Improving Heterogeneous Treatment Effect Estimation with Data Borrowing under Covariates Mismatch

Samhita Pal

Joint work with Dr. Amir Asiaee and Dr. Jared Huling

Vanderbilt University Medical Center

ENAR 2026

Overview

Introduction

Motivation: Heterogeneous Treatment Effects in Trials

- Heterogeneous treatment effects (HTEs) describe how individual patient characteristics $\boldsymbol{X}$ modify the causal effect of treatment.

Motivation: Heterogeneous Treatment Effects in Trials

- Heterogeneous treatment effects (HTEs) describe how individual patient characteristics $\boldsymbol{X}$ modify the causal effect of treatment.
- Randomized controlled trials (RCTs) are the gold standard for causal inference.

Motivation: Heterogeneous Treatment Effects in Trials

- Heterogeneous treatment effects (HTEs) describe how individual patient characteristics $\boldsymbol{X}$ modify the causal effect of treatment.
- Randomized controlled trials (RCTs) are the gold standard for causal inference.
- However, **most RCTs are powered for average effects** rather than the fine-grained heterogeneity that guides individualized decisions.

Motivation: Heterogeneous Treatment Effects in Trials

- Heterogeneous treatment effects (HTEs) describe how individual patient characteristics $\boldsymbol{X}$ modify the causal effect of treatment.
- Randomized controlled trials (RCTs) are the gold standard for causal inference.
- However, **most RCTs are powered for average effects** rather than the fine-grained heterogeneity that guides individualized decisions.
- **Integrating information from large observational studies (OS)** has the potential to sharpen HTE estimation in RCTs.

Motivation: Heterogeneous Treatment Effects in Trials

- Heterogeneous treatment effects (HTEs) describe how individual patient characteristics $\boldsymbol{X}$ modify the causal effect of treatment.
- Randomized controlled trials (RCTs) are the gold standard for causal inference.
- However, **most RCTs are powered for average effects** rather than the fine-grained heterogeneity that guides individualized decisions.
- **Integrating information from large observational studies (OS)** has the potential to sharpen HTE estimation in RCTs.
- **Problem:** Integration of RCT and OS is not straight forward since
  - OS are vulnerable to confounding;
  - OS and RCT often observe different (overlapping) sets of covariates.

RCT and OS integration

- **Goal:** Efficient and consistent estimation of the conditional average treatment effect (CATE) in the trial population,

$$\tau^r(\boldsymbol{x}) = \mathbb{E}\big\{ Y(1) - Y(-1) \mid \boldsymbol{X} = \boldsymbol{x},\ S = r \big\},$$

where $Y(a)$ is the potential outcome under treatment $a \in \{-1, 1\}$ and $S = r$ indicates the RCT.

RCT and OS integration

- **Goal:** Efficient and consistent estimation of the conditional average treatment effect (CATE) in the trial population,

$$\tau^r(\boldsymbol{x}) = \mathbb{E}\big\{ Y(1) - Y(-1) \mid \boldsymbol{X} = \boldsymbol{x},\ S = r \big\},$$

  where $Y(a)$ is the potential outcome under treatment $a \in \{-1, 1\}$ and $S = r$ indicates the RCT.
- **Solution:** Develop a family of calibrated regression estimators:
  - **R-OSCAR**: borrows from both OS arms to improve efficiency over the RCT-only baseline RACER when covariates are shared;

RCT and OS integration

- **Goal:** Efficient and consistent estimation of the conditional average treatment effect (CATE) in the trial population,

$$\tau^r(\boldsymbol{x}) = \mathbb{E}\big\{ Y(1) - Y(-1) \mid \boldsymbol{X} = \boldsymbol{x},\ S = r \big\},$$

  where $Y(a)$ is the potential outcome under treatment $a \in \{-1, 1\}$ and $S = r$ indicates the RCT.
- **Solution:** Develop a family of calibrated regression estimators:
  - **R-OSCAR**: borrows from both OS arms to improve efficiency over the RCT-only baseline RACER when covariates are shared;
  - **MR-OSCAR**: additionally leverages OS-only covariates under covariate mismatch to further improve on both RACER and R-OSCAR.

Innovation

- Prior work has mostly focused on borrowing **control arms** to reduce the variance of the treatment effect.
- Existing methods often assume that **CATEs are transportable** between populations.
- **Gap in the literature:** methods for efficient CATE estimation in an RCT that
  - reliably leverage **both treatment arms** from the OS, and
  - relax the **transportability** assumption.

RACER: a method for estimating the CATE from RCT data

Pseudo-outcomes: turning CATE into a regression problem

- Idea: construct a pseudo-outcome whose conditional mean is the CATE in the RCT, then estimate that conditional mean by regression.
- Given augmentation $m(X)$ and randomization probabilities $\pi_A(X)$, define

$$\tau_m(X, A, Y) = \frac{A\{Y - m(X)\}}{\pi_A(X)}.$$

Pseudo-outcomes: turning CATE into a regression problem

- Idea: construct a pseudo-outcome whose conditional mean is the CATE in the RCT, then estimate that conditional mean by regression.
- Given augmentation $m(X)$ and randomization probabilities $\pi_A(X)$, define

$$\tau_m(X, A, Y) = \frac{A\{Y - m(X)\}}{\pi_A(X)}.$$

- **Key fact 1:** in the RCT, $\mathbb{E}[\tau_m(X, A, Y) \mid X] = \tau^r(X)$.

Pseudo-outcomes: turning CATE into a regression problem

- Idea: construct a pseudo-outcome whose conditional mean is the CATE in the RCT, then estimate that conditional mean by regression.
- Given augmentation $m(X)$ and randomization probabilities $\pi_A(X)$, define

$$\tau_m(X, A, Y) = \frac{A\{Y - m(X)\}}{\pi_A(X)}.$$

- **Key fact 1:** in the RCT, $\mathbb{E}[\tau_m(X, A, Y) \mid X] = \tau^r(X)$.
- **Key fact 2:** under SUTVA, conditional ignorability and weak positivity, $\arg\min_m \mathrm{var}[\tau_m(X, A, Y) \mid X] = \pi^r_{-1}(x)\,\mu^r_{+1}(x) \; + \; \pi^r_{+1}(x)\,\mu^r_{-1}(x)$

Pseudo-outcomes: turning CATE into a regression problem

- Idea: construct a pseudo-outcome whose conditional mean is the CATE in the RCT, then estimate that conditional mean by regression.
- Given augmentation $m(X)$ and randomization probabilities $\pi_A(X)$, define

$$\tau_m(X, A, Y) = \frac{A\{Y - m(X)\}}{\pi_A(X)}.$$

- **Key fact 1:** in the RCT, $\mathbb{E}[\tau_m(X, A, Y) \mid X] = \tau^r(X)$.
- **Key fact 2:** under SUTVA, conditional ignorability and weak positivity, $\arg\min_m \text{var}[\tau_m(X, A, Y) \mid X] = \pi^r_{-1}(x)\,\mu^r_{+1}(x) \; + \; \pi^r_{+1}(x)\,\mu^r_{-1}(x)$ = conditional mean outcome (CMO).

Pseudo-outcomes: turning CATE into a regression problem

- Idea: construct a pseudo-outcome whose conditional mean is the CATE in the RCT, then estimate that conditional mean by regression.
- Given augmentation $m(X)$ and randomization probabilities $\pi_A(X)$, define

$$\tau_m(X, A, Y) = \frac{A\{Y - m(X)\}}{\pi_A(X)}.$$

- **Key fact 1:** in the RCT, $\mathbb{E}[\tau_m(X, A, Y) \mid X] = \tau^r(X)$.
- **Key fact 2:** under SUTVA, conditional ignorability and weak positivity, $\arg\min_m \text{var}[\tau_m(X, A, Y) \mid X] = \pi^r_{-1}(x)\,\mu^r_{+1}(x) \; + \; \pi^r_{+1}(x)\,\mu^r_{-1}(x)$ = conditional mean outcome (CMO).
- **Key fact 3:** for any x, the true CATE $\tau(x)$ minimizes the following conditional risk: $\tau(x) = \arg\min_f E_{Y,A}[(\tau_m(x, A, Y) - f(x))^2 \mid X = x]$.

Estimation as Empirical Risk Minimizer.

$$\hat{\tau}^r(\cdot) = \arg\min_{f \in \mathcal{F}} \frac{1}{n^r} \sum_{i=1}^{n^r} \left\{ \tau_m(X_i^r, A_i^r, Y_i^r) - f(X_i^r) \right\}^2.$$

$$m(x) := \hat{\mu}^r(x) = \sum_a \pi^r_{-a}(x) \hat{\mu}^r_a(x), \quad \hat{\mu}^r_a(x) \in \mathcal{M}^r_a,$$

- At the population level the minimizer is $\tau^r(\cdot)$, so this is a standard regression of pseudo-outcomes on X.
- Crucially, the estimator's risk depends on the choice of augmentation function m.
- Our key insight is that the risk minimizing choice is the CMO.
- This shows that a more accurate estimation of the CMO (as m) reduces CATE estimation error.

R-OSCAR: A Method for Borrowing Data from an OS for CATE Estimation with Perfect Covariate Alignment

- Since the CMO is a weighted average of $\mu^r_{+1}(x)$ and $\mu^r_{-1}(x)$, improving its estimation requires accurate estimation of these per-arm outcomes.

R-OSCAR: **R**obust **O**bservational **S**tudies for **C**MO-**A**ugmented **R**CT

- Since the CMO is a weighted average of $\mu^r_{+1}(x)$ and $\mu^r_{-1}(x)$, improving its estimation requires accurate estimation of these per-arm outcomes.
- R-OSCAR allows us to improve estimates of $\mu^r_a(x)$, and thus the CMO, by appropriately correcting for systematic differences between OS and RCT data, without introducing bias in the CATE due to confounding.

R-OSCAR: **R**obust **O**bservational **S**tudies for **C**MO-**A**ugmented **R**CT

- Since the CMO is a weighted average of $\mu^r_{+1}(x)$ and $\mu^r_{-1}(x)$, improving its estimation requires accurate estimation of these per-arm outcomes.
- R-OSCAR allows us to improve estimates of $\mu^r_a(x)$, and thus the CMO, by appropriately correcting for systematic differences between OS and RCT data, without introducing bias in the CATE due to confounding.
- A direct but potentially suboptimal approach would set $m(x) := \sum_a \pi^r_{-a}(x)\hat{\mu}^o(x, a)$, where $\hat{\mu}^o(x, a)$ is an outcome mean estimate derived from OS data and applied to RCT covariates.

R-OSCAR: Robust Observational Studies for CMO-Augmented RCT

- Since the CMO is a weighted average of $\mu^r_{+1}(x)$ and $\mu^r_{-1}(x)$, improving its estimation requires accurate estimation of these per-arm outcomes.
- R-OSCAR allows us to improve estimates of $\mu^r_a(x)$, and thus the CMO, by appropriately correcting for systematic differences between OS and RCT data, without introducing bias in the CATE due to confounding.
- A direct but potentially suboptimal approach would set $m(x) := \sum_a \pi^r_{-a}(x)\hat{\mu}^o(x,a)$, where $\hat{\mu}^o(x,a)$ is an outcome mean estimate derived from OS data and applied to RCT covariates.
- But, $\forall a, x : \mu^r_a(x) \neq \mu^o(x,a)$.

R-OSCAR: Robust Observational Studies for CMO-Augmented RCT

- Since the CMO is a weighted average of $\mu^r_{+1}(x)$ and $\mu^r_{-1}(x)$, improving its estimation requires accurate estimation of these per-arm outcomes.
- R-OSCAR allows us to improve estimates of $\mu^r_a(x)$, and thus the CMO, by appropriately correcting for systematic differences between OS and RCT data, without introducing bias in the CATE due to confounding.
- A direct but potentially suboptimal approach would set $m(x) := \sum_a \pi^r_{-a}(x)\hat{\mu}^o(x,a)$, where $\hat{\mu}^o(x,a)$ is an outcome mean estimate derived from OS data and applied to RCT covariates.
- But, $\forall a, x : \mu^r_a(x) \neq \mu^o(x,a)$.
- To account for this bias, we introduce an outcome mean discrepancy defined as $\delta_a(x) \equiv \mu^r_a(x) - \mu^o(x,a) \implies \mu^r_a(x) = \delta_a(x) + \mu^o(x,a)$.

R-OSCAR Stage 1: OS Outcome Model

Input: OS data $\{(\boldsymbol{X}_j, A_j, Y_j)\}_{j=1}^{n_o}$.

For each arm $a \in \{-1, +1\}$:

$$\underbrace{\hat{\mu}_a^o(\boldsymbol{x})}_{\text{estimate OS predictor}} = \arg\min_{\mu \in \mathcal{M}_a^o} \frac{1}{n_a^o} \sum_{\substack{i \in OS: \\ A_i = a}} \big(Y_i - \mu(\boldsymbol{X}_i)\big)^2$$

- Learns outcome-covariate relationship from the large OS
- May be biased for the RCT population due to distributional differences
- In practice: ridge regression, random forest, or neural network
- **Next:** Learn a discrepancy function from RCT data to correct the OS predictions.

R-OSCAR Stage 2: Outcome Calibration

$$\underbrace{\hat{\delta}_a^t}_{\text{RCT correction / transport}} \in \arg\min_{d \in \mathcal{D}_a} \frac{1}{n_a^r} \sum_{\substack{i \in RCT: \\ A_i = a}} \Big(Y_i - \underbrace{\hat{\mu}_a^o(\boldsymbol{X}_i)}_{\substack{\text{OS fit} \\ \text{evaluated on RCT X}}} - d(\boldsymbol{X}_i) \Big)^2 + \underbrace{\mathcal{P}_{\text{cal}}(d)}_{\substack{\text{shrink discrepancy} \\ \text{toward 0}}}$$

The calibrated prediction: $\underbrace{\hat{\mu}_a^{\text{cal}}(\boldsymbol{x})}_{\text{OS + correction}} = \hat{\mu}_a^o(\boldsymbol{x}) + \underbrace{\hat{\delta}_a^t(\boldsymbol{x})}_{\text{add discrepancy}}$

- If OS model is nearly correct for RCT: $\underbrace{\hat{\delta}_a^t \approx 0}_{\text{strong borrowing}}$
- If OS model is poor: $\underbrace{\hat{\delta}_a^t}_{\text{absorbs mismatch}}$

R-OSCAR Stages 3-4: CMO and CATE Calibration I

Stage 3: Form the calibrated CMO and pseudo-outcomes:

$$\underbrace{\hat{m}(\boldsymbol{x})}_{\text{baseline mean (CMO)}} = \sum_a \pi^r_{-a}(\boldsymbol{x})\, \hat{\mu}^{\text{cal}}_a(\boldsymbol{x})$$

$$\underbrace{\hat{\psi}_i}_{\text{RCT pseudo-outcome}} = \frac{A_i\big(Y_i - \hat{m}(\boldsymbol{X}_i)\big)}{\pi^r_{A_i}(\boldsymbol{X}_i)}$$

Stage 4: The preliminary CATE is

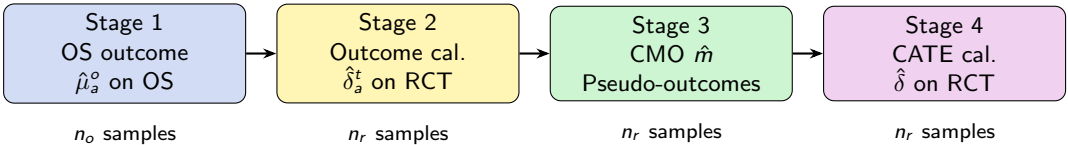
$$\underbrace{\tilde{\tau}(\boldsymbol{x})}_{\text{difference in calibrated means}} = \sum_a a\, \hat{\mu}^{\text{cal}}_a(\boldsymbol{x})$$

R-OSCAR Stages 3-4: CMO and CATE Calibration II

Estimate a CATE correction:

$$\underbrace{\hat{\delta}}_{\text{final RCT calibration}} \in \arg\min_{d \in \mathcal{D}} \frac{1}{n_r} \sum_{i \in RCT} \left(\hat{\psi}_i - \tilde{\tau}(\boldsymbol{X}_i) - d(\boldsymbol{X}_i)\right)^2 + \mathcal{P}_\tau(d)$$

$$\underbrace{\hat{\tau}_{\text{R-OSCAR}}(\boldsymbol{x})}_{\text{final CATE}} = \tilde{\tau}(\boldsymbol{x}) + \underbrace{\hat{\delta}(\boldsymbol{x})}_{\text{add correction}}$$



MR-OSCAR: A Method for Borrowing Data from an OS for CATE Estimation under Covariate Mismatch

Two Data Sources, Three Covariate Blocks

RCT ($S = r$, small n_r):

- Randomized treatment $A \in \{-1, +1\}$
- Observes $(\boldsymbol{U}, \boldsymbol{Z})$ but **not** $\boldsymbol{V}$

OS ($S = o$, large n_o):

- Confounded treatment assignment
- Observes $(\boldsymbol{Z}, \boldsymbol{V})$ but **not** $\boldsymbol{U}$

$\boldsymbol{U} \in \mathbb{R}^{p_u}$ RCT only

$\boldsymbol{Z} \in \mathbb{R}^{p_z}$ Shared

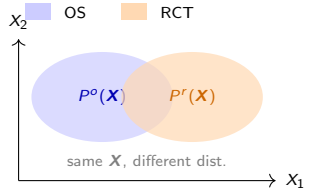
$\boldsymbol{V} \in \mathbb{R}^{p_v}$ OS only

Observed Data

$$\text{RCT:} \quad \{(\boldsymbol{U}_i, \boldsymbol{Z}_i, A_i, Y_i)\}_{i=1}^{n_r}$$
$$\text{OS:} \quad \{(\boldsymbol{Z}_j, \boldsymbol{V}_j, A_j, Y_j)\}_{j=1}^{n_o}$$

Covariate Mismatch $\neq$ Covariate Shift

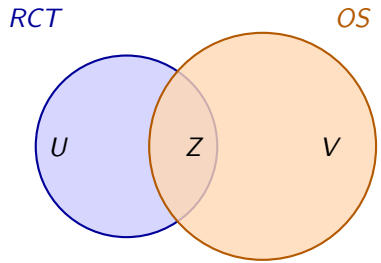
Covariate Shift



Same covariates observed in both sources;
$P^o(\boldsymbol{X}) \neq P^r(\boldsymbol{X})$

Issue: distribution changes, not variables.

Covariate Mismatch (our setting)



Issue: different feature spaces

The Core Challenge

Outcome models trained on $(\boldsymbol{Z}, \boldsymbol{V})$ in the OS cannot be evaluated on RCT units where $\boldsymbol{V}$ is missing.

The CATE Target

Our goal is to estimate the conditional average treatment effect in the RCT population:

CATE Definition

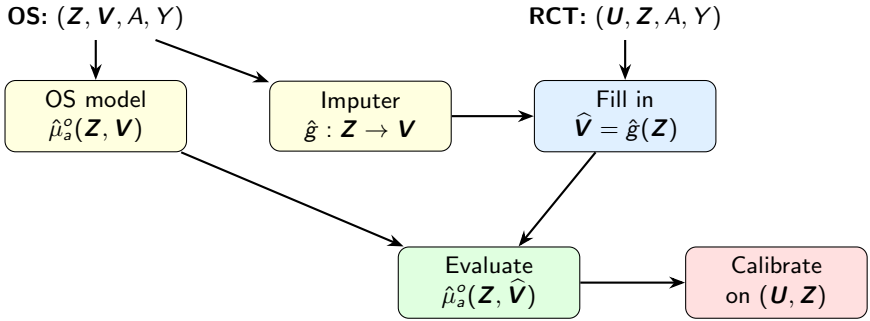
$$\tau^r(\boldsymbol{u}, \boldsymbol{z}) = \mathbb{E}\big[Y(1) - Y(-1) \mid \boldsymbol{U} = \boldsymbol{u}, \boldsymbol{Z} = \boldsymbol{z}, S = r\big]$$

Key points:

- The CATE is defined on $(\boldsymbol{U}, \boldsymbol{Z})$—the covariates available in the RCT
- It **marginalizes over $\boldsymbol{V}$**, since $\boldsymbol{V}$ is never observed in the trial
- Causal identification comes entirely from randomization $(A \perp\!\!\!\perp Y(a) \mid \boldsymbol{U}, \boldsymbol{Z}, S = r)$
- The OS enters only as an *auxiliary predictor*—it does not provide causal identification

MR-OSCAR: **M**ismatch-aware **R**obust **O**bservational **S**tudies for **C**MO-**A**ugmented **R**CT

Idea: Predict the missing $\boldsymbol{V}$ from the shared $\boldsymbol{Z}$, then run R-OSCAR as if $\boldsymbol{V}$ were observed.



Numerical Results

Simulation setup

- Two sources with partially overlapping covariates:
  $\boldsymbol{X}_{\text{RCT}} = (\boldsymbol{U}, \boldsymbol{Z}), \boldsymbol{X}_{\text{OS}} = (\boldsymbol{Z}, \boldsymbol{V}), p_{\text{tot}} = 100.$
- Sample sizes: $n_r = 300$ (RCT) and $n_o = 1000$ (OS), unless noted.
- Treatment:
  - RCT: completely randomized, $\Pr(A = 1) = 0.5$.
  - OS: logistic assignment using a subset of $\boldsymbol{Z}$; treated proportion $\approx 1/3$.
- Methods compared:
  - **RACER**: RCT-only CATE via outcome regression + calibration on $(\boldsymbol{U}, \boldsymbol{Z})$.
  - **SR-OSCAR**: borrows from OS using only shared covariates $\boldsymbol{Z}$.
  - **MR-OSCAR**: augments RCT with imputed $\widehat{\boldsymbol{V}}$ learned from $(\boldsymbol{Z}, \boldsymbol{V})$ in the OS and calibrates to the RCT.

Simulation results: when is borrowing helpful?

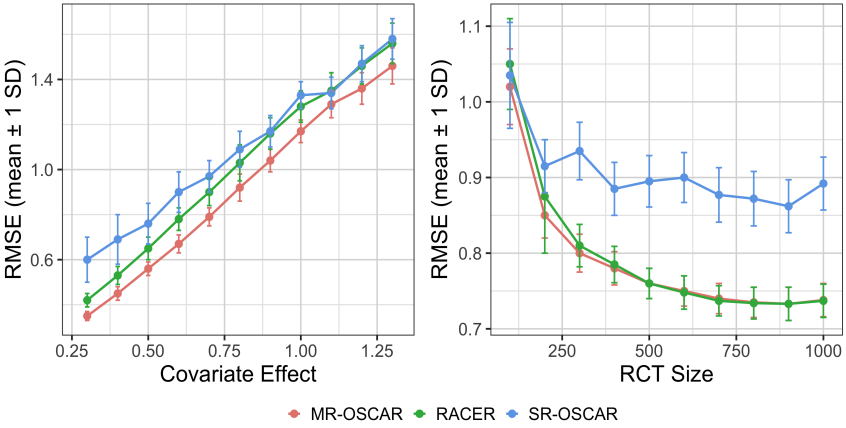
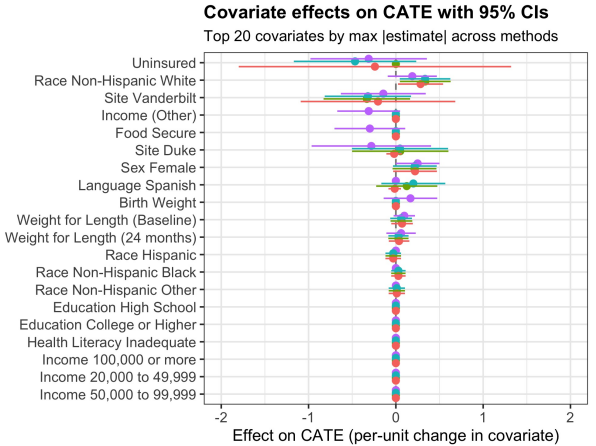


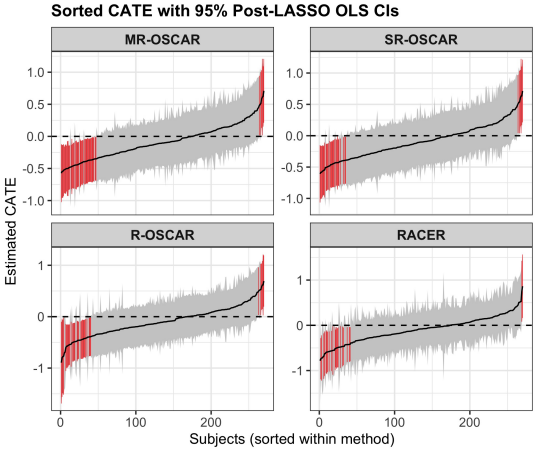
Figure: **Left:** *Mean RMSE (points) with* ± 1 *SD (bars) versus the effect size of the OS-only covariates* $\boldsymbol{V}$. **Right:** *Mean RMSE versus RCT sample size* n_r *(OS fixed at* $n_o = 1000$*).*

Greenlight Plus study: design and covariate mismatch

- Pediatric obesity prevention trial (**Greenlight Plus**):
  - Outcome: weight-for-length z-score at 12 months.
  - Trial covariates include demographics, baseline anthropometrics, socioeconomic and literacy measures.
- External EHR dataset:
  - Control patients only ($A = 0$); rich baseline covariates including **insurance status** V.
- We deliberately impose covariate mismatch:
  - Mask binary insurance V in the RCT and treat it as OS-only.
  - Use EHR controls to learn $\Pr(V = 1 \mid Z)$ and impute $\widehat{V}$ in the trial (MR-OSCAR).
- Methods compared on these data:
  - **RACER**: RCT-only baseline using all observed trial covariates.
  - **SR-OSCAR**: R-OSCAR applied using only shared covariates Z.
  - **MR-OSCAR**: imputation-augmented version using $(Z, \widehat{V})$ in the RCT.
  - **R-OSCAR (oracle)**: uses (Z, V) in both EHR and RCT; represents a gold-standard scenario where insurance is observed in the trial.

Greenlight Plus: CATE uncertainty and covariate effects





Conclusion

Papers

ArXiv R-OSCAR paper

Google Scholar profile