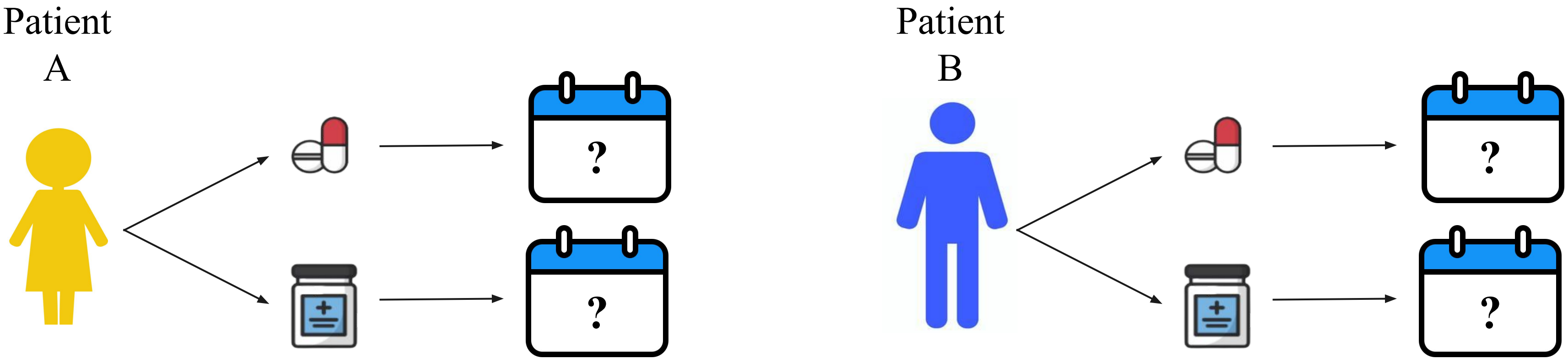
42nd International Conference of Machine Learning - ICML 2025

TECHNION
Israel Institute
of Technology

TEL AVIV UNIVERSITY

Motivation

Which cancer treatment would lead to longer survival for each patient?

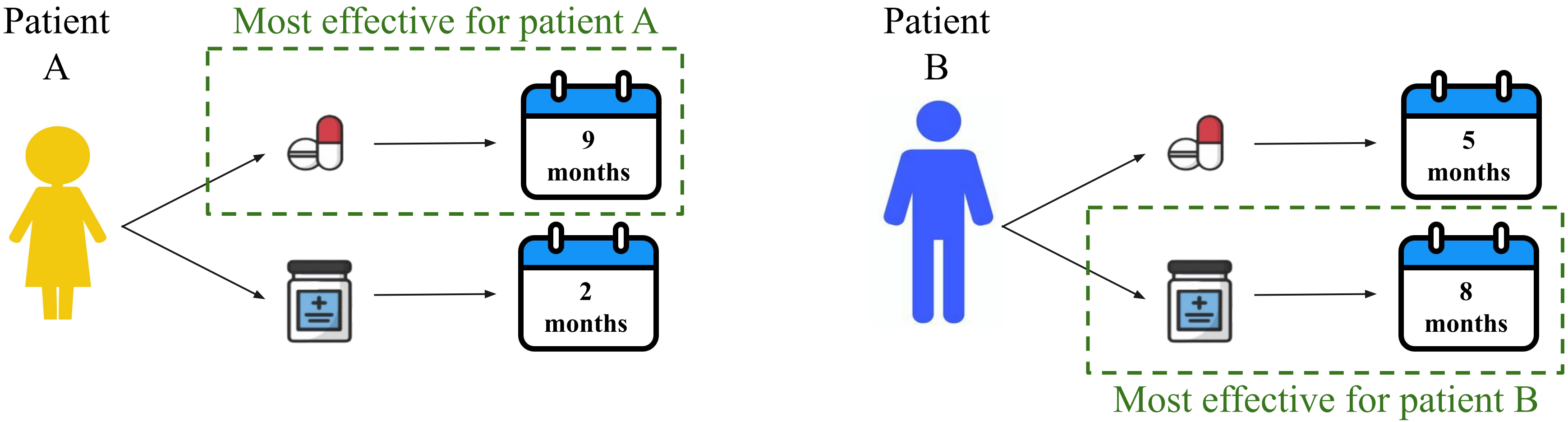


TECHNION Israel Institute of Technology

TEL AVIV UNIVERSITY

Motivation

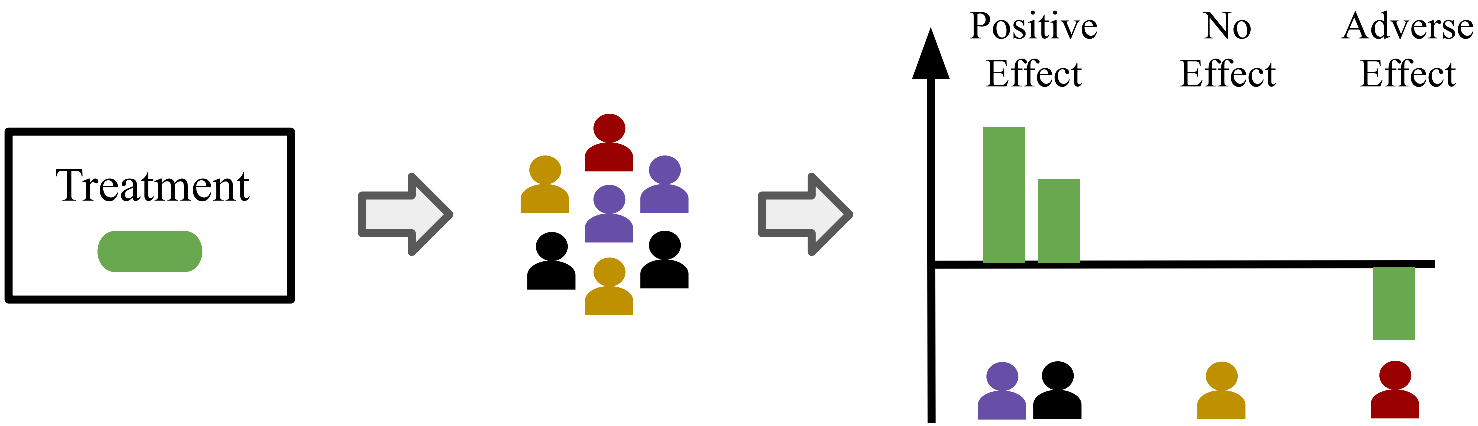
Which cancer treatment would lead to longer survival for each patient?



Knowing this is a key step toward **the selection of the most effective treatment for each individual**.

Heterogeneous Treatment Effects

Heterogeneous Treatment Effects (HTE) describe the way treatments impact different subgroups.



Technion Israel Institute of Technology

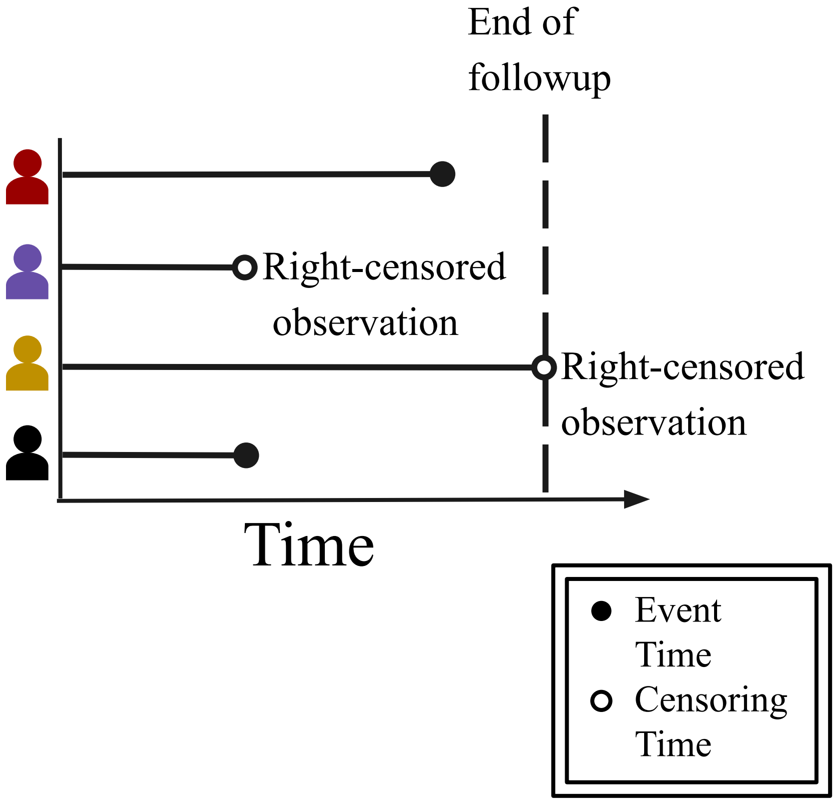
Tel Aviv University

Time to Event Analysis

Time-to-event (TTE) analysis is used when the **outcome is time until a specific event**.

Right-censored observations:

- Exact event-time is unknown.
- Event-free until a certain time.



Estimand

Our goal is to estimate the HTE in TTE data,

defined by

$$\tau(x) = E\left\{g(\widetilde{T}_i^0) - g(\widetilde{T}_i^1) | X_i = x\right\}$$

for a known function *g(·)*.

Notation

N iid observations

$X_i \in \mathbb{R}^p$ - vector of *p* covariates of observation *i*

$\widetilde{T}_i \in \mathbb{R}_+$ - event time

$W_i \in \{0,1\}$ - treatment assignment

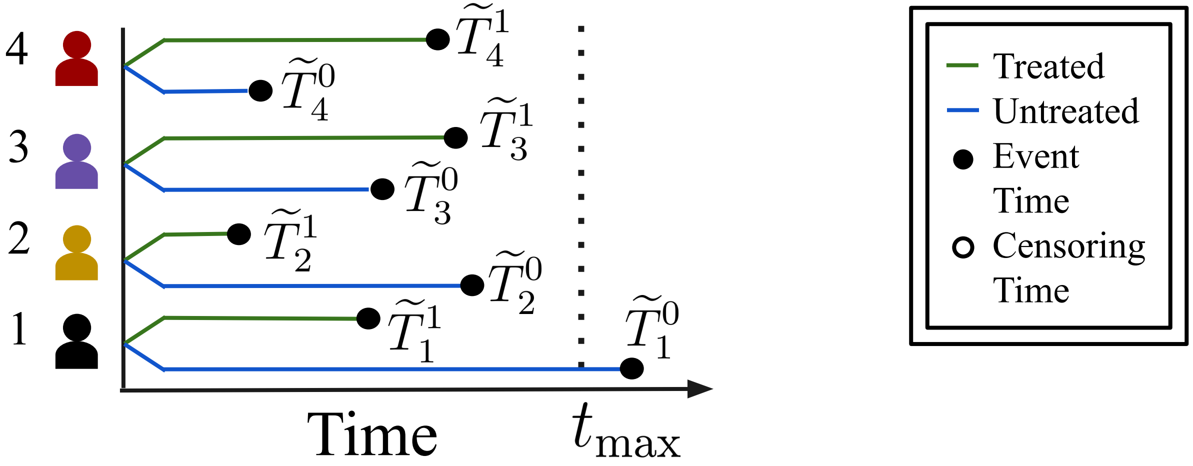
$\{\widetilde{T}_i^0, \widetilde{T}_i^1\}$ - potential outcomes under $W_i = 0$ and $W_i = 1$

Examples of the HTE for specific functions *g(·)*

g(·)	$\tau(\cdot)$
$g(\widetilde{T}_i) = \min(\widetilde{T}_i, h)$	difference in restricted mean survival time (RMST)
$g(\widetilde{T}_i) = I(\widetilde{T}_i \geq h)$	differences in survival functions

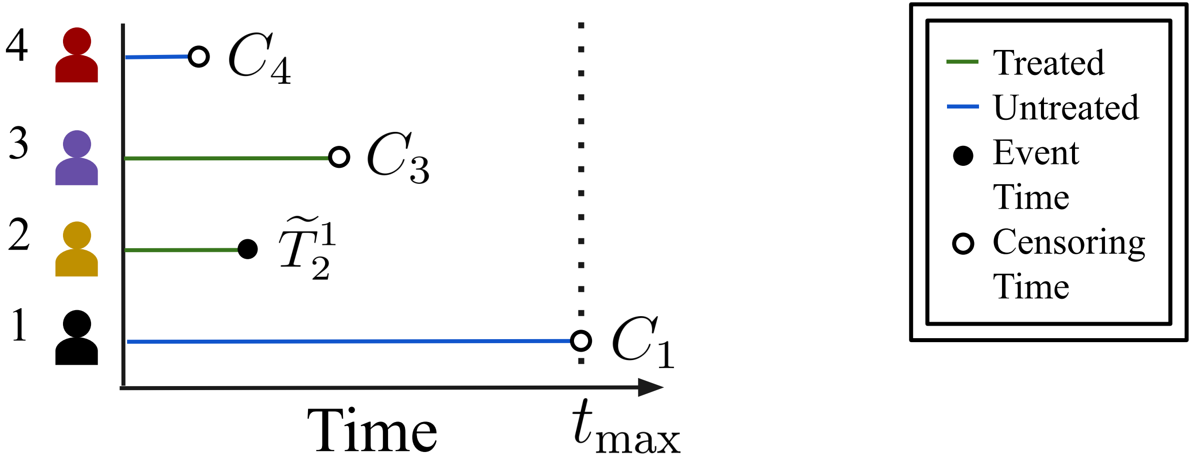
Ideal Data

Each sample would include **both potential outcomes** - under treatment and under no treatment.



Observed Data

Each sample receives a **single treatment, with right-censored observations** providing only a lower bound on survival under that arm.



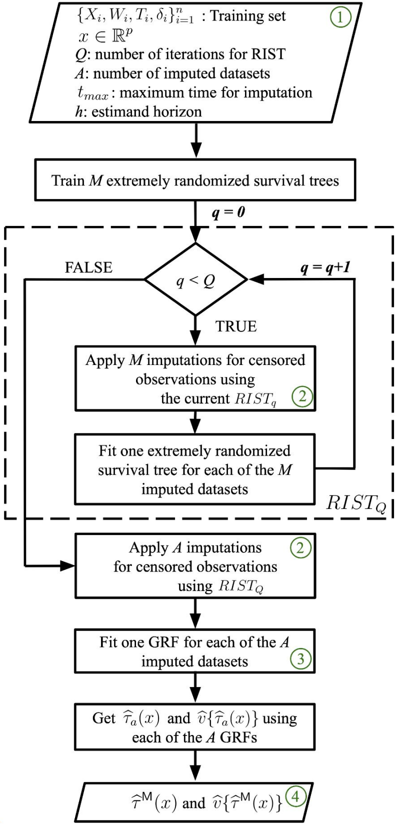
Existing Methods

Existing methods, such as causal survival forests (CSF) with inverse probability weighting and a doubly robust estimator [1], **struggle under heavy censoring and cannot handle instrumental variables (IV).**

[1] Cui, Y., Kosorok, M. R., Sverdrup, E., Wager, S., and Zhu, R. ***Estimating heterogeneous treatment effects with right-censored data via causal survival forests***. Journal of the Royal Statistical Society Series B: Statistical Methodology, 85(2):179–211, May 2023.

MISTR - An Overview

MISTR is a non-parametric multiple imputation based algorithm for estimating HTE in TTE data.



TECHNION Israel Institute of Technology

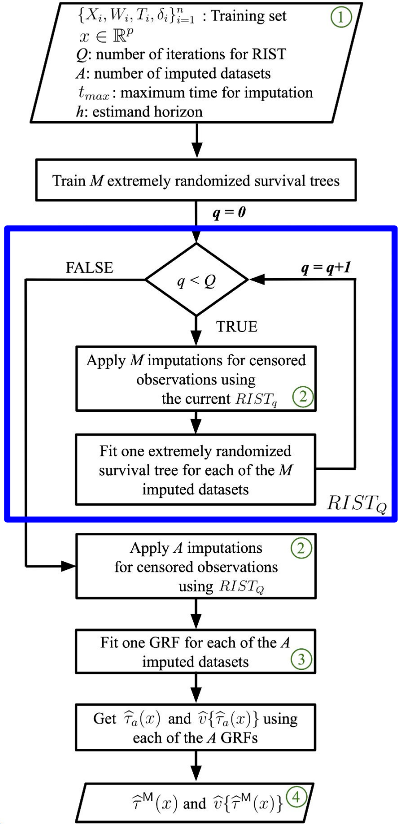
TEL AVIV UNIVERSITY

MISTR - An Overview

MISTR is a non-parametric multiple imputation based algorithm for estimating HTE in TTE data.

MISTR utilizes:

- Recursively Imputed Survival Trees (RIST) [2]



[2] Zhu, R. and Kosorok, M. R. ***Recursively Imputed Survival Trees***. Journal of the American Statistical Association, 107(497):331–340, March 2012.

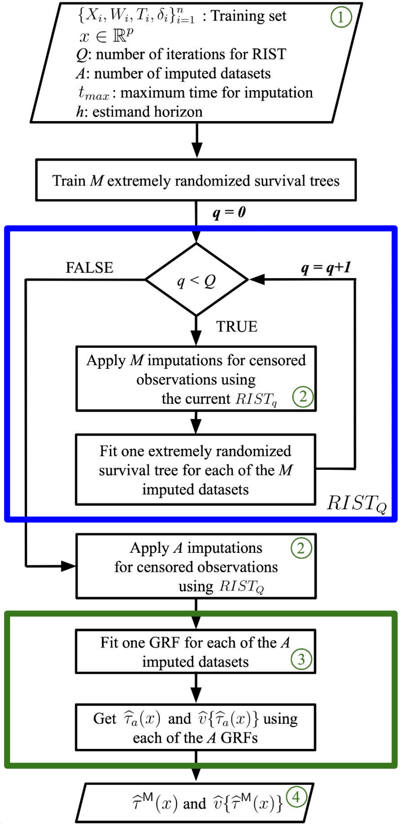
MISTR - An Overview

MISTR is a non-parametric multiple imputation based algorithm for estimating HTE in TTE data.

MISTR utilizes

- Recursively Imputed Survival Trees (RIST) [2]
- Generalized Random Forest (GRF) [3]

to overcome existing limitations.

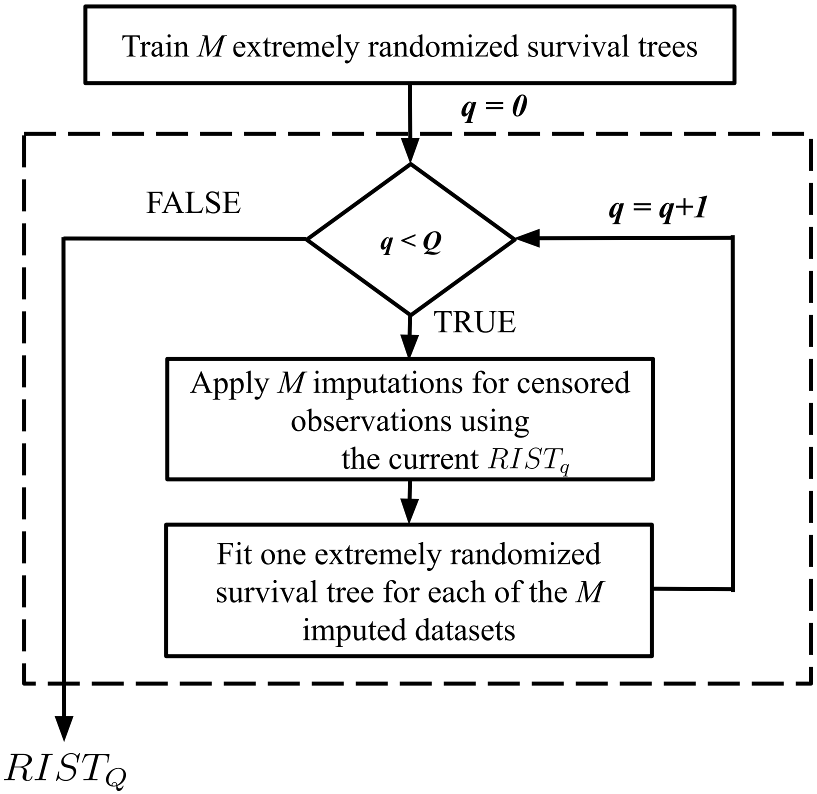


[2] Zhu, R. and Kosorok, M. R. ***Recursively Imputed Survival Trees***. Journal of the American Statistical Association, 107(497):331–340, March 2012.

[3] Athey, S., Tibshirani, J., and Wager, S. ***Generalized random forests***. The Annals of Statistics, 47(2), April 2019.

Q-RIST

Q-RIST [2] is a forest of M extremely randomized survival trees, trained over Q recursive steps.



[2] Zhu, R. and Kosorok, M. R. ***Recursively Imputed Survival Trees***. Journal of the American Statistical Association, 107(497):331–340, March 2012.

[3] Athey, S., Tibshirani, J., and Wager, S. ***Generalized random forests***. The Annals of Statistics, 47(2), April 2019.

Q-RIST

Q-RIST [2] is a forest of M extremely randomized survival trees, trained over Q recursive steps.

It is used for **estimating the conditional survival function while extracting more information from censored data**.

$$\boxed{\Pr(\widetilde{T}_i > t \mid X_i, W_i, C_i, \widetilde{T}_i > C_i) = \frac{\Pr(\widetilde{T}_i > t \mid X_i, W_i)}{\Pr(\widetilde{T}_i > C_i \mid X_i, W_i, C_i)}}$$

$X_i \in \mathbb{R}^p$ - vector of p covariates of observation i

$\widetilde{T}_i \in \mathbb{R}_+$ - event time

$C_i \in \mathbb{R}_+$ - censoring time

$W_i \in \{0, 1\}$ - treatment assignment

[2] Zhu, R. and Kosorok, M. R. ***Recursively Imputed Survival Trees***. Journal of the American Statistical Association, 107(497):331–340, March 2012.

[3] Athey, S., Tibshirani, J., and Wager, S. ***Generalized random forests***. The Annals of Statistics, 47(2), April 2019.

GRF

GRF [3] can be used for **estimating HTE** in non-censored data.

$$\tau(x) = E\left\{g(\widetilde{T}_i^0) - g(\widetilde{T}_i^1)|X_i = x\right\}$$

Heterogeneity is captured through similarity weights, defined as the fraction of trees in which a given training and test sample share the same leaf.

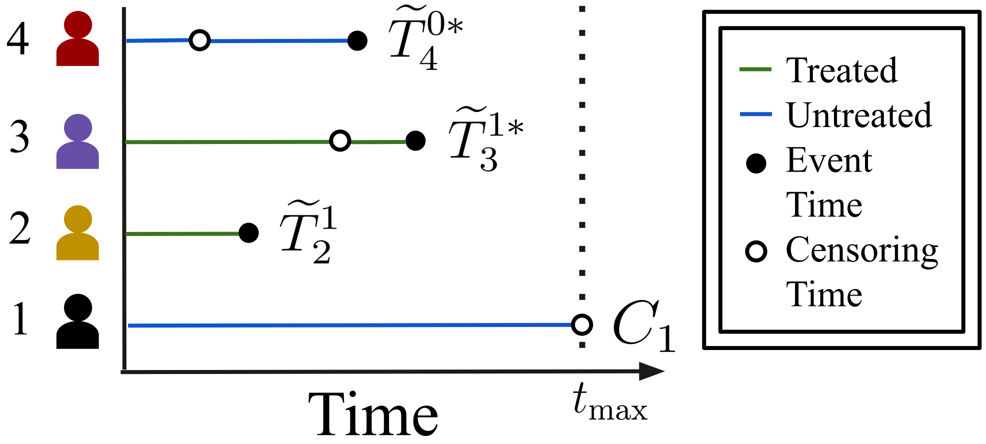
[2] Zhu, R. and Kosorok, M. R. ***Recursively Imputed Survival Trees***. Journal of the American Statistical Association, 107(497):331–340, March 2012.
[3] Athey, S., Tibshirani, J., and Wager, S. ***Generalized random forests***. The Annals of Statistics, 47(2), April 2019.

MISTR

Initially, RIST is employed to impute event times for censored observations.

The imputed and non-censored observations are merged to form complete uncensored datasets.

Observed Data with a Single Imputation



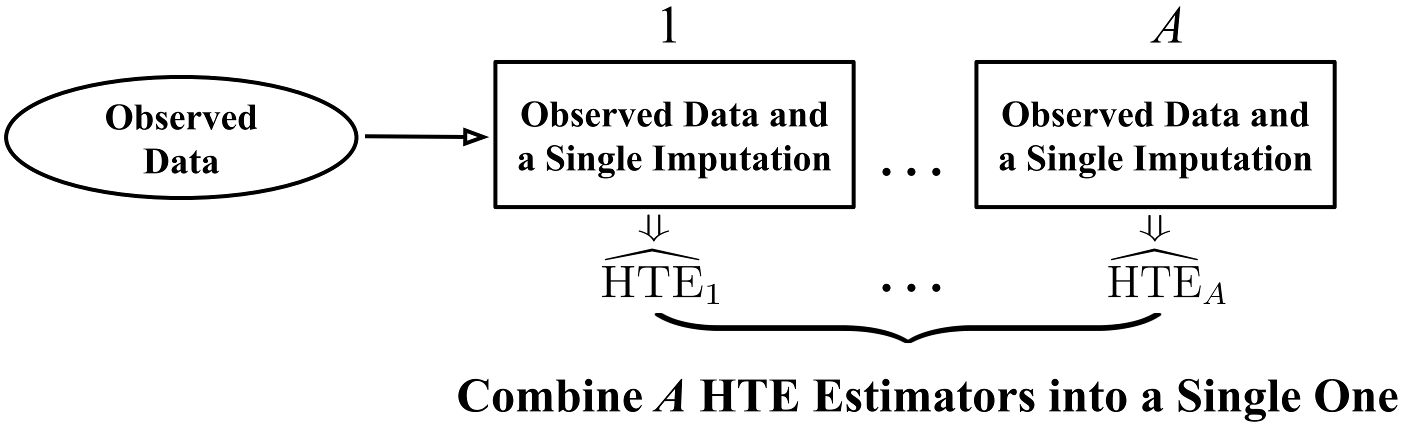
[2] Zhu, R. and Kosorok, M. R. ***Recursively Imputed Survival Trees***. Journal of the American Statistical Association, 107(497):331–340, March 2012.

[3] Athey, S., Tibshirani, J., and Wager, S. ***Generalized random forests***. The Annals of Statistics, 47(2), April 2019.

MISTR

GRF is then used to produce A HTE estimates, each based on different imputed event times.

Our proposed estimate is the average of these A HTE estimates.

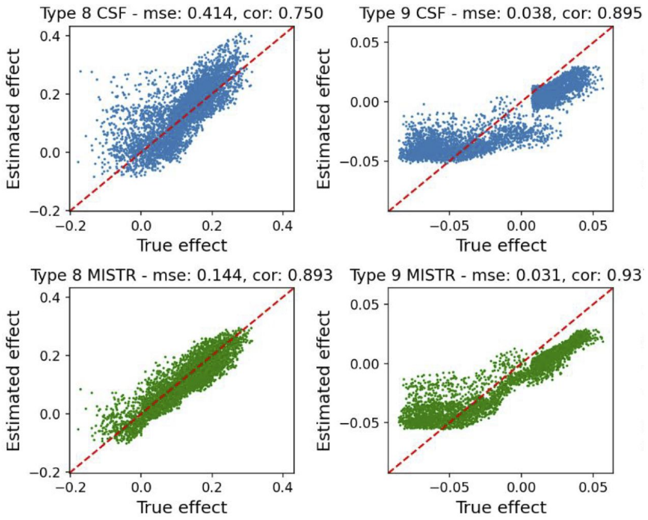


TECHNION Israel Institute of Technology

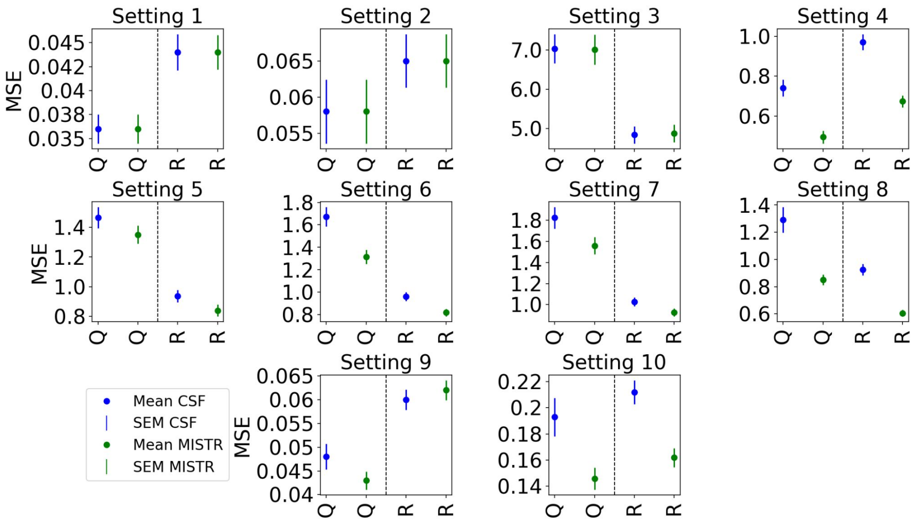
TEL AVIV UNIVERSITY

Simulations

We show that MISTR outperforms prior methods under heavy censoring.



True vs Estimated effect Calculated over one random test sample of 5000 observations.



Mean squared error (MSE) on randomly sampled test set (R) and on covariates quantiles test set (Q).

Simulations

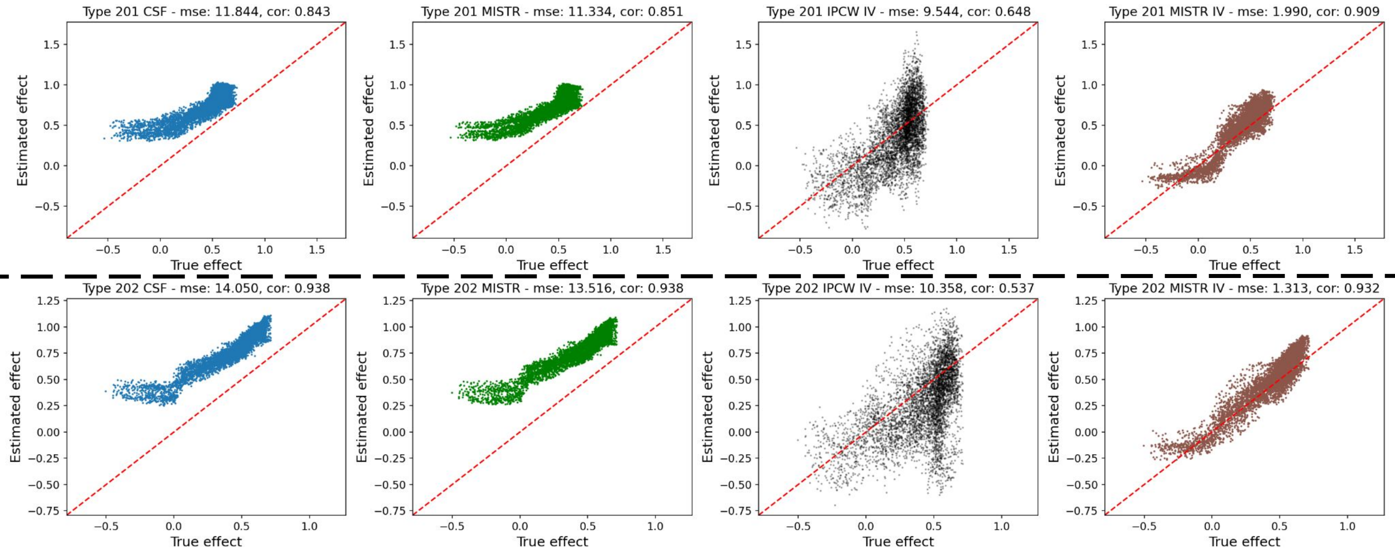
MISTR can be extended to the setting of instrumental variable (MISTR-IV).

Non-IV methods

IV methods

Type 201

Type 202



True vs Estimated effect Calculated over one random test sample of 5000 observations.

TECHNION Israel Institute of Technology

TEL AVIV UNIVERSITY

Use-Case: HIV Clinical Trial

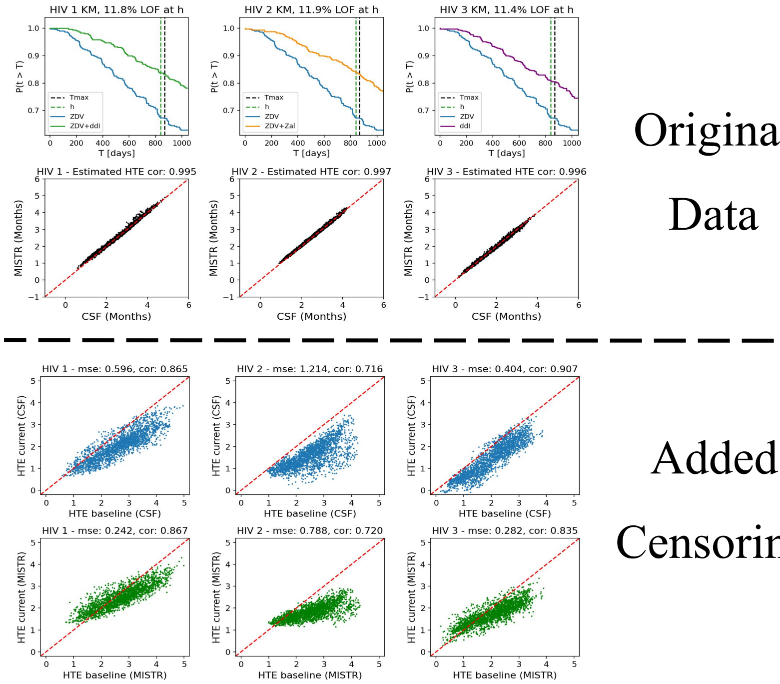
Which treatment results in longer survival with no AIDS progression?

$N = 2{,}139$ HIV-infected patients

$p = 12$ covariates

4 treatment groups

We analyze the original data and the data with added censoring.



Original Data

Added Censoring

TECHNION Israel Institute of Technology

TEL AVIV UNIVERSITY

Use-Case: HIV Clinical Trial

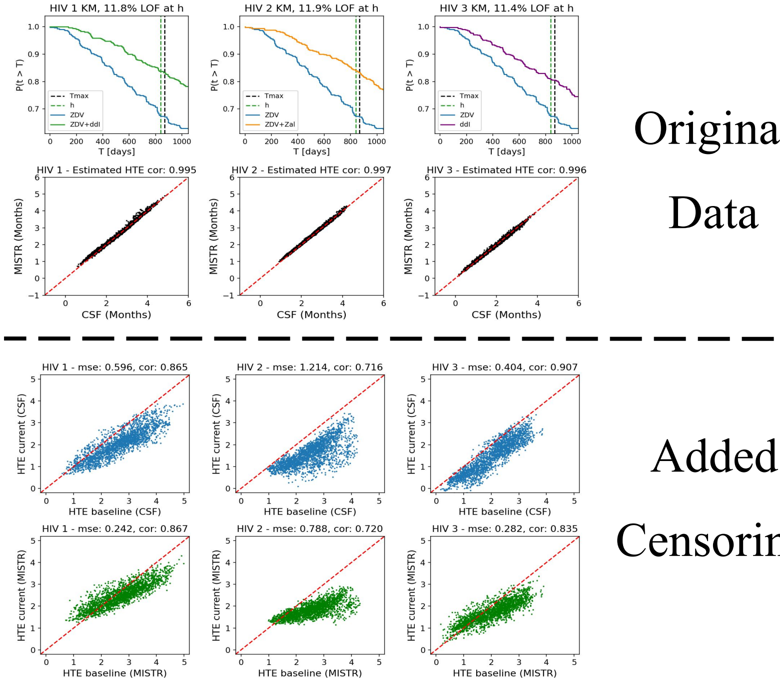
Which treatment results in longer survival with no AIDS progression?

$N = 2{,}139$ HIV-infected patients

$p = 12$ covariates

4 treatment groups

- **Similar results when censoring rate is low.**
- **MISTR outperforms CSF at high censoring rates.**



Original Data

Added Censoring

Summary

This work presents **MISTR - a novel non-parametric approach for estimating HTE and its variance in TTE data**:

Summary

This work presents **MISTR - a novel non-parametric approach for estimating HTE and its variance in TTE data**:

- **MISTR expands the range of cases** that can be effectively addressed.

Summary

This work presents **MISTR - a novel non-parametric approach for estimating HTE and its variance in TTE data**:

- **MISTR expands the range of cases** that can be effectively addressed.
- **MISTR outperforms other existing approaches**, especially in heavy censoring rates.

Summary

This work presents **MISTR - a novel non-parametric approach for estimating HTE and its variance in TTE data**:

- **MISTR expands the range of cases** that can be effectively addressed.
- **MISTR outperforms other existing approaches**, especially in heavy censoring rates.
- **MISTR can incorporate an instrumental variable.**

Summary

This work presents **MISTR - a novel non-parametric approach for estimating HTE and its variance in TTE data**:

- **MISTR expands the range of cases** that can be effectively addressed.
- **MISTR outperforms other existing approaches**, especially in heavy censoring rates.
- **MISTR can incorporate an instrumental variable.**

Additional details, simulations, and a real-world use case with IV are included in the paper.

Thank You!

Tomer Meir
Technion - Israel Institute of Technology

tomer1812@gmail.com

Prof. Uri Shalit
Department of Statistics and Operations Research, Tel Aviv University

urishalit@tauex.tau.ac.il

Prof. Malka Gorfine
Department of Statistics and Operations Research, Tel Aviv University

gorfinem@tauex.tau.ac.il

Further details are available at:

Full paper: https://arxiv.org/abs/2502.01575

Project Github: https://github.com/tomer1812/mistr