Multi-Study Causal Forest (MCF): Improving the estimation of heterogeneous treatment effects using auxiliary data

Presenter: Ashwini Venkatasubramaniam, GSK
(Co-author: Julian Wolfson, UMN)

PSI Conference, Belfast
16/06/2026

Motivation

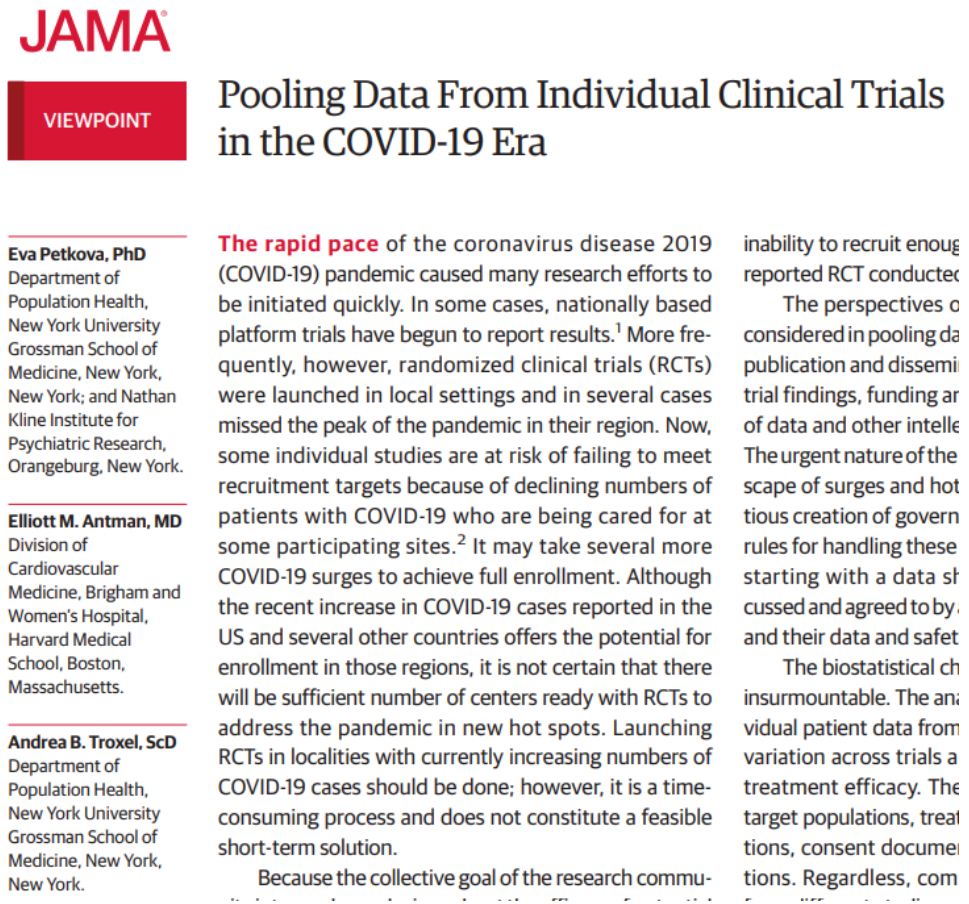
❑ One size does not fit all:

- Moving away from the assumption that a treatment affects all patients in the same way.
- Account for patient characteristics driving differences in response to treatment.

❑ Use of multiple studies:

- Increased availability of large multi-site datasets
- Legislative guidance is available

Landscape is ripe for leveraging data to estimate causal effects.



JAMA

VIEWPOINT

Pooling Data From Individual Clinical Trials in the COVID-19 Era

Eva Petkova, PhD
Department of Population Health, New York University Grossman School of Medicine, New York, New York; and Nathan Kline Institute for Psychiatric Research, Orangeburg, New York.

Elliott M. Antman, MD
Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts.

Andrea B. Troxel, ScD
Department of Population Health, New York University Grossman School of Medicine, New York, New York.

The rapid pace of the coronavirus disease 2019 (COVID-19) pandemic caused many research efforts to be initiated quickly. In some cases, nationally based platform trials have begun to report results.[1] More frequently, however, randomized clinical trials (RCTs) were launched in local settings and in several cases missed the peak of the pandemic in their region. Now, some individual studies are at risk of failing to meet recruitment targets because of declining numbers of patients with COVID-19 who are being cared for at some participating sites.[2] It may take several more COVID-19 surges to achieve full enrollment. Although the recent increase in COVID-19 cases reported in the US and several other countries offers the potential for enrollment in those regions, it is not certain that there will be sufficient number of centers ready with RCTs to address the pandemic in new hot spots. Launching RCTs in localities with currently increasing numbers of COVID-19 cases should be done; however, it is a time-consuming process and does not constitute a feasible short-term solution.

Because the collective goal of the research commu-

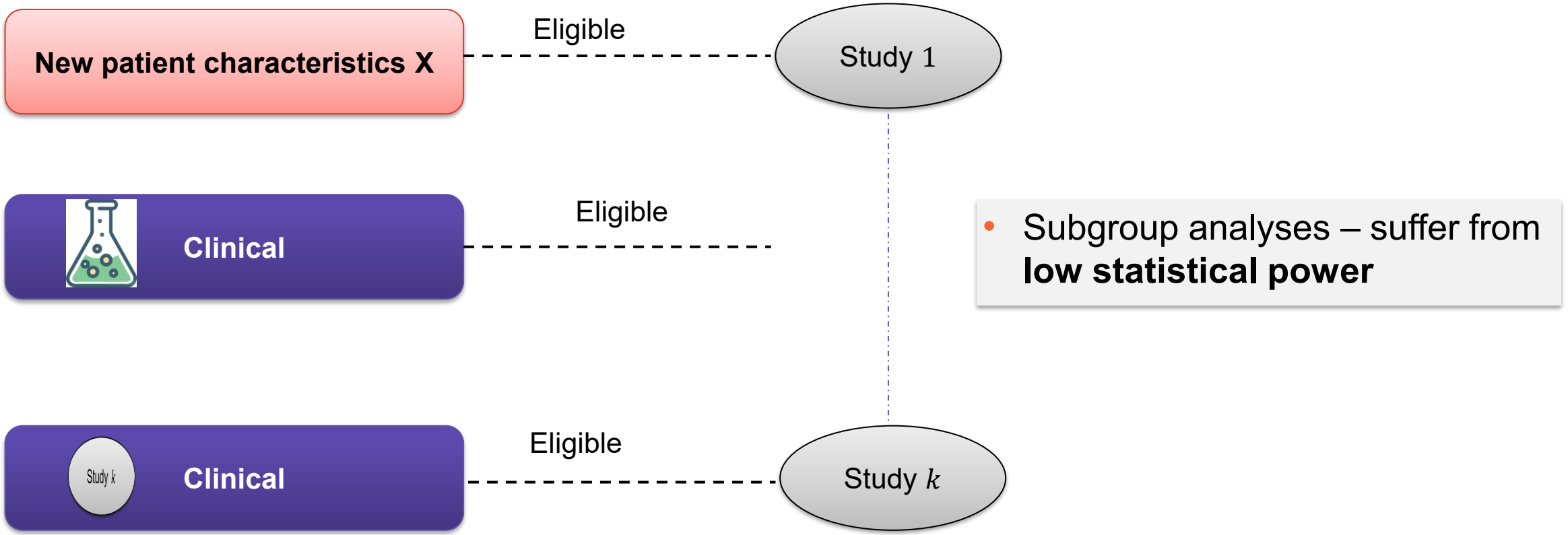
inability to recruit enoug
reported RCT conducte

The perspectives o
considered in pooling da
publication and dissemi
trial findings, funding ar
of data and other intelle
The urgent nature of the
scape of surges and hot
tious creation of govern
rules for handling these
starting with a data sh
cussed and agreed to by
and their data and safet

The biostatistical ch
insurmountable. The ana
vidual patient data from
variation across trials a
treatment efficacy. The
target populations, treat
tions, consent docume
tions. Regardless, com

How is the use of multiple studies relevant for heterogeneous treatment effect estimation?

Significantly larger sample sizes are required to detect treatment effect heterogeneity (TEH) than what is needed to detect an average treatment effect (ATE)



- Subgroup analyses – suffer from **low statistical power**

Exchangeability

Many existing borrowing approaches rely on assumption of marginal exchangeability for direct borrowing of data

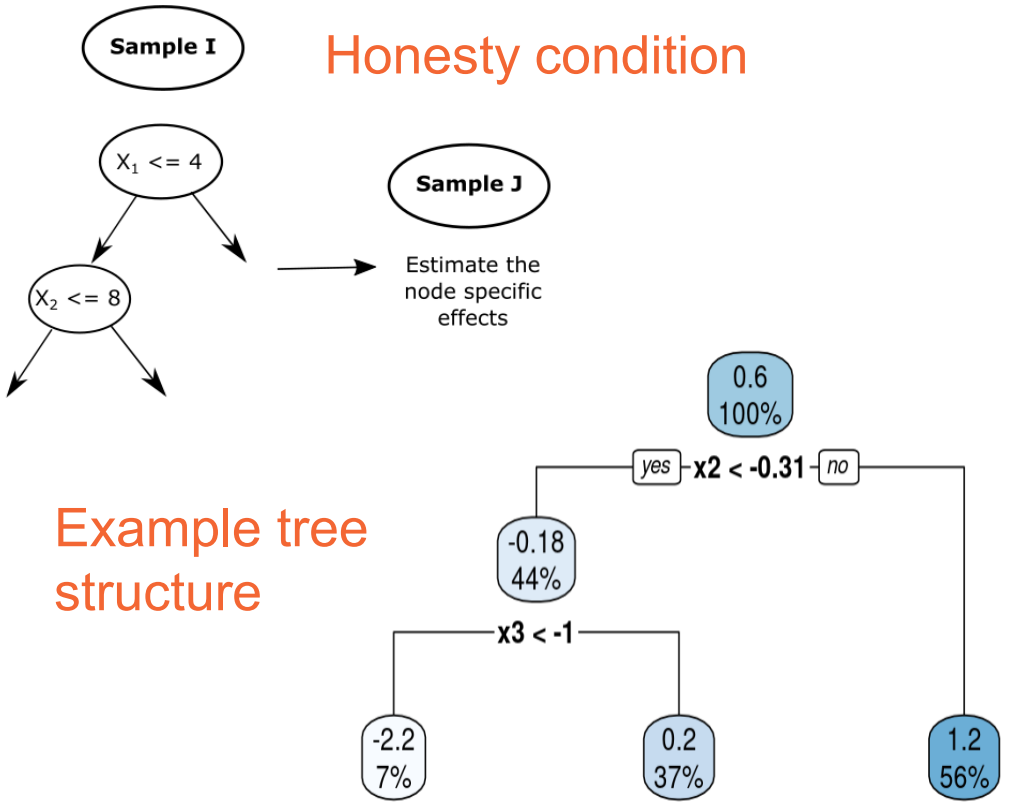
1. Bayesian perspective: Extent of borrowing is determined by evaluating the consistency between the **primary** and **auxiliary** data sources.
2. Exchangeability may not hold in the **presence of treatment effect heterogeneity.**
    - BDB methods: Borrow strength for the **average treatment effect (ATE)**.
        - **Marginal** exchangeability: ATE between historical studies and the primary data are similar enough to be combined
        - **Covariate-adjusted** exchangeability: Studies are exchangeable after conditioning on a set of observed covariates.
    - A **linear relationship** is assumed!

Flexible modelling of treatment effect heterogeneity: Single-study methods

Many nonparametric methods for **single study scenarios** have been developed including R Learner, DR-Learner, Causal Forest and Bayesian Causal Forest.

Causal Forest:

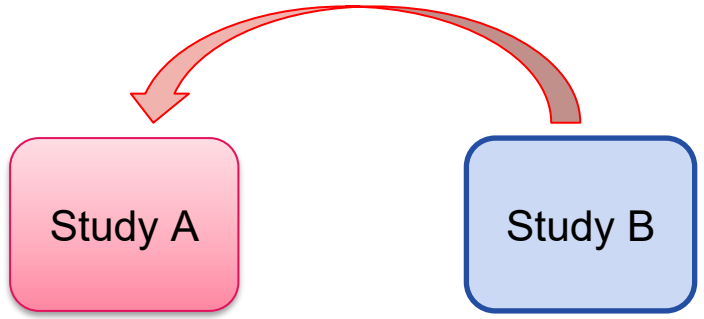
- Use of **Honest Trees**: Separate sample for building tree structure and estimating the treatment effect
- Explicitly optimise for **treatment effect heterogeneity**
- Generate estimates of **Conditional average treatment effects (CATEs).**



Goal: Borrow from auxiliary data source in the presence of varying sources of treatment effect heterogeneity

Includes both within study and between study heterogeneity

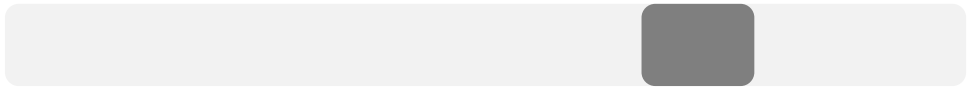
1. Type 1: **Within study heterogeneity:**
   - Differences *within* a given study
2. Type 2: **Between study heterogeneity:**
   - Differences *across* various studies
3. Seek **to improve the accuracy of CATE estimates** by using data from Study B:
   - If Study B is **very similar** to Study A, greater information is borrowed.
   - Else the information in Study B is given lower weight



All of Study A – use as the base trial.

Defining Exchangeability

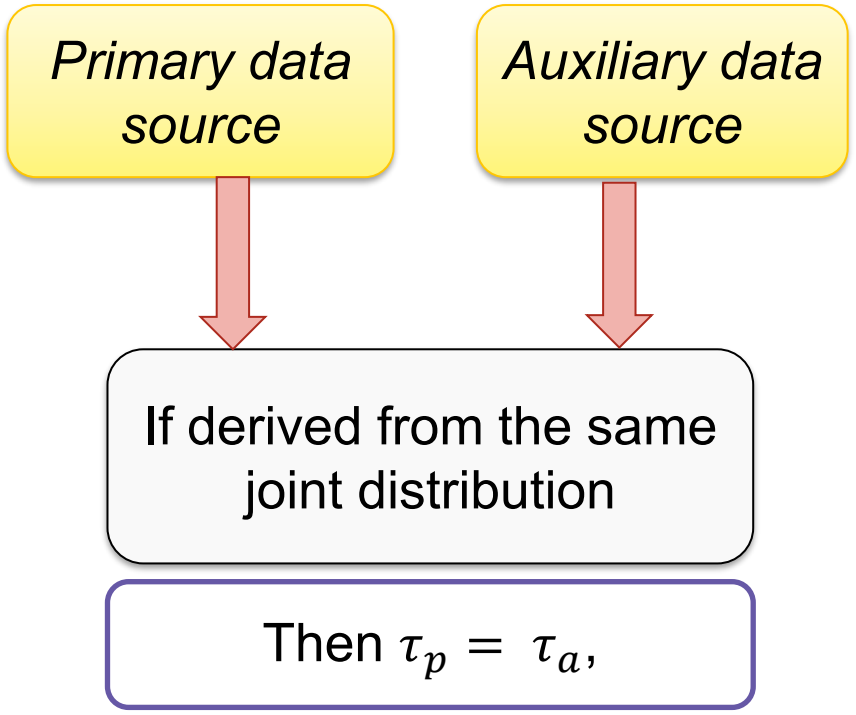
In the presence of treatment effect heterogeneity: Between study heterogeneity



Range of variation between studies

Perfectly homogeneous

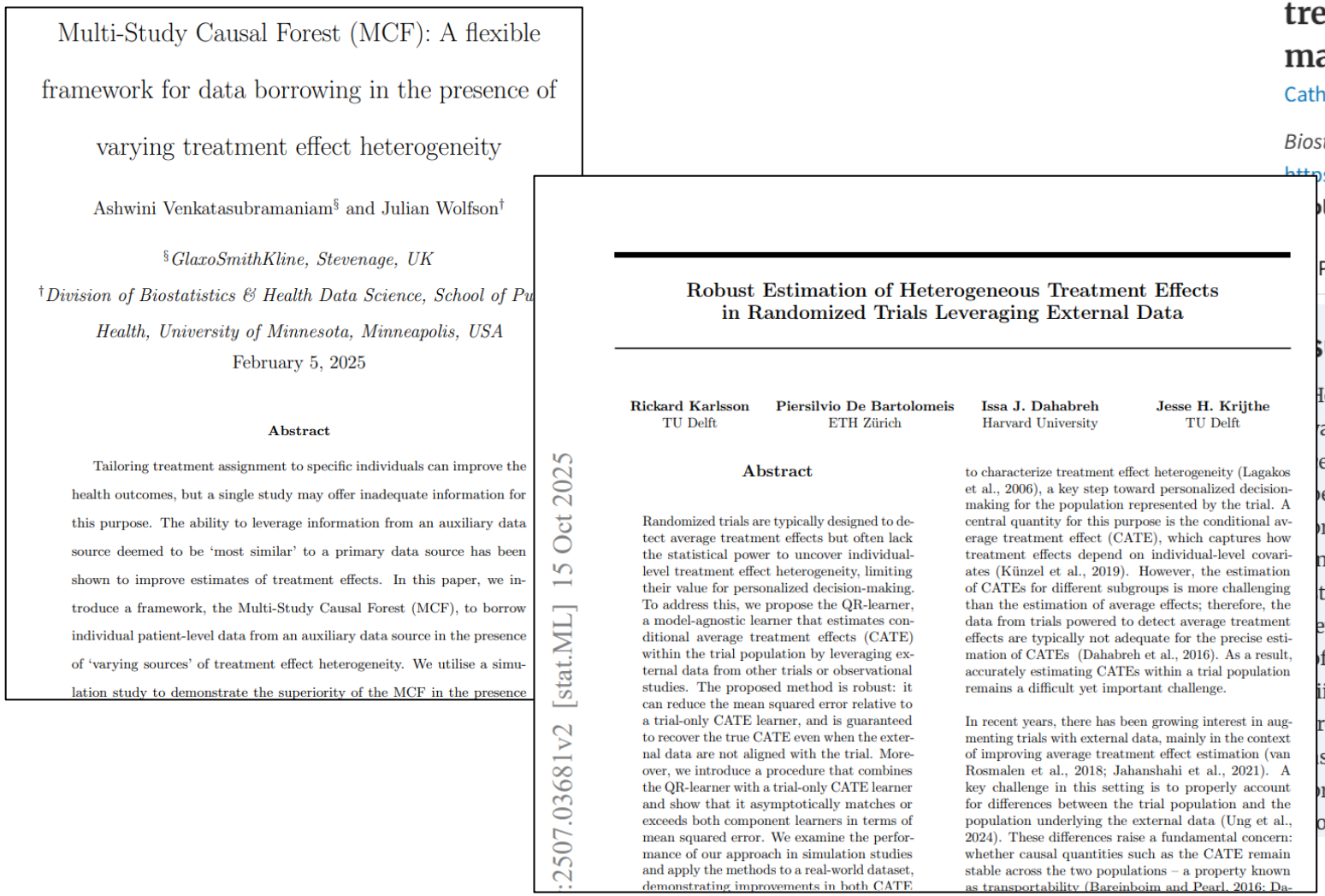
Substantial heterogeneity



If CATEs in the primary and auxiliary data sources are exchangeable – borrow from the auxiliary data source

Growing interest in such methods in the literature

Data borrowing or pooling across multiple studies



Multi-Study Causal Forest (MCF): A flexible framework for data borrowing in the presence of varying treatment effect heterogeneity

Ashwini Venkatasubramaniam[§] and Julian Wolfson[†]

[§]*GlaxoSmithKline, Stevenage, UK*

[†]*Division of Biostatistics & Health Data Science, School of Pu*

Health, University of Minnesota, Minneapolis, USA

February 5, 2025

Abstract

Tailoring treatment assignment to specific individuals can improve the health outcomes, but a single study may offer inadequate information for this purpose. The ability to leverage information from an auxiliary data source deemed to be 'most similar' to a primary data source has been shown to improve estimates of treatment effects. In this paper, we introduce a framework, the Multi-Study Causal Forest (MCF), to borrow individual patient-level data from an auxiliary data source in the presence of 'varying sources' of treatment effect heterogeneity. We utilise a simulation study to demonstrate the superiority of the MCF in the presence

Robust Estimation of Heterogeneous Treatment Effects in Randomized Trials Leveraging External Data

Rickard Karlsson TU Delft

Piersilvio De Bartolomeis ETH Zürich

Issa J. Dahabreh Harvard University

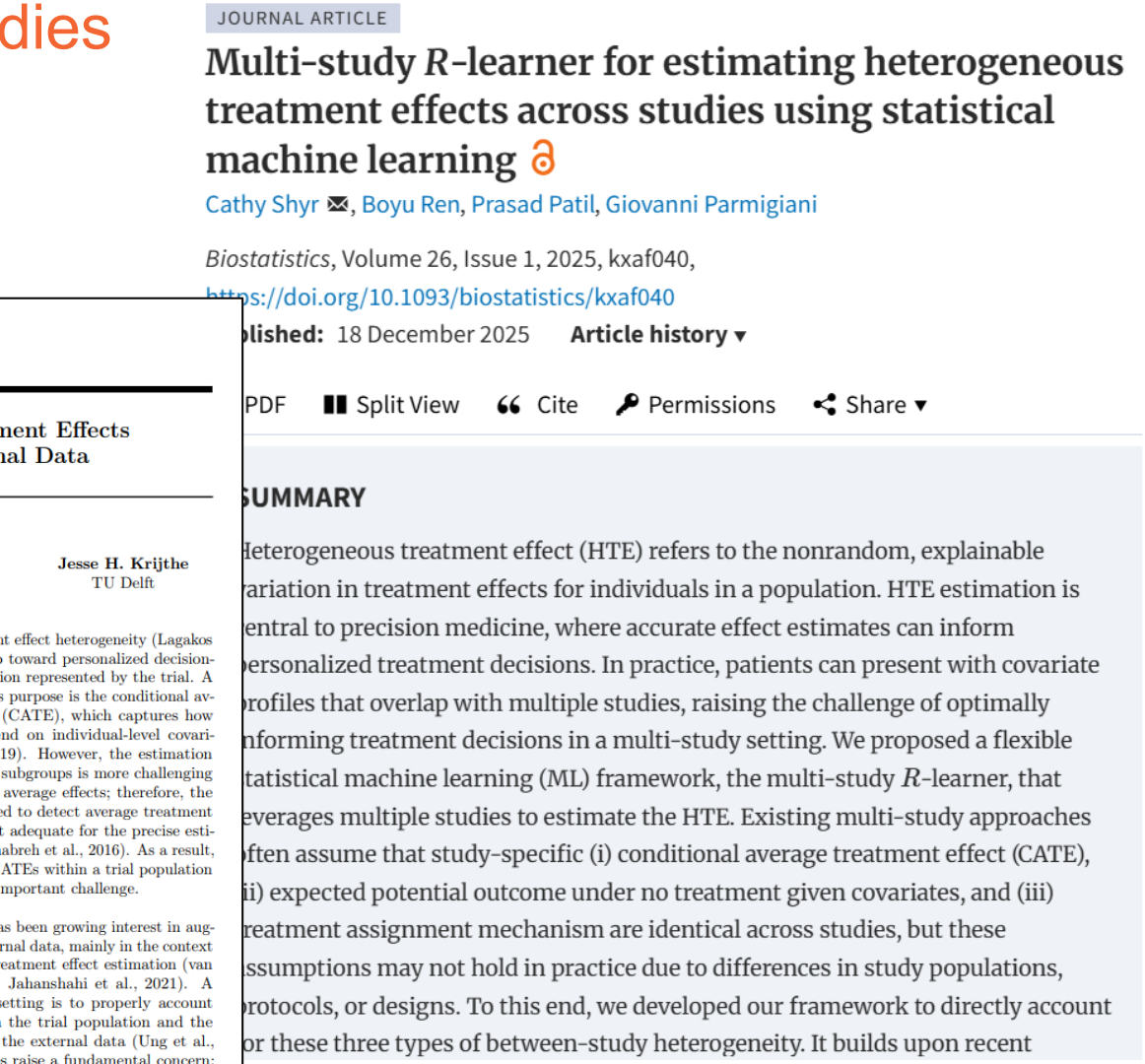
Jesse H. Krijthe TU Delft

Abstract

Randomized trials are typically designed to detect average treatment effects but often lack the statistical power to uncover individual-level treatment effect heterogeneity, limiting their value for personalized decision-making. To address this, we propose the QR-learner, a model-agnostic learner that estimates conditional average treatment effects (CATE) within the trial population by leveraging external data from other trials or observational studies. The proposed method is robust: it can reduce the mean squared error relative to a trial-only CATE learner, and is guaranteed to recover the true CATE even when the external data are not aligned with the trial. Moreover, we introduce a procedure that combines the QR-learner with a trial-only CATE learner and show that it asymptotically matches or exceeds both component learners in terms of mean squared error. We examine the performance of our approach in simulation studies and apply the methods to a real-world dataset, demonstrating improvements in both CATE

to characterize treatment effect heterogeneity (Lagakos et al., 2006), a key step toward personalized decision-making for the population represented by the trial. A central quantity for this purpose is the conditional average treatment effect (CATE), which captures how treatment effects depend on individual-level covariates (Künzel et al., 2019). However, the estimation of CATEs for different subgroups is more challenging than the estimation of average effects; therefore, the data from trials powered to detect average treatment effects are typically not adequate for the precise estimation of CATEs (Dahabreh et al., 2016). As a result, accurately estimating CATEs within a trial population remains a difficult yet important challenge.

In recent years, there has been growing interest in augmenting trials with external data, mainly in the context of improving average treatment effect estimation (van Rosmalen et al., 2018; Jahanshahi et al., 2021). A key challenge in this setting is to properly account for differences between the trial population and the population underlying the external data (Ung et al., 2024). These differences raise a fundamental concern: whether causal quantities such as the CATE remain stable across the two populations – a property known as transportability (Bareinboim and Pearl, 2016; Da-



JOURNAL ARTICLE

Multi-study *R*-learner for estimating heterogeneous treatment effects across studies using statistical machine learning

Cathy Shyr ✉, Boyu Ren, Prasad Patil, Giovanni Parmigiani

Biostatistics, Volume 26, Issue 1, 2025, kxaf040,

https://doi.org/10.1093/biostatistics/kxaf040

lished: 18 December 2025 **Article history ▾**

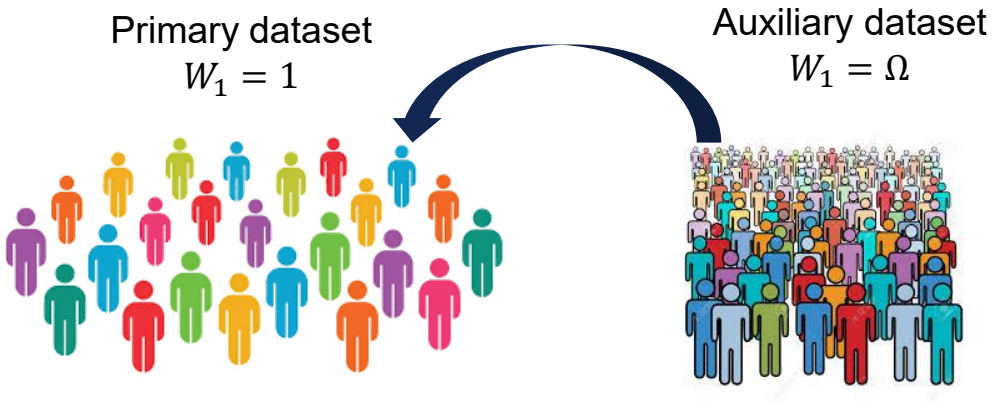
PDF Split View Cite Permissions Share ▾

SUMMARY

Heterogeneous treatment effect (HTE) refers to the nonrandom, explainable variation in treatment effects for individuals in a population. HTE estimation is central to precision medicine, where accurate effect estimates can inform personalized treatment decisions. In practice, patients can present with covariate profiles that overlap with multiple studies, raising the challenge of optimally informing treatment decisions in a multi-study setting. We proposed a flexible statistical machine learning (ML) framework, the multi-study *R*-learner, that leverages multiple studies to estimate the HTE. Existing multi-study approaches often assume that study-specific (i) conditional average treatment effect (CATE), (ii) expected potential outcome under no treatment given covariates, and (iii) treatment assignment mechanism are identical across studies, but these assumptions may not hold in practice due to differences in study populations, protocols, or designs. To this end, we developed our framework to directly account for these three types of between-study heterogeneity. It builds upon recent

Multi- Study Causal Forest - Algorithm

1. Fit separate forests over the Primary dataset CF_1, the Auxiliary dataset CF_2 and the Combined dataset CF_3.
2. Evaluate **similarity** between the CATEs predicted over primary data
3. Compute propensity scores to **weigh individual patients** in the auxiliary dataset – most similar to the Primary dataset.
4. MCF: Use the **product of the above two weights** $W_3 = W_1 * W_2$



Overall similarity is evaluated using the **correlation coefficient** over CATEs

Individual patients are weighted using **propensity scores**

Translate to a simulation study to enable comparison

Goal: How **best to borrow** from the auxiliary dataset – in a manner that the information used is closest to the primary dataset?

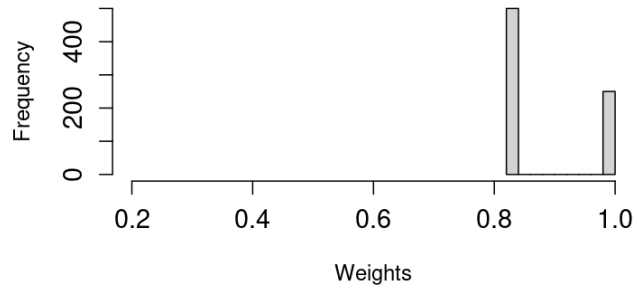
- Split the Primary dataset (n = 500):
  - Training (50%) and Test dataset (50%)
- Auxiliary dataset (n = 500)
- Evaluate performance of the method – using RMSE

Simulation Parameters

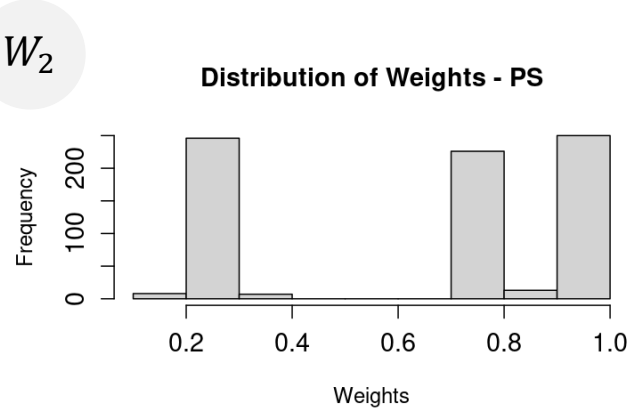
- Number of covariates: 10
- Differences in **underlying CATE functions**.
- **Different propensity score models** between the primary and auxiliary datasets
  - Primary data: Driven by X_1; Auxiliary data: Driven by $X_2^2 + X_3 + X_4$
- **Presence of correlation** among covariates – 0.2

Weights computed for the Multi-study Causal Forest

W_1

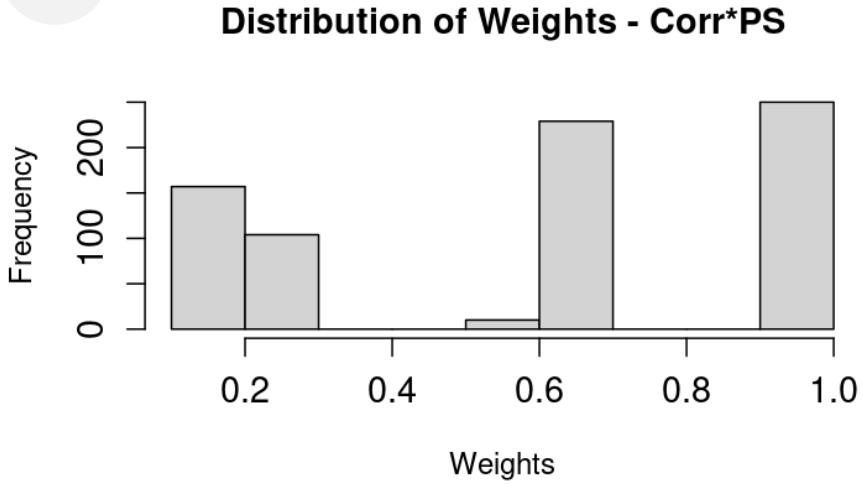


W_2



Multi study Causal Forest

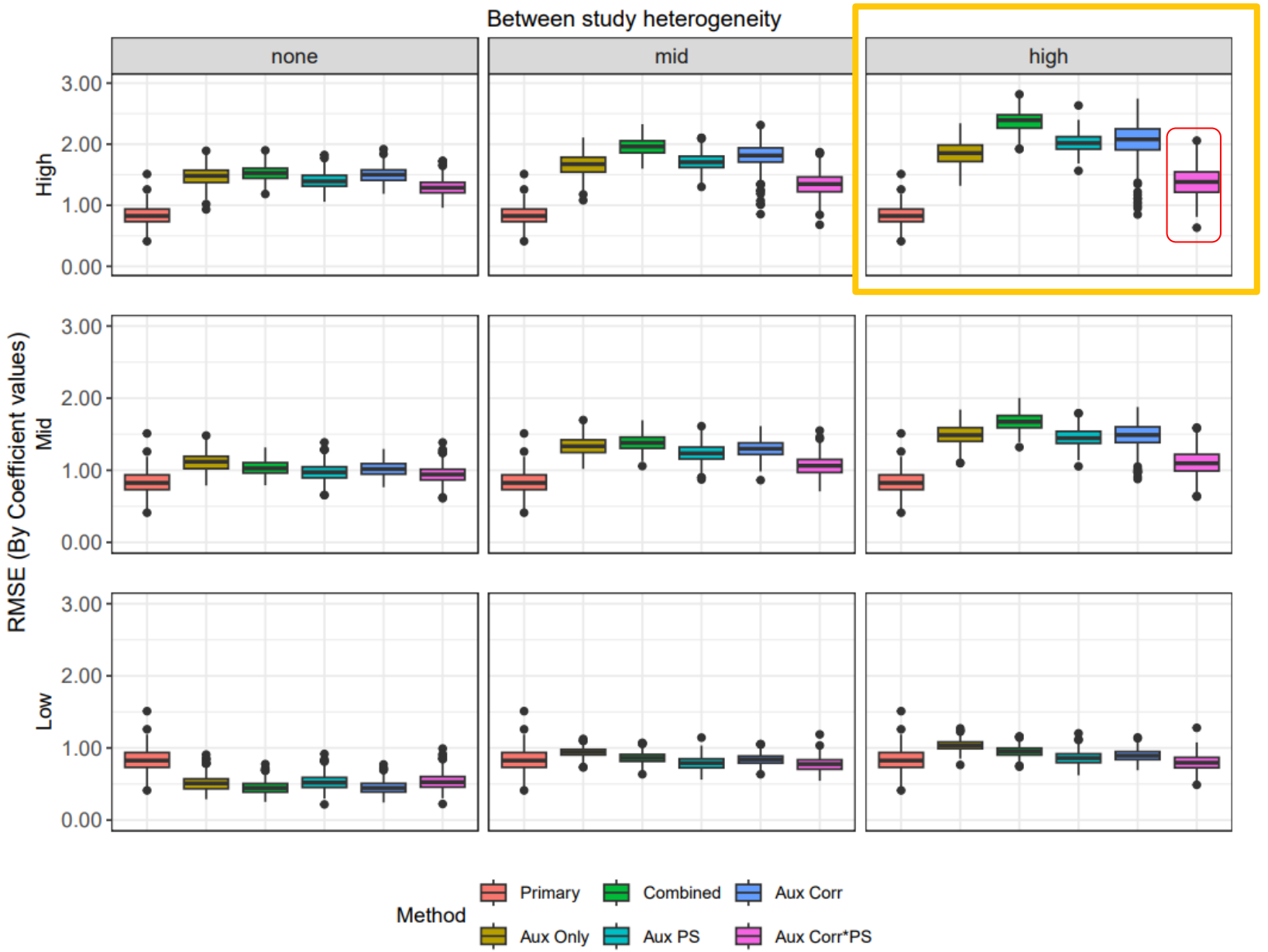
W_3



Key question: Is it possible to borrow from an auxiliary data source – such that it is most “similar” to a “target” data source?

Results

- Same parameters for primary and auxiliary dataset
  - Minimal difference in RMSE
- Different parameters for primary and auxiliary dataset
  - MCF corresponds to RMSE - closest to primary dataset.



Presence of correlation = 0.2

Conclusions

Summary:

- The MCF is a flexible approach and is able to identify:
    1. When CATE functions are exchangeable
    2. When there are differences between CATE functions
- Determines the extent to which data should be borrowed from an auxiliary data source.

Future steps:

- Discuss scenarios where the data sources contain different number of covariates or in the presence of missingness.
- Offer solutions when data is only available on a single arm in the auxiliary study.

gsk.com