Detecting critical treatment effect bias even in small subgroups

Piersilvio De Bartolomeis, Javier Abad, Konstantin Donhauser, Fanny Yang

Department of Computer Science, ETH Zürich

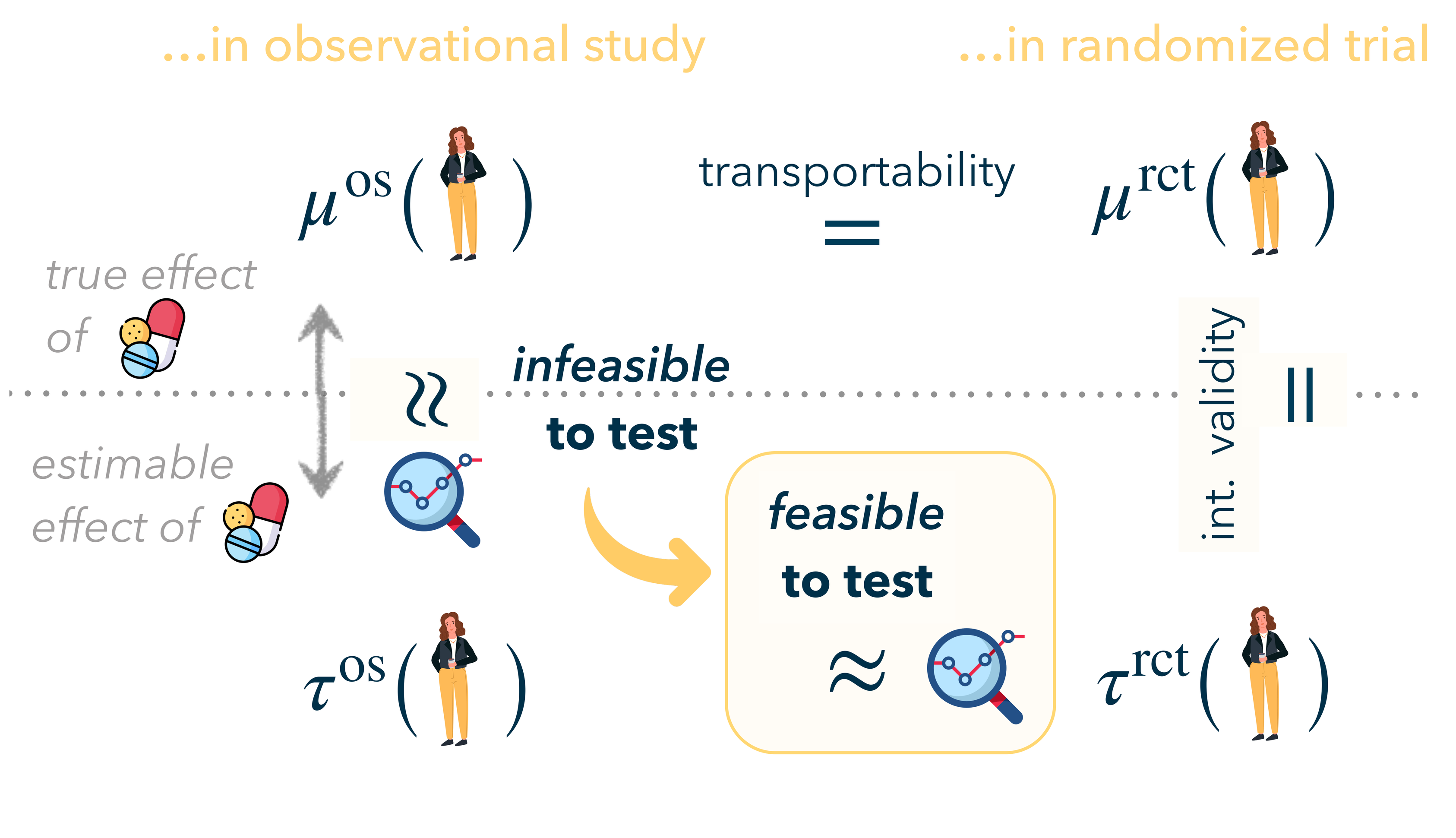
Problem Setting

- $\mathbb{P}^\diamond$ over $(X, U, Y(0), Y(1), Y, T)$ for $\diamond \in \{\text{rct}, \text{os}\}$
- We observe $D_\diamond = \{(X_i, Y_i, T_i)\}_{i=1}^n$ sampled i.i.d from $\mathbb{P}^\diamond$

Trade-off between randomized and observational data:

- $\mathbb{P}^{\text{rct}}$ satisfies internal validity: $T \perp\!\!\!\perp (Y(1), Y(0))$
  - we can estimate $\mu^{\text{rct}}(X) := \mathbb{E}_{\mathbb{P}^{\text{rct}}}[Y(1) - Y(0) | X]$
  - but the support of $\mathbb{P}_X^{\text{rct}}$ is limited (e.g. no children)
- $\mathbb{P}^{\text{os}}$ covers a broader population: $\text{supp}(\mathbb{P}_X^{\text{rct}}) \subset \text{supp}(\mathbb{P}_X^{\text{os}})$
  - but many sources of bias $\implies$ $\mu^{\text{os}}(X)$ is not identifiable

Can we detect when observational data does not allow reliable inference?



Null Hypothesis

- **Goal**: Given a tolerance function δ, test for

$$H_0 : |\tau^{\text{rct}}(X) - \tau^{\text{os}}(X)| \leq \delta(X), \quad \mathbb{P}_X^{\text{rct}} - \text{a.s.}$$

- **How**: We can test if for some $g^\star : \mathbb{R}^d \to [-1, 1]$ in $\mathcal{G}$

$$H_0^{\mathcal{G}} : \underbrace{\tau^{\text{rct}}(X) - \tau^{\text{os}}(X) + g^\star(X)\delta(X)}_{:=\psi_{g^\star}(X)} = 0, \quad \mathbb{P}_X^{\text{rct}} - \text{a.s.}$$

  - but: $g^\star$ is unknown, and thus we cannot use the standard kernel conditional moment tests[1,2]

Test Statistic

- **Idea:** The oracle cross U-statistic is asymptotically normal[3]

$$\hat{\mathbb{H}}^2(\psi_{g^\star}) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=n+1}^{2n} \psi_{g^\star}(X_i) k(X_i, X_j) \psi_{g^\star}(X_j)$$

with $X_1, \ldots, X_{2n}$ i.i.d. from $\mathbb{P}^{\text{rct}}$ and bounded kernel k

- We can compute an asymptotically valid C.I. for

$$\mathbb{H}^2_{\text{OPT}} := \min_{g \in \mathcal{G}} \left| \frac{\sqrt{n}\, \hat{\mathbb{H}}^2(\psi_g)}{\hat{\sigma}\left(\hat{\mathbb{H}}^2(\psi_g)\right)} \right| \leq \left| \frac{\sqrt{n}\, \hat{\mathbb{H}}^2(\psi_{g^\star})}{\hat{\sigma}\left(\hat{\mathbb{H}}^2(\psi_{g^\star})\right)} \right| \to |\mathcal{N}(0,1)|$$

  - but: ψ_g depends on τ^{os}, which is estimated from data

Theoretical guarantees: Asymptotic validity

Assume that

$$n_{\text{os}} \gg n \quad \text{and} \quad \|\tau^{\text{os}} - \hat{\tau}^{\text{os}}\|_{L^2(\mathbb{P}^{\text{rct}})} = O_{\mathbb{P}^{\text{os}}}(n_{\text{os}}^{-1/2})$$

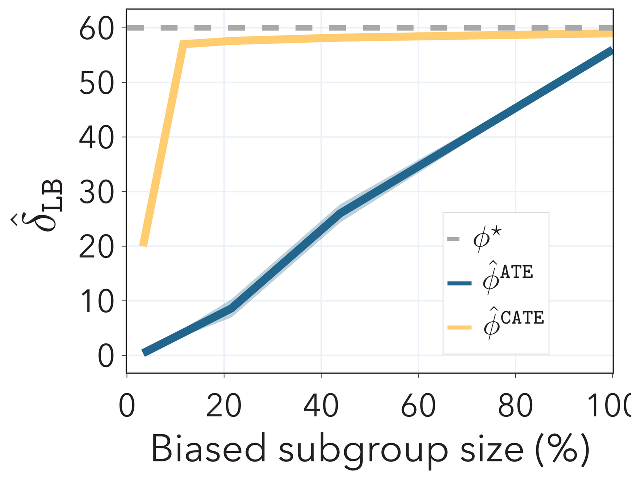
Then, under weak regularity conditions, we have

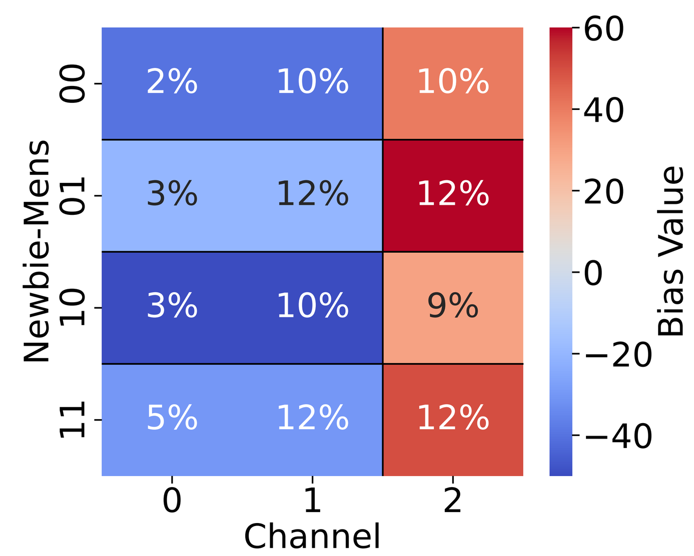
$$\frac{\sqrt{n}\, \hat{\mathbb{H}}^2(\hat{\psi}_{g^\star})}{\hat{\sigma}\left(\hat{\mathbb{H}}^2(\hat{\psi}_{g^\star})\right)} \to \mathcal{N}(0,1), \quad \text{as} \quad n \to \infty \quad \text{and} \quad n_{\text{os}} \to \infty$$

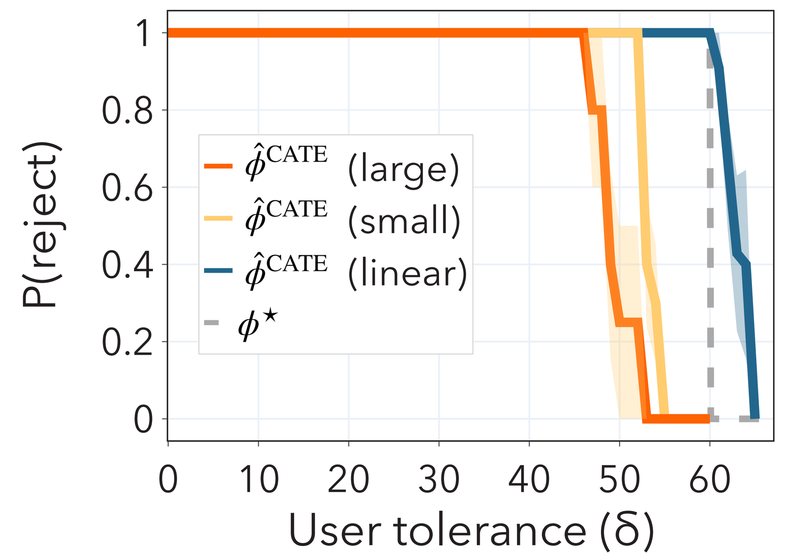
Hence, $\hat{\phi}_\alpha := \mathbb{I}\{\hat{\mathbb{H}}^2_{\text{OPT}} \geq z_{\alpha/2}\}$ is asympt. valid at level α

Semi-synthetic Experiments

- **Data:** MineThatData Email[4]
  - $X :=$ customer data
  - $T :=$ exposure to ads
  - $Y :=$ dollars spent
- Set tolerance $\delta(x) = \delta$ for all x
- Estimate $\hat{\delta}_{\text{LB}} := \inf_\delta\{\delta : \hat{\phi}(\alpha) = 0\}$







Real-world Evaluation

- **Data:** Women's Health Initiative
  - $T :=$ hormone therapy (HT)
  - $Y :=$ coronary heart disease
- **Question**: Is there enough bias to explain away the benefits of HT in young women?
- **Ground truth**: No! (from established medical knowledge)
- **Our strategy**:
  1. Estimate $\hat{\delta}_{\text{CT}} := \mathbb{E}_{\mathbb{P}^{\text{os}}}[\tau^{\text{os}}(X) \mid X_{\text{age}} \leq 60]$
  2. Reject study if $\hat{\delta}_{\text{LB}} \geq \hat{\delta}_{\text{CT}}$

Statistical tests	$\hat{\phi}^{\text{CATE}}$	$\hat{\phi}^{\text{ATE}}$	$\hat{\phi}^{\text{CATE}}_{\delta=0}$	$\hat{\phi}^{\text{ATE}}_{\delta=0}$
$\hat{\delta}_{\text{CT}}$	0.32	0.32	0.32	0.32
$\hat{\delta}_{\text{LB}}$	**0.25**	0.11	✗	✗
Reject the study	0	0	1	1

References

1. Muandet et al. 2020. Kernel conditional moment test via maximum moment restriction.
2. Hussain et al. 2023. Falsification of internal and external validity in observational studies via conditional moment restrictions.
3. Kim and Ramdas 2024. Dimension-agnostic inference using cross U-statistics.
4. Kevin Hillstrom 2008. The MineThatData e-mail analytics and data mining challenge