Flexible Machine Learning Estimation of Conditional Average Treatment Effects

A Blessing and a Curse.

Richard A. J. Post[1]

[1]Department of Mathematics and Computer Science
Eindhoven University of Technology

Amsterdam Causality Meeting, April 2024

Based on: R.A.J. Post, I.L. van den Heuvel, M. Petkovic and E.R. van den Heuvel, Epidemiology, 35(1), p 32-40, 2024.

TU/e

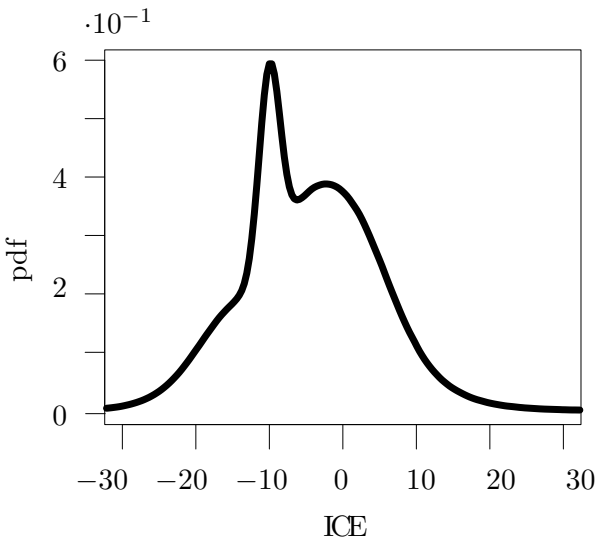
: r.a.j.post@tue.nl

Overview

TU/e

Causal effect heterogeneity

- Causal effects may seriously vary across individuals.



- There is the hope that the growing availability of health data presents an opportunity to make precision medicine a clinical reality (Kosorok and Laber, 2019).

TU/e

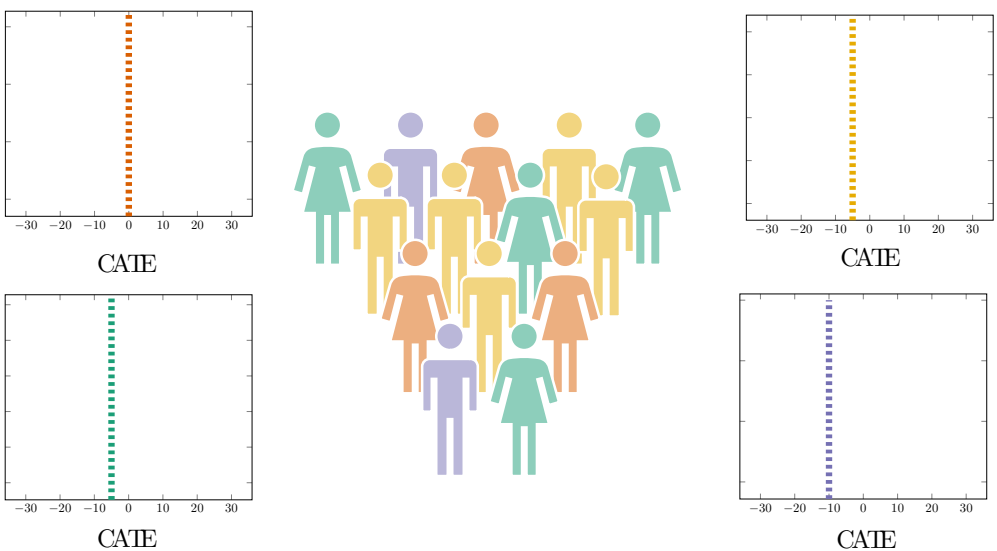
Conditional Average Treatment Effect

- The individual treatment effect (ITE) cannot be observed due to the fundamental problem of causal inference (Holland, 1986)
  - → For each individual, we can only observe the outcome under one level of the treatment.
- To study causal effect heterogeneity, the conditional average treatment effect (CATE) is often estimated (Hernán and Robins, 2020, Chapter 4).

TU/e

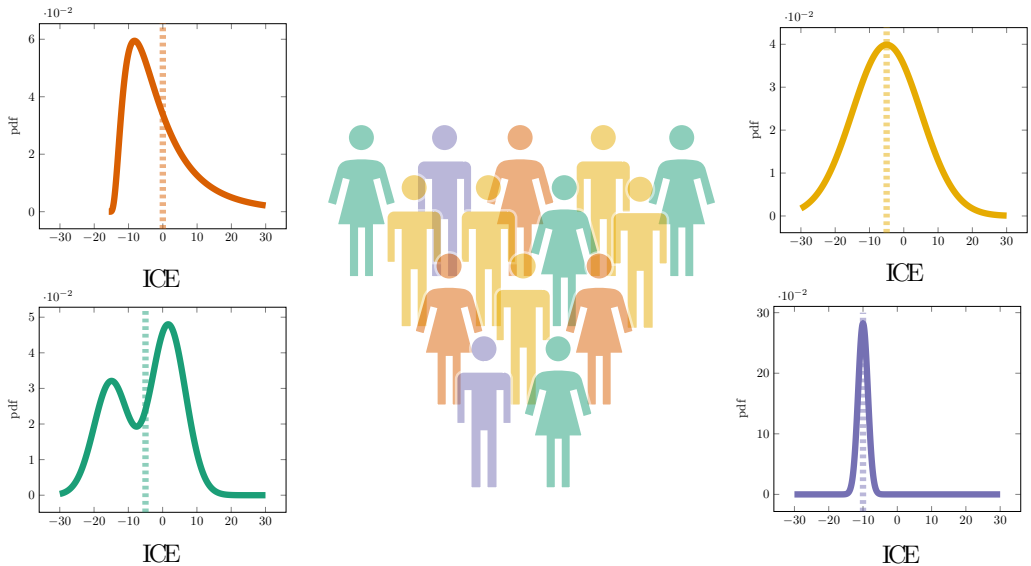
Conditional Average Treatment Effect

- Machine learning methods have been adapted to estimate CATEs to mimic individual causal effects (Caron et al., 2022).
  - → The CATEs may be described by complex functions of feature levels $\boldsymbol{x}$ (=blessing).



ITE vs CATE

- The individualized CATEs might not be very representative for an ITE in case of remaining effect heterogeneity (Hand, 1992; Kravitz et al., 2004; Greenland et al., 2019).



- The individualized CATE is sometimes perceived as equivalent to the ITE (see, e.g. Lu et al. (2018))
  → Assumes conditional between-individual homogeneity (=curse).

Potential outcomes

- Factual outcomes of individual i
  - → Y_i equals the outcome
  - → A_i equals the treatment assignment

- Probability distributions of factual and counterfactual outcomes are defined in terms of the potential outcome framework (Rubin, 1974).
  - → Y_i^a equals the potential outcome under an intervention on the treatment to level a (counterfactual when $A_i \neq a$, and equivalent to $Y \mid do(A = a)$).
  - → For an arbitrary i: a draw from the outcomes in a universe where everyone is exposed to a.
  - → $Y_i^1 - Y_i^0$ equals the individual treatment effect (ITE) of the binary exposure
  - → The joint distribution of (Y_i^1, Y_i^0) cannot be studied directly as a result of *the fundamental problem of causal inference* (Holland, 1986).

Causal assumptions

Assumption 1

Causal consistency

$$Y_i = Y_i^{A_i}$$

Assumption 2

Conditional exchangeability

$$A \perp\!\!\!\perp Y^0, Y^1 \mid \boldsymbol{X}$$

- The set of features $\boldsymbol{X}$ will contain modifiers (as we are interested in causal effect heterogeneity) and confounders that are necessary to obtain independence.
- For a feature L that is only a confounder but not a modifier (on the additive scale) $\forall l$: $\mathbb{E}[Y^1 - Y^0 \mid L = l, \widetilde{\boldsymbol{X}} = \widetilde{\boldsymbol{x}}] = \mathbb{E}[Y^1 - Y^0 \mid \widetilde{\boldsymbol{X}} = \widetilde{\boldsymbol{x}}]$

Assumption 3

Positivity

$$\forall \boldsymbol{x}\colon\ 0 < \mathbb{P}(A = 1 \mid \boldsymbol{X}{=}\boldsymbol{x}) < 1$$

Observed outcome distribution

- By causal consistency, we use the parameterization

$$Y_i = Y_i^0 + b_i A_i, \tag{1}$$

where b_i is the ITE of individual i, so that $Y_i^1 = Y_i^0 + b_i$.

→ The conditional mean of b_i given features $\boldsymbol{X}_i$ equals the CATE $\tau(\boldsymbol{X}_i)$, where $\tau(\boldsymbol{x}) = \mathbb{E}[Y^1 - Y^0 \mid \boldsymbol{X} = \boldsymbol{x}]$.

- For our purposes, it helps to rewrite Equation (1) as

$$Y_i = \theta_0(\boldsymbol{X}_i) + N_{Yi} + (\tau(\boldsymbol{X}_i) + U_{1i})\, A_i, \tag{2}$$

→ The ITE can be divided into $\tau(\boldsymbol{X}_i)$ and the zero-mean individual deviation from the CATE that is referred to as U_{1i}.

→ The individual Y^0 has also been rewritten as the sum of its conditional expectation $\theta_0(\boldsymbol{x}) = \mathbb{E}[Y_i^0 \mid \boldsymbol{X}_i{=}\boldsymbol{x}]$ and zero mean deviation from this expectation N_{Yi}.

Running example: Framingham Heart Study

- We simulated data based on the Framingham Heart Study (FHS) and focus on the heterogeneity in the effect of non-alcoholic fatty liver disease on a clinical precursor to heart failure (Chiu et al., 2020).
- We will simulate the following cause-effect relations

$$
\begin{aligned}
A_i &= \mathbb{1}\left\{ \frac{\exp(\alpha_0 + \alpha_{\text{SBP}} X_{\text{SBP},i} + \alpha_{\text{sex}} X_{\text{sex},i})}{1 + \exp(\alpha_0 + \alpha_{\text{SBP}} X_{\text{SBP},i} + \alpha_{\text{sex}} X_{\text{sex},i})} > N_{Ai} \right\} \qquad (3) \\
Y_i^0 &= \beta_0 + \beta_{\text{sex}} X_{\text{sex},i} + \beta_{\text{SBP}} X_{\text{SBP},i} + N_{Yi} \\
Y_i^1 &= Y_i^0 + (\boldsymbol{\tau_0} + \boldsymbol{\tau}_{\text{sex}} \boldsymbol{X}_{\text{sex},\boldsymbol{i}} + \boldsymbol{\tau}_{\text{SBP}} \boldsymbol{X}_{\text{SBP},\boldsymbol{i}} + U_{1i}),
\end{aligned}
$$

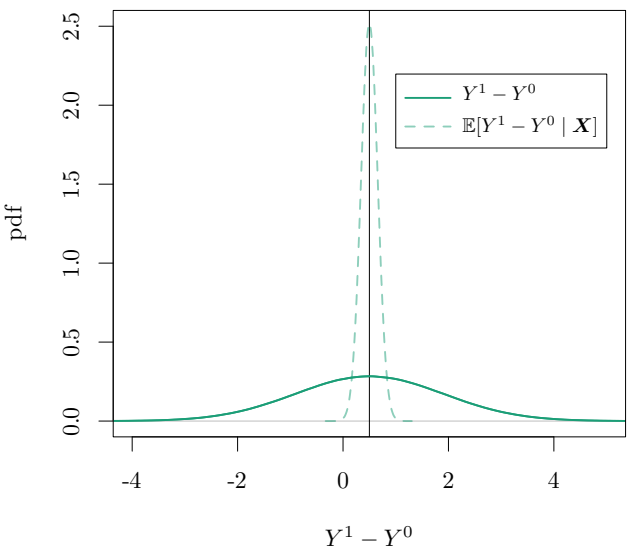
→ $\boldsymbol{U_{1i}} \perp\!\!\!\perp \boldsymbol{N_{Yi}}$, $X_{\text{sex},i} \sim \text{Ber}(p)$, $X_{\text{SBP},i} \sim \mathcal{N}(0,1)$, $U_{1i} \sim \mathcal{N}(0, \sigma_1^2)$, $N_{Yi} \sim \mathcal{N}(0, \sigma_0^2)$, and $N_{Ai} \sim \text{Uni}[0,1]$.

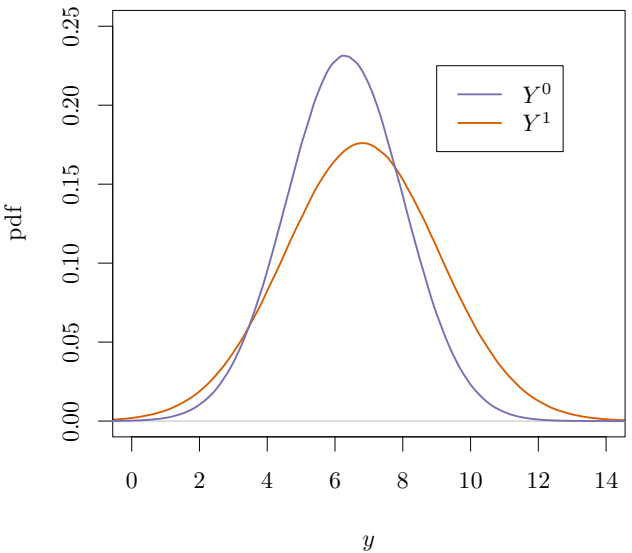
→ There is **no unmeasured confounding**, i.e. $N_{Ai} \perp\!\!\!\perp N_{Yi}, U_{1i}$ so that $A_i \perp\!\!\!\perp (Y_i^1, Y_i^0) \mid X_{\text{sex},i}, X_{\text{SBP},i}$.

→ X_0, is a measured variable correlated with the level of the individual modifier U_1, $(U_1, X_0)^T \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma_1^2 & \rho\delta\sigma_1^2 \\ \rho\delta\sigma_1^2 & \delta^2\sigma_1^2 \end{pmatrix}\right)$.

Running example

- Parameters are derived from the subset of the FHS participants ($n = 2356$) as used by Chiu et al. (2020).
- The ATE $\mathbb{E}[Y^1 - Y^0] = 0.5$, SD of the ITE $\sqrt{\text{var}(Y^1 - Y^0)} = 1.41$ and positive effect probability (PEP) $\mathbb{P}(Y^1 - Y^0 > 0) = 0.64$





CATE estimation: R-learner

- Since the actual causal effects are not observed, defining an appropriate loss function is not straightforward, and using machine learning methods to study causal effect heterogeneity is challenging (Athey and Imbens, 2016).
- R-learners (Nie and Wager, 2020) also known as 'double machine learning' algorithms (Chernozhukov et al., 2018; Bach et al., 2022) appropriately adjust for measured confounding.

- R-learners focus on the relation between normalized outcome and treatment assignment as originally used by Robinson (1988)

$$Y_i - m(\boldsymbol{X}_i) = (A_i - e(\boldsymbol{X}_i))\,\widetilde{\tau}(\boldsymbol{X}_i) + \widetilde{N}_{Yi}, \tag{4}$$

where $m(\boldsymbol{x}) = \mathbb{E}[Y \mid \boldsymbol{X}{=}\boldsymbol{x}]$, $e(\boldsymbol{x}) = \mathbb{E}[A \mid \boldsymbol{X}{=}\boldsymbol{x}]$ and
$\forall \boldsymbol{x}, a\colon \mathbb{E}\left[\widetilde{N}_{Yi} \mid A_i{=}a, \boldsymbol{X}_i{=}\boldsymbol{x}\right] = 0.$

CATE estimation: R-learner

Using parameterization (2) one can derive, for $\boldsymbol{X}_i = \boldsymbol{x}$,

$$
\begin{aligned}
&Y_i - m(\boldsymbol{x}) = \\
&(\theta_0(\boldsymbol{x}) + N_{Yi} + (\tau(\boldsymbol{x}) + U_{1i})\, A_i) - (\theta_0(\boldsymbol{x}) + \mathbb{E}\left[(\tau(\boldsymbol{x}) + U_{1i})\, A_i \mid \boldsymbol{X}_i{=}\boldsymbol{x}\right]) \\
&= (A_i - e(\boldsymbol{x}))\, \tau(\boldsymbol{x}) + (A_i U_{1i} - \mathbb{E}\left[A_i U_{1i} \mid \boldsymbol{X}_i{=}\boldsymbol{x}\right] + N_{Yi})\,.
\end{aligned}
\tag{5}
$$

- If $\mathbb{E}[U_1 + N_Y \mid A{=}1, \boldsymbol{X}{=}\boldsymbol{x}] \neq \mathbb{E}[N_Y \mid A{=}0, \boldsymbol{X}{=}\boldsymbol{x}]$, then $\widetilde{\tau}(\boldsymbol{x}) \neq \tau(\boldsymbol{x})$.
- In absence of unmeasured confounding, i.e. $A \perp\!\!\!\perp N_Y, U_1$, for $\boldsymbol{X}_i = \boldsymbol{x}$, $\widetilde{\tau}(\boldsymbol{x}) = \tau(\boldsymbol{x})$!

CATE estimation: CRF algorithm

The generalized random forest implementation of the CRF as implemented in the R-package `grf` (Athey et al., 2019)

1. Predict $m(\boldsymbol{x}_i)$ and $e(\boldsymbol{x}_i)$ for each i by fitting two separate regression forests consisting of honest trees, each tree is fitted on a random subsample of half the sample size.
   - → A Honest tree uses different subsamples for constructing the tree and making predictions (Wager and Athey, 2018).
   - → Out-of-bag predictions $\hat{m}^{-i}(\boldsymbol{x}_i)$ and $\hat{e}^{-i}(\boldsymbol{x}_i)$ are only based on those trees that did not use individual i for training.
2. Create the centered outcomes $\widetilde{Y}_i = Y_i - \hat{m}^{-i}(\boldsymbol{x}_i)$ and $\widetilde{A}_i = A_i - \hat{e}^{-i}(\boldsymbol{x}_i)$
3. Grow B trees for $\widetilde{Y}$
   - → For each tree, a random subsample $\mathcal{I}$ of the available data is taken
   - → Subsequently, the subsample is again divided into $\mathcal{J}_1$ and $\mathcal{J}_2$. The honest decision tree is only fitted on $\mathcal{J}_1$ and optimizes the heterogeneity in the effect of $\widetilde{A}$ on $\widetilde{Y}$ between the different nodes using gradient-based approximations of treatment-effect estimates in candidate children notes, see Athey et al. (2019, Section 2.3).

4 For a set of for a new set of features $\boldsymbol{x}$, similarity weights are estimated per tree b, α_{bj}, and are non-zero (and equal) for those elements of $\mathcal{J}_2$ that fall in the same leaf as $\boldsymbol{x}$, and are averaged over all trees to obtain α_j.

→ When $\exists i$: $\boldsymbol{x}_i{=}\boldsymbol{x}$ then $\forall j{\neq}i$: $\alpha_j(\boldsymbol{x}_i)$ are created by averaging similarity weights over trees where $i \notin \mathcal{J}_1$.

5 $\widetilde{\tau}(\boldsymbol{x})$ is estimated as

$$\hat{\widetilde{\tau}}(\boldsymbol{x}) = \frac{\sum_{i=1}^{n} \alpha_i(\boldsymbol{x}) \left(Y_i - \hat{m}^{-i}(\boldsymbol{x}_i)\right) \left(A_i - \hat{e}^{-i}(\boldsymbol{x}_i)\right)}{\sum_{i=1}^{n} \alpha_i(\boldsymbol{x}) \left(A_i - \hat{e}^{-i}(\boldsymbol{x}_i)\right)}, \tag{6}$$

see Athey and Wager (2019) for more details.

6 The ATE is estimated using the augmented inverse probability weighting (AIPW) estimator (Robins and Rotnitzky, 1995) and equals

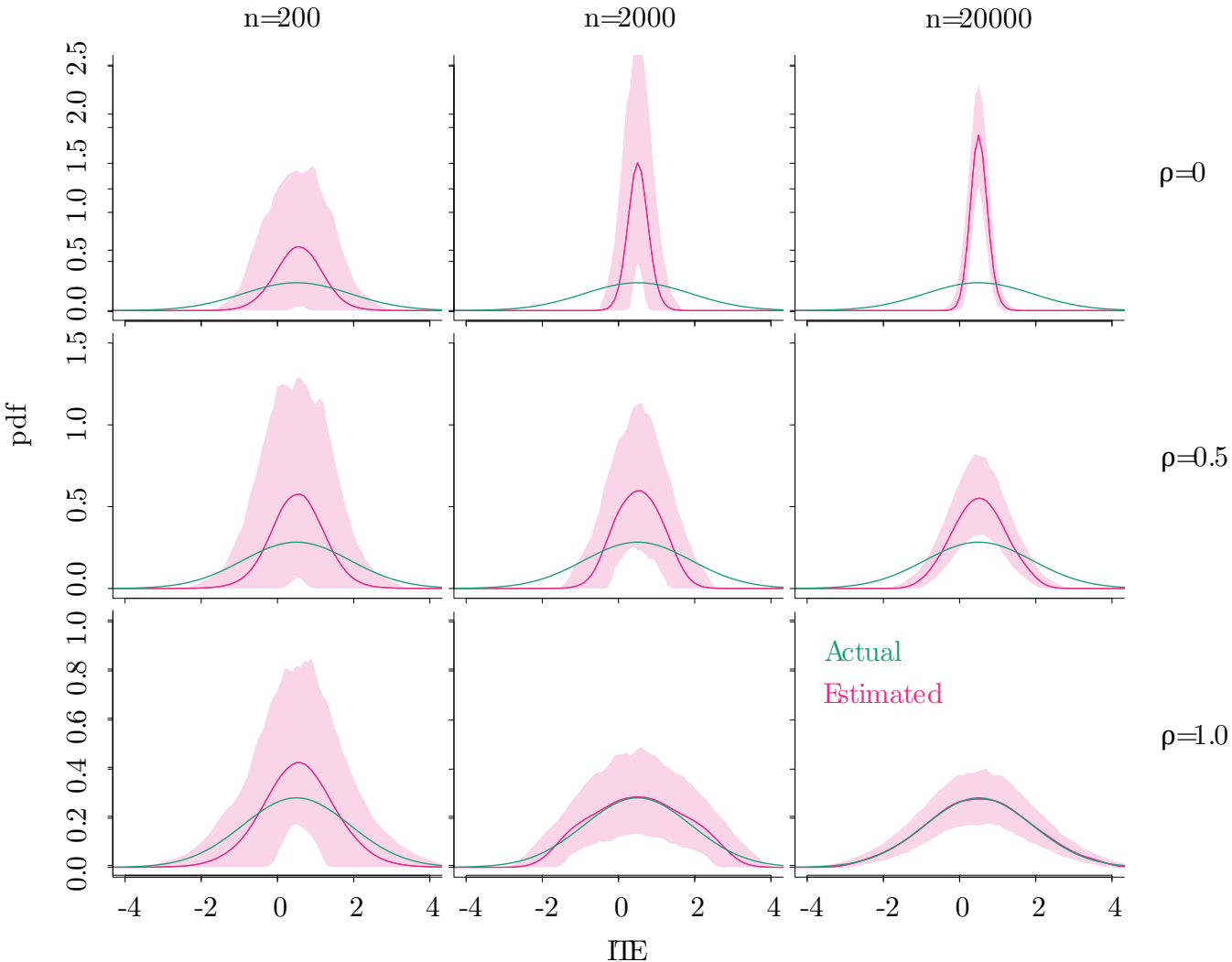
$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\widetilde{\tau}}(\boldsymbol{x}_i) + A_i\frac{Y_i - \hat{\mu}_{1i}(\boldsymbol{x}_i)}{\hat{e}^{-i}(\boldsymbol{x}_i)} - (1-A_i)\frac{Y_i - \hat{\mu}_{0i}(\boldsymbol{x}_i)}{1-\hat{e}^{-i}(\boldsymbol{x}_i)}\right), \tag{7}$$

where $\hat{\mu}_{1i}(\boldsymbol{x}_i) = \hat{m}^{-i}(\boldsymbol{x}_i) + (1-\hat{e}^{-i}(\boldsymbol{x}_i))\hat{\widetilde{\tau}}(\boldsymbol{x}_i)$ and $\hat{\mu}_{0i}(\boldsymbol{x}_i) = \hat{m}^{-i}(\boldsymbol{x}_i) - \hat{e}^{-i}(\boldsymbol{x}_i)\hat{\widetilde{\tau}}(\boldsymbol{x}_i)$.

Running example - results CRF

We fit a CRF to the simulated data, to estimate the $(X_{\text{sex}}, X_{\text{SBP}}, X_0)$-CATE for each individual. We vary the sample size, $n \in \{200, 2000, 20000\}$, and the correlation between the unmeasured modifier U_1 and the measured X_0, $\rho \in \{0, 0.25, 0.5, 0.75, 1\}$, while fixing $\delta = 2$.

- Per simulation, we fit a CRF to estimate the $(X_{\text{sex}}, X_{\text{SBP}}, X_0)$-CATE for each individual and compute the empirical standard deviation (SD) and positive effect probability (PEP), $\mathbb{P}(Y^1 - Y^0 > 0)$, of the estimated CATE distribution, and corresponding 95% confidence intervals (CIs) based on 1000 bootstrap samples.
- Based on 1000 simulations, we derive the bias, MSE, and coverage for the ATE, SD and PEP of the ITE distribution based on the CATE distribution.

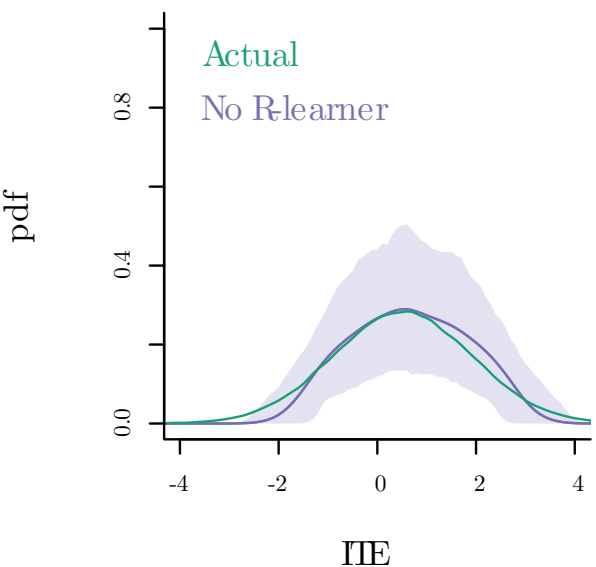

n=200
n=2000
n=20000
ρ=0
ρ=0.5
ρ=1.0
pdf
IIE
Actual
Estimated

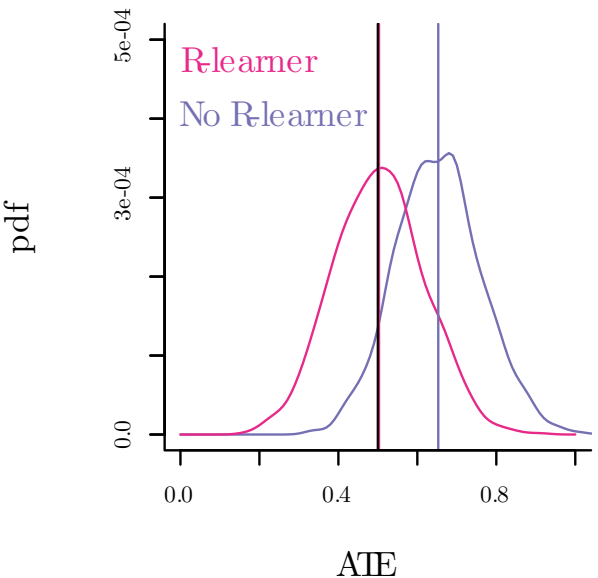
		Bias			MSE			Coverage		
ρ	n	ATE	SD	PEP	ATE	SD	PEP	ATE	SD	PEP
0	200	0.05	-0.97	0.18	0.20	0.97	0.08	0.95	0.02	0.83
0	2000	-0.00	-1.18	0.33	0.02	1.39	0.11	0.94	0.00	0.20
0	20000	0.00	-1.18	0.35	0.00	1.40	0.12	0.95	0.00	0.00
0.25	200	0.01	-0.94	0.16	0.20	0.91	0.07	0.94	0.03	0.86
0.25	2000	0.00	-1.07	0.28	0.02	1.15	0.08	0.94	0.00	0.29
0.25	20000	-0.00	-1.02	0.26	0.00	1.04	0.07	0.95	0.00	0.00
0.50	200	0.02	-0.87	0.14	0.19	0.80	0.06	0.95	0.07	0.86
0.50	2000	0.01	-0.82	0.15	0.02	0.69	0.03	0.95	0.00	0.56
0.50	20000	0.00	-0.71	0.12	0.00	0.51	0.02	0.94	0.00	0.01
0.75	200	0.04	-0.73	0.12	0.16	0.58	0.04	0.95	0.19	0.87
0.75	2000	0.01	-0.53	0.06	0.01	0.29	0.01	0.95	0.00	0.85
0.75	20000	0.00	-0.38	0.05	0.00	0.15	0.00	0.94	0.00	0.51
1.00	200	0.04	-0.56	0.09	0.14	0.37	0.03	0.95	0.37	0.88
1.00	2000	0.01	-0.20	0.01	0.01	0.05	0.00	0.96	0.42	0.93
1.00	20000	0.00	-0.04	0.00	0.00	0.00	0.00	0.95	0.76	0.94

Intermezzo: Confounder adjustment is crucial

- Sex is no modifier but only a confounder: $\tau_{\text{sex}} = 0$
- $n = 2000$, $\rho = 1$

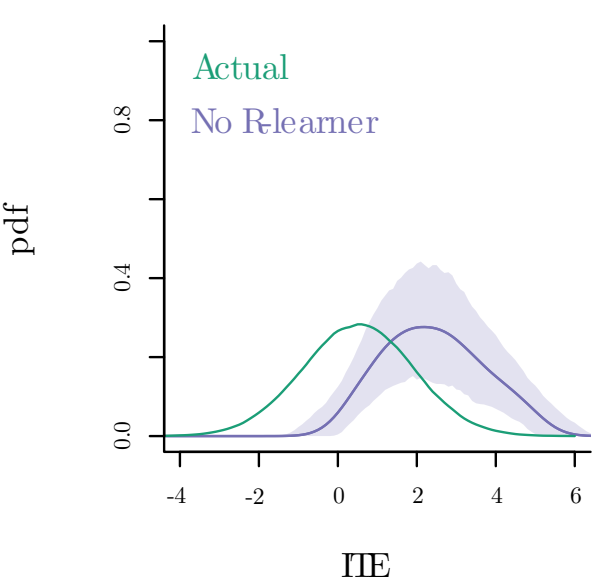
	Bias		
	ATE	SD	PEP
R-learner	0.00	-0.20	0.01
No R-learner	0.15	-0.22	0.04

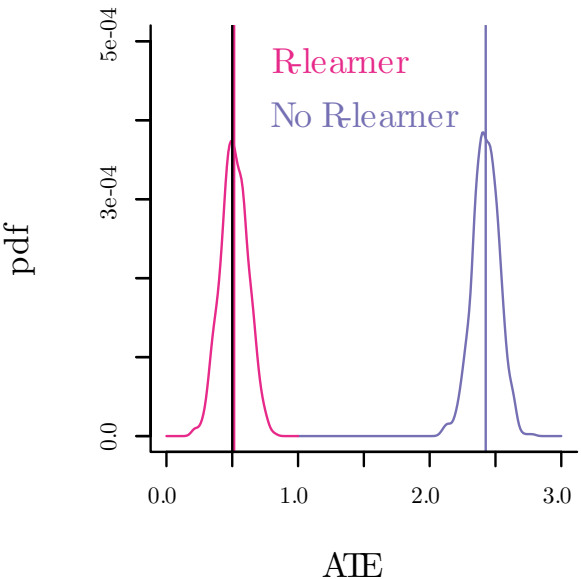




Intermezzo: Confounder adjustment is crucial

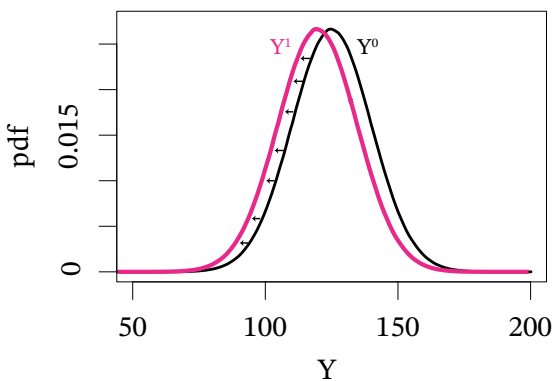
- Sex is no modifier but only a confounder: $\tau_{\text{sex}} = 0$
- $n = 2000$, $\rho = 1$
- Stronger confounder ($\beta_{\text{sex}} = 3.2$ and $\alpha_{\text{sex}} = 3$):

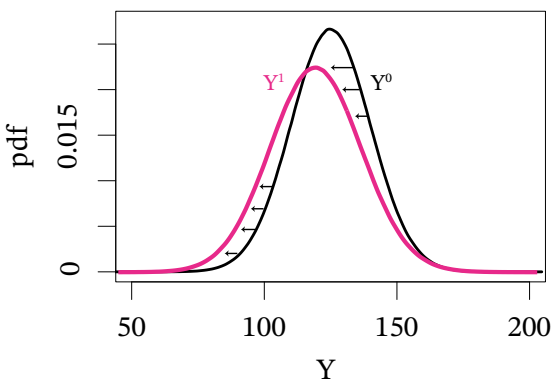




Going beyond CATEs

- If there is no remaining effect variability given $\boldsymbol{X}$, then the individualized CATE equals the individual TE.
- Then, the (conditional) Y^1 distribution equals the Y^0 distribution shifted by a constant so that the conditional variances should be equal.





- Otherwise, we would like to estimate the conditional (on $\boldsymbol{X}$) variance of the ITE next to the CATE.

Identification of the conditional ITE variance

- For the parameterization in Equation (2), the variance of $Y^1 - Y^0 \mid \boldsymbol{X}{=}\boldsymbol{x}$ equals $\sigma_1^2(\boldsymbol{x}) = \mathbb{E}[(U_1)^2 \mid \boldsymbol{X}{=}\boldsymbol{x}]$.
- For $h(\boldsymbol{x}_i) = \mathbb{E}[(Y_i)^2 \mid \boldsymbol{X}_i{=}\boldsymbol{x}_i]$, in the absence of unmeasured confounding,

$$(Y_i)^2 - h(\boldsymbol{x}) = (A_i - e(\boldsymbol{x}))\Delta(\boldsymbol{x}_i) + \gamma_i,$$

where $\forall \boldsymbol{x}$: $\mathbb{E}[\gamma_i \mid A{=}1, \boldsymbol{X}{=}\boldsymbol{x}] = \mathbb{E}[\gamma_i \mid A{=}0, \boldsymbol{X}{=}\boldsymbol{x}] = 0$,

$$\Delta(\boldsymbol{x}) = \left(\tau(\boldsymbol{x})^2 + \sigma_1^2(\boldsymbol{x}) + 2\tau(\boldsymbol{x})\theta_0(\boldsymbol{x}) + 2\mathbb{E}[N_Y U_1 \mid \boldsymbol{X}_i{=}\boldsymbol{x}]\right),$$

- Via estimation of $\tau(\boldsymbol{x})$ and $\theta_0(\boldsymbol{x})$, a R-learner can be used to estimate

$$\widetilde{\sigma_1}^2(\boldsymbol{x}) = \sigma_1^2(\boldsymbol{x}) + 2\mathbb{E}[N_Y U_1 \mid \boldsymbol{X}{=}\boldsymbol{x}].$$

- $\widetilde{\sigma_1}^2(\boldsymbol{x})$ represents the sum of the conditional variance of the ITE and twice the conditional covariance of Y^0 and the ITE.
- However, as a result of the fundamental problem of causal inference, $\mathbb{E}[N_Y U_1 \mid \boldsymbol{X}=\boldsymbol{x}]$ is not identifiable.
  - → So, we cannot estimate $\sigma_1^2(\boldsymbol{x})$ without an additional (cross-world) assumption on the joint distribution of Y^0 and $Y^1 - Y^0$.
- Under conditional independent effect deviation (CIED), $\widetilde{\sigma_1}^2(\boldsymbol{x}) = \sigma_1^2(\boldsymbol{x})$, and the conditional variance becomes identifiable.

Assumption 4 (CIED)

Conditional independent effect deviation

$$Y^1 - Y^0 \perp\!\!\!\perp Y^0 \mid \boldsymbol{X}=\boldsymbol{x}$$

CIED example: Clopidogrel

- The antiplatelet medicine Clopidogrel reduces the risk of stroke and myocardial infarction in individuals with acute coronary syndrome.
  - → The effect depends on its conversion to an active metabolite which is accomplished by the cytochrome P450 2C19 (CYP2C19) enzyme (Lee et al., 2022).

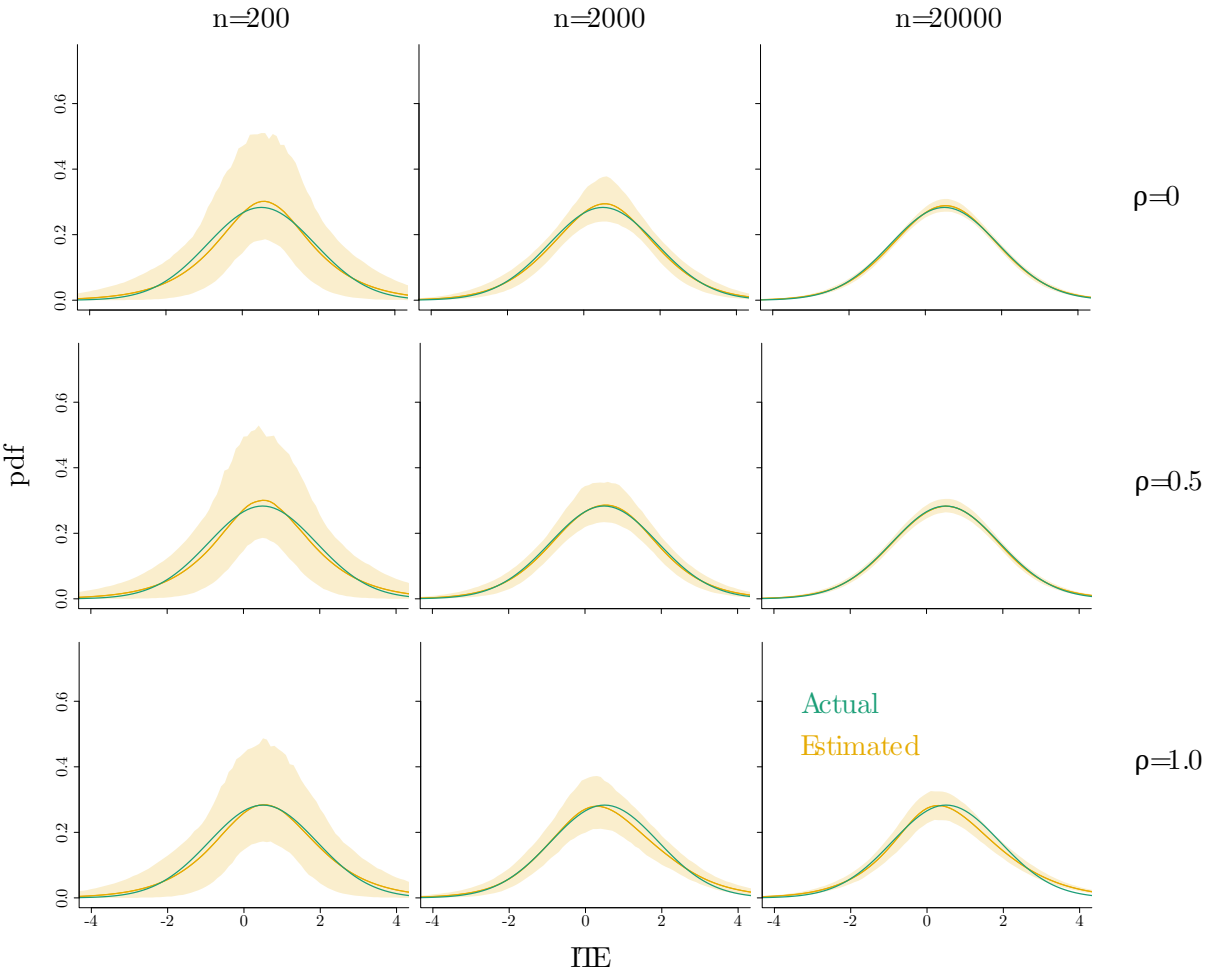
- For individuals with a CYP2C19 gene mutation, the drug is known to have a reduced antiplatelet effect; effect heterogeneity.
- There is no reason to believe that the phenotype affects platelet aggregation in the absence of the drug.

CIED remarks

- In cases where Y^0 is still expected to inform on the value of $Y^1 - Y^0$ given the levels of $\boldsymbol{X}$, the identification assumption does not apply.
- Similar to Assumption 2 [*Conditional Exchangeability*], this causal assumption **cannot be verified with data** as it concerns unmeasured features that affect both Y^0 and $Y^1 - Y^0$ and should be judged by experts in the field of application.
- In contrast to Conditional Exchangeability, there is no reason that guarantees that CIED holds in a randomized experiment.

Running example - results extended CRF

- ATE estimate remains the same as for the original CRF.
- The SD of the effect in the full population can be derived from the CATEs and conditional variances.
- Only when the conditional ITE distribution can be well approximated with a Gaussian distribution the distribution of $Y^1 - Y^0 \mid \boldsymbol{X}{=}\boldsymbol{x}$ is identified by the CATE and the conditional variance.
  - → For illustration, we will assume the Gaussianity of the conditional ITE distribution in our example so that we can use the extended CRF to estimate the ITE distribution from the simulated datasets. The PEP is estimated as $n^{-1} \sum_{i=1}^{n} \mathbb{P}(Z_i > 0)$, where $Z_i \sim \mathcal{N}\left(\hat{\widetilde{\tau}}(\boldsymbol{x}_i), \max\left\{0, \hat{\widetilde{\sigma_1}}^2(\boldsymbol{x})\right\}\right)$.


n=200
n=2000
n=20000
ρ=0
ρ=0.5
ρ=1.0
pdf
IIE
Actual
Estimated
0.0
0.2
0.4
0.6
-4
-2
0
2
4

Results Extended CRF

		Bias			MSE			Coverage		
ρ	n	ATE	SD	PEP	ATE	SD	PEP	ATE	SD	PEP
0	200	0.05	-0.06	0.04	0.20	0.25	0.02	0.95	0.90	0.91
0	2000	-0.00	0.04	0.01	0.02	0.02	0.00	0.94	0.93	0.90
0	20000	0.00	0.03	0.01	0.00	0.00	0.00	0.95	0.89	0.82
0.25	200	0.01	-0.08	0.03	0.20	0.29	0.02	0.94	0.88	0.92
0.25	2000	0.00	0.04	0.01	0.02	0.02	0.00	0.94	0.93	0.90
0.25	20000	-0.00	0.04	0.00	0.00	0.00	0.00	0.95	0.86	0.86
0.50	200	0.02	-0.07	0.03	0.19	0.27	0.02	0.95	0.90	0.92
0.50	2000	0.01	0.05	0.01	0.02	0.02	0.00	0.95	0.94	0.91
0.50	20000	0.00	0.05	0.00	0.00	0.00	0.00	0.94	0.80	0.92
0.75	200	0.04	-0.07	0.03	0.16	0.26	0.02	0.95	0.92	0.92
0.75	2000	0.01	0.07	-0.00	0.01	0.03	0.00	0.95	0.91	0.94
0.75	20000	0.00	0.07	-0.00	0.00	0.01	0.00	0.94	0.71	0.93
1.00	200	0.04	-0.03	0.02	0.14	0.26	0.02	0.95	0.91	0.92
1.00	2000	0.01	0.09	-0.01	0.01	0.03	0.00	0.96	0.90	0.93
1.00	20000	0.00	0.09	-0.01	0.00	0.01	0.00	0.95	0.56	0.89

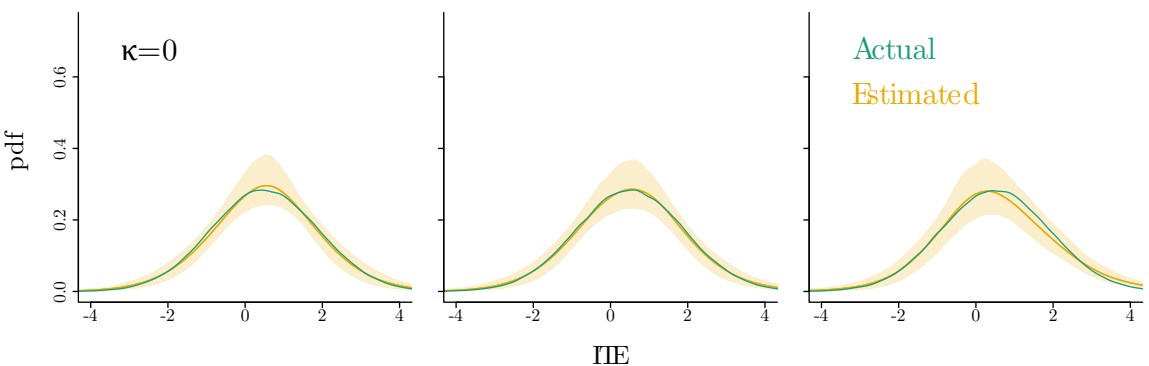
CIED violation

Cause-effect relations in Equation (3), can be generalized with

$$\begin{pmatrix} N_Y \\ U_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \kappa\sigma_0\sigma_1 \\ \kappa\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} \right),$$

so that the original example is obtained for $\kappa = 0$. We set σ_1 so that the variance of Y^1 is the same as in the main example.

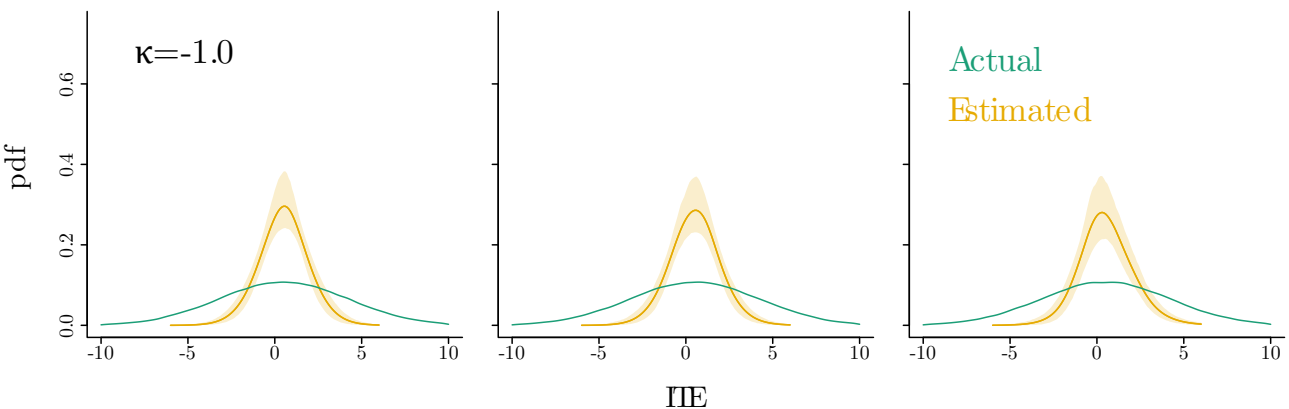
- For $n = 2000$ and $\kappa = 0$ (so that CIED applies)



The estimand of the extended CRF is $\widetilde{\sigma_1}^2(\boldsymbol{x}) = \sigma_1^2(\boldsymbol{x}) + 2\mathbb{E}[N_Y U_1 \mid \boldsymbol{X}{=}\boldsymbol{x}]$, so that for $\kappa \neq 0$: $\widetilde{\sigma_1}^2(\boldsymbol{x}) \neq \sigma_1^2(\boldsymbol{x})$.

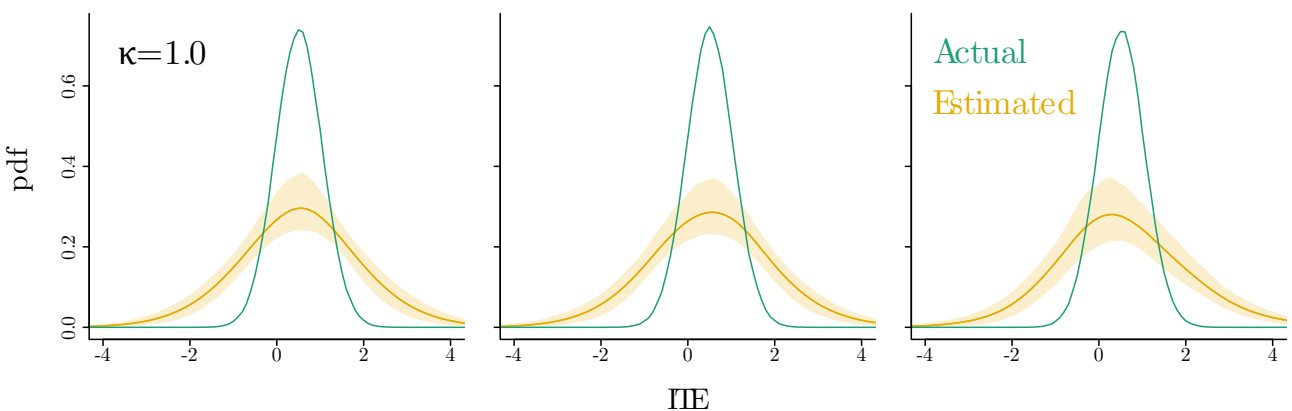
$\kappa < 0$

- $n = 2000$



$\kappa > 0$

- $n = 2000$



Concluding remarks

- The individualized CATE can still seriously differ from the ITE.
  - → Remaining effect heterogeneity beyond heterogeneity in the CATEs can result in a lack of generalizability (Seamans et al., 2021).
- Under remaining effect homogeneity, the conditional second moments of the treated and the controls should be similar.
- To estimate the (conditional) variance of the ITE, we need to assume how the ITE and Y^0 are correlated.
  - → For example, by assuming conditional independent effect deviation.
- As an example, we have extended the CRF algorithm (Athey et al., 2019) also to estimate the difference in conditional variance between treated and controls.
- The identification assumptions cannot be verified with data and should be based on expert knowledge.

Any questions?

: r.a.j.post@tue.nl

References I

Susan Athey and Guido W. Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016. doi: 10.1073/pnas.1510489113.

Susan Athey and Stefan Wager. Estimating Treatment Effects with Causal Forests: An Application. *Observational studies*, 5(2):37–51, 2019. doi: 10.1353/obs.2019.0001.

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *Annals of Statistics*, 47(2):1179–1203, 2019. doi: 10.1214/18-AOS1709.

Philipp Bach, Victor Chernozhukov, Malte S. Kurz, and Martin Spindler. Doubleml - an object-oriented implementation of double machine learning in python. *Journal of Machine Learning Research*, 23(53):1–6, 2022.

Alberto Caron, Gianluca Baio, and Ioanna Manolopoulou. Estimating individual treatment effects using non-parametric regression models: A review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 185(3):1115–1149, 2022. doi: 10.1111/rssa.12824.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James M. Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 01 2018. doi: 10.1111/ectj.12097.

References II

Laura S. Chiu, Alison Pedley, Joseph M. Massaro, Emelia J. Benjamin, Gary F. Mitchell, David D. McManus, Jayashri Aragam, Ramachandran S. Vasan, Susan Cheng, and Michelle T. Long. The association of non-alcoholic fatty liver disease and cardiac structure and function—framingham heart study. *Liver International*, 40(10):2445–2454, 2020. doi: 10.1111/liv.14600.

Sander Greenland, Michael P. Fay, Erica H. Brittain, Joanna H. Shih, Dean A. Follmann, Erin E. Gabriel, and James M. Robins. On Causal Inferences for Personalized Medicine: How Hidden Causal Assumptions Led to Erroneous Causal Claims About the D-Value. *The American Statistician*, 74(3):243–248, 2019. doi: 10.1080/00031305.2019.1575771.

David J. Hand. On comparing two treatments. *The American Statistician*, 46(3):190–192, 1992. doi: 10.1080/00031305.1992.10475881.

Miguel A Hernán and James M. Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, Boca Raton, Florida, 1st edition, 2020.

Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986. doi: 10.1080/01621459.1986.10478354.

Michael R. Kosorok and Eric B. Laber. Precision medicine. *Annual Review of Statistics and Its Application*, 6(1):263–286, 2019. doi: 10.1146/annurev-statistics-030718-105251.

R.L. Kravitz, N. Duan, and J. Braslow. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Quarterly*, 82(4):661–687, 2004. doi: 10.1111/j.0887-378X.2004.00327.x.

References III

Craig R. Lee, Jasmine A. Luzum, Katrin Sangkuhl, Roseann S. Gammal, Marc S. Sabatine, Charles Michael Stein, David F. Kisor, Nita A. Limdi, Yee Ming Lee, Stuart A. Scott, Jean-Sébastien Hulot, Dan M. Roden, Andrea Gaedigk, Kelly E. Caudle, Teri E. Klein, Julie A. Johnson, and Alan R. Shuldiner. Clinical pharmacogenetics implementation consortium guideline for cyp2c19 genotype and clopidogrel therapy: 2022 update. *Clinical Pharmacology & Therapeutics*, 112(5):959–967, 2022. doi: 10.1002/cpt.2526.

Min Lu, Saad Sadiq, Daniel J. Feaster, and Hemant Ishwaran. Estimating Individual Treatment Effect in Observational Data Using Random Forest Methods. *Journal of Computational and Graphical Statistics*, 27(1):209–219, 2018. doi: 10.1080/10618600.2017.1356325.

Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 09 2020. doi: 10.1093/biomet/asaa076.

James M. Robins and Andrea Rotnitzky. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90 (429):122–129, 1995. doi: 10.2307/2291135.

Peter M Robinson. Root-n-consistent semiparametric regression. *Econometrica*, 56(4): 931–954, 1988. doi: 10.2307/1912705.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. doi: 10.1037/h0037350.

References IV

Marissa J. Seamans, Hwanhee Hong, Benjamin Ackerman, Ian Schmid, and Elizabeth A. Stuart. Generalizability of subgroup effects. *Epidemiology*, 32(3):389–392, 2021. doi: 10.1097/EDE.0000000000001329.

Stefan Wager and Susan Athey. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523): 1228–1242, 2018. doi: 10.1080/01621459.2017.1319839.

Non-Gaussian conditional effect distributions

We let $U_1 \overset{d}{=} \exp(X) - \mu$, where $X \sim \mathcal{N}(0, \sigma^2)$ and

$$\sigma = \sqrt{\log\left(0.5\sqrt{4\sigma_1^2 + 1} + 0.5\right)} \tag{8}$$

$$\mu = \exp(0 + 0.5\sigma^2), \tag{9}$$

- $X_3 \mid U_1{=}u_1 \sim \mathcal{N}\left(u_1\delta\rho, \delta^2\sigma_1^2(1-\rho^2)\right)$

For this setting, Assumption 4 is valid so that the extended CRF can be used to estimate the SD. However, assuming Gaussianity of the conditional ITE, will result in a bias of the PEP when ρ equals 0 or 0.5.

Non-Gaussian conditional effect distributions

- $n = 2000$