An Applied Researcher's Guide to ITT Effects from Multisite (blocked) Individually Randomized Trials: Estimands, Estimators, and Estimates

UC Berkeley Simons Institute – Feb, 2022

Luke Miratrix (Harvard Grad School of Ed.),
Michael J. Weiss (MDRC), and
Brit Henderson (MDRC)

Thanks to Mike Weiss for making most of these slides for an initial presentation

mdrc
BUILDING KNOWLEDGE
TO IMPROVE SOCIAL POLICY

I'm planning an evaluation that will randomize 6,000 people within 20 sites. Can you write the part of the analysis plan that describes the estimation model for the overall average ITT effect?

ALEX

O.K.

Why don’t you do your own job?

My collaborator, who initiated this adventure

MIKE

Here's the estimator I have in mind:

$$Y = \sum_{j=1}^{20} \alpha_j * Site_j + \beta * T + \varepsilon$$

<u>Where</u>:

Y = Outcome of interest

$Site_j$ = Set to 1 if person was at site *j* and 0 otherwise

T = Set to 1 if person assigned to treatment and 0 otherwise

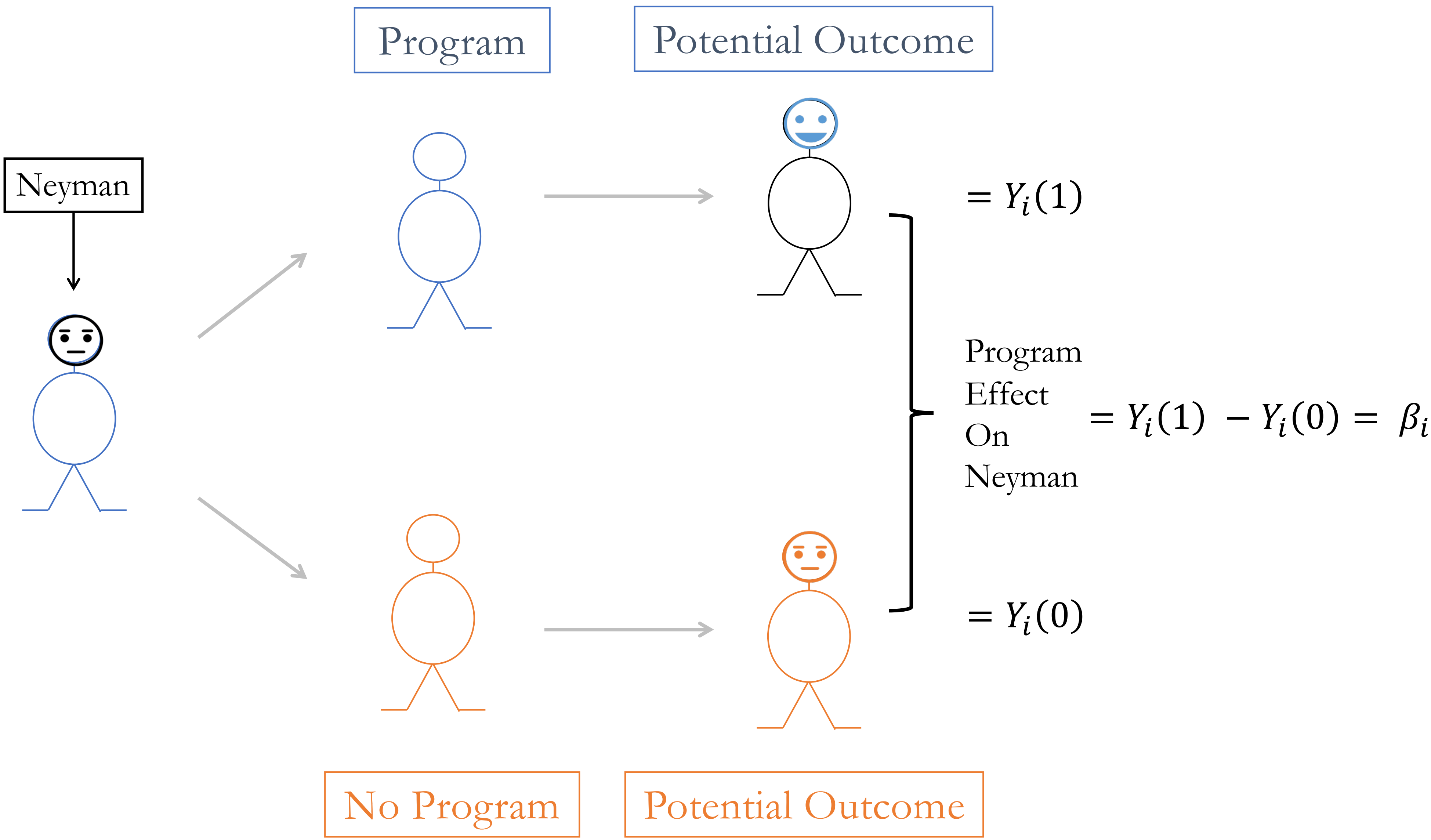
ε = error, assumed i.i.d. normal

ALEX

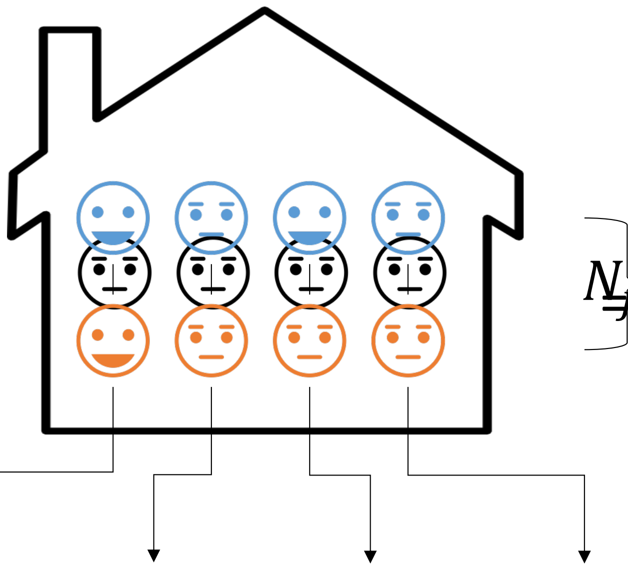
ESTIMAND

What is the target of inference?

Program Effect for a Person


Program
Potential Outcome
Neyman
$= Y_i(1)$
Program
Effect
On
Neyman
$= Y_i(1) - Y_i(0) = \beta_i$
$= Y_i(0)$
No Program
Potential Outcome

Average Program Effect for a Group of People at a Site



$N_j = 4$ = # of persons at site j

$$\beta_{\cdot j} = \sum_{i=1}^{N_j} \frac{1}{N_j} \beta_{ij}$$

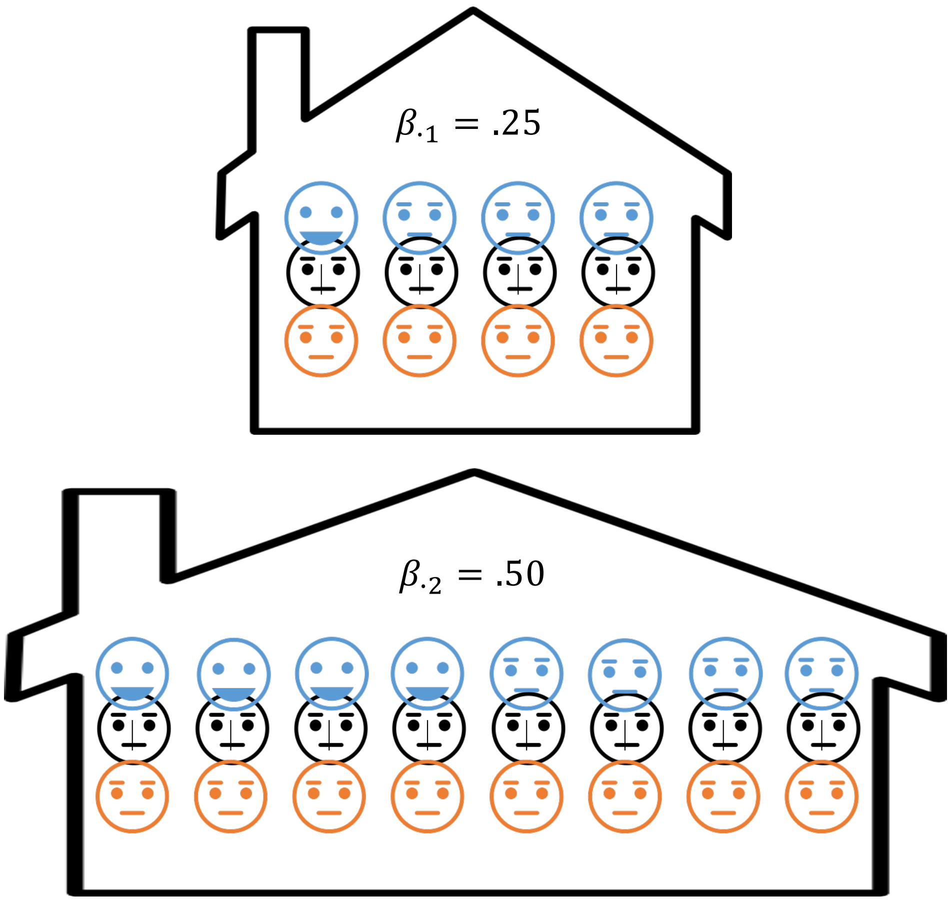
$$= \frac{1}{4} * 0 + \frac{1}{4} * 0 + \frac{1}{4} * 1 + \frac{1}{4} * 0 = 0.25$$

<u>Where</u>:

$\beta_{\cdot j}$ = average ITT effect at site j

β_{ij} = ITT effect for person i at site j

Estimand: Effect for Average **Person or Site**?



$$\beta_{Person} = 0.416$$

$$\beta_{Site} = 0.375$$

Estimand: Effect for Average **Person or Site**?

Target Unit		
	Person	$\beta_{person} = \sum_{j=1}^{J} \frac{N_j}{N} \beta_{\cdot j}$
	Site	$\beta_{site} = \sum_{j=1}^{J} \frac{1}{J} \beta_{\cdot j}$

Where:

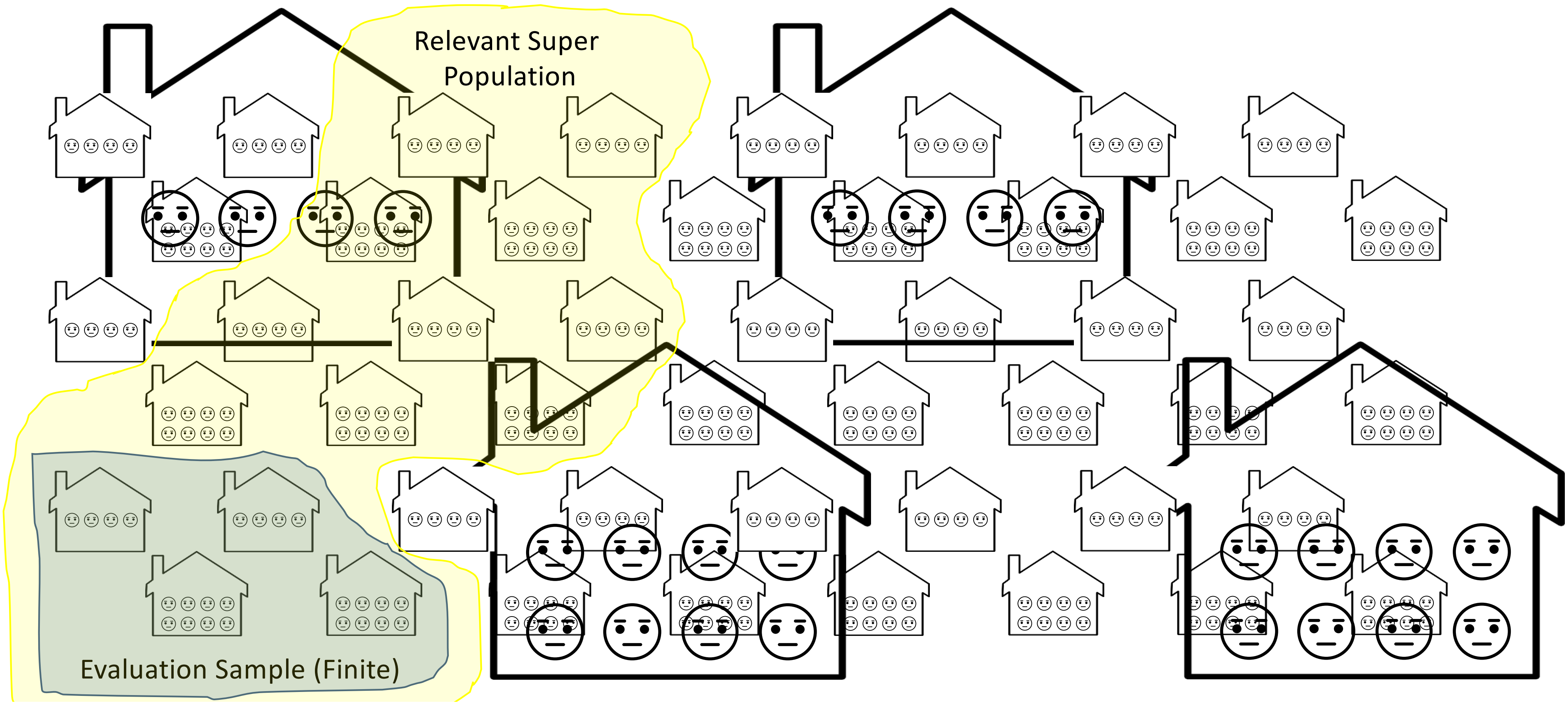
J = # of sites in study

N = total # of persons

N_j = # of persons at site j

$\beta_{\cdot j}$ = average ITT effect at site j

Estimand: Effect for **Finite or Super** Population?

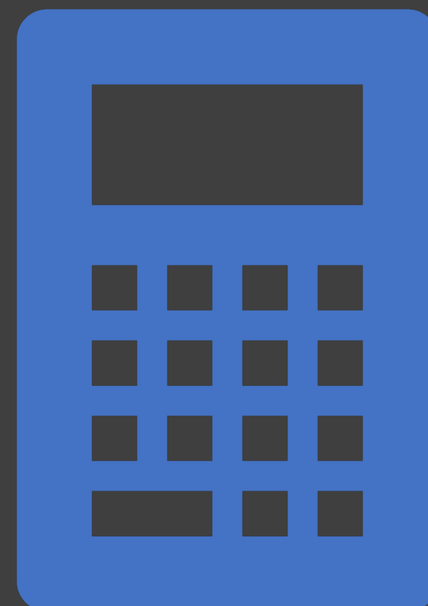


Four Estimands

		Target Population	
		Finite	Super
Target Unit	Person	$\beta_{FP-person} = \sum_{j=1}^{J} \frac{N_j}{N} \beta_{.j}$	$\beta_{SP-person} = \sum_{j=1}^{J^*} \frac{N_j^*}{N^*} \beta_{.j}^*$
	Site	$\beta_{FP-site} = \sum_{j=1}^{J} \frac{1}{J} \beta_{.j}$	$\beta_{SP-site} = \sum_{j=1}^{J^*} \frac{1}{J^*} \beta_{.j}^*$

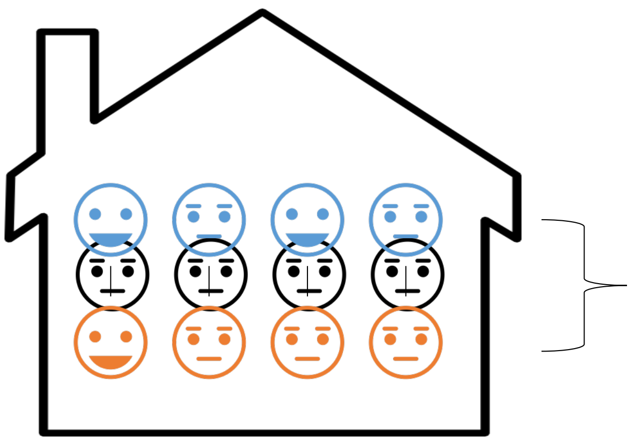
ESTIMATORS

Rule for calculating an estimate based on observed data



Convenient way to describe estimators of β

$$\hat{\beta} = \sum_{j=1}^{J} w_j * \hat{\beta}_{\cdot j}$$



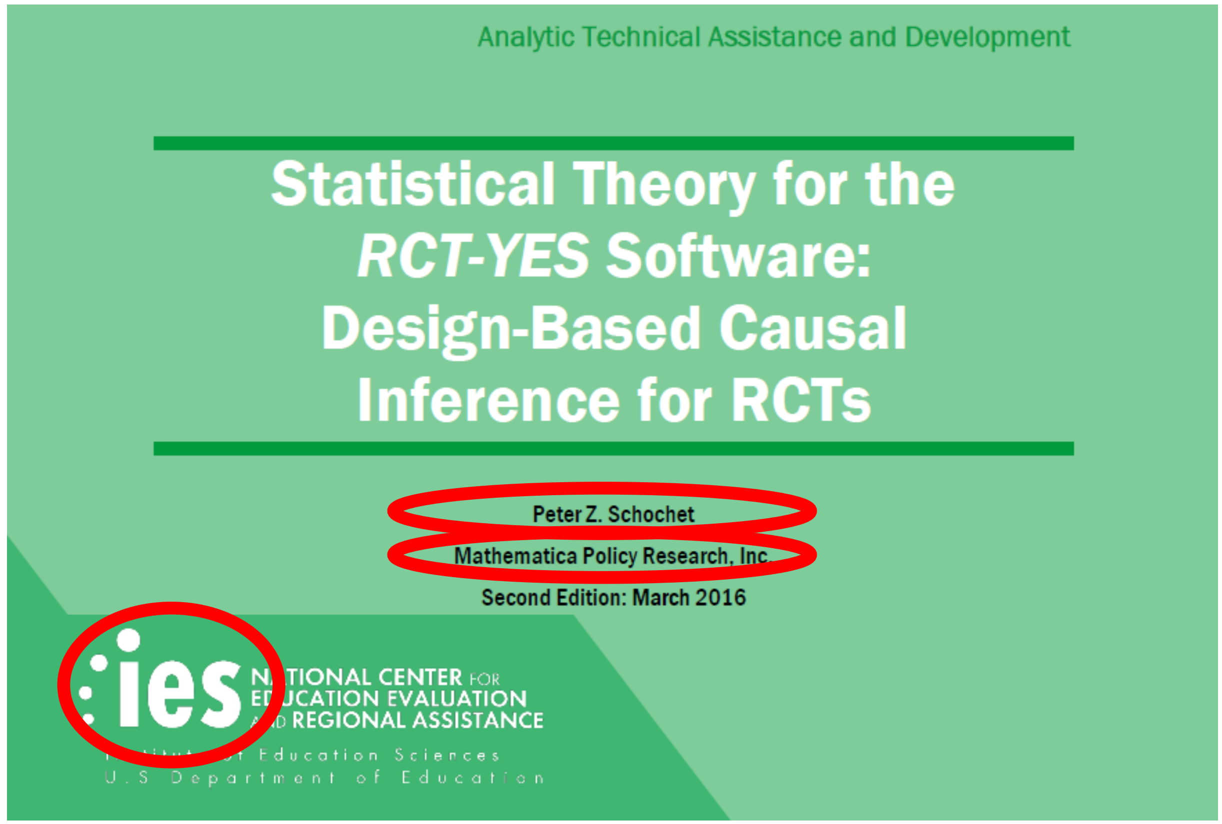
$\hat{\beta}_{\cdot j}$ = *estimated* average ITT effect at site j = 0.66

(simple difference-in-means estimator)

Classes of Estimators

1. Design Based
2. Linear Regression
3. Multilevel Modeling

1. Design-based Estimators (4 of them)


Analytic Technical Assistance and Development

Statistical Theory for the *RCT-YES* Software: Design-Based Causal Inference for RCTs

Peter Z. Schochet

Mathematica Policy Research, Inc.

Second Edition: March 2016

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE

Institute of Education Sciences

U.S. Department of Education

Some nice features of design-based estimators

- Simple
- Clear connection to estimands
- Unbiased
- Specialized software designed for RCTs and for easy use

Design-based estimators of β

Name	Estimator	w_j	Estimand
Design Based – person $\hat{\beta}_{DB-person}$	$\sum_{j=1}^{J} \frac{N_j}{N} \hat{\beta}_{\cdot j}$	$w_j \propto N_j$	β_{person}
Design Based – site $\hat{\beta}_{DB-site}$	$\sum_{j=1}^{J} \frac{1}{J} \hat{\beta}_{\cdot j}$	$w_j \propto 1$	β_{site}

Where:

N_j = number of people at site *j* in sample

N = number of people in full sample

$\hat{\beta}_{\cdot j}$ = *estimated* average ITT effect at site *j* (simple difference-in-means estimator)

J = number of sites in sample

2. Linear Regression (8 of them)

- See lots of text books

The linear regression estimators

	Our Name	Model	w_j	Estimand
Fixed Effects	(1) Fixed Effects (FE) $\hat{\beta}_{FE}$	$Y = \sum_{j=1}^{J} \alpha_j * Site_j + \beta * T + \varepsilon$	$w_j \propto N_j T_j (1 - T_j)$	$\beta_{FP-person}$
	(2) FE - heteroskedastic robust $\hat{\beta}_{FE-Het}$	"	"	$\beta_{FP-person}$
	(3) FE - cluster robust $\hat{\beta}_{FE-CR}$	"	"	$\beta_{SP-person}$
	(4) FE – club sandwich $\hat{\beta}_{Club}$	"	"	$\beta_{SP-person}$
Weighted Regression	(5) FE - person-weights $\hat{\beta}_{FE-weight-person}$	" $\omega_{ij}^{person} = T_{ij}\left(\frac{T_{..}}{T_{.j}}\right) + (1 - T_{ij})\left(\frac{1 - T_{..}}{1 - T_{.j}}\right)$	$w_j \propto N_j$	$\beta_{FP-person}$
	(6) FE - site-weights $\hat{\beta}_{FE-weight-site}$	" $\omega_{ij}^{Site} = \left[T_{ij}\left(\frac{T_{..}}{T_{.j}}\right) + (1 - T_{ij})\left(\frac{1 - T_{..}}{1 - T_{.j}}\right)\right]\left[\frac{\frac{N}{J}}{N_j}\right]$	$w_j \propto 1$	$\beta_{FP-site}$
Fully Interacted	(7) FE - w/ interactions - person $\hat{\beta}_{FE-inter-person}$	$Y = \sum_{j=1}^{J} \alpha_j * Site_j + \sum_{j=1}^{J} \beta_j * Site_j * T + \varepsilon$	$w_j \propto N_j$	$\beta_{FP-person}$
	(8) FE - w/ interactions - site $\hat{\beta}_{FE-inter-site}$	"	$w_j \propto 1$	$\beta_{FP-site}$

3. Multilevel Models (3 of them)

Journal of Research on Educational Effectiveness

ISSN: 1934-5747 (Print) 1934-5739 (Online) Journal homepage: http://www.tandfonline.com/loi/uree20

Using Multisite Experiments to Study Cross-Site Variation in Treatment Effects: A Hybrid Approach With Fixed Intercepts and a Random Treatment Coefficient

Howard S. Bloom, Stephen W. Raudenbush, Michael J. Weiss & Kristin Porter

The multilevel model estimators

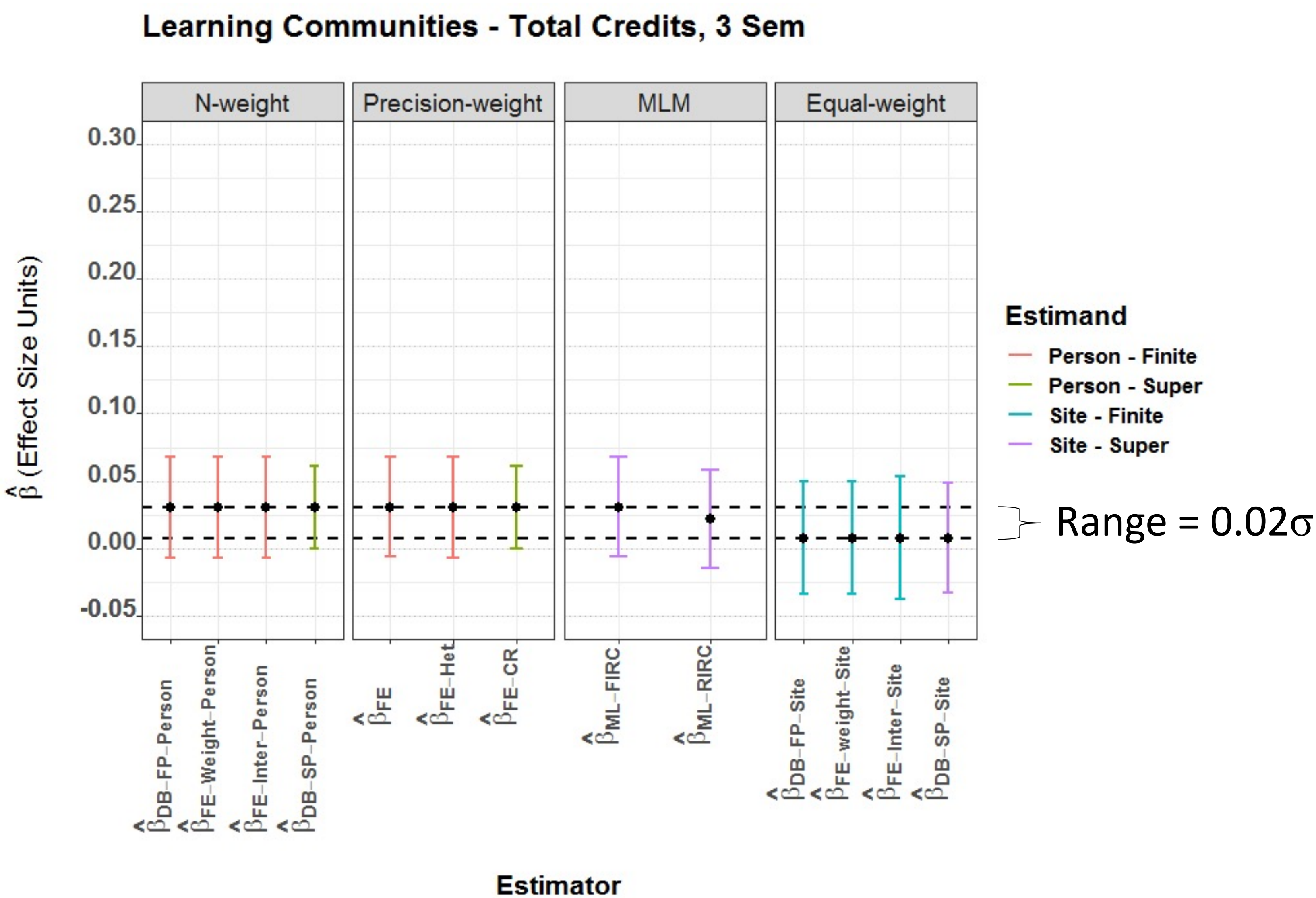
Our Name	Model	w_j	Estimand
1. Fixed Intercepts Random Treatment Coefficient (FIRC) $\hat{\beta}_{ML-FIRC}$	Level 1: $Y_{ij} = \alpha_{.j} + \beta_{.j}T_{ij} + e_{ij}$ Level 2: $\alpha_{.j} = \alpha_{.j}$ $\beta_{.j} = \beta + b_j$ Where: $e_{ij} \sim N(0, \sigma^2)$ $b_j \sim N(0, \tau_b^2)$	$w_j \propto \left[\hat{\tau} + \frac{\hat{\sigma}^2}{N_j T_{.j}(1 - T_{.j})}\right]^{-1}$	$\beta_{SP-site}$
2. Random Intercept Random Treatment Coefficient (RIRC) $\hat{\beta}_{ML-RIRC}$	Level 1: $Y_{ij} = \alpha_{.j} + \beta_{.j}T_{ij} + e_{ij}$ Level 2: $\alpha_{.j} = \alpha + a_j$ $\beta_{.j} = \beta + b_j$ Where: $e_{ij} \sim N(0, \sigma^2)$ $\begin{pmatrix} a_j \\ b_j \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_a^2 & \tau_{ab} \\ \tau_{ab} & \tau_b^2 \end{pmatrix}\right]$	Unsure	$\beta_{SP-site}$
2. Random intercept, constant coefficient ($\hat{\beta}_{ML-RICC}$)	Level 1: $Y_{ij} = \alpha_{.j} + \beta T_{ij} + e_{ij}$ Level 2: $\alpha_{.j} = \alpha + a_j$ Where: $\alpha_j \sim N(0, \tau_\alpha^2)$	Basically like fixed effects model	Surprise! $\beta_{FP-person}$

ESTIMATES

The 12 Studies

Early Childhood-Element. School	Middle School-High School	Post-secondary Education	Labor Market Programs
Head Start Impact Study	Enhanced Reading Opportunity mdrc	Learning Communities mdrc	Welfare-to-Work Programs mdrc
After School – Reading Program mdrc	Career Academies mdrc	Performance-based Scholarships mdrc	
After School – Math Program mdrc	Communities in Schools	Encouraging Summer Enrollment (1) mdrc	
	Early College H.S.	Encouraging Summer Enrollment (2) mdrc	

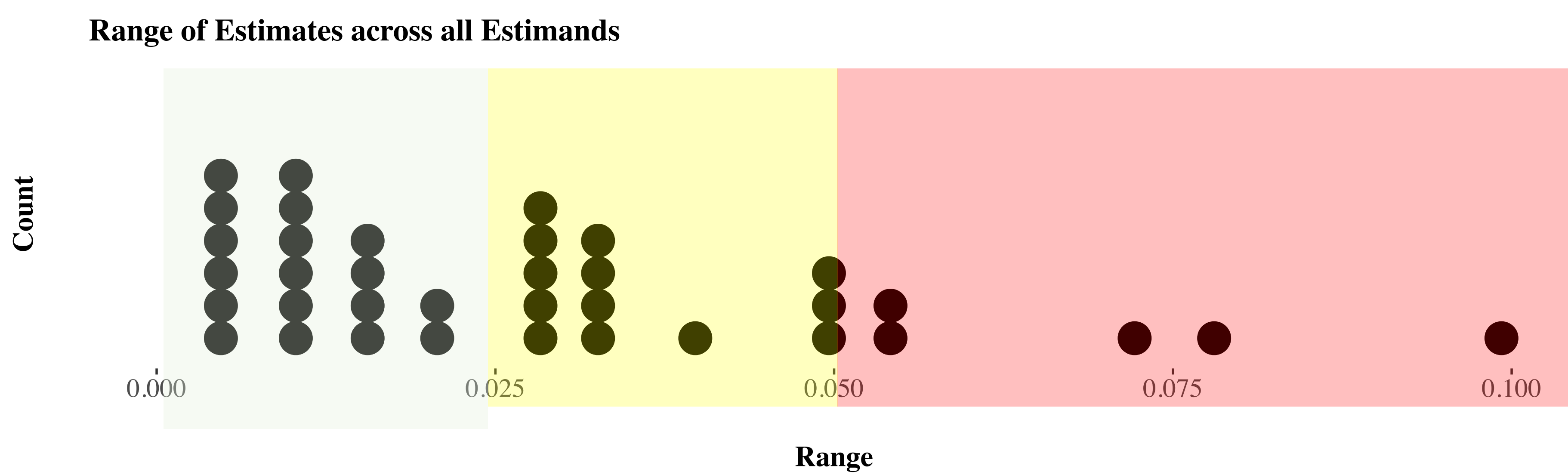
RQ1:
Does choice of estimand/estimator matter for the estimate of β?


Learning Communities - Total Credits, 3 Sem
N-weight
Precision-weight
MLM
Equal-weight
β̂ (Effect Size Units)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
Estimand
Person - Finite
Person - Super
Site - Finite
Site - Super
Range = 0.02σ
β̂DB−FP−Person
β̂FE−Weight−Person
β̂FE−Inter−Person
β̂DB−SP−Person
β̂FE
β̂FE−Het
β̂FE−CR
β̂ML−FIRC
β̂ML−RIRC
β̂DB−FP−Site
β̂FE−weight−Site
β̂FE−Inter−Site
β̂DB−SP−Site
Estimator

Meh

OK, I'm paying attention

Yikes?

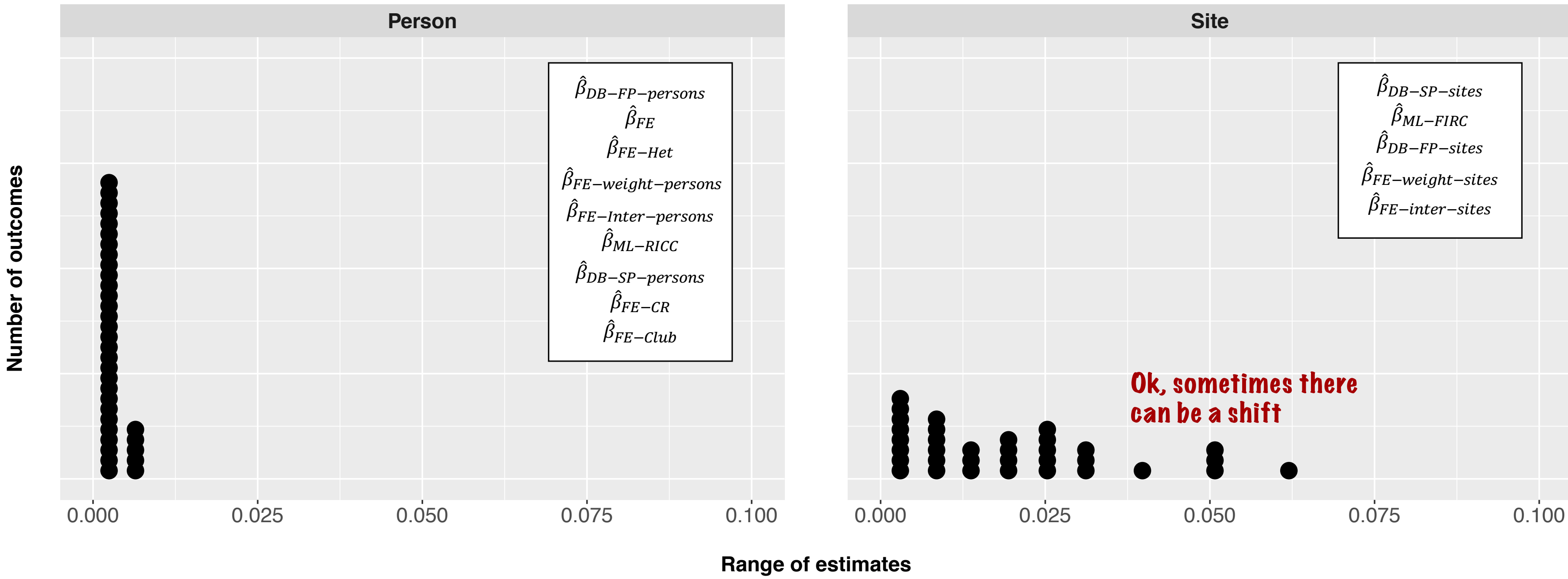


Impacts in effect size units.

Notes: Each dot represents a single outcome for a single study.

The x-axis is the range of point estimates ($max[\hat{\beta}] - min[\hat{\beta}]$) across all 14 estimators, in effect size units.

Group estimators by whether they are person or site targeting



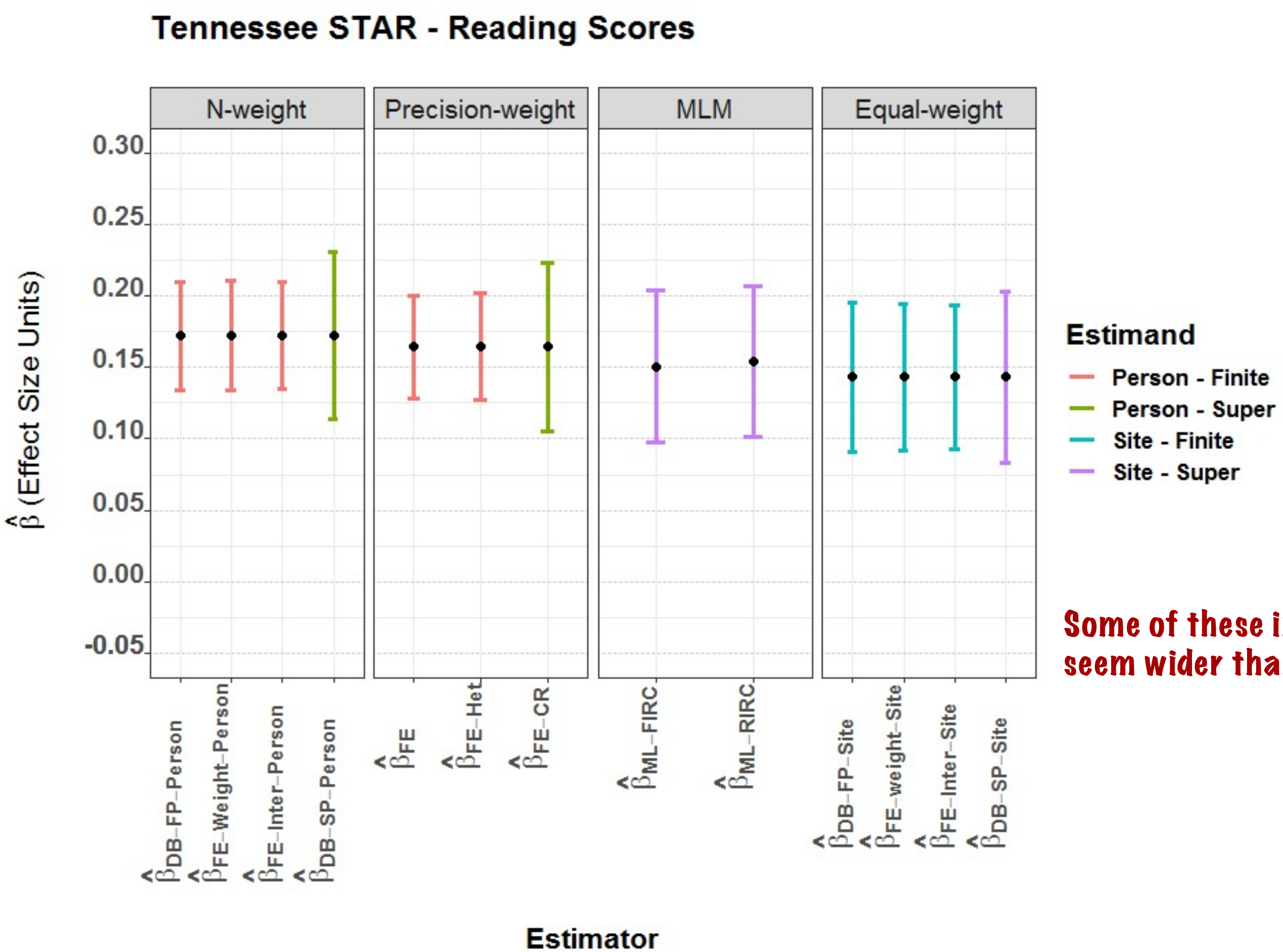
The person weighting ones are all basically the same. Site, less so.

Notes: Each dot represents a single outcome for a single study.

The x-axis is the range of point estimates ($max[\hat{\beta}] - min[\hat{\beta}]$) across all 14 estimators, in effect size units.

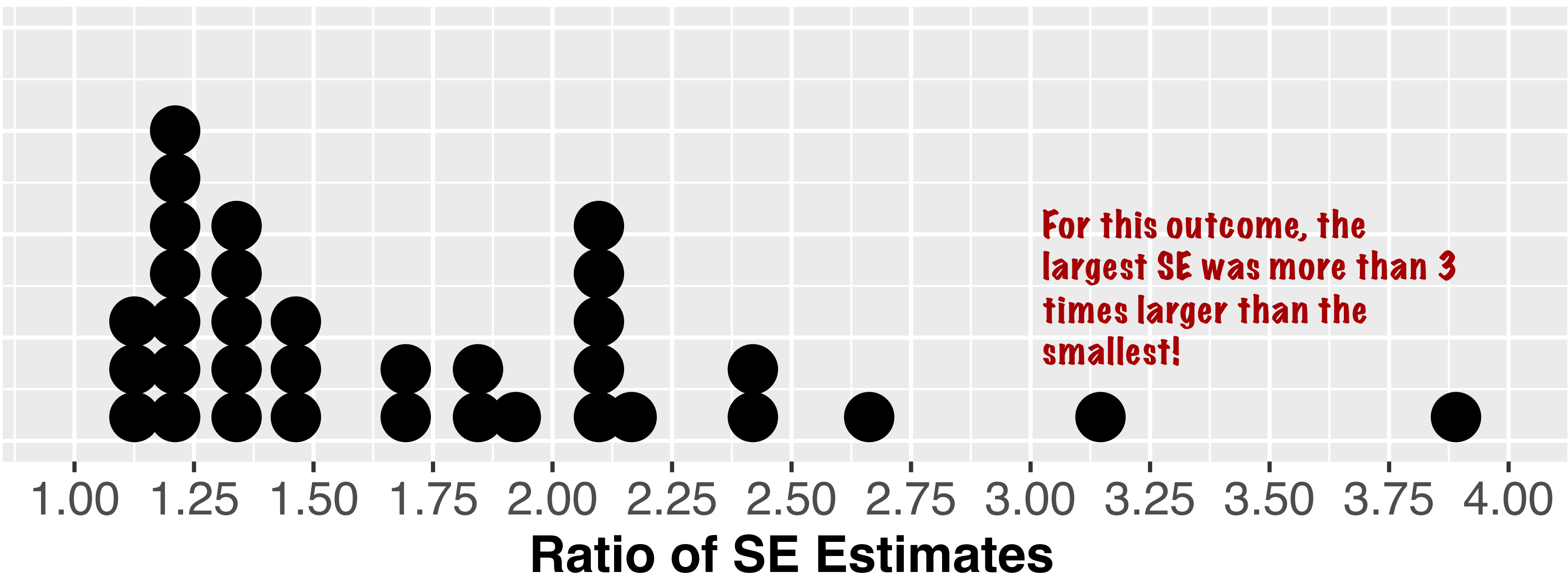
RQ2:

Does choice of estimand/estimator matter for the estimate of $\mathbf{SE}(\hat{\beta})$?

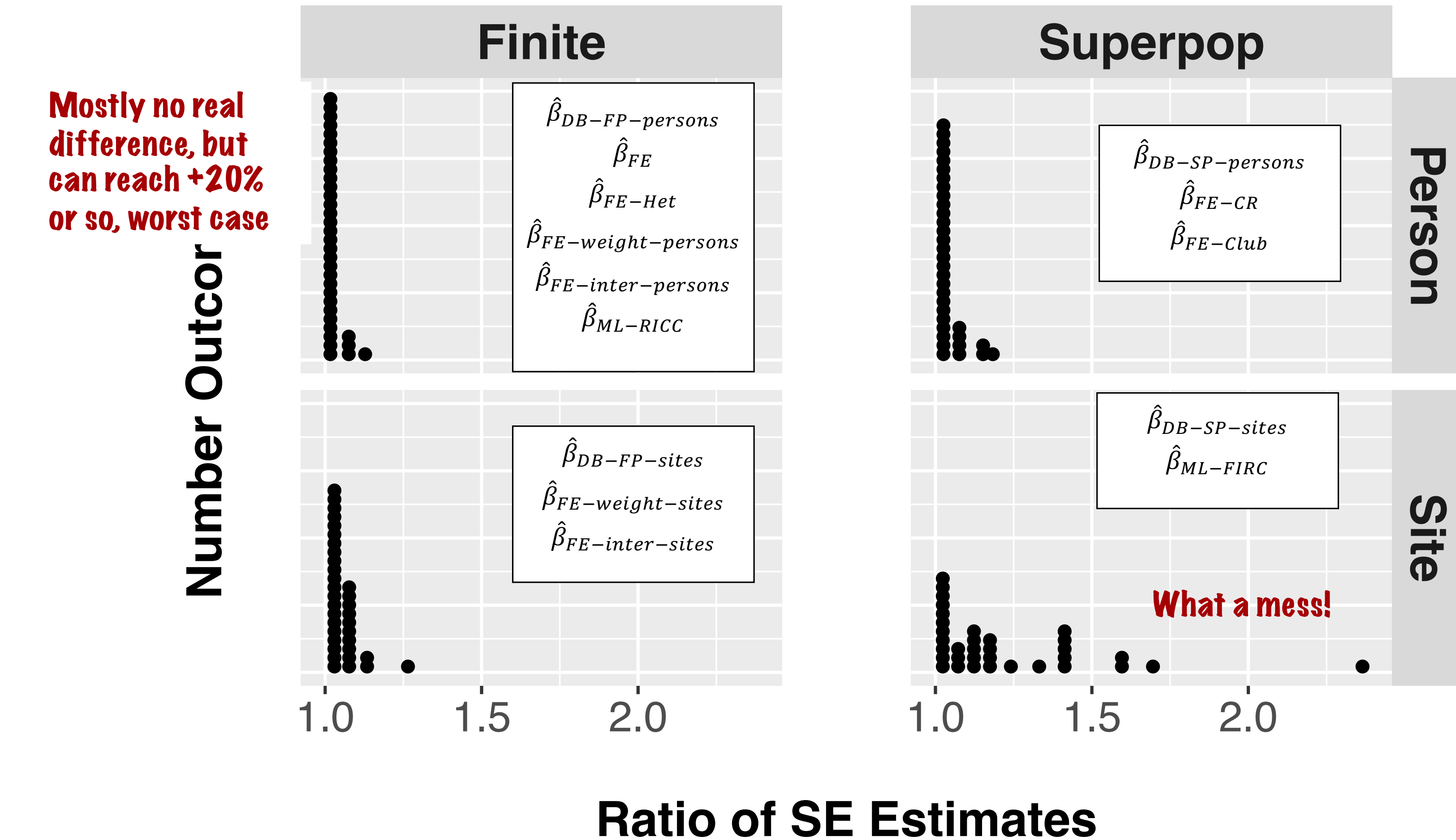


Some of these intervals seem wider than others

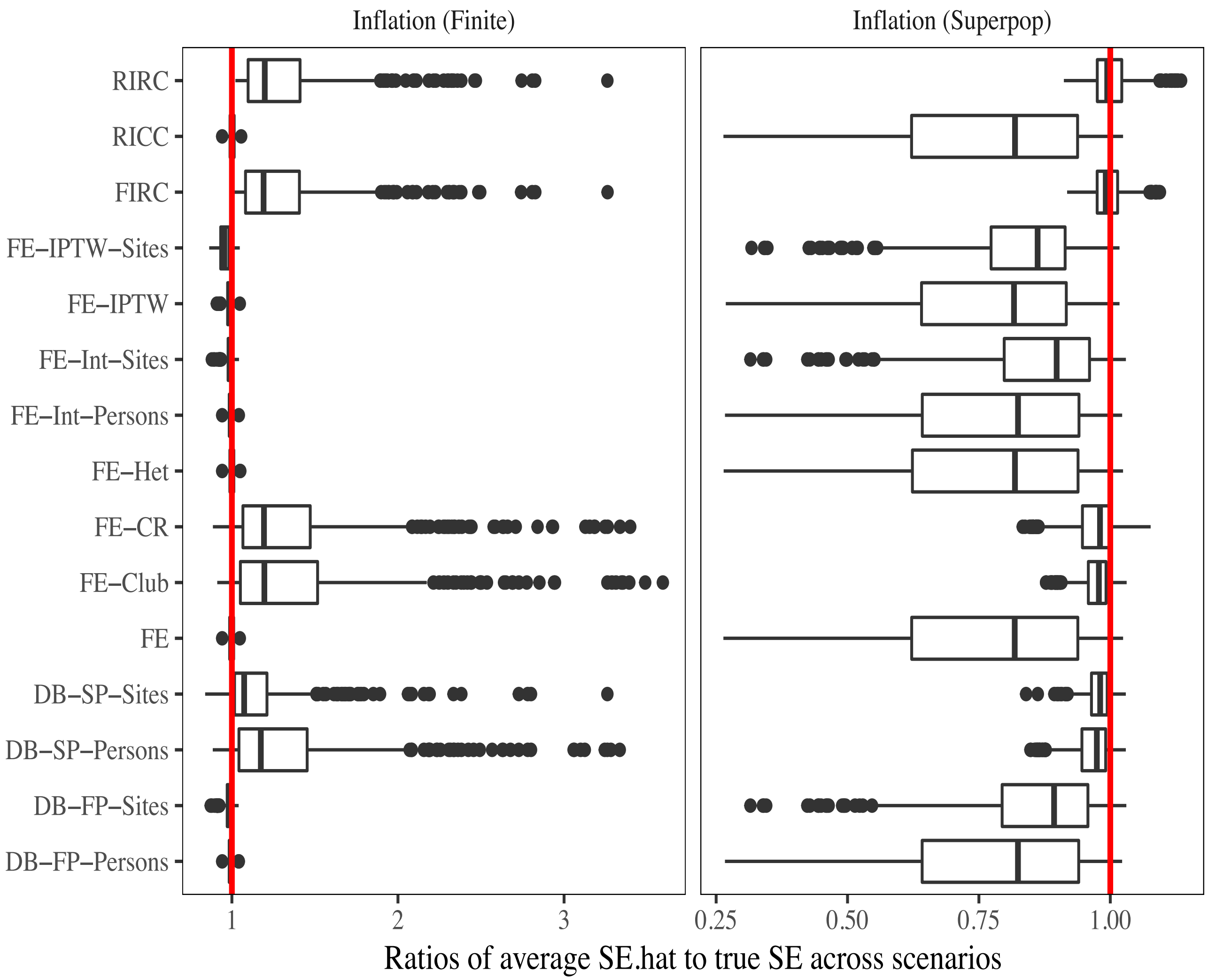
The largest estimated SE can be a lot different than the smallest



Notes: Each dot represents a single outcome for a single study.
The x-axis is the ratio of largest to smallest estimated SE ($max[\widehat{SE}(\hat{\beta})]/min[\widehat{SE}(\hat{\beta})]$) across all estimators, in effect size units.



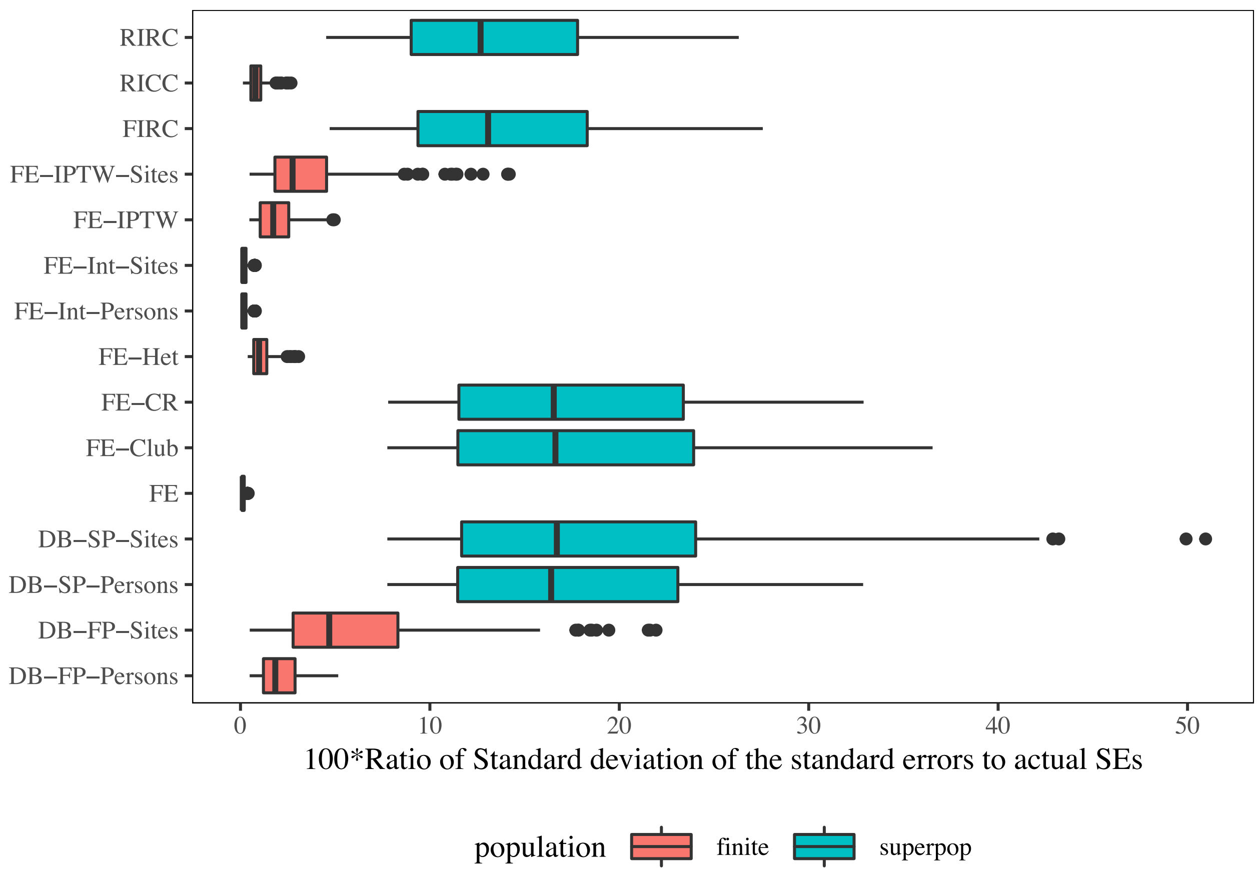
Notes: Each dot represents a single outcome for a single study.
The x-axis is the ratio of largest to smallest estimated SE ($max[\widehat{SE}(\hat{\beta})]/min[\widehat{SE}(\hat{\beta})]$) across all 13 estimators, in effect size units.



Are the Standard Error estimates calibrated?

Generally yes, if you are in the right framework.

Boxplots show calibration across simulation scenarios



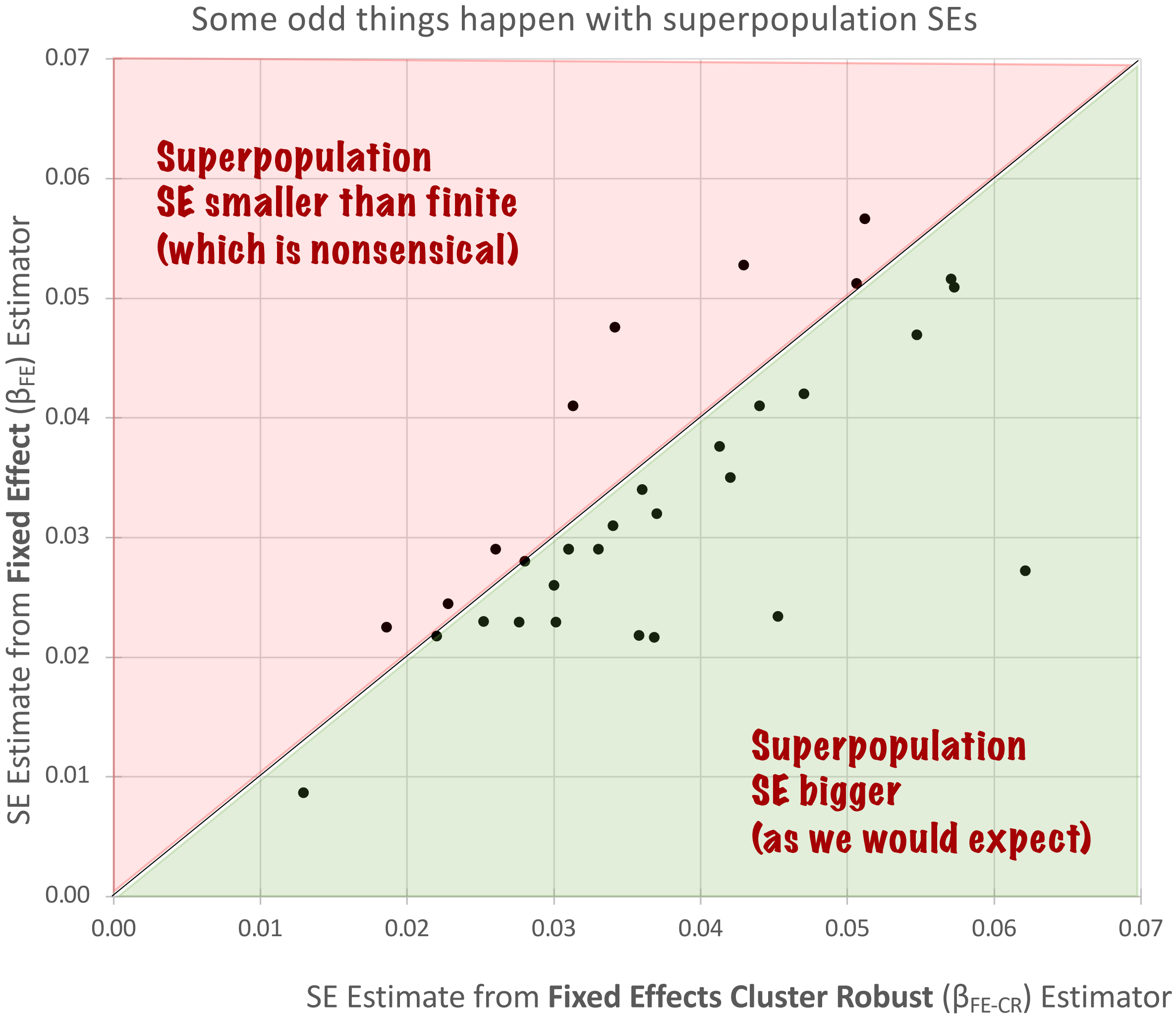
How well can we estimate our standard errors anyway?

For superpopulation, not well.

For finite, almost perfectly in some cases.

Site average estimation does have a price, as usual.

Superpopulation SE estimates vary a lot, causing trouble.

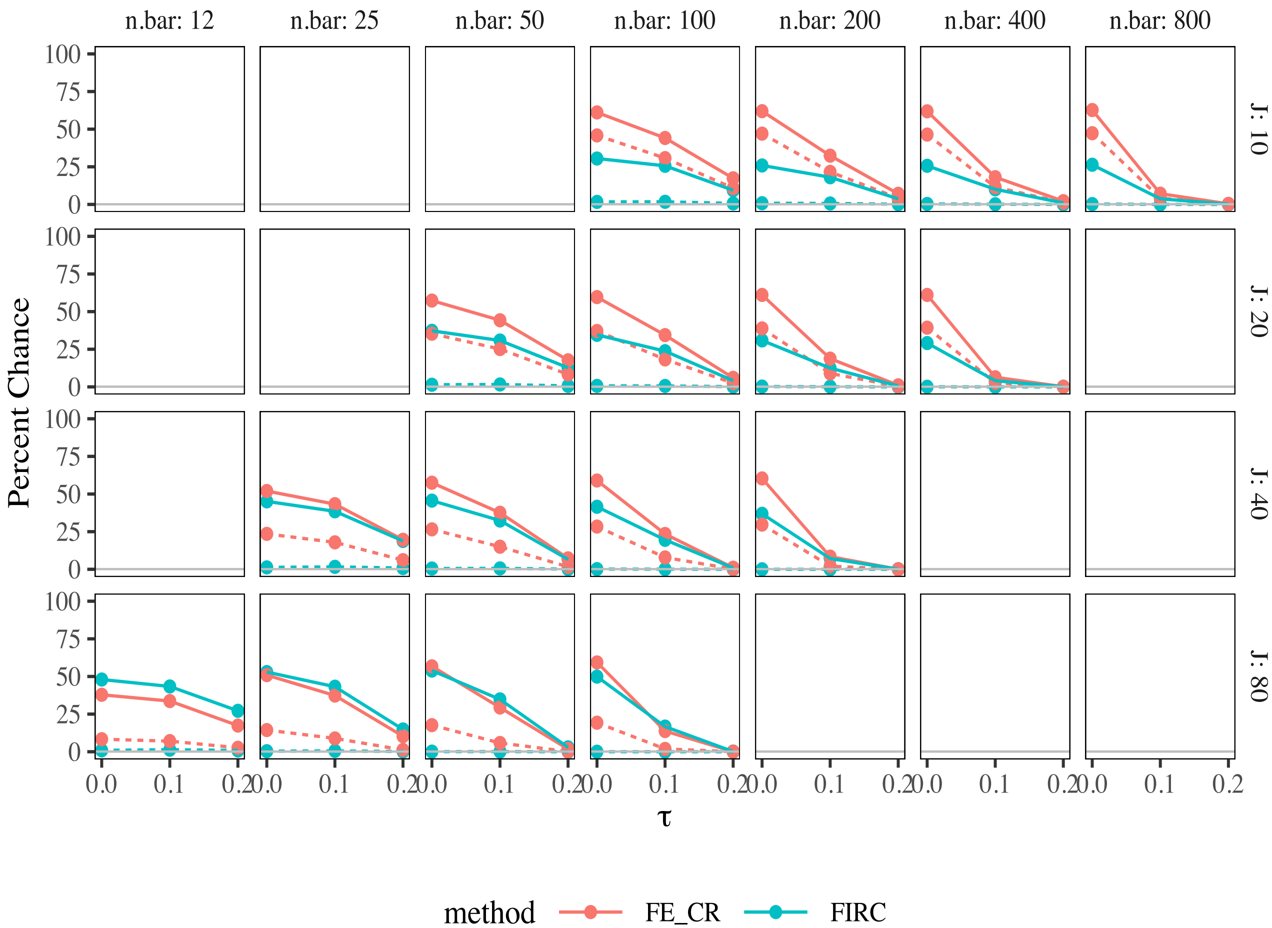


Cluster robust FE targets superpop person weighted. We expect the SEs to be LARGER than the finite person weighted

This often does not happen.

(The other superpopulation estimators suffer the same.)

Superpopulation estimated SEs are lower than finite estimated SEs quite often



From Simulation:

How often do we get a smaller finite population standard error than a superpopulation one?

Not infrequently.

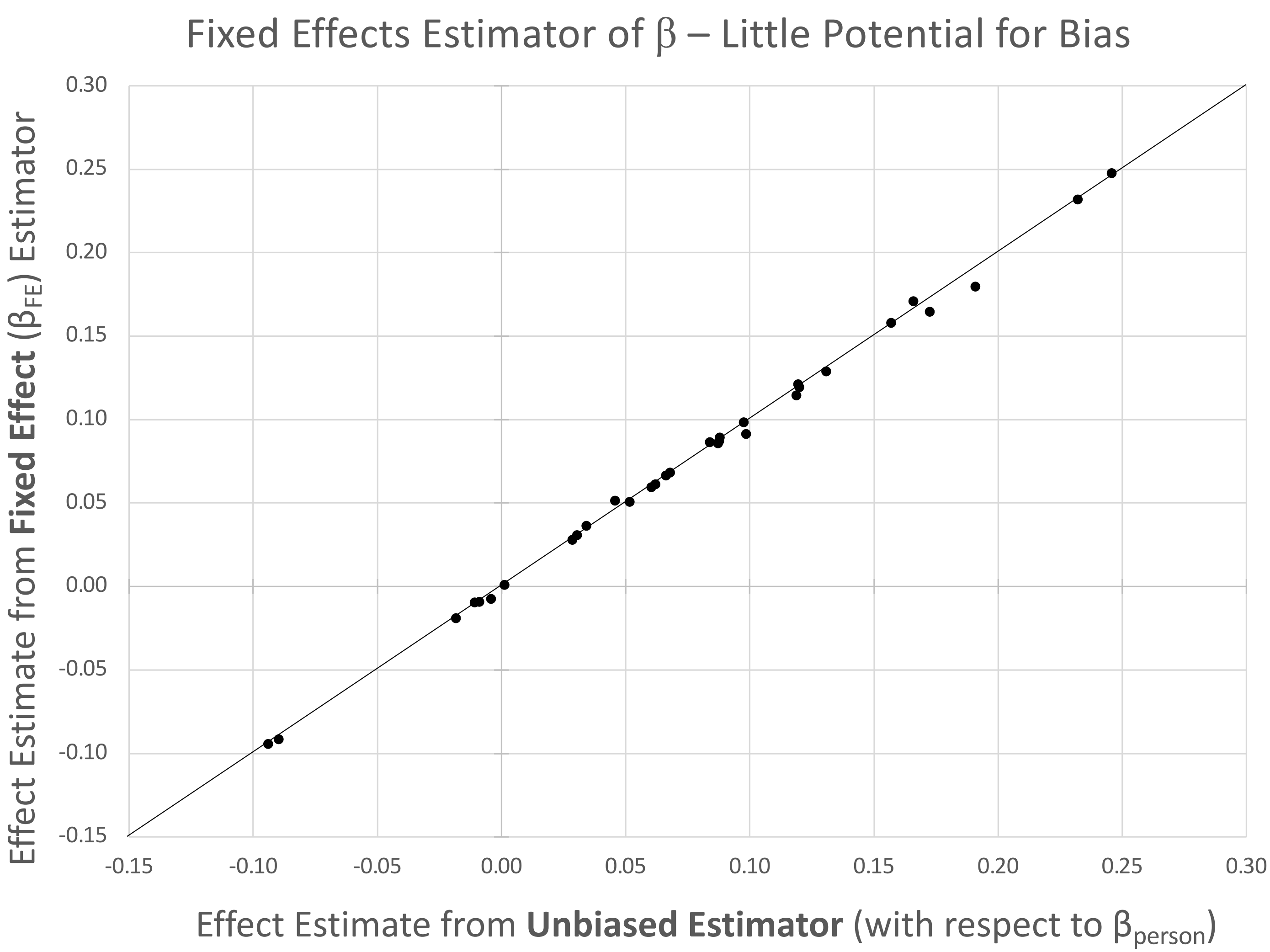
Instability of all estimators at the superpopulation level in the face of cross site impact variation makes life difficult.

RQ3:
Bias Precision Trade-off?

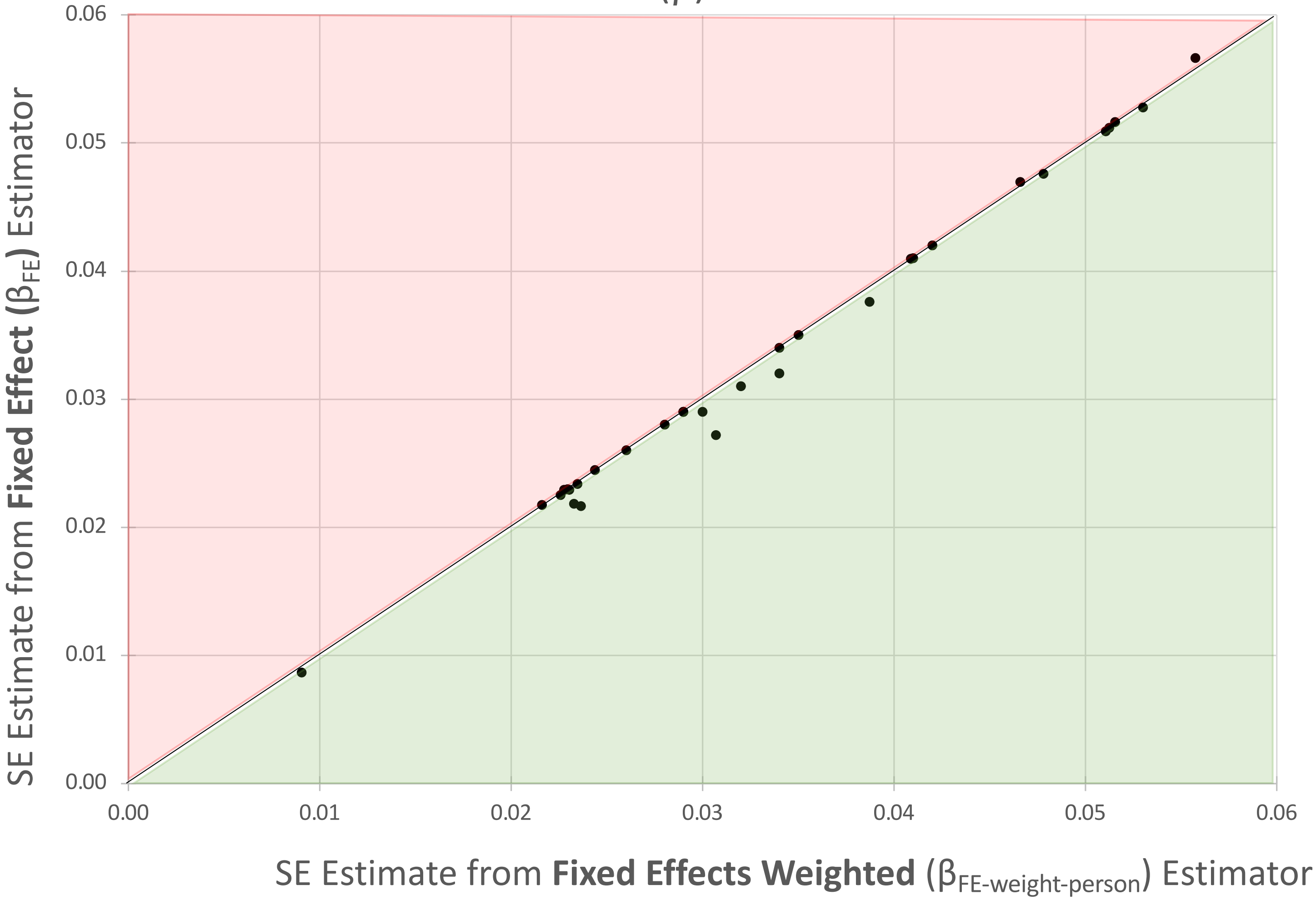
RQ3: Bias Precision Trade-off?

Part I: The estimand of $\beta_{FP-person}$

Unbiased vs. Fixed Effects models

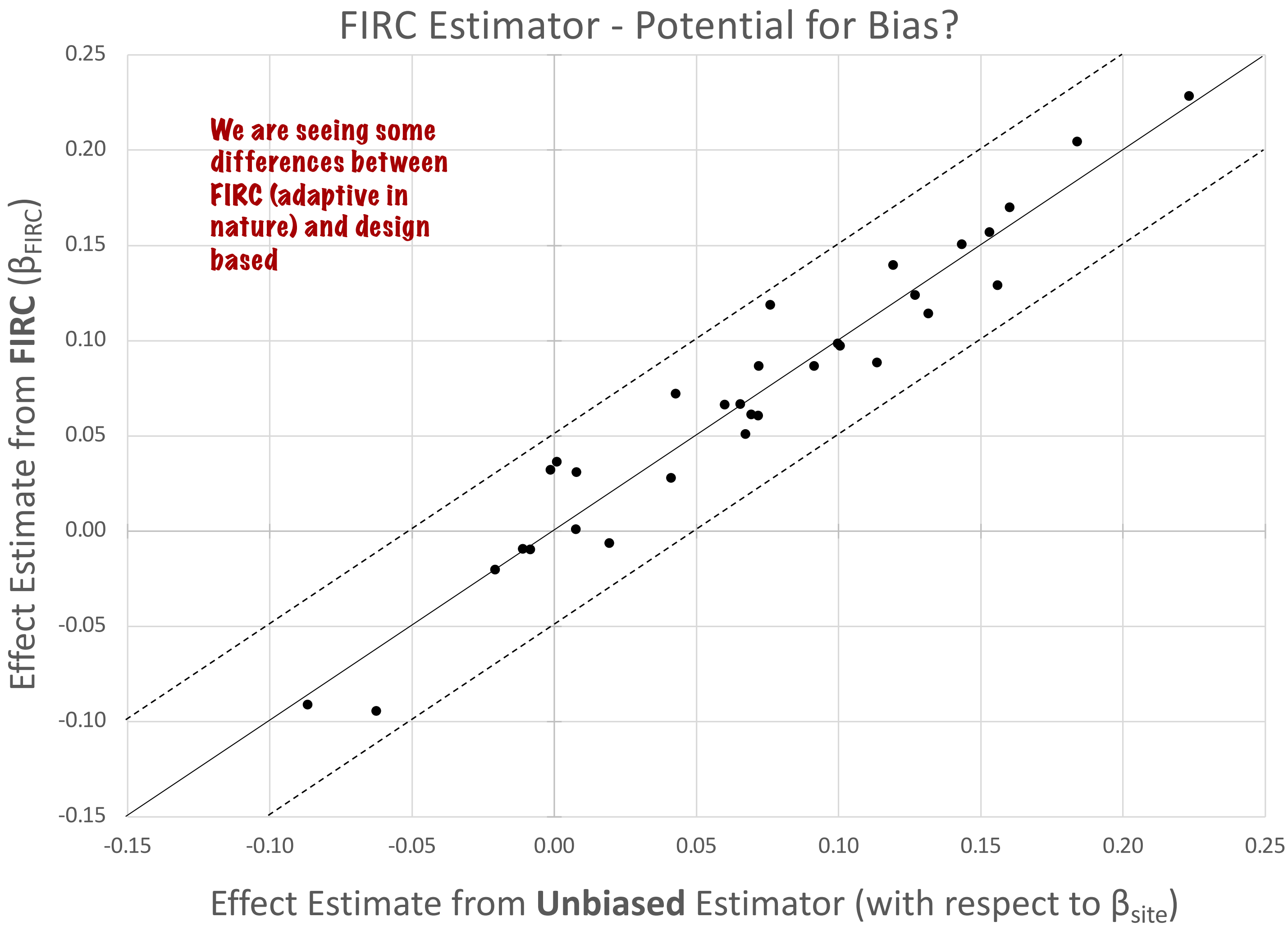

Fixed Effects Estimator of β – Little Potential for Bias
Effect Estimate from Fixed Effect (βFE) Estimator
Effect Estimate from Unbiased Estimator (with respect to βperson)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
0.25
0.30

Fixed Effects Estimator of SE(β) - Limited Precision Gains

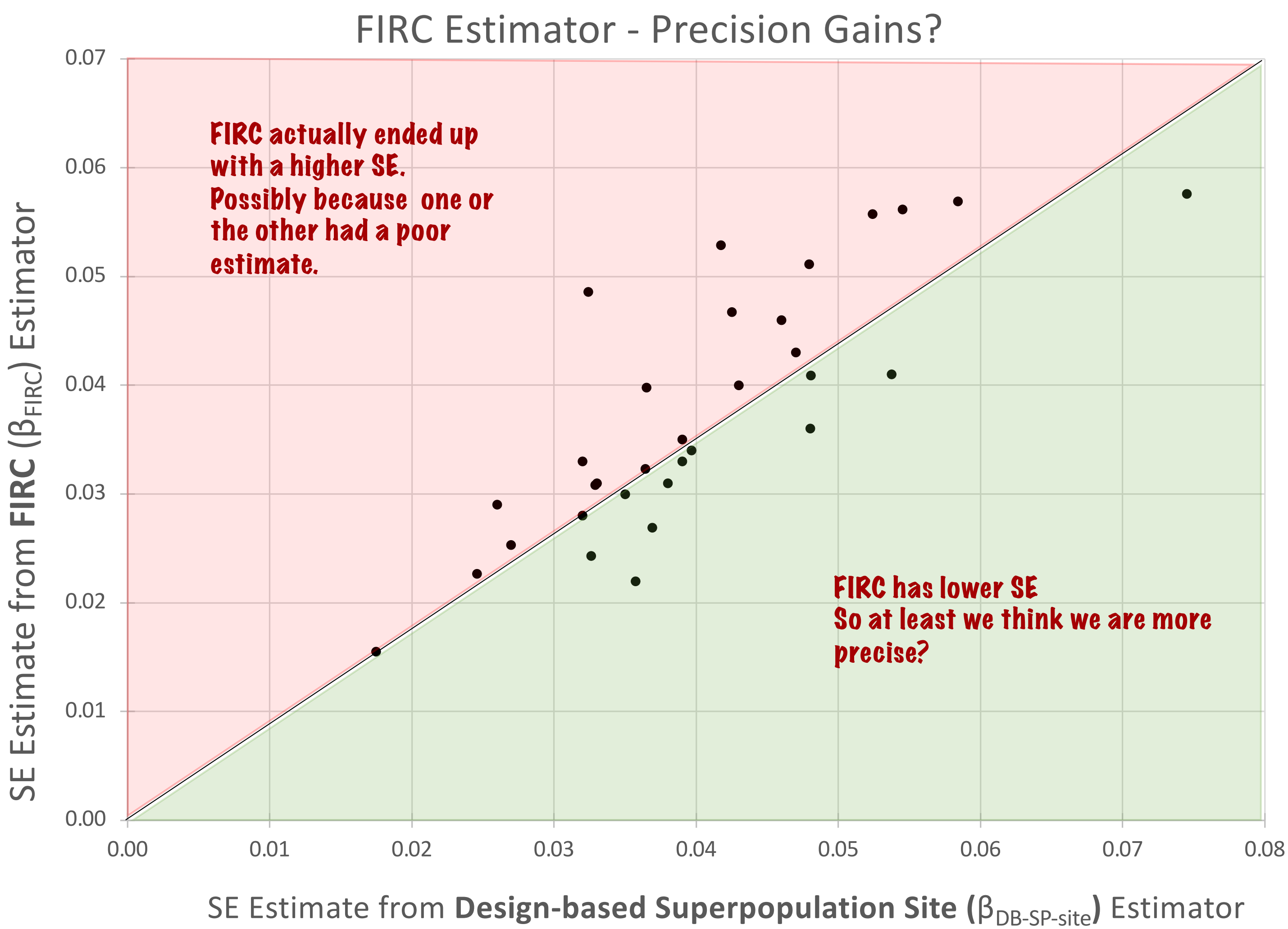


RQ3: Bias Precision Trade-off?
Part II: The estimand of $\beta_{SP-site}$

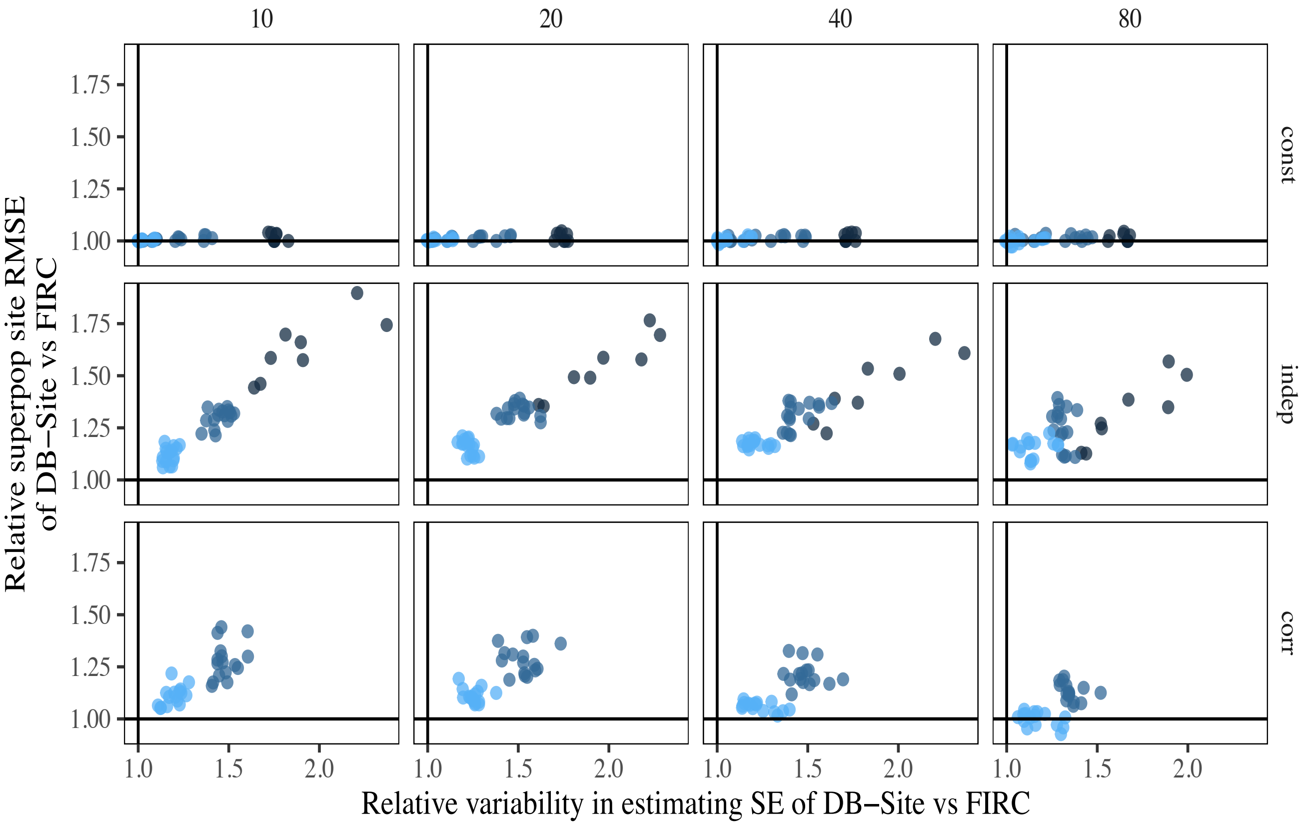
Unbiased Design Based vs. FIRC



FIRC Estimator - Potential for Bias?
We are seeing some differences between FIRC (adaptive in nature) and design based
Effect Estimate from FIRC (β_{FIRC})
Effect Estimate from Unbiased Estimator (with respect to β_{site})
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
0.25

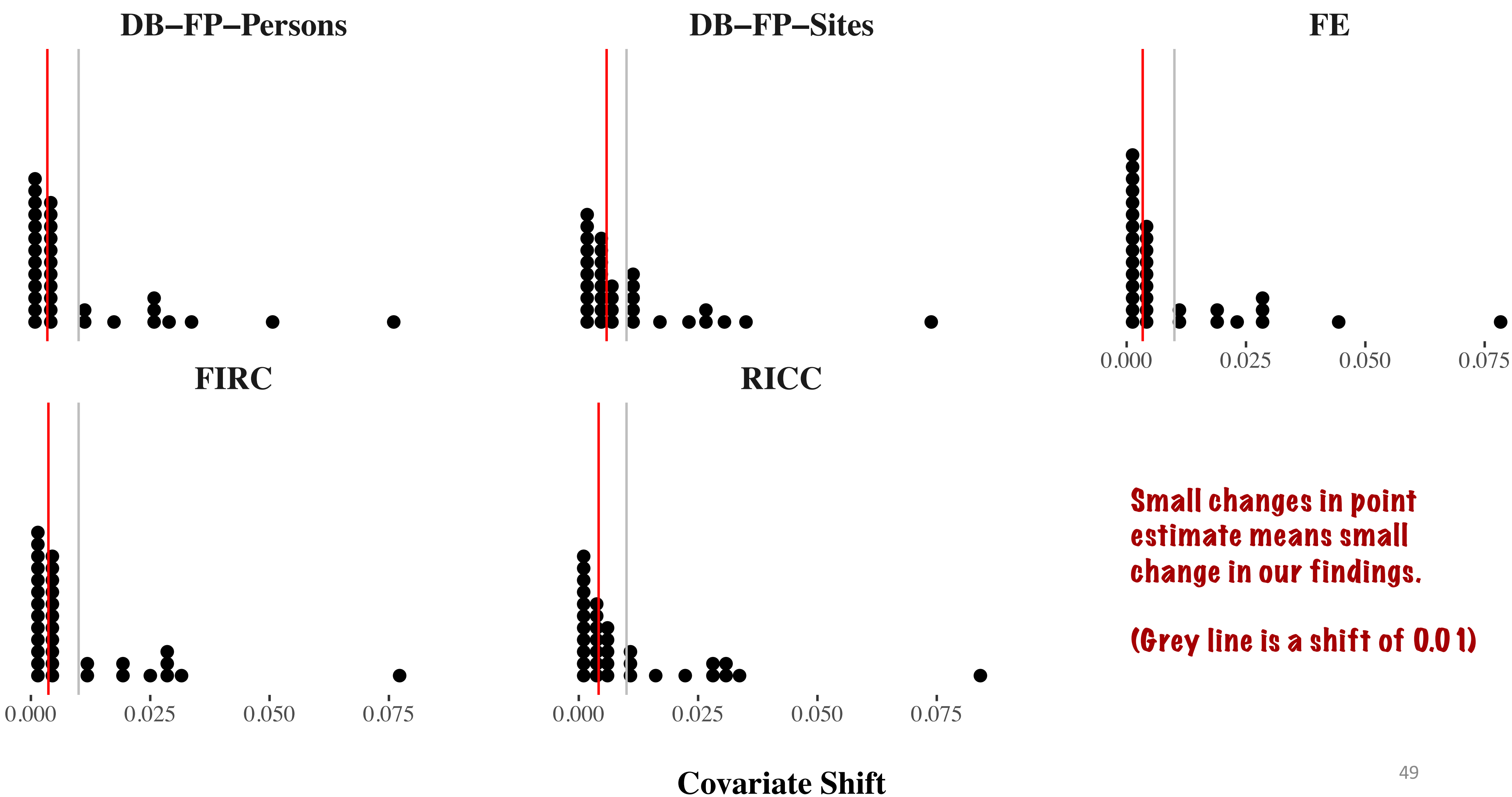

FIRC Estimator - Precision Gains?
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
FIRC actually ended up with a higher SE.
Possibly because one or the other had a poor estimate.
FIRC has lower SE
So at least we think we are more precise?
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
SE Estimate from **FIRC** (β_{FIRC}) Estimator
SE Estimate from **Design-based Superpopulation Site** ($\beta_{DB\text{-}SP\text{-}site}$) Estimator

Simulation Commentary: Infinite Site is hard and the $\beta_{SP-person}$ is a troublemaker



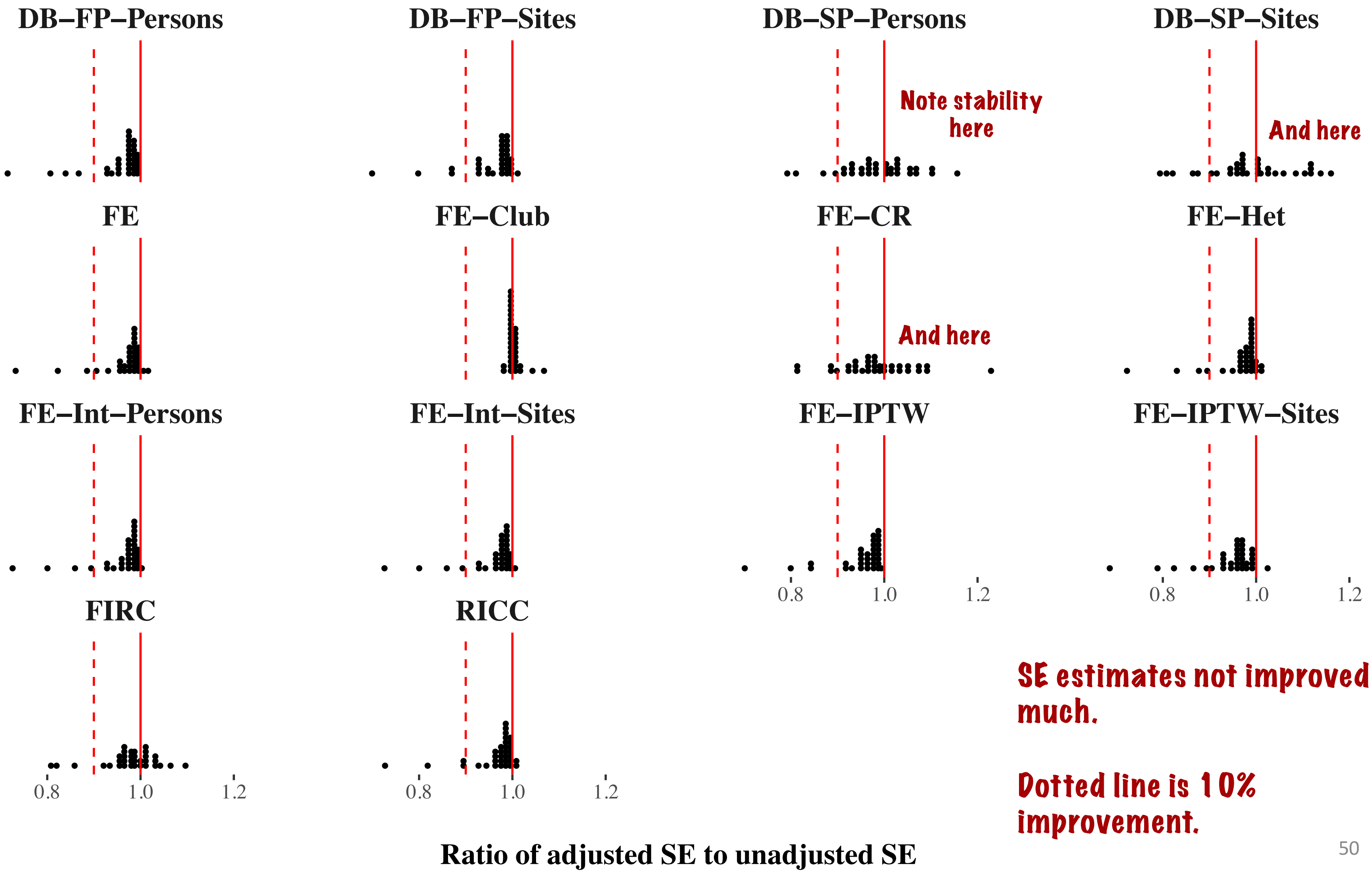
- FIRC is adaptive, clinging to fixed effect fairly strongly unless there is a large amount of cross site variation.
- That being said, FIRC is quite unstable.
- Unbiased approaches are even more unstable, however.
- Across of all simulation scenarios we consider, the RMSE of FIRC was higher than DB-SP-Site in only 2% of them.

And what about covariate adjustment?



Small changes in point estimate means small change in our findings.

(Grey line is a shift of 0.01)



SE estimates not improved much.

Dotted line is 10% improvement.

Two additional resources with our paper
(3 papers for the price of one?)

A) Technical appendix gives overview of all estimators with some details and notes on their use

B) Multifactor simulation appendix explores estimator performance under hypothetical MLM DGP

The stand we take

- Estimand choice matters.
- β estimator choice matters for site-super estimand, otherwise not much
- $\mathrm{SE}(\hat{\beta})$ estimator matters for site estimands, much less for person estimands
- The superpopulation site estimators differ the most, and are the most unstable (difficult).

Thank you

Luke Miratrix
lmiratrix@g.harvard.edu

Michael Weiss
Michael.Weiss@mdrc.org

Brit Henderson
Brit.Henderson@mdrc.org

Thanks to Mike Weiss for making most of these slides for an initial presentation