Targeted Maximum Likelihood Estimation; Do we have a free lunch?

Sanne Roels, Bob Murray
Statistical Modelling and Methodology and Neuroscience Biostatistics

June 17th 2024

Johnson&Johnson

Motivation

Adjusting for the influence of covariates in RCTs

Why do it?

- Inclusion of prognostic baseline factors improves precision for treatment effect estimates in RCTs.
- **FDA guidance(MAY 2023):** propose robust estimators for linear and non-linear models to obtain **marginal** (unconditional) effect estimates **with increased precision** by **incorporating covariates**.
- Adjustment for baseline covariates can improve precision if/when
  - Baseline covariates are prognostic for the outcome
  - There are slight imbalances between the 2 exposure groups with respect to these baseline covariates

Adjusting for the influence of covariates in RCTs

Some things that hold us back

- **Knowledge of the relationship between covariates and outcome**
- **Specifying the correct model**
- **Obtaining a marginal treatment effect (non-collapsibility under non-linear models)**
- **Targeted Maximum Likelihood Estimation (TMLE) promises to address all these concerns…**

Obtaining a Marginal Treatment Effect with Causal Inference

Use a model for the outcome

Standardization

Y = outcome

X = covariates

A = treatment (1 = active; 0 = control)

- Interested in a **causal, marginal** treatment effect: $E(Y^{a="Active"}) - E(Y^{a="Control"})$

Pt ID	A	X	Y	$\widehat{Y}^{a=1}$	$\widehat{Y}^{a=0}$
1	0	Smoke	73	77	42.5
2	0	Smoke	78	77	42.5
3	0	Smoke	82	77	42.5
4	0	Don't Smoke	95	98.5	64
5	0	Don't Smoke	100	98.5	64
6	0	Smoke	43	77	42.5
7	0	Smoke	40	77	42.5
8	0	Don't Smoke	57	98.5	64
9	0	Don't Smoke	65	98.5	64
10	0	Don't Smoke	72	98.5	64

$$E\left[\widehat{Y}^{a=1}\right] - E\left[\widehat{Y}^{a=0}\right] = 87.75 - 53.25 = 34.5$$

Outcome model must be correct to get consistent estimates

Rubin, D. Causal inference using Potential Outcomes. Journal of the American Statistical Association. 2005, 100, 322-331

Use a model for treatment assignment

Inverse Probability Weighting

Y = outcome

X = covariates

A = treatment (1 = active; 0 = control)

- Interested in a **causal, marginal** treatment effect: $E(Y^{a="Active"}) - E(Y^{a="Control"})$

Pt ID	A	X	Y	PS	Weight
1	1	Smoke	73	0.6	1.66
2	1	Smoke	78	0.6	1.66
3	1	Smoke	82	0.6	1.66
4	1	Don't Smoke	95	0.4	2.5
5	1	Don't Smoke	100	0.4	2.5
6	0	Smoke	43	0.4	2.5
7	0	Smoke	40	0.4	2.5
8	0	Don't Smoke	57	0.6	1.66
9	0	Don't Smoke	65	0.6	1.66
10	0	Don't Smoke	72	0.6	1.66

Treatment assignment model must be correct to get consistent estimates

J&J

Hernan --Hernan & Robins (2024). What If. Retrieved via https://www.hsph.harvard.edu/miguel-hernan/wp-content/uploads/sites/1268/2024/04/hernanrobins_WhatIf_26apr24.pdf

Can we use both?

Targeted Maximum Likelihood Estimation (TMLE)

- TMLE allows us to utilise both outcome model and treatment assignment model to obtain consistent estimates
- How does it work?
  1. Use outcome model to obtain initial estimates $\hat{Y}$ under both treatments ($\hat{Y}^{a=0}$ & $\hat{Y}^{a=1}$)
  2. Update $\hat{Y}$ by modelling with information from the treatment assignment model so it becomes $\hat{Y}^{*}$
     - Specifically, we fit the model $\mathrm{Y} = \hat{Y} + \boldsymbol{\varepsilon} H(A, X)$, where $H(A, X)$ is a function of the propensity score

$$\text{ATE} = \frac{1}{n}\sum_{i=1}^{n} Y^{*}(1, X_i) - Y^{*}(0, X_i), \text{ with } X_i \text{ baseline covariate information}$$

➔ Estimation of parameters can be via flexible, data-adaptive SuperLearner ensemble ML algorithms

➔ Inference via Influence Curves or bootstrap approach and is **consistent if either model for outcome or treatment assignment is correct** –doubly robust approach

➔ Targets a hypothetical estimand

van der Laan, Mark J. and Rubin, Daniel. "Targeted Maximum Likelihood Learning" *The International Journal of Biostatistics*, vol. 2, no. 1, 2006.

Simulation Setup

Targeted Maximum Likelihood Estimation (TMLE)

Questions we want to answer:

1. What are the operating characteristics of TMLE in a randomised clinical trial setting
2. How does it perform vs other 'typical estimators' and is this performance influenced by
    1. Sample size
    2. Strength of association between X and Y
    3. Effect size

Estimators under comparison

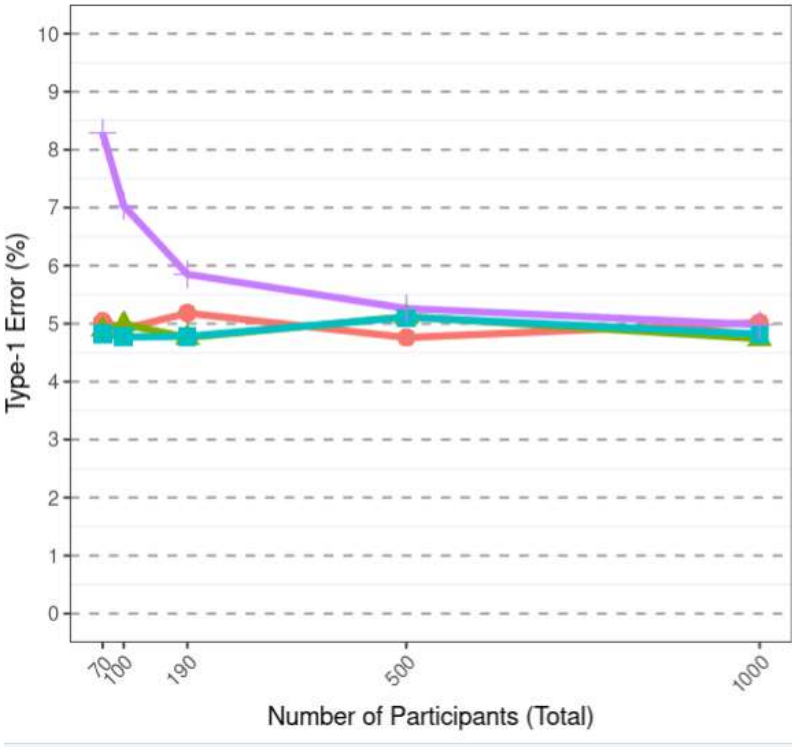
1. ANCOVA No covariates
2. ANCOVA incl. 1 highly correlated covariate
3. ANCOVA incl. 2 highly correlated covariates
4. ANCOVA incl. all correlated covariates
5. Basic implementation of TMLE

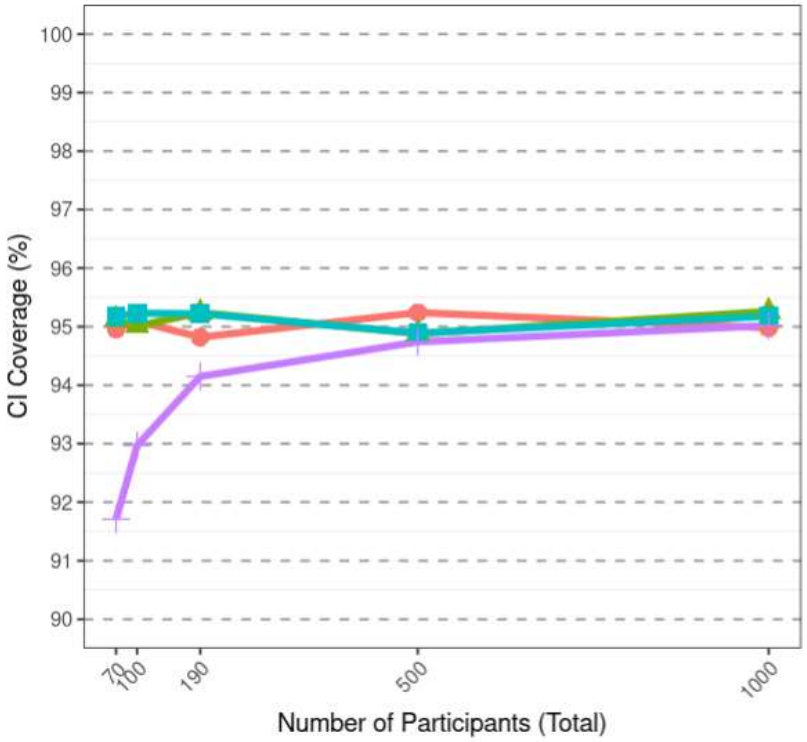
Data generation

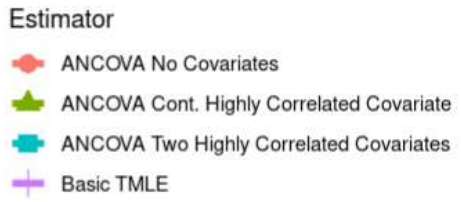
- Two counter factual outcomes $Y^{a=1}$, $Y^{a=0}$ and 10 covariates simulated from MVN distribution
- 5 covariates have no relationship with outcome
- 3 have slight relationship
- 2 have strong relationship
- 1:1 randomisation
- Varying sample size
- No missing data

Operating Characteristics under Null Hypothesis

Using Influence Curve Inference

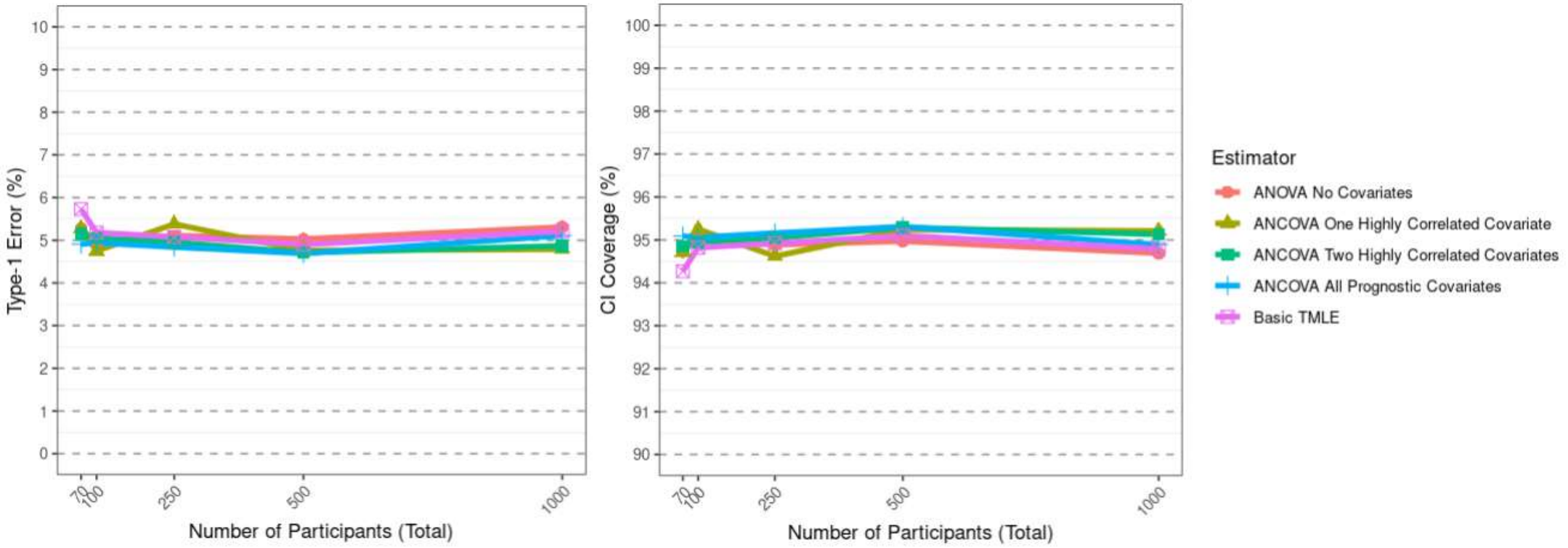






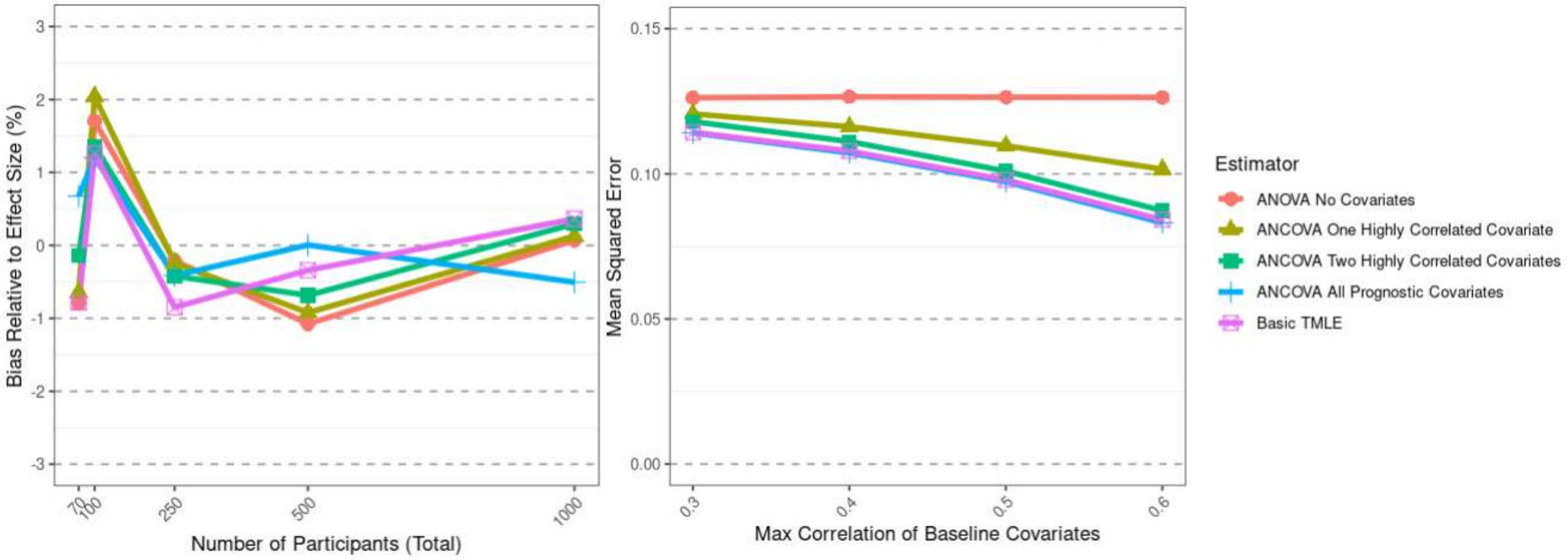
Operating Characteristics under Null Hypothesis

Using Bootstrap Inference



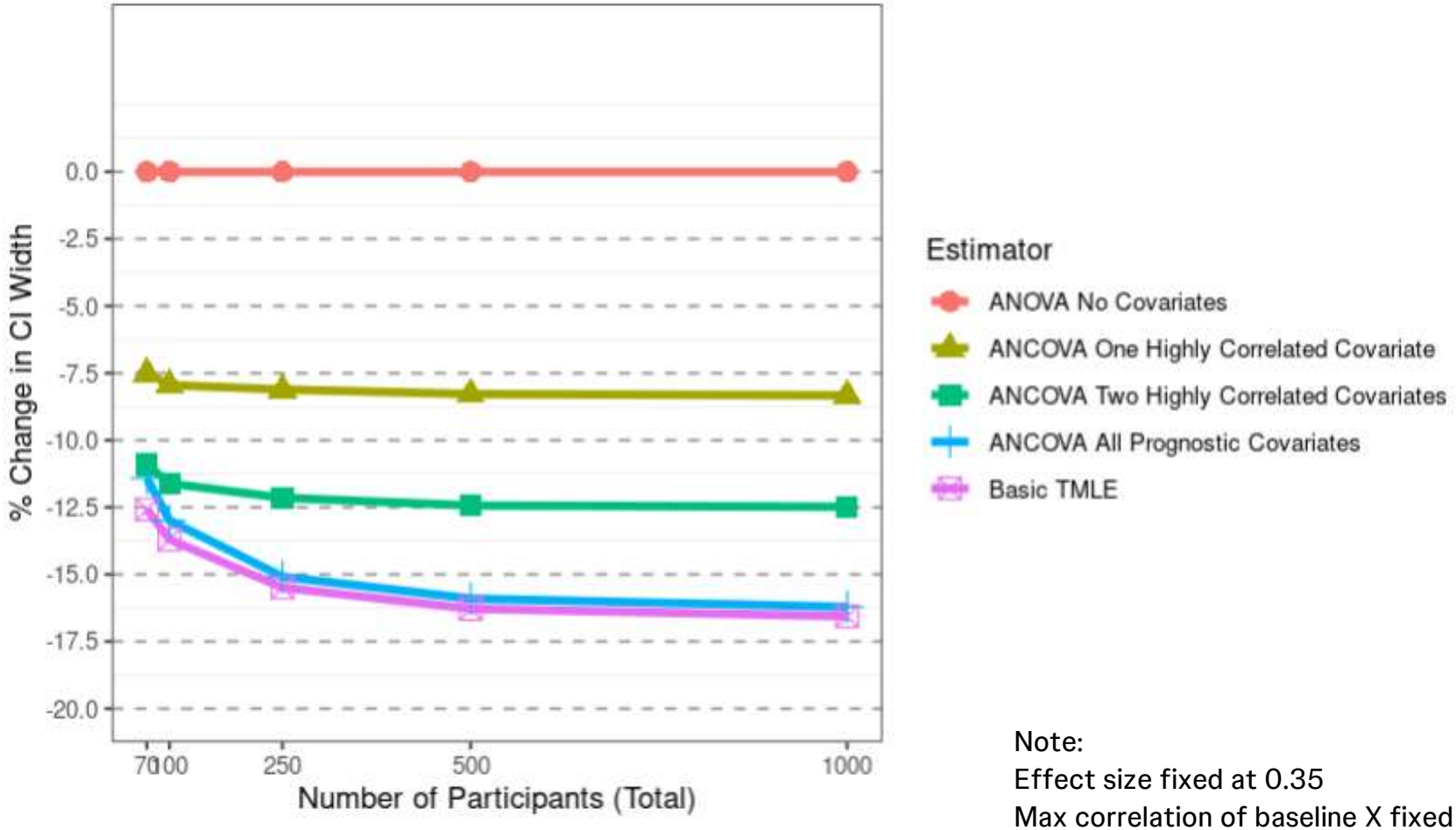
Operating Characteristics under Alternative Hypothesis

Using Bootstrap Inference



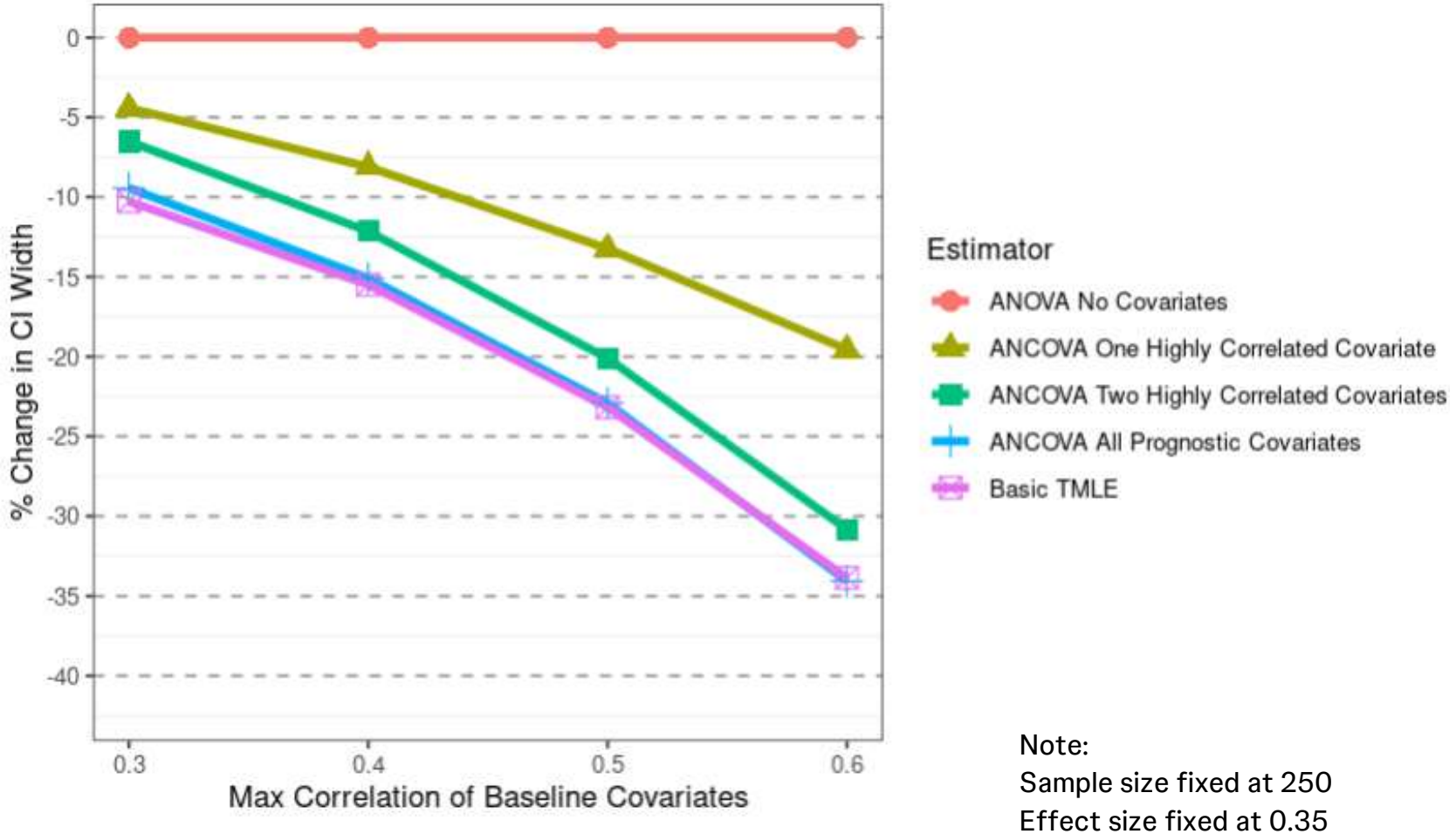
Note:
Effect size fixed at 0.35
Max correlation of baseline X fixed at 0.4 (left plot)
Sample size fixed at 250 (right plot)

How does sample size influence performance?



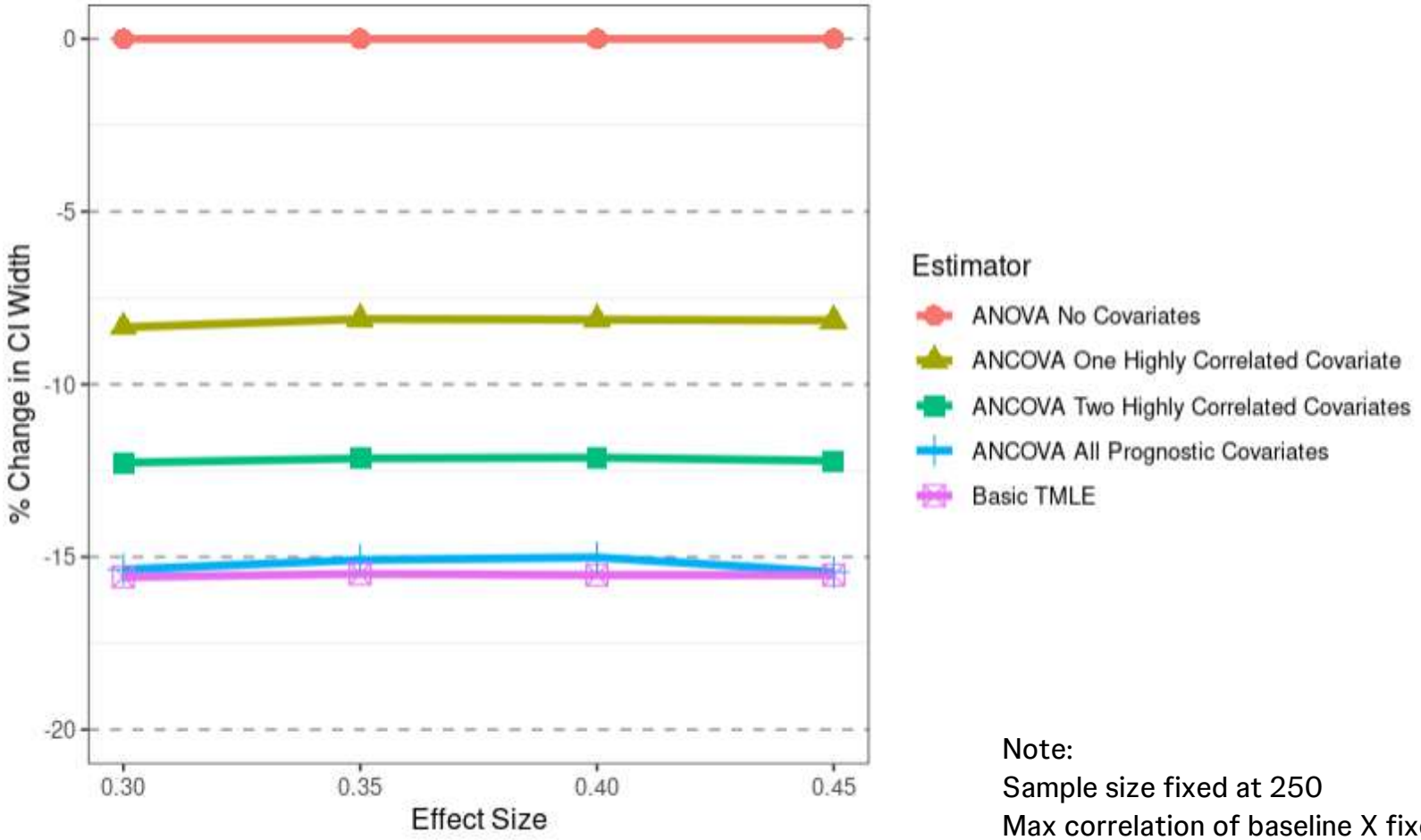
Note:
Effect size fixed at 0.35
Max correlation of baseline X fixed at 0.4

How does strength of covariate set influence performance?



Note:
Sample size fixed at 250
Effect size fixed at 0.35

How does effect size influence performance?



Note:
Sample size fixed at 250
Max correlation of baseline X fixed at 0.4

Conclusions

Summary

- TMLE has:
  - Low bias, MSE and well controlled type-1 error and coverage (using bootstrap CIs)
  - Performance on a par with 'the correct model' and superior to scenarios where you miss an important covariate in your modelling
- When using it, you don't need to worry about
  - Specifying what covariates need to be in the model
  - The functional relationship between covariates and outcome
  - Chance imbalance between treatment arms in covariates of interest
- It will always target a marginal treatment effect

Recommendations

- Avoid inference curve-based inference when sample size is small (N< 500). Best to run a simulation study first if sample size is very small (N< 100)
- Make sure to add SL.mean to your super learner library for treatment allocation
- Base TMLE limited in scope for complex estimands; Longitudinal TMLE is more flexible
- Be ready to pre-specify your prognostic variables
- Regulatory acceptance TBC

Conclusions

Opportunities for further work

- Explore TMLE in the context of
  - Effect modification
  - Binary outcomes
  - Simple missing data scenarios (using TMLE)
  - Complex missing data scenarios (using longitudinal TMLE)
  - Time to event endpoints (using surv TMLE)

Key References

Hernán, M.A. & Robins, J.M. (2020). ”Causal Inference: What If”. Boca Raton: *Chapman & Hall/CRC*.

van der Laan, Mark J. and Rubin, Daniel. "Targeted Maximum Likelihood Learning" *The International Journal of Biostatistics*, vol. 2, no. 1, 2006.

Gruber, Susan and van der Laan, Mark J., "Targeted Maximum Likelihood Estimation: A Gentle Introduction" (August 2009). *U.C. Berkeley Division of Biostatistics Working Paper Series.* Working Paper 252.

Tackney, M.S., Morris, T., White, I. *et al.* ”A comparison of covariate adjustment approaches under model misspecification in individually randomized trials.” *Trials* **24**, 14 (2023).

U.S. Food and Drug Administration. “Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products; Guidance for Industry.” Retrieved via *https://www.fda.gov/media/148910/download*

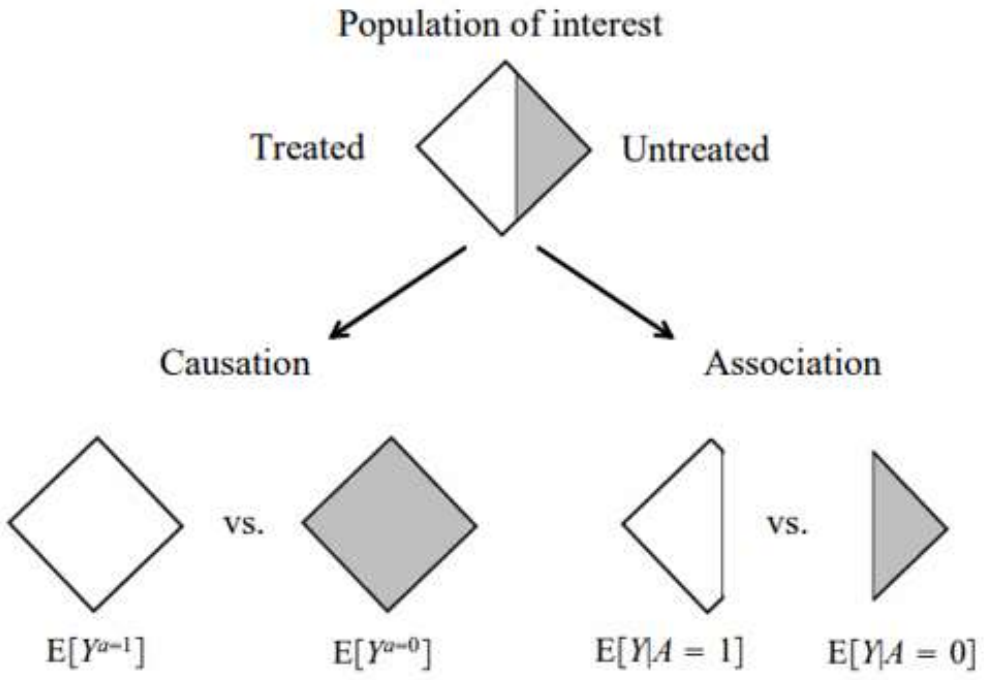
Causal Inference

Causation vs. Association

Causal Inference
Causal inference is explicit about both the causal question and the assumptions underlying the data analysis, and data generating model enabling the causal inferential conclusions.

Causation
Compares the same population under 2 different exposure/treatment/ intervention levels

Association
Compares 2 disjoint subsets of the population, as determined by their exposure/treatment/ intervention level



J&J

Hernan and Robins (2024). What if? Introduction: less casual on causal inferences.
Fig. 1.1 from Chapter 1 of What If?

Fundamental problem of Causal Inference

On the **individual level** only the response to one of the two possible exposures is observed.

Disease progression study

Did receiving therapy (A=1) prevent me from developing a worsened condition (Y=1) within 6 months?

What ***actually*** happened

- I received treatment and did not develop the worsened condition.
- My actual exposure was $A = 1$.
- My observed outcome was $Y \rightarrow Y^1 = 0$ (known).

What ***would have*** happened (contrary to fact)

- Had I not gotten treatment, would I have developed worsened condition?
- My counterfactual exposure is $A = 0$.
- My counterfactual outcome was Y^0 (unknown).

Identifying Assumptions

Exchangeability

Full or Marginal Exchangeability: Y^0 and Y^1 are independent of the treatment A.

$$Y^0, Y^1 \perp\!\!\!\perp A$$

Conditional Exchangeability: Given pre-treatment covariates X, Y^0 and Y^1 are independent of the treatment A.

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

→ among units with the same L values, think of A as being randomly assigned

→ guaranteed to hold in an ideal randomized experiment, without intercurrent events/LTFU, by virtue of randomization

also referred to in the literature as:

- "ignorability"
- "no unmeasured confounders" assumption

Violations

→ In an RCT, if A denotes the *randomized treatment assignment*, marginal or conditional exchangeability could be assumed depending on how randomization was done.

→ In an RCT, if A denotes the *actual treatment received*, marginal or conditional exchangeability may or may not hold.

Identifying Assumptions

Positivity

Full Positivity: The chance of receiving a treatment is not 0 or 1.

$$0 < P(A = a) < 1, \forall a$$

Conditional Positivity: Given pre-treatment covariates X, the chance of receiving treatment is not 0 or 1.

$$0 < P(A = a \mid L = l) < 1, \forall a, \forall l$$

Violations

→ treatment assignment is **not deterministic**, marginally or conditionally within the set of covariates

→ if positivity does not hold, we could have a situation where someone has nobody to be compared to (i.e., nobody "like them" who got the other treatment)