Improving the precision of instrumental variable estimators with post-stratification

Nicole Pashley (Rutgers University)
Joint work with Luke Keele (University of Pennsylvania) & Luke Miratrix (Harvard University)

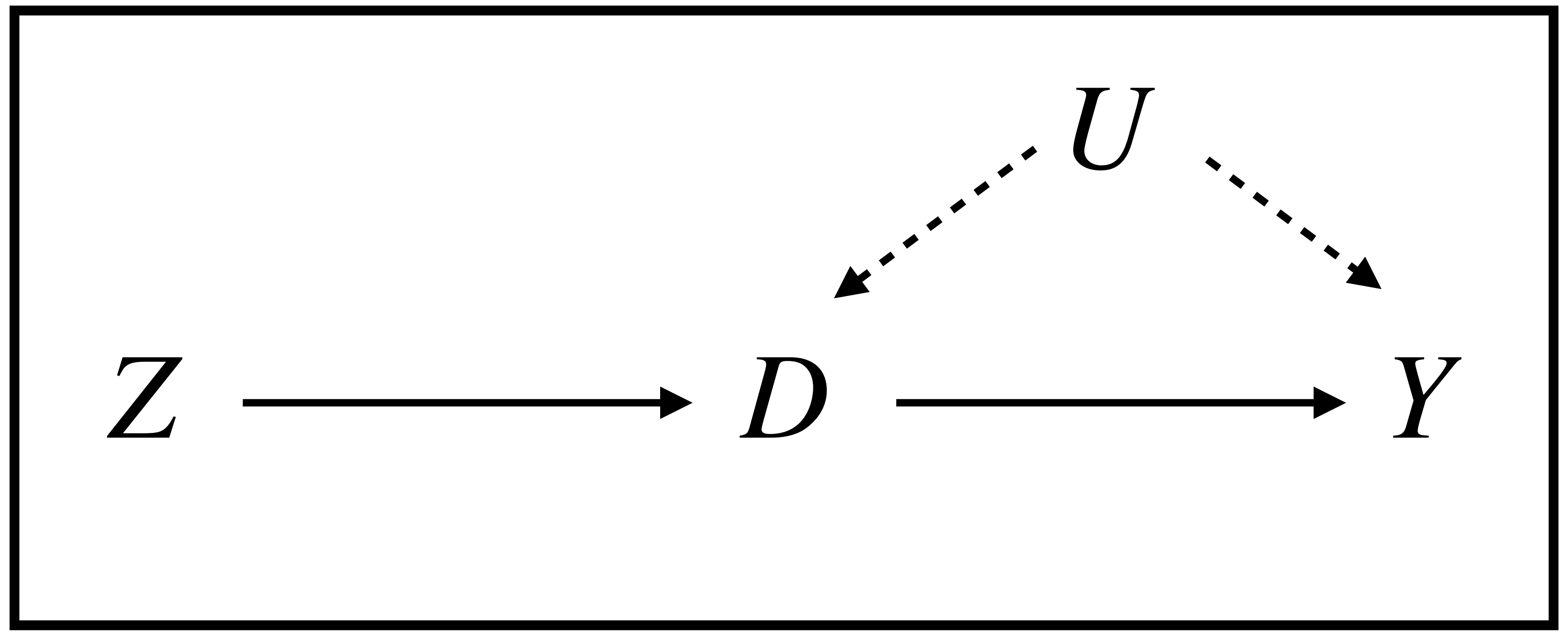
DAE 2024

Outline

- Noncompliance and instrumental variables: Introduction
- Post-stratification with IV
- Simulations

Noncompliance problem

- We have a sample of N participants and we are able to run an experiment randomly assigning (complete randomization) the treatment to individuals
    - $Z_i \in \{0,1\}$: Assignment of unit i (random)
- But individuals don’t always do what they were assigned to do
    - $D_i(z) \in \{0,1\}$: Treatment uptake of unit i if assigned to treatment z (fixed)
- How do we learn about the effect of treatment *uptake* on the outcome?
    - $Y_i(z)$: Potential outcome of unit i if assigned to treatment z (fixed)



Noncompliance problem

- **Principal stratification**: stratify units based on uptake profiles (i.e., on pair of $D_i(z)$ values).

Do what they're told

Refuse to take treatment

	$D_i(1) = 1$	$D_i(1) = 0$
$D_i(0) = 0$	Complier	Never-taker
$D_i(0) = 1$	Always-taker	Defier

Do the opposite of what they are told

Take treatment no matter what

Goal: Learn about the effect of treatment uptake among compliers

Instrumental variables analysis

$$\begin{tikzcd} & & U \arrow[dl, dashed] \arrow[dr, dashed] & \\ Z \arrow[r] & D \arrow[rr] & & Y \end{tikzcd}$$

- Instrumental variables (IV) assumptions (Angrist, Imbens, & Rubin, 1996):
  - **Effective randomization:** Z independent of $\{D(1), D(0)\}$ and $\{Y(1), Y(0)\}$
  - **Exclusion restriction**: $Y_i(1) = Y_i(0)$ if $D_i(1) = D_i(0)$
  - **Monotonicity**: $D_i(1) \geq D_i(0) \implies$ no defiers.
  - **Relevance:** Nonzero effect of Z on D (i.e., there are some compliers)

Instrumental variables analysis

$$
\begin{tikzcd}
 & & U \arrow[dl, dashed] \arrow[dr, dashed] & \\
Z \arrow[r] & D \arrow[rr] & & Y
\end{tikzcd}
$$

- Under the prior assumptions….

Average potential outcome among compliers

$$\text{ITT} = \bar{Y}(1) - \bar{Y}(0) = \frac{1}{N}\sum_{i=1}^{N} Y_i(1) - \frac{1}{N}\sum_{i=1}^{N} Y_i(0) = \pi_c(\bar{Y}_c(1) - \bar{Y}_c(0))$$

Intent-to-treat (ITT) effect of Z on Y

Proportion compliers

$$\pi_c = \bar{D}(1) - \bar{D}(0) = \frac{1}{N}\sum_{i=1}^{N} D_i(1) - \frac{1}{N}\sum_{i=1}^{N} D_i(0)$$

ITT of Z on D

- Under these assumptions, can write the **Complier Average Causal Effect (CACE)** as ratio of intent-to-treat (ITT) of Z on Y over the proportion of compliers

$$\tau = \bar{Y}_c(1) - \bar{Y}_c(0) = \frac{\bar{Y}(1) - \bar{Y}(0)}{\bar{D}(1) - \bar{D}(0)} = \frac{\text{ITT}}{\pi_c}$$

- Consistent estimator for the **Complier Average Causal Effect (CACE):**

$$\tau = \frac{\bar{Y}^{obs}(1) - \bar{Y}^{obs}(0)}{\bar{D}^{obs}(1) - \bar{D}^{obs}(0)} = \frac{\widehat{\text{ITT}}}{\hat{f}}$$

Strong vs weak instrument

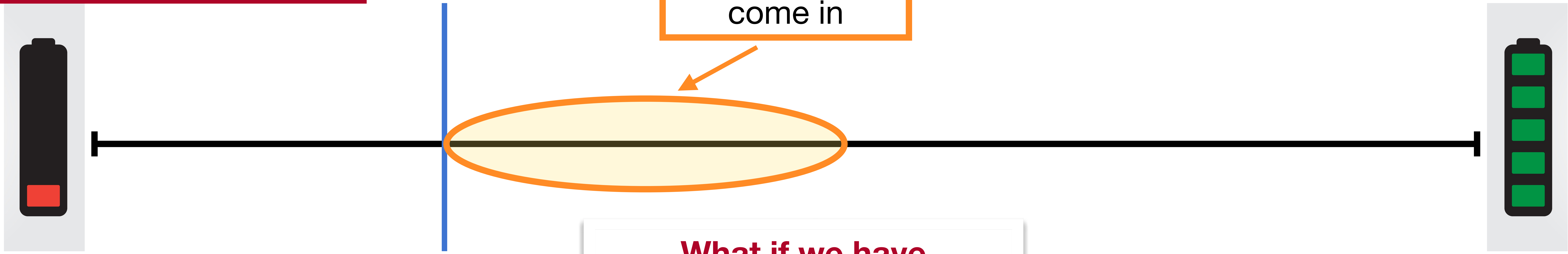
Weak instrument
Instrument has nonzero but weak effect on treatment (Bound, Jaeger, & Baker, 1995; Staiger & Stock, 1997)

I.e., Weak correlation between Z and D

Strong instrument
Instrument has strong effect on treatment

I.e., Strong correlation between Z and D

Where we come in



Pass weak instrument test

What if we have information that helps predict whether you are a complier?

Outline

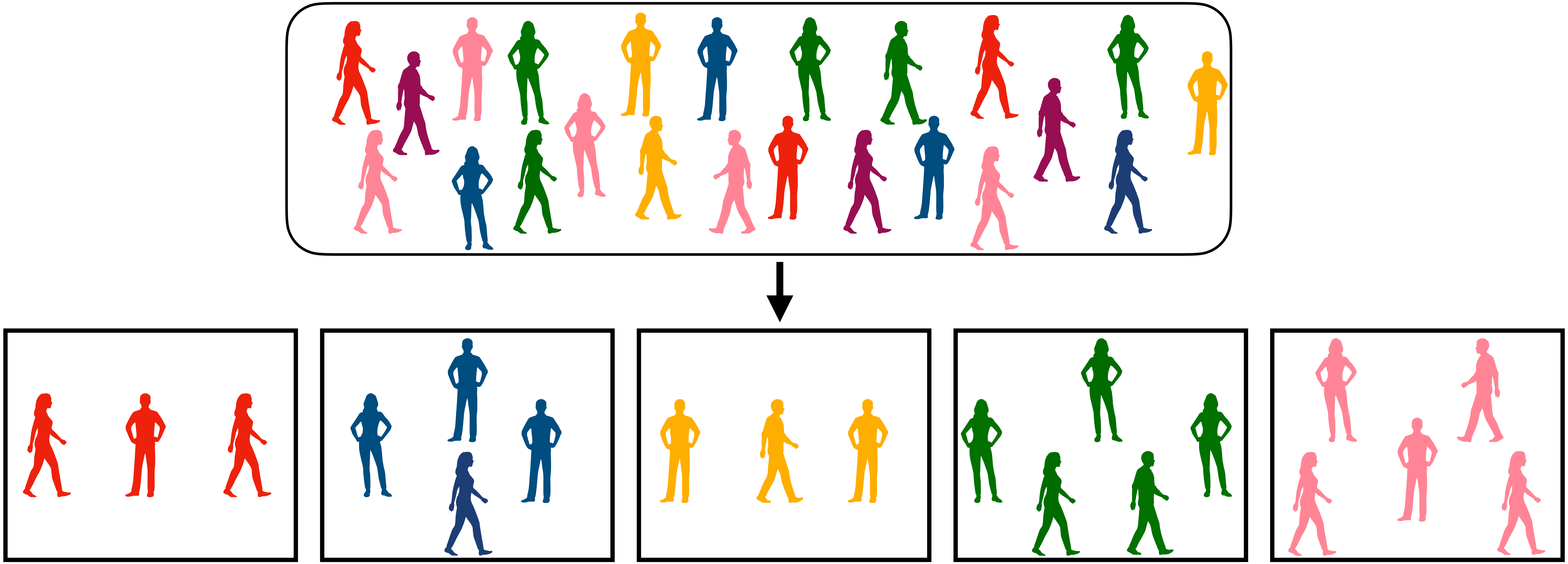
- Noncompliance and instrumental variables: Introduction
- Post-stratification with IV
- Simulations

Post-stratification with IV

- **Post-stratification:** After randomization (**post**), break units up into groups (**strata**). Analyze the experiment *as-if* it had been block-randomized, with strata as blocks. (Holt & Smith, 1979; McHugh & Matts, 1983; Miratrix, Sekhon, & Yu, 2013)

Within-stratum IV with post-stratification

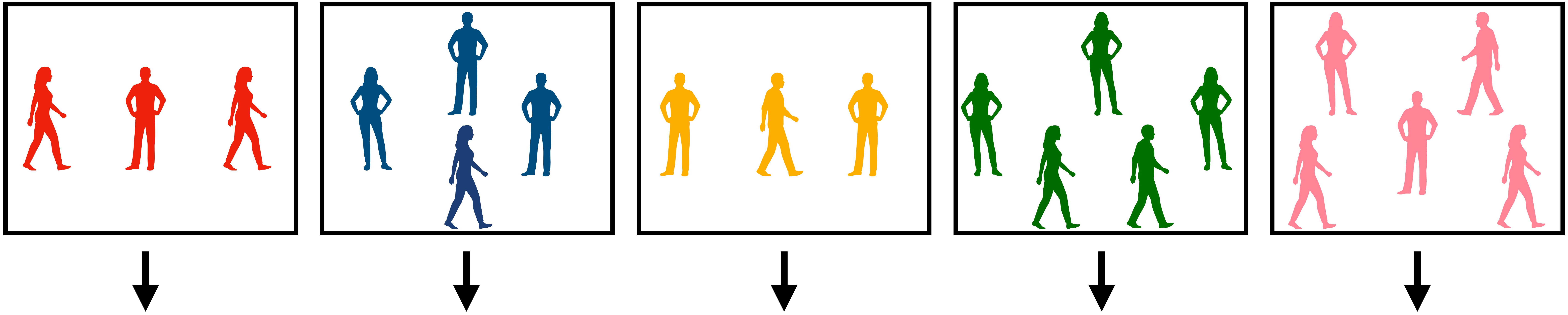
- **Within-stratum IV with post-stratification** ($\widehat{\tau}_{\text{IV-w}}$)**:**
    1. After randomization, break units up into **strata** (based on some covariate).



Within-stratum IV with post-stratification

- **Within-stratum IV with post-stratification** ($\hat{\tau}_{\text{IV-w}}$)**:**

  2. Perform IV analysis within each stratum.



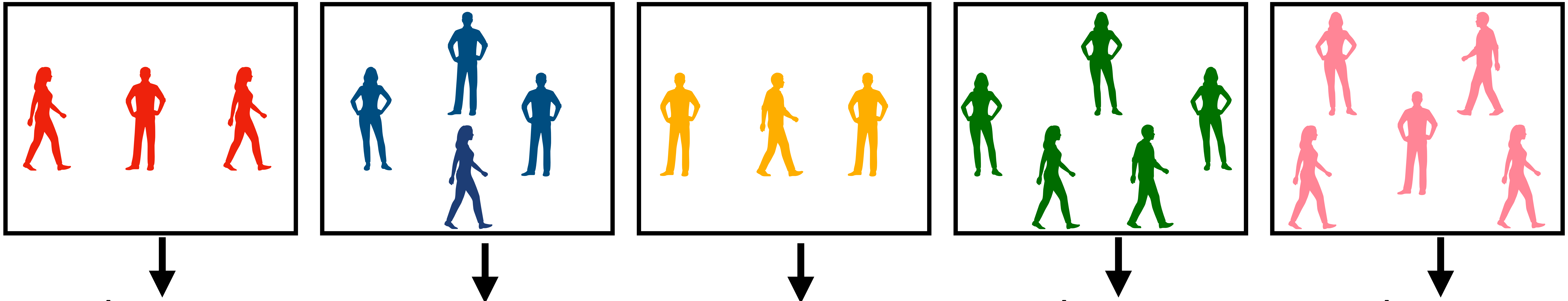
$\hat{\tau}_1$ $\hat{\tau}_2$ $\hat{\tau}_3$ $\hat{\tau}_4$ $\hat{\tau}_5$

$\hat{\tau}_g$: Estimated CACE in stratum g

Within-stratum IV with post-stratification

- **Within-stratum IV with post-stratification** ($\hat{\tau}_{\text{IV-w}}$):

3. Aggregate IV estimates, weighting by estimated # of compliers in stratum.



$$\frac{\hat{f}_1 N_1}{\sum_{k=1}^{5} \hat{f}_k N_k}\hat{\tau}_1 + \frac{\hat{f}_2 N_2}{\sum_{k=1}^{5} \hat{f}_k N_k}\hat{\tau}_2 + \frac{\hat{f}_3 N_3}{\sum_{k=1}^{5} \hat{f}_k N_k}\hat{\tau}_3 + \frac{\hat{f}_4 N_4}{\sum_{k=1}^{5} \hat{f}_k N_k}\hat{\tau}_4 + \frac{\hat{f}_5 N_5}{\sum_{k=1}^{5} \hat{f}_k N_k}\hat{\tau}_5$$

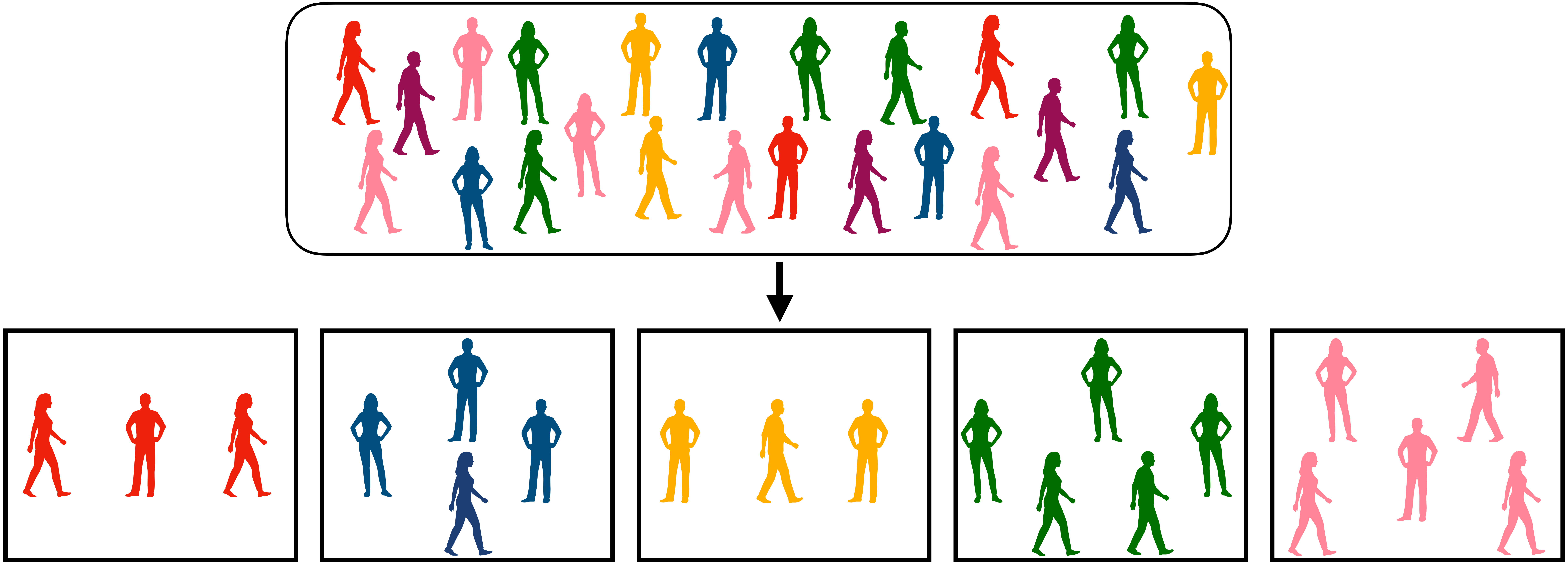
$\hat{f}_g$: Estimated proportion of compliers in stratum g

N_g: Number of units in stratum g

Alternative: Across-strata post-stratification with IV

- **Across-strata post-stratification with IV ($\widehat{\tau}_{\text{IV-a}}$):**

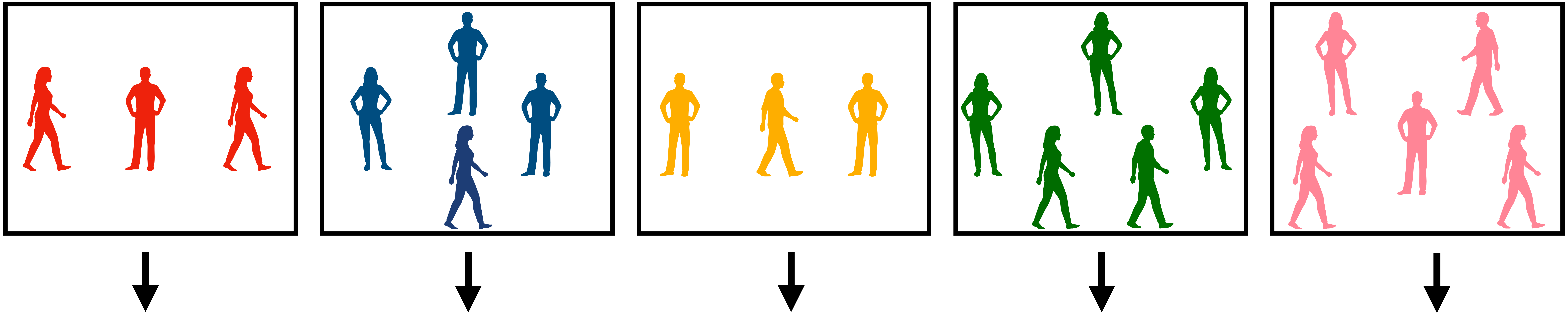
    1. After randomization, break units up into **strata** (based on some covariate).



Alternative: Across-strata post-stratification with IV

- **Across-strata post-stratification with IV ($\hat{\tau}_{\text{IV-a}}$):**

  2. Create a post-stratified estimator of the ITT, weighting by number of units in each stratum ($\widehat{\text{ITT}}_{\text{PS}}$).

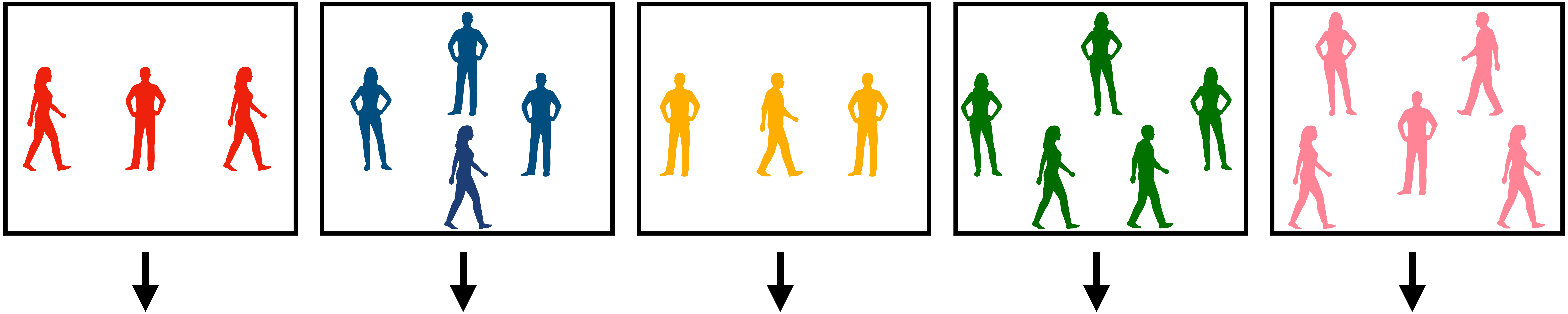


$$\frac{N_1}{N}\widehat{\text{ITT}}_1 + \frac{N_2}{N}\widehat{\text{ITT}}_2 + \frac{N_3}{N}\widehat{\text{ITT}}_3 + \frac{N_4}{N}\widehat{\text{ITT}}_4 + \frac{N_5}{N}\widehat{\text{ITT}}_5$$

Alternative: Across-strata post-stratification with IV

- **Across-strata post-stratification with IV** ($\hat{\tau}_{\text{IV-a}}$)**:**

    3. Create a post-stratified estimator of proportion of compliers, weighting by number of units in each stratum ($\hat{f}_{\text{PS}}$).



$$\frac{N_1}{N}\hat{f}_1 + \frac{N_2}{N}\hat{f}_2 + \frac{N_3}{N}\hat{f}_3 + \frac{N_4}{N}\hat{f}_4 + \frac{N_5}{N}\hat{f}_5$$

Post-stratification with IV

- **Post-stratified IV estimators:**

$$\hat{\tau}_{\text{IV-w}} = \sum_{g=1}^{G} \frac{\hat{f}_g N_g}{\sum_{k=1}^{G} \hat{f}_k N_k} \hat{\tau}_g = \sum_{g=1}^{G} \frac{\hat{f}_g N_g}{\sum_{k=1}^{G} \hat{f}_k N_k} \frac{\widehat{\text{ITT}}_g}{\hat{f}_g}$$

$$\hat{\tau}_{\text{IV-a}} = \frac{\widehat{\text{ITT}}_{\text{PS}}}{\hat{f}_{\text{PS}}} = \frac{\sum_{g=1}^{G} \frac{N_g}{N} \widehat{\text{ITT}}_g}{\sum_{g=1}^{G} \frac{N_g}{N} \hat{f}_g}$$

- If there are no strata with 0 estimated compliers, these estimators will be the same.
- Otherwise, $\hat{\tau}_{\text{IV-w}}$ will **drop** those strata estimated to have 0 compliers
    - This helps stabilize the estimator

Potential Benefits of Post-stratification (without dropping strata)

1. **Increased (asymptotic) precision** (under some regularity conditions)
   - Post-stratification will increase precision if within strata variability of $Y_i(z) - \tau D_i(z)$ is small and across strata variability of $Y_i(z) - \tau D_i(z)$ is large
2. **Improved estimators for standard error**
3. **Decreased bias**
   - Biased depends on **variance** of $\hat{f}$ (or $\hat{f}_{\text{PS}}$) and **covariance** between $\hat{f}$ (or $\hat{f}_{\text{PS}}$) and estimated $\widehat{\text{ITT}}$ (or $\widehat{\text{ITT}}_{\text{PS}}$)

We did math!
Equations and technical details available in manuscript

OK theory is great…but what about in practice?

Stratifying on covariates predictive of…	Performance
Outcome	Successful at reducing variance whether we keep all strata or not
Compliance	Does not lead to significant reduction in variance **unless** we drop strata with low compliance

Other estimators: DSF, DSS, DSS0

- **Idea:** Drop strata that we believe contain a small proportion of compliers
- Use $\widehat{\tau}_{\text{IV-w}}$ but drop strata according to the rule…

Name	Drop any stratum that...
Drop-Small-F (DSF)	fails the weak instrument test (based on the F statistic >10 for importance of ***Z*** in predicting ***D***)
Drop-Small-Strata (DSS)	has less than 2% estimated compliers
Drop-Small-Strata-less-than-0 (DSS0)	has 0 or less estimated compliers (can get negative estimates with two-sided noncompliance)

Other estimators: PWIV

- **Precision-Weighted-(IV) (PWIV)**: Use $\widehat{\tau}_{\text{IV-w}}$ but weight by (Bloom) estimated inverse variance of strata,

$$\widehat{\tau}_{\text{PWIV}} = \frac{1}{c} \sum_{g=1}^{G} \frac{\widehat{f}_g^2}{\widehat{\text{var}}(\widehat{ITT}_g)} \widehat{\tau}_g$$

Normalizing constant

Depends on size of strata (inversely) and variability within strata

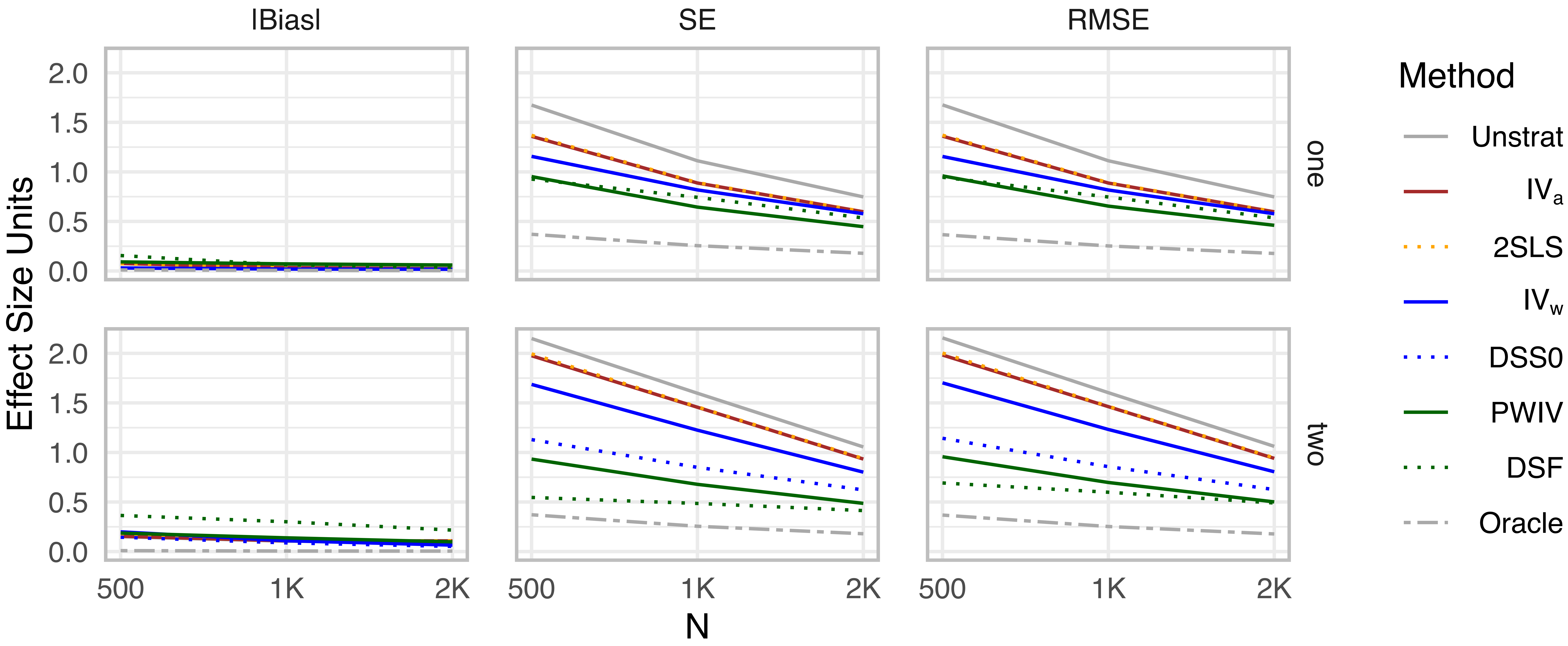
- All of these alternative estimators present a **bias-variance tradeoff**

Outline

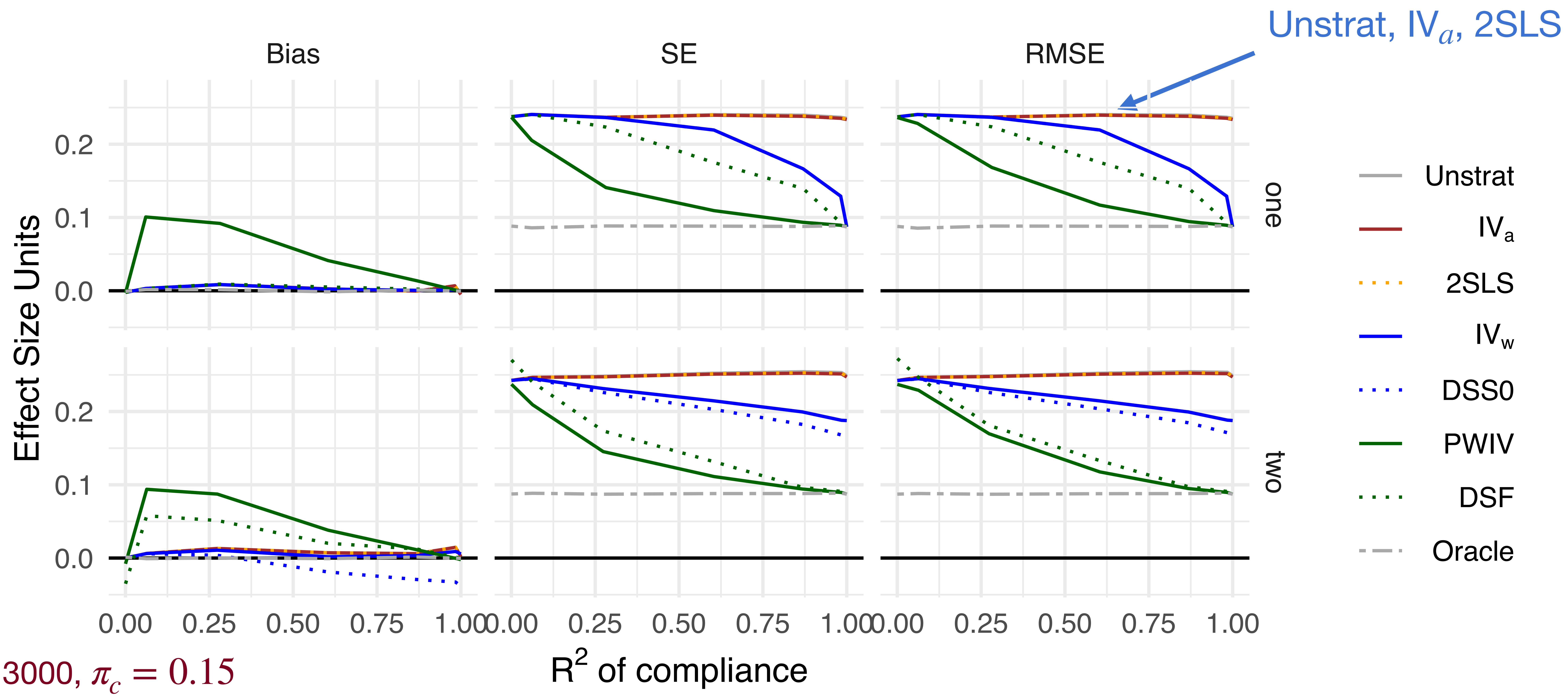
- Noncompliance and instrumental variables: Introduction
- Post-stratification with IV
- Simulations

Results: Overall comparison

Averaging over: Whether stratification is predictive of compliance or outcomes, proportion of compliers, how different compliers and noncompliers are, heterogeneity of treatment effects



Results: Performance based on compliance prediction



- N = 3000, $\pi_c = 0.15$
- Stratifying by compliance
- Stratifying by size of treatment effect among compliers (compliance and CACE correlated)

Take-aways

- **Post-stratification** can help improve **precision and stability** of IV estimators if…
  - We stratify on a covariate predictive of compliance **and either** (a) drop strata with low or no compliers (b) deliberately re-weight strata to increase precision
  - We stratify on a covariate predictive of outcome
- **Post-stratification** can also aid in **bias reduction** but the effects on SE appear more significant
- Standard two-stage least squares analyses with covariates that are predictive of compliance will fail to take advantage of the benefits of compliance knowledge

Thank you!

Luke J. Keele

lukekeele.com

Luke W. Miratrix

scholar.harvard.edu/lmiratrix

Nicole E. Pashley

sites.google.com/view/npashley/

Paper: Pashley, N. E., Keele, L., & Miratrix, L. W. (2023). Improving instrumental variable estimators with post-stratification. arXiv preprint arXiv:2303.10016.
https://arxiv.org/abs/2303.10016

References

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), 444-455.

Baiocchi, M., Small, D. S., Lorch, S., & Rosenbaum, P. R. (2010). Building a stronger instrument in an observational study of perinatal care for premature infants. *Journal of the American Statistical Association*, *105*(492), 1285-1296.

Baiocchi, M., Small, D. S., Yang, L., Polsky, D., & Groeneveld, P. W. (2012). Near/far matching: a study design approach to instrumental variables. *Health Services and Outcomes Research Methodology*, *12*(4), 237-253.

Bound, J., Jaeger, D., and Baker, R. (1995). Problems with intrustmental variables esti- mation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association, 90(430):443–450.

Gerber, A. S. and Green, D. P. (2000). The effects of personal canvassing, telephone calls, and direct mail on voter turnout: A field experiment. American Political Science Review, 94(3):653–663.

Green, D. P., Gerber, A. S., and Nickerson, D. W. (2003). Getting out the vote in lo- cal elections: Results from six door-to-door canvassing experiments. Journal of Politics, 65(4):1083–1096.

Holt, D., & Smith, T. F. (1979). Post stratification. Journal of the Royal Statistical Society Series A: Statistics in Society, 142(1), 33-46.

References

Keele, L., & Morgan, J. W. (2016). How strong is strong enough? Strengthening instruments through matching and weak instrument tests. *The Annals of Applied Statistics*, *10*(2), 1086-1106.

Kleiber & Zeileis (2008). Applied Econometrics with R. New York: Springer-Verlag. ISBN 978-0-387-77316-2. URL: https://CRAN.R-project.org/package=AER

McHugh, R., & Matts, J. (1983). Post-stratification in the randomized clinical trial. *Biometrics*, 217-225.

Miratrix, L. W., Sekhon, J. S., and Yu, B. (2013). Adjusting treatment effect estimates by post-stratification in randomized experiments. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75(2):369–396.

Pashley, N. E., & Miratrix, L. W. (2022). Block what you can, except when you shouldn't. *Journal of Educational and Behavioral Statistics*, *47*(1), 69-100.

(SPOT)light Investigators (2010). (SPOT)light study website.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. Econometrica, 65:557–586.

Zubizarreta, J. R., Small, D. S., Goyal, N. K., Lorch, S., & Rosenbaum, P. R. (2013). Stronger instruments via integer programming in an observational study of late preterm birth outcomes. *The Annals of Applied Statistics*, 25-50.

Connections to matching and IV literature

Baiocchi, Small, Lorch, & Rosenbaum, 2010
Baiocchi, Small, Yang, Polsky,& Groeneveld, 2012
Keele & Morgan, 2016
Zubizarreta, Small, Goyal, Lorch, & Rosenbaum, 2013

Connections to matching and IV literature

- Baiocchi, Small, Lorch, & Rosenbaum, 2010
- Baiocchi, Small, Yang, Polsky,& Groeneveld, 2012
- Keele & Morgan, 2016
- Zubizarreta, Small, Goyal, Lorch, & Rosenbaum, 2013

Key equations

- Asymptotic variance of IV estimator: asyVar($\hat{\tau}_{\text{IV}}$) $= \dfrac{1}{\pi_c^2}\text{var}\left(\widehat{\text{ITT}} - \dfrac{\text{ITT}}{\pi_c}\hat{f}\right)$

- Asymptotic variance of post-stratified IV estimator:

$$\text{asyVar}(\hat{\tau}_{\text{IV-a}}) = \frac{1}{\pi_c^2}\text{var}_{\text{PS}}(\widehat{\text{ITT}}_{\text{PS}}) + \tau^2\frac{1}{\pi_c^2}\text{var}_{\text{PS}}(\hat{f}_{\text{PS}}) - 2\tau\frac{1}{\pi_c^2}\text{cov}_{\text{PS}}(\widehat{\text{ITT}}_{\text{PS}}, \hat{f}_{\text{PS}})$$

$$= \frac{1}{\pi_c^2}\text{var}_{\text{PS}}\left(\widehat{\text{ITT}}_{\text{PS}} - \tau\hat{f}_{\text{PS}}\right)$$

Blocking style variance

- Bias:

$$E\left[\frac{\widehat{\text{ITT}}}{\hat{f}}\right] - \tau \approx \frac{1}{\pi_c^2}\left[\tau\text{var}(\hat{f}) - \text{Cov}(\widehat{\text{ITT}}, \hat{f})\right]$$

$$= \frac{1}{\pi_c^2(N-1)}\left[\frac{\pi_n((1-p)\pi_c + \pi_a)}{p(1-p)}(\bar{Y}_n(0) - \bar{Y}_c(0)) + \frac{\pi_a(p\pi_c + \pi_n)}{p(1-p)}(\bar{Y}_c(1) - \bar{Y}_a(1))\right]$$

Potential benefits of post-stratification (without dropping strata)

1. Increased (asymptotic) precision (under some regularity conditions)

- Asymptotic variance of IV estimator: $\text{asyVar}(\hat{\tau}_{\text{IV}}) = \dfrac{1}{\pi_c^2} \boxed{\text{var}\left(\widehat{\text{ITT}} - \tau\hat{f}\right)}$

Standard Neyman variance for a completely randomized experiment with potential outcomes

$P_i(z) = Y_i(z) - \tau D_i(z)$

[i.e., var $\left(\bar{P}^{obs}(1) - \bar{P}^{obs}(0)\right)$]

Potential benefits of post-stratification (without dropping strata)

1. Increased (asymptotic) precision (under some regularity conditions)

- Asymptotic variance of IV estimator: $\text{asyVar}(\hat{\tau}_{\text{IV}}) = \frac{1}{\pi_c^2}\text{var}\left(\widehat{\text{ITT}} - \tau\hat{f}\right)$

- Asymptotic variance of post-stratified IV estimator:

$$\text{asyVar}(\hat{\tau}_{\text{IV-a}}) = \frac{1}{\pi_c^2}\boxed{\text{var}_{\text{Block}}\left(\widehat{\text{ITT}}_{\text{PS}} - \tau\hat{f}_{\text{PS}}\right)}$$

Standard Neyman variance for block-randomized experiment with potential outcomes

$P_i(z) = Y_i(z) - \tau D_i(z)$

[i.e., $\text{var}_{\text{Block}}\left(\bar{P}^{obs}(1) - \bar{P}^{obs}(0)\right)$]

Potential benefits of post-stratification (without dropping strata)

1. **Increased (asymptotic) precision** (under some regularity conditions)
   - **Take away:** Post-stratification will increase precision if within stratum variability of $P_i(z) = Y_i(z) - \tau D_i(z)$ is small and across strata variability of $P_i(z) = Y_i(z) - \tau D_i(z)$ is large
     - Based blocking/post-stratification theory (Miratrix, Sekhon, & Yu, 2013; Pashley & Miratrix, 2022)
   - Note: Additional boost if we can drop entire strata of noncompliers

We did math!
Full equations and all technical details available in manuscript

Potential benefits of post-stratification (without dropping strata)

1. **Increased (asymptotic) precision** (under some regularity conditions)
    - **Take away:** Post-stratification will increase precision if within stratum variability of $P_i(z) = Y_i(z) - \tau D_i(z)$ is small and across strata variability of $P_i(z) = Y_i(z) - \tau D_i(z)$ is large
        - Based blocking/post-stratification theory (Miratrix, Sekhon, & Yu, 2013; Pashley & Miratrix, 2022)
    - Note: Additional boost if we can drop entire strata of noncompliers
2. **Improved estimators for standard error**

Potential benefits of post-stratification (without dropping strata)

3. Decreased bias

- In math:

$$E\left[\frac{\widehat{\text{ITT}}}{\hat{f}}\right] - \tau \approx \frac{1}{\pi_c^2}\left[\tau \text{var}(\hat{f}) - \text{Cov}(\widehat{\text{ITT}}, \hat{f})\right]$$

$$\approx \frac{1}{\pi_c^2(N-1)}\left[\frac{\pi_n((1-p)\pi_c + \pi_a)}{p(1-p)}(\bar{Y}_n(0) - \bar{Y}_c(0)) + \frac{\pi_a(p\pi_c + \pi_n)}{p(1-p)}(\bar{Y}_c(1) - \bar{Y}_a(1))\right]$$

- Biased depends on **variance** of $\hat{f}$ (or $\hat{f}_{\text{PS}}$) and **covariance** between $\hat{f}$ (or $\hat{f}_{\text{PS}}$) and estimated $\widehat{\text{ITT}}$ (or $\widehat{\text{ITT}}_{\text{PS}}$)
- (Approximate) Direction of bias depends on differences between complier and noncomplier average potential outcomes: $\bar{Y}_c(1) - \bar{Y}_a(1)$ and $\bar{Y}_n(0) - \bar{Y}_c(0)$
- Note: Dropping strata with some compliers will introduce bias

Why is stratifying on imperfect compliance not helpful?

- We will increase precision if we can reduce the variability of the quantity

$P_i(z) = Y_i(z) - \tau D_i(z)$ within strata

  - Stratifying based on predicted compliance (mostly) targets the quantity $D_i(z)$
  - If CACE τ is modest in size, this will not reduce variability much

- Consider the across-strata post-stratification estimator:

$$\hat{\tau}_{\text{IV-a}} = \frac{\widehat{\text{ITT}}_{\text{PS}}}{\hat{f}_{\text{PS}}} = \frac{\sum_{g=1}^{G} \frac{N_g}{N} \widehat{\text{ITT}}_g}{\sum_{g=1}^{G} \frac{N_g}{N} \hat{f}_g}$$

If we don't drop strata, their contribution to the numerator is in proportion to their size — strata with low/no compliance are adding noisy estimates of (close to) 0!

Simulation set-up

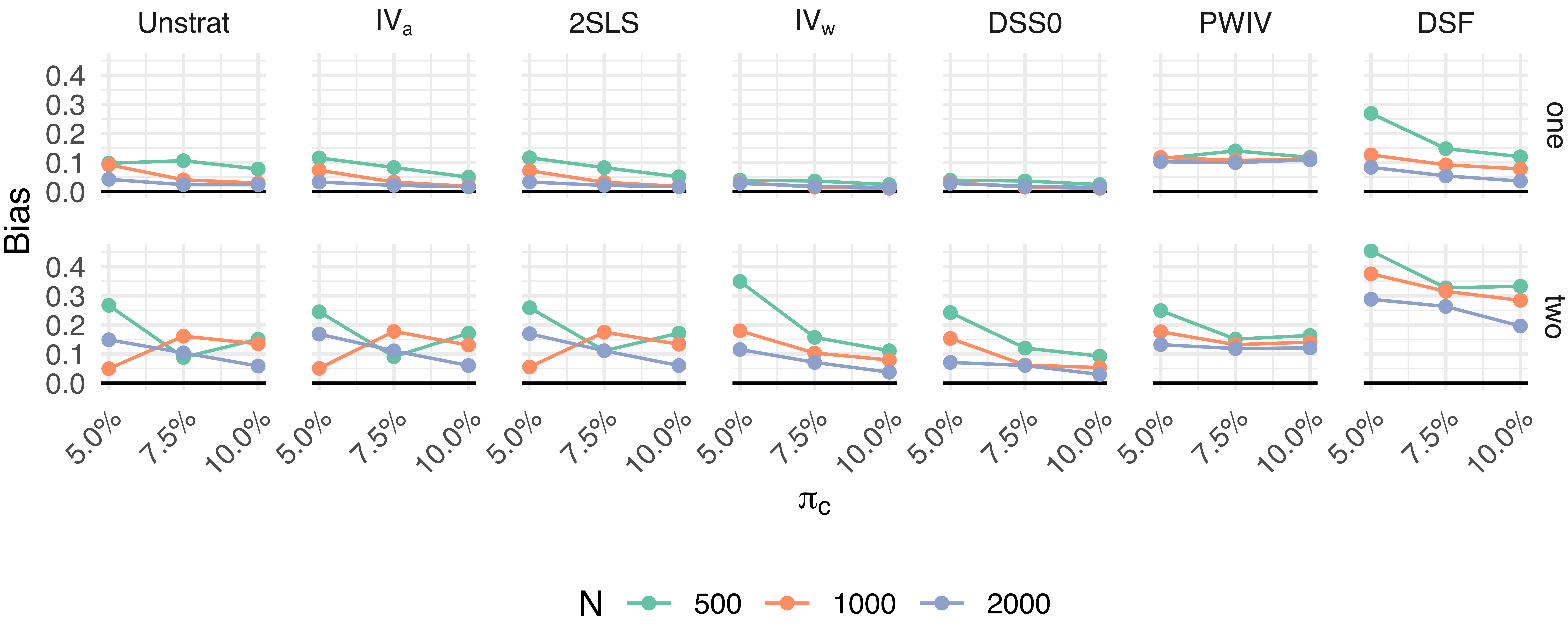
- Simulations varied…

1. Overall size of the experiment (N = 500, 1000, 2000).
2. The overall proportion of compliers (5%, 7.5%, and 10%).
3. One-sided or two-sided noncompliance
4. Whether stratum membership predicts complier status or not.
5. Whether stratum membership predicts the outcome or not.
6. Whether the mean of the never-takers is below, equal to, or above the mean of the compliers' control potential outcomes.
7. Whether the treatment impact is different across strata, or constant.

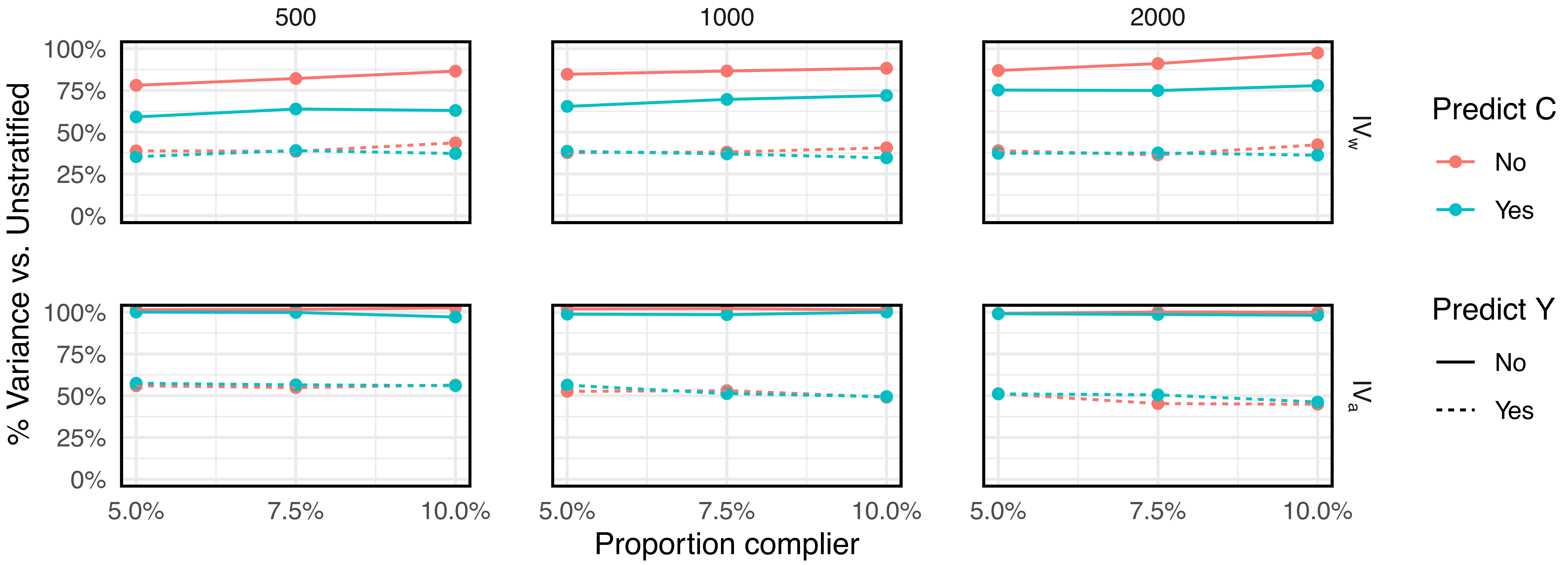
Simulation set-up

- Comparing the following CACE estimators:
  - ✴ Unstratified IV estimator (UNSTRAT)
  - ✴ $\widehat{\tau}_{\text{IV-w}}$ (IV-w)
  - ✴ $\widehat{\tau}_{\text{IV-a}}$ (IV-a)
  - ✴ Drop-Small-F (DSF)
  - ✴ Drop-Small-Strata (DSS)
  - ✴ Drop-Small-Strata-less-than-0 (DSS0)
  - ✴ Precision-Weighting-(IV) (PWIV)
  - ✴ Two-stage least squares, implemented with `ivreg` (Kleiber & Zeileis, 2008) in `R` with stratifying covariate included (2SLS)
  - ✴ Oracle (difference in means for compliers)

Results: Overall comparison



Results: Comparing variances based on stratification variable



Empirical examples

- We consider two **get-out-the-vote (GOTV)** experiments
- **Goal:** Learn about strategies to increase voter turnout
- **Gerber and Green (2000)**: 3 interventions (door-to-door canvasing, phone calls, and mailers) but just focus on door-to-door canvassing.
    - Stratify on age, vote in 1996, household size
- **Green, Gerber, and Nickerson (2003):** Intervention is door-to-door canvassing
    - Stratify based on “turf” which is a geographic region
- **Noncompliance**: Some people assigned a canvasser didn’t open their door

Gerber & Green (2000): Results

Method	CACE estimate	SE estimate	% SE	N
UNSTRAT	0.084	0.0275	100	23,450
IV_w	0.095	0.0238	86.2	23,450
IV_a	0.095	0.0238	86.2	23,450
DSS	0.095	0.0238	86.2	23,450
PWIV	0.092	0.0226	82.0	23,450
DSF	0.095	0.0238	86.2	23,450

Green et al. (2003): Results

Method	CACE estimate	SE estimate	% SE	N
UNSTRAT	0.082	0.023	100	18,661
IV_w	0.055	0.022	96	18,661
IV_a	0.055	0.022	96	18,661
DSS	0.055	0.022	96	18,661
PWIV	0.042	0.020	86	18,661
DSF	0.047	0.022	97	16,010