Design considerations for subgroup analyses in cluster-randomized trials

Brian D. Williamson, PhD

Kaiser Permanente Washington Health Research Institute

17 July, 2024

https://bdwilliamson.github.io/#talks

Acknowledgments

This work was done in collaboration with:

Yates Coley
KAISER PERMANENTE.
Kaiser Permanente Washington
Health Research Institute

Clarissa Hsu
KAISER PERMANENTE.
Kaiser Permanente Washington
Health Research Institute

Courtney McCracken
KAISER PERMANENTE.
CRE Center for Research and Evaluation

Andrea Cook
KAISER PERMANENTE.
Kaiser Permanente Washington
Health Research Institute

Heterogeneity of treatment effects: what and why

Heterogeneity of treatment effects:

- differential effect of exposure/intervention across subgroups
- does not necessarily imply a causal or biological interpretation

Heterogeneity of treatment effects: what and why

Heterogeneity of treatment effects:

- differential effect of exposure/intervention across subgroups
- does not necessarily imply a causal or biological interpretation

If HTE is present:

- resources can be prioritized to maximize benefit and reduce risks [Knol and VanderWeele (2012)]
- can determine whether intervention remediates or exacerbates health disparities [Petkovic et al. (2020)]

Case study: ENSPIRE trial

ENSPIRE: Engaging Staff to Improve COVID-19 Vaccination Response at Long-term Care Facilities

Background:

- Rates of COVID-19 disease and death higher in long-term care than general population [Chidambaram (2022)]
- Lag in COVID-19 vaccinations among long-term care workers
- Long-term care (LTC) workers often low-wage, represent diverse social, cultural, racial, and ethnic backgrounds

Case study: ENSPIRE trial

ENSPIRE: Engaging Staff to Improve COVID-19 Vaccination Response at Long-term Care Facilities

Background:

- Rates of COVID-19 disease and death higher in long-term care than general population [Chidambaram (2022)]
- Lag in COVID-19 vaccinations among long-term care workers
- Long-term care (LTC) workers often low-wage, represent diverse social, cultural, racial, and ethnic backgrounds

Design:

- Enrolled 40 LTC centers, encompassing approximately 4000 staff members
- Recruited from 2 US states (Georgia and Washington)
- Recruited from urban (Atlanta or Seattle) and suburban/rural locations

Case study: ENSPIRE trial

Primary outcomes:

1. LTC center COVID-19 booster rate (cluster-level)
   - chosen due to concerns about data capture

2. LTC care staff net promoter score (individual-level)

Intervention: booster promotion materials co-designed with LTC staff

"Usual care"
(Standard materials)

vs.

"Co-design"
(Materials designed by teams of staff)

Case study: ENSPIRE trial

Co-design: tailoring messages to specific audiences

- In ENSPIRE, focused on language, cultural, and ethnic affinity groups
- Measured race and ethnicity allowing 57 total options

Case study: ENSPIRE trial

Co-design: tailoring messages to specific audiences

- In ENSPIRE, focused on language, cultural, and ethnic affinity groups
- Measured race and ethnicity allowing 57 total options

HTE analysis natural for interventions, particularly co-designed:

- Planned analyses based on state, urban vs suburban/rural, baseline booster uptake
- Self-reported race and ethnicity both interesting and important, however:
  - neither sites nor other entity had comprehensive and accurate race and ethnicity data for staff
  - staff survey would only capture a portion of staff members
  - also, didn't know what this would look like with a cluster-level outcome!

Individual–level data structure

Data on n_c staff members in each cluster $c = 1, \ldots, K$:

- A_{ci}: indicator of intervention assignment (cluster–level)
- $Z_{ci} \in \mathbb{R}^q$: cluster–level subgroup variables
- $W_{ci} \in \mathbb{R}^p$: baseline covariates
- Y_{ci}: binary outcome

HTE with individual–level outcomes

Suppose we are interested in HTE by a single binary variable W_{ci}.

Target estimand: the difference in the risk difference based on levels of W_{ci}.

Estimate using linear regression:

$$E(Y_{ci} \mid A_{ci} = a, Z_{ci} = z, W_{ci} = w) = \beta_0 + \beta_1 a + \beta_2 z + \beta_3 w + \beta_4 aw.$$

HTE with individual–level outcomes

Suppose we are interested in HTE by a single binary variable W_{ci}.

Target estimand: the difference in the risk difference based on levels of W_{ci}.

Estimate using linear regression:

$$E(Y_{ci} \mid A_{ci} = a, Z_{ci} = z, W_{ci} = w) = \beta_0 + \beta_1 a + \beta_2 z + \beta_3 w + \beta_4 a w.$$

Under standard assumptions, β_4 quantifies HTE.

Cluster–level data structure

In many settings, we observe cluster–level data:

- $\overline{Y}_c := \frac{1}{n_c} \sum_{i=1}^{n_c} Y_{ci}$: cluster-level outcome proportion

- A_c: indicator of intervention assignment

- $Z_c \in \mathbb{R}^q$: cluster-level subgroup variables

- $X_c \in \mathbb{R}^p$: (possibly aggregated) baseline covariates

Cluster–level data structure

In many settings, we observe cluster–level data:

- $\overline{Y}_c := \frac{1}{n_c} \sum_{i=1}^{n_c} Y_{ci}$: cluster–level outcome proportion

- A_c: indicator of intervention assignment
- $Z_c \in \mathbb{R}^q$: cluster–level subgroup variables
- $X_c \in \mathbb{R}^p$: (possibly aggregated) baseline covariates

Possible options for aggregating an individual–level covariate W_{cj}:

- $X_{cj} = V_{cj} - \frac{1}{K} \sum_{c=1}^{K} V_{cj}$, where $V_{cj} = \frac{1}{n_c} \sum_{i=1}^{n_c} W_{cji}$
- $X_{cj} = I\left(V_{cj} > t\right)$ for threshold t

Aggregated data: complications

Individual–level effects may not be identifiable based on aggregate data

- Often referred to as ecological bias [Wakefield (2008)]

- Exposure–outcome association can be biased: [Greenland & Morgenstern (1989)]
    - by aggregated confounders
    - by aggregated HTE variables

Aggregated data: complications

Individual–level effects may not be identifiable based on aggregate data

- Often referred to as ecological bias [Wakefield (2008)]

- Exposure–outcome association can be biased: [Greenland & Morgenstern (1989)]
  - by aggregated confounders
  - by aggregated HTE variables

Compositional data cause challenges in interpretation:

- Arise when aggregating mutually–exclusive binary variables

- Changing the value of one proportion necessarily changes at least one other

HTE with cluster–level outcomes

Now, outcome $\overline{Y}_c$ is a proportion; interest in HTE by single variable X_c.

Target estimand: difference in difference of proportions based on values of X_c.

Estimate using linear regression:

$$E(\overline{Y}_c \mid A_c = a, Z_c = z, X_c = x) = \alpha_0 + \alpha_1 a + \alpha_2 z + \alpha_3 x + \alpha_4 ax.$$

HTE with cluster–level outcomes

Now, outcome $\overline{Y}_c$ is a proportion; interest in HTE by single variable X_c.

Target estimand: difference in difference of proportions based on values of X_c.

Estimate using linear regression:

$$E(\overline{Y}_c \mid A_c = a, Z_c = z, X_c = x) = \alpha_0 + \alpha_1 a + \alpha_2 z + \alpha_3 x + \alpha_4 ax.$$

Under standard assumptions, α_4 quantifies HTE.

HTE with cluster-level outcomes

Now, outcome $\overline{Y}_c$ is a proportion; interest in HTE by single variable X_c.

Target estimand: difference in difference of proportions based on values of X_c.

Estimate using linear regression:

$$E(\overline{Y}_c \mid A_c = a, Z_c = z, X_c = x) = \alpha_0 + \alpha_1 a + \alpha_2 z + \alpha_3 x + \alpha_4 ax.$$

Under standard assumptions, α_4 quantifies HTE.

If W represents a single binary variable, then aggregation determines interpretation:

- Using mean–centered proportion: difference in difference for 1–unit increase from the across–cluster mean proportion
- Using threshold: difference in difference comparing "high" to "low"

HTE with cluster-level outcomes

Now, outcome $\overline{Y}_c$ is a proportion; interest in HTE by single variable X_c.

Target estimand: difference in difference of proportions based on values of X_c.

Estimate using linear regression:

$$E(\overline{Y}_c \mid A_c = a, Z_c = z, X_c = x) = \alpha_0 + \alpha_1 a + \alpha_2 z + \alpha_3 x + \alpha_4 ax.$$

Under standard assumptions, α_4 quantifies HTE.

If W represents a single binary variable, then aggregation determines interpretation:

- Using mean–centered proportion: difference in difference for 1–unit increase from the across–cluster mean proportion
- Using threshold: difference in difference comparing "high" to "low"

If W is compositional, this is further complicated:

- how are the categories transformed to enable model fitting?
- what does a one–unit change imply?

Simulation setup

Data mimic ENSPIRE trial:

- Sample $K \in \{30, 40, 80\}$ clusters within two regions

- Clusters are urban or suburban/rural

- Randomization constrained to balance on region, urban vs suburban/rural

- Each center has 100 staff members, who each:
  - have record of receiving COVID-19 booster vaccine

  - have self-reported race (recorded as single binary variable)

- Define cluster-level booster rate $\overline{Y}_c$

- Define two aggregations of self-reported race:
  - center-level proportion

 - indicator of whether proportion > 0.5

Simulation setup

Data mimic ENSPIRE trial:

- Sample $K \in \{30, 40, 80\}$ clusters within two regions

- Clusters are urban or suburban/rural

- Randomization constrained to balance on region, urban vs suburban/rural

- Each center has 100 staff members, who each:
    - have record of receiving COVID-19 booster vaccine

    - have self-reported race (recorded as single binary variable)

- Define cluster-level booster rate $\overline{Y}_c$

- Define two aggregations of self-reported race:
    - center-level proportion

    - indicator of whether proportion > 0.5

Fit both individual-level and cluster-level regression models

Tested for HTE by self-reported race, investigated type I error and power

Results: type I error

Type I error for HTE by self-reported race

Interaction variable(s)
- ● Self-reported race
- ▲ Self-reported race and urban status
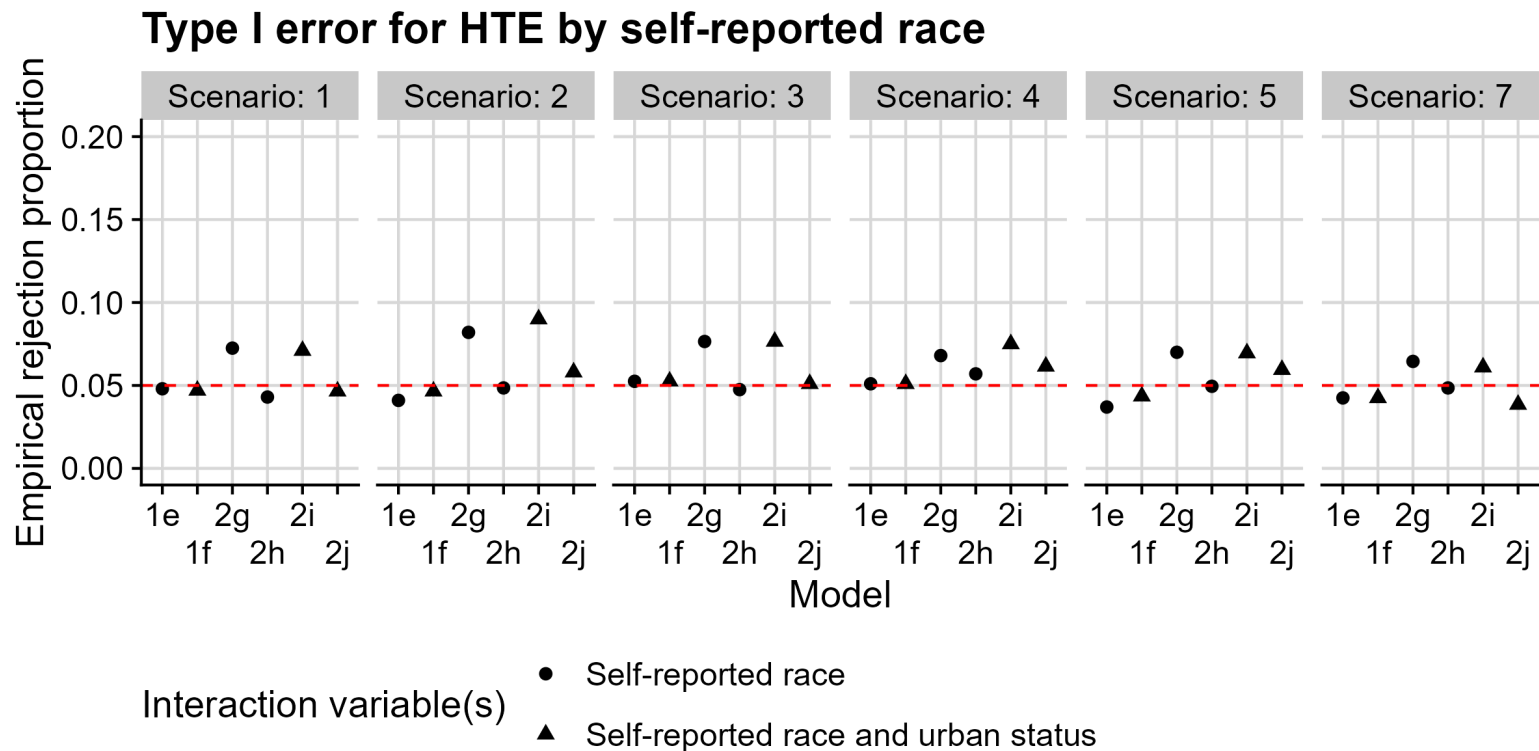
Scenarios encode presence of true differences in booster rate or HTE:

- 1, 2, 5: no difference in booster rate by self–reported race, no HTE
- 3, 4, 7: difference in booster rate by self–reported race, no HTE
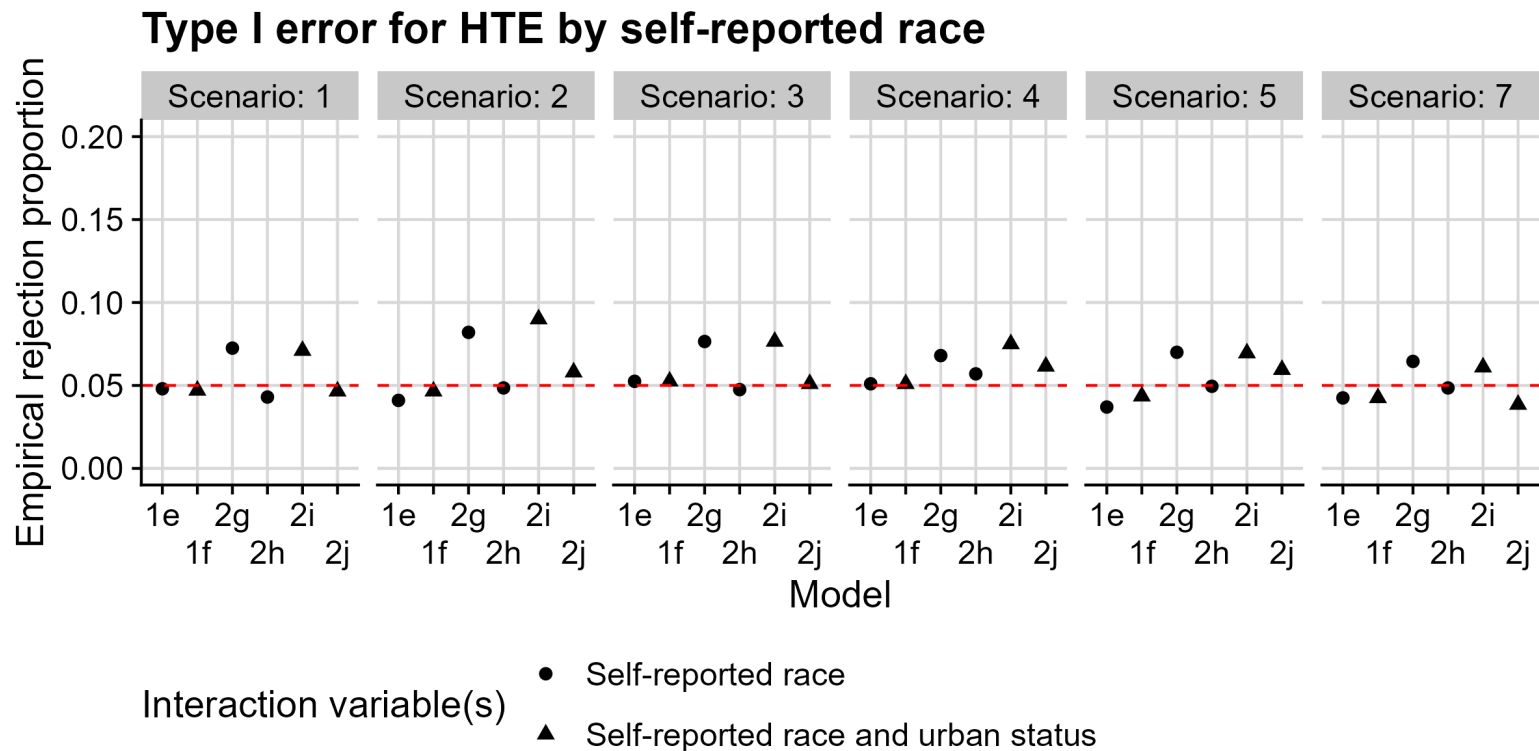
13

Results: type I error

Type I error for HTE by self-reported race

Interaction variable(s)
● Self-reported race
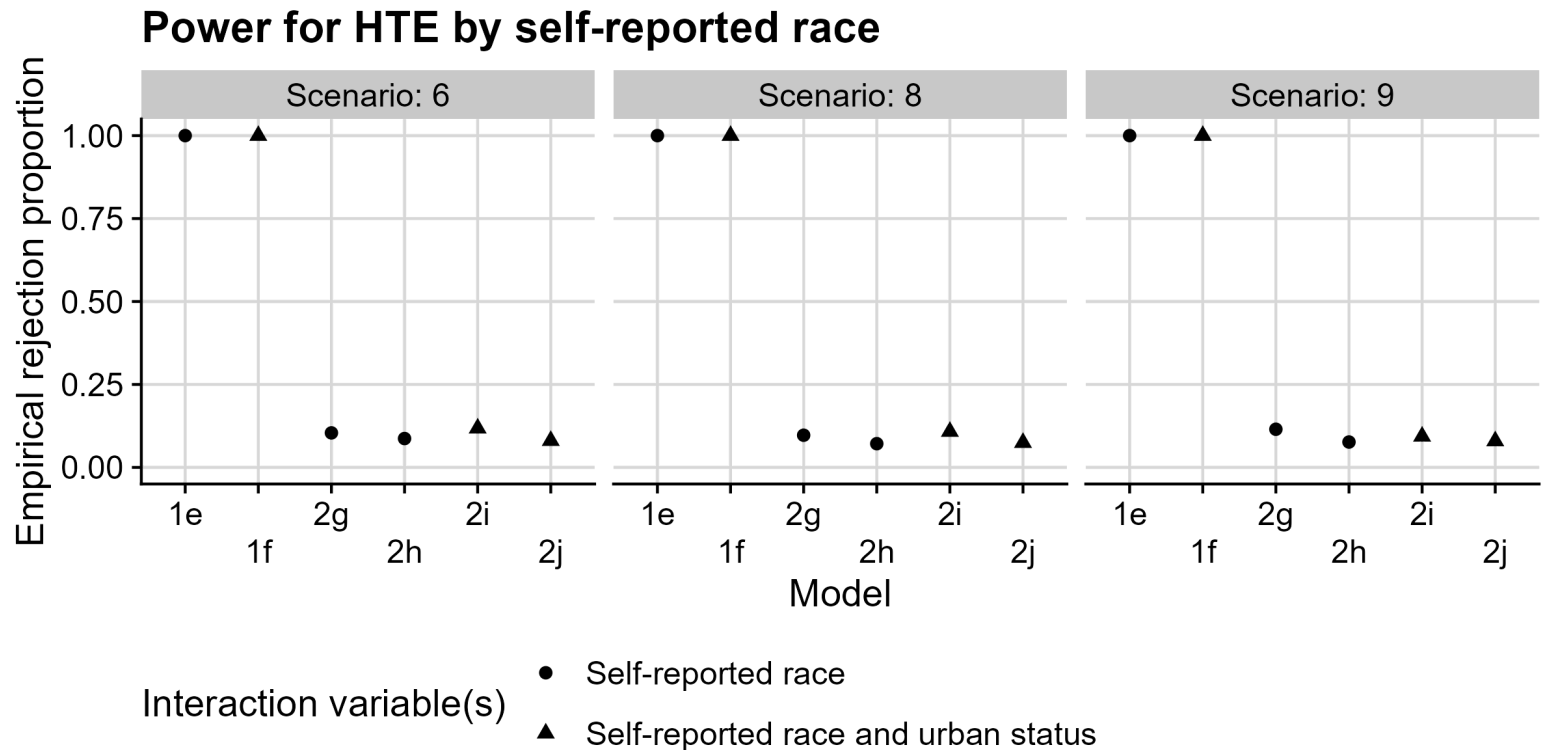▲ Self-reported race and urban status

Models encode individual– (1) vs cluster–level (2), adjustment and interaction variables:

- 1e, 1f: adjust for self–reported race, allow HTE by self–reported race (1f also by urban)

- 2g, 2i: adjust for center–level race proportion, allow HTE by self–reported race (2i also by urban)

- 2h, 2j: adjust for proportion > 0.5, allow HTE by self–reported race (2j also by urban)

14

Results: power

Power for HTE by self-reported race

Interaction variable(s)
- ● Self-reported race
- ▲ Self-reported race and urban status

Scenarios encode presence of true differences in booster rate or HTE:

- 6: difference in booster rate by self–reported race, HTE by self–reported race
- 8: difference in booster rate by urban and self–reported race, HTE by self–reported race
- 9: difference in booster rate by urban and self–reported race, HTE by urban and self–reported race

Results: power
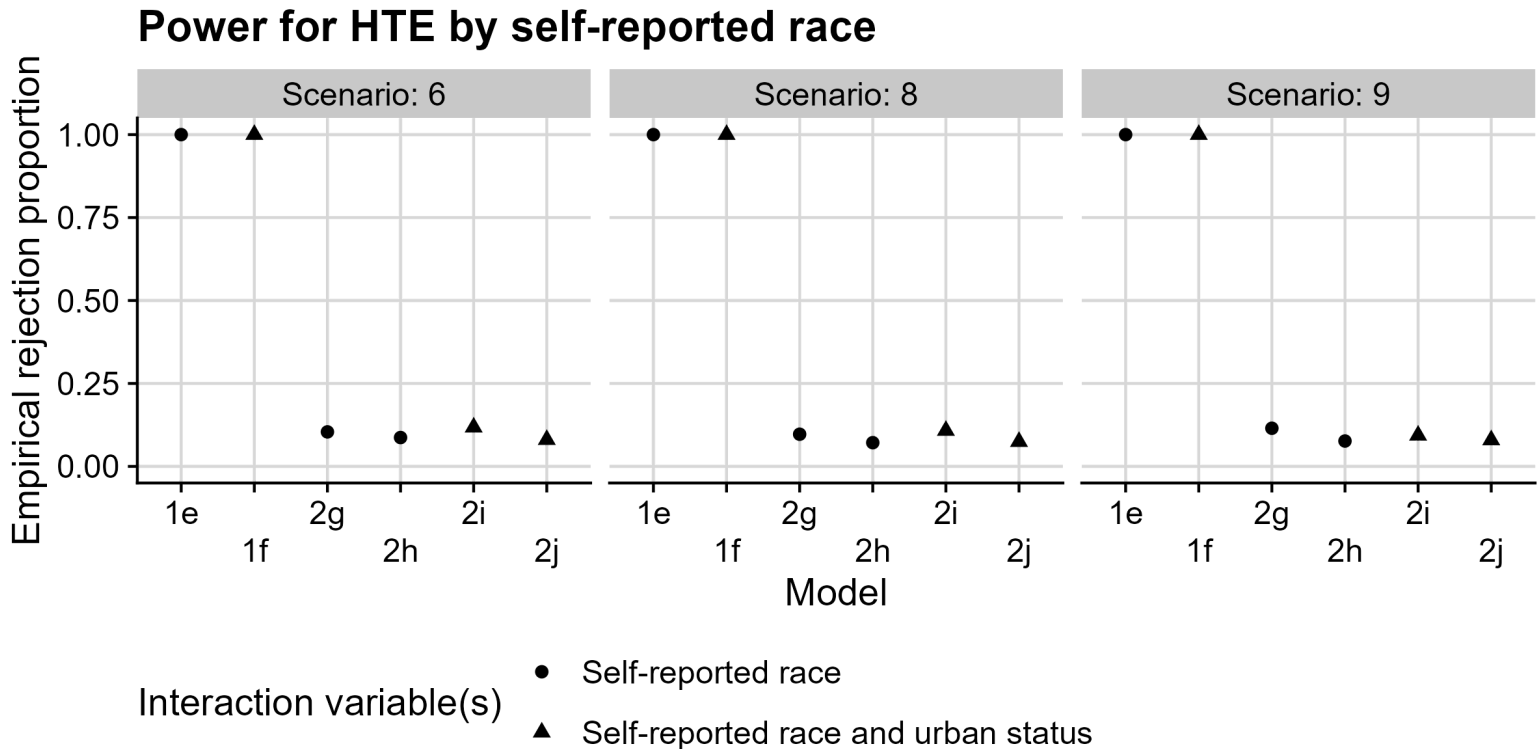
Power for HTE by self-reported race

Models encode individual– (1) vs cluster–level (2), adjustment and interaction variables:

- 1e, 1f: adjust for self–reported race, allow HTE by self–reported race (1f also by urban)

- 2g, 2i: adjust for center–level race proportion, allow HTE by self–reported race (2i also by urban)

- 2h, 2j: adjust for proportion > 0.5, allow HTE by self–reported race (2j also by urban)

Closing thoughts

High power, controlled type I error to detect HTE in the individual–level analysis

- where possible, this approach is key for identifying an remedying disparities

Low power to detect HTE by an aggregated individual–level predictor

- even in simple setting with only two possible categories!
- likely exacerbated in more realistic settings
- elevated type I error when using proportion in some settings

However, meaningful estimated differences across groups can still be acted upon

Power calculations should aggregate simulated individual–level data, rather than simulating cluster–level data directly.

For more details, including more realistic simulation settings:

Williamson BD, Coley RY, Hsu C, McCracken CE, Cook AJ (2023). Considerations for subgroup analyses in cluster–randomized trials based on individual–level predictors. *Prevention Science*.

References

- Knol MJ and VanderWeele TJ (2012). Recommendations for presenting analyses of effect modification and interaction. *International Journal of Epidemiology*.

- Petkovic et al. (2020). Reporting of health equity considerations in cluster and individually randomized trials. *Trials*.

- Chidambaram P (2022). Over 200,000 residents and staff in long-term care facilities have died from COVID-19. *Kaiser Family Foundation*.

- Wakefield J (2008). Ecologic studies revisited. *Annual Review of Public Health*.

- Greenland S & Morgenstern H (1989). Ecological bias, confounding, and effect modification. *International Journal of Epidemiology*.