PENNSTATE

Combining Experimental and Non-Experimental Design in Causal Inference

Kari Lock Morgan
Department of Statistics
Penn State University

Rao Prize Conference
May 12th, 2017

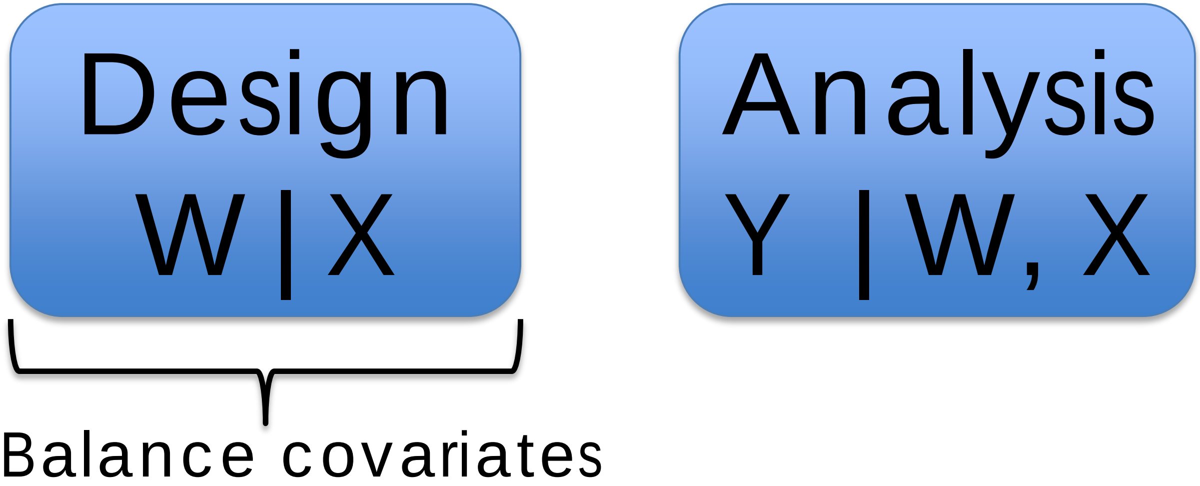
A Tribute to Don

- "Design trumps analysis"
- Motivated by a real study
- Experimental design & rerandomization
- Observational study & propensity scores
- Rubin causal model & potential outcomes
- Educational testing (AP scores)
- (Missing data)
- (Noncompliance)

"Design trumps Analysis"

"For Objective Causal Inference, Design trumps Analysis" – Rubin 2008

X = covariates, W = treatment, Y = outcome(s)

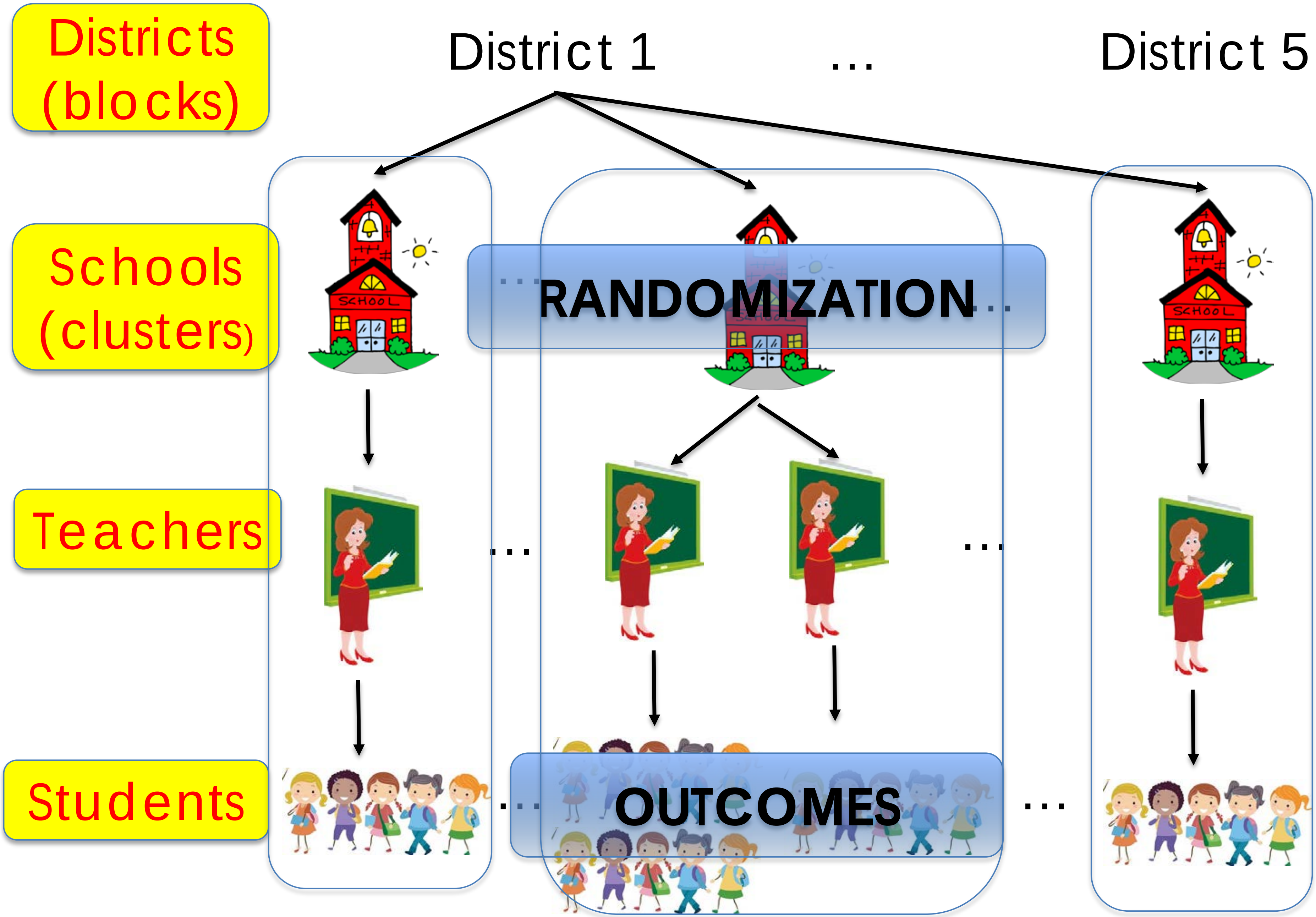


As much as possible should be done *without* observed outcomes!

Knowledge in Action

- Goal: estimate *causal effect* of Knowledge in Action (KIA) (a form of project-based learning) in AP classes on AP scores and other outcomes
- Part 1 (“Efficacy Study”): randomize schools to KIA or control; compare outcomes after 1 year
- Part 2 (“Maturation Study”): continue to follow schools another year (experimental & observational)

*In this talk I'll just focus on one district



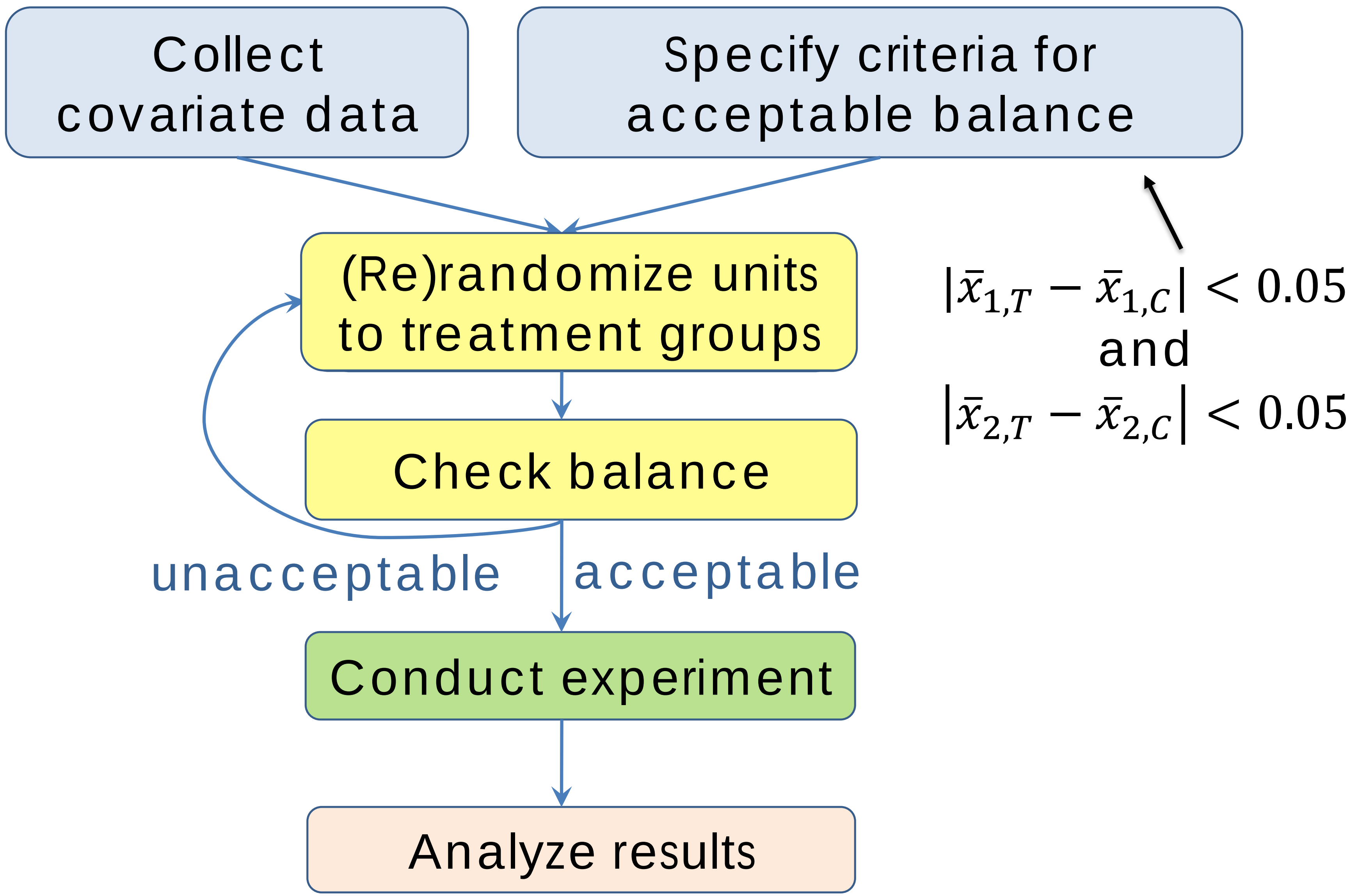
Covariates

- Covariates available at randomization:
  - School covariates (e.g. Title 1 status, type, etc.)
  - Teacher covariates (e.g. years of experience)
  - Previous student (class) covariates:
    - Race/ethnicity → X_2
    - Poverty status → X_2
    - Parental education
    - PSAT scores → X_1
    - 8th grade standardized test scores → X_1
    - Total number of students
    - Number of students who took the AP exam

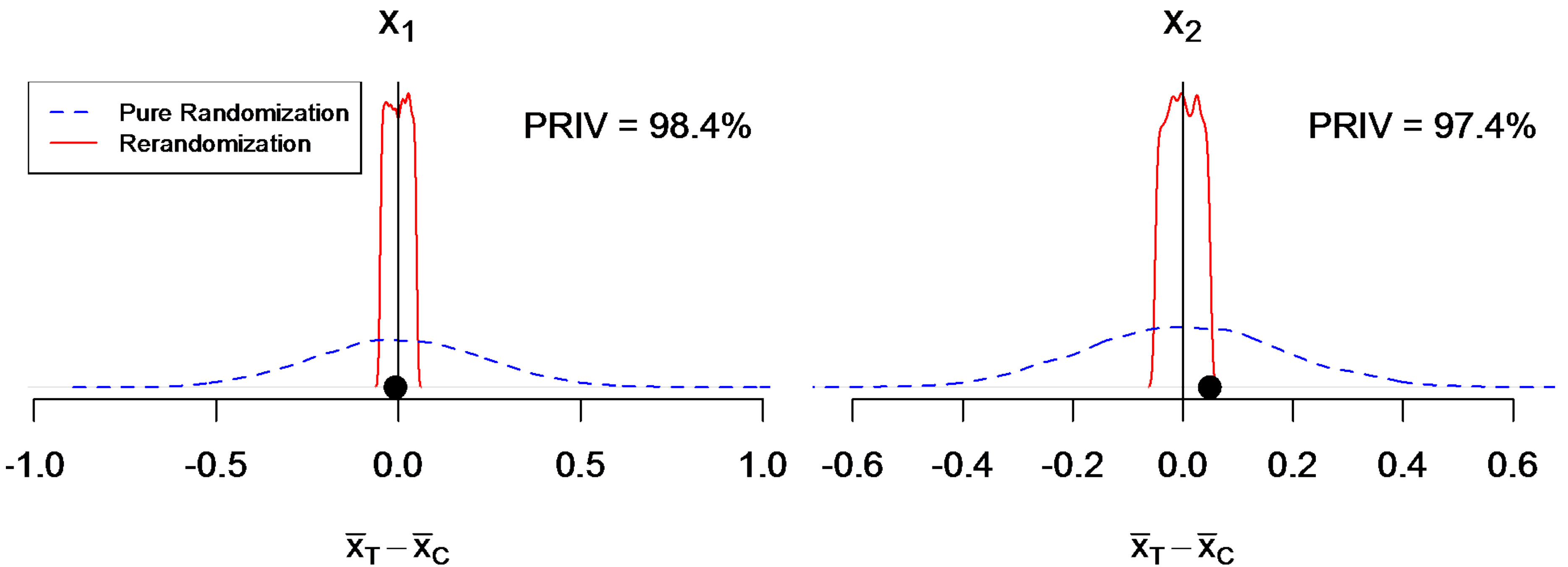
2 covariates used for randomization

- If covariates are available, we should use them when we randomize!

Rerandomization



Covariate Balance: Empirical



Percent reduction in variance:

$$PRIV = \frac{var(\bar{x}_{j,T} - \bar{x}_{j,C}) - var(\bar{x}_{j,T} - \bar{x}_{j,C} \mid rerand.)}{var(\bar{x}_{j,T} - \bar{x}_{j,C})}$$

Covariate Balance: Theoretical

Suppose

- $\bar{x}_{j,T} - \bar{x}_{j,C} \sim$ Normal for $j \in 1 \ldots k$
- $x_1 \perp x_2 \perp \ldots \perp x_k$
- Rerandomize if $\left|\bar{x}_{j,T} - \bar{x}_{j,C}\right| \geq a_j$ for $j \in 1 \ldots k$

Then the PRIV for x_j is

$p_{x_1} = 0.984$

$p_{x_2} = 0.973$

$$p_{z_j} = 1 - 2\left[\frac{\gamma\left(\frac{3}{2}, \frac{a_j^2}{2var(x_j)\left({}^1/_{n_T} + {}^1/_{n_C}\right)}\right)}{\gamma\left(\frac{1}{2}, \frac{a_j^2}{2var(x_j)\left({}^1/_{n_T} + {}^1/_{n_C}\right)}\right)}\right],$$

where $\gamma(b, c) \equiv \int_0^c y^{b-1} e^{-y} dy$.

Outcome PRIV

- If rerandomization is equal percent variance reducing (EPVR), then PRIV for the ***outcome*** difference in means is

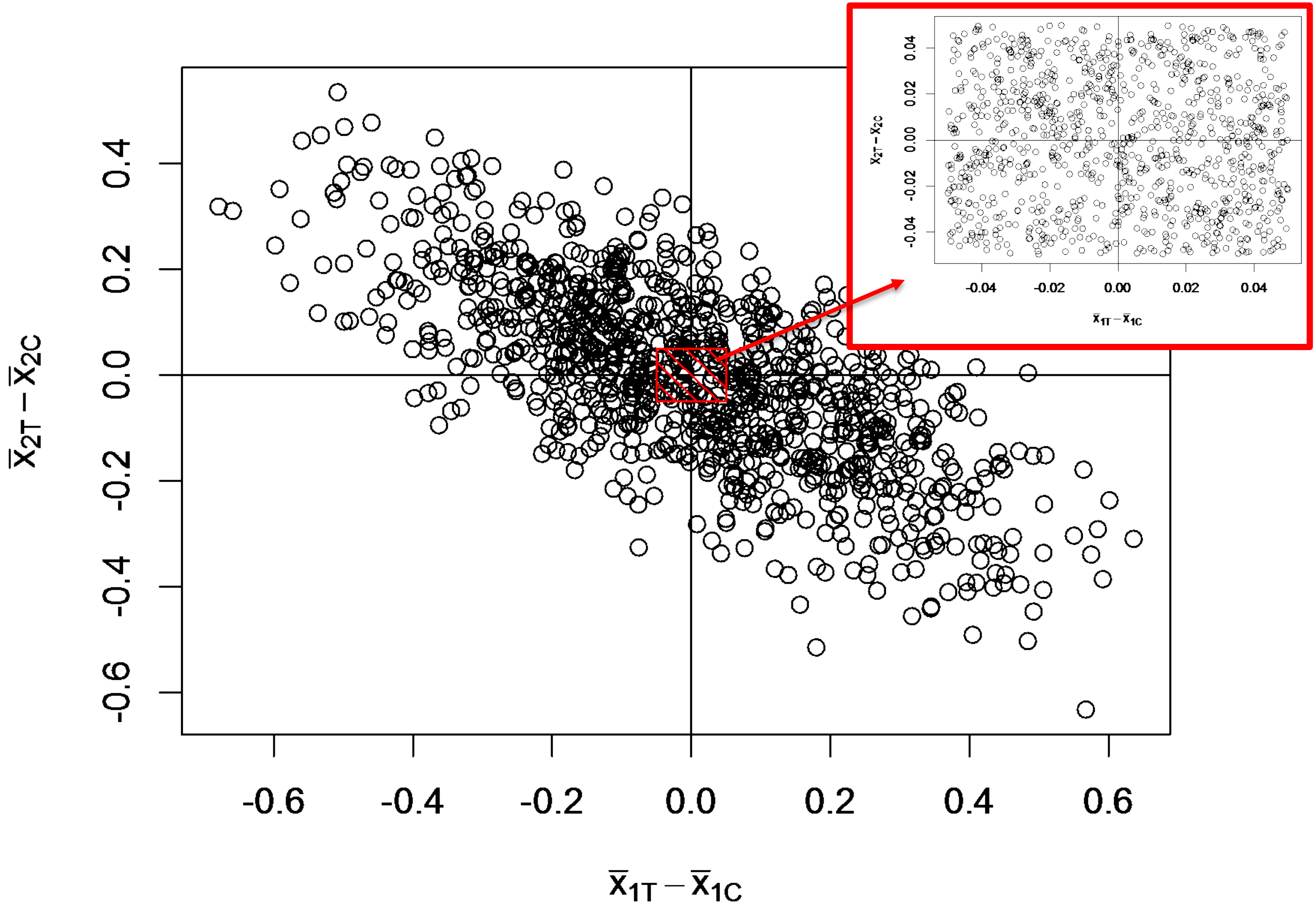
$$PRIV_Y = R^2 \times PRIV_X$$

- Here, $R^2 \approx 0.75$ and $PRIV_X \approx 98\%$, so

$$PRIV_Y \approx 0.75 \times 0.98 = 74\%$$

- Precision increases by a factor of $\frac{1}{(1-0.74)} = 3.85$
- *Equivalent to almost quadrupling n!!!* (Effective sample size goes from 76 to 293!)
- NOTE: This is TRUE variance! Need randomization-based inference to reflect this…

Correlational Structure



Affine Invariance

<u>Affine invariance:</u> rerandomization stays the same for any affine transformation *a* + ***bx***

If rerandomization criterion is affinely invariant and ***x*** is ellipsoidally symmetric…

1. $\mathrm{E}(\bar{\boldsymbol{X}}_T - \bar{\boldsymbol{X}}_c \mid \text{rerand.}) = \mathrm{E}(\bar{\boldsymbol{X}}_T - \bar{\boldsymbol{X}}_c) = \mathbf{0}$

=> Rerandomization leads to unbiased estimates for any linear function of ***x***

2. $\mathrm{cov}(\bar{\boldsymbol{X}}_T - \bar{\boldsymbol{X}}_c \mid \text{rerand.}) \propto \mathrm{cov}(\bar{\boldsymbol{X}}_T - \bar{\boldsymbol{X}}_c)$

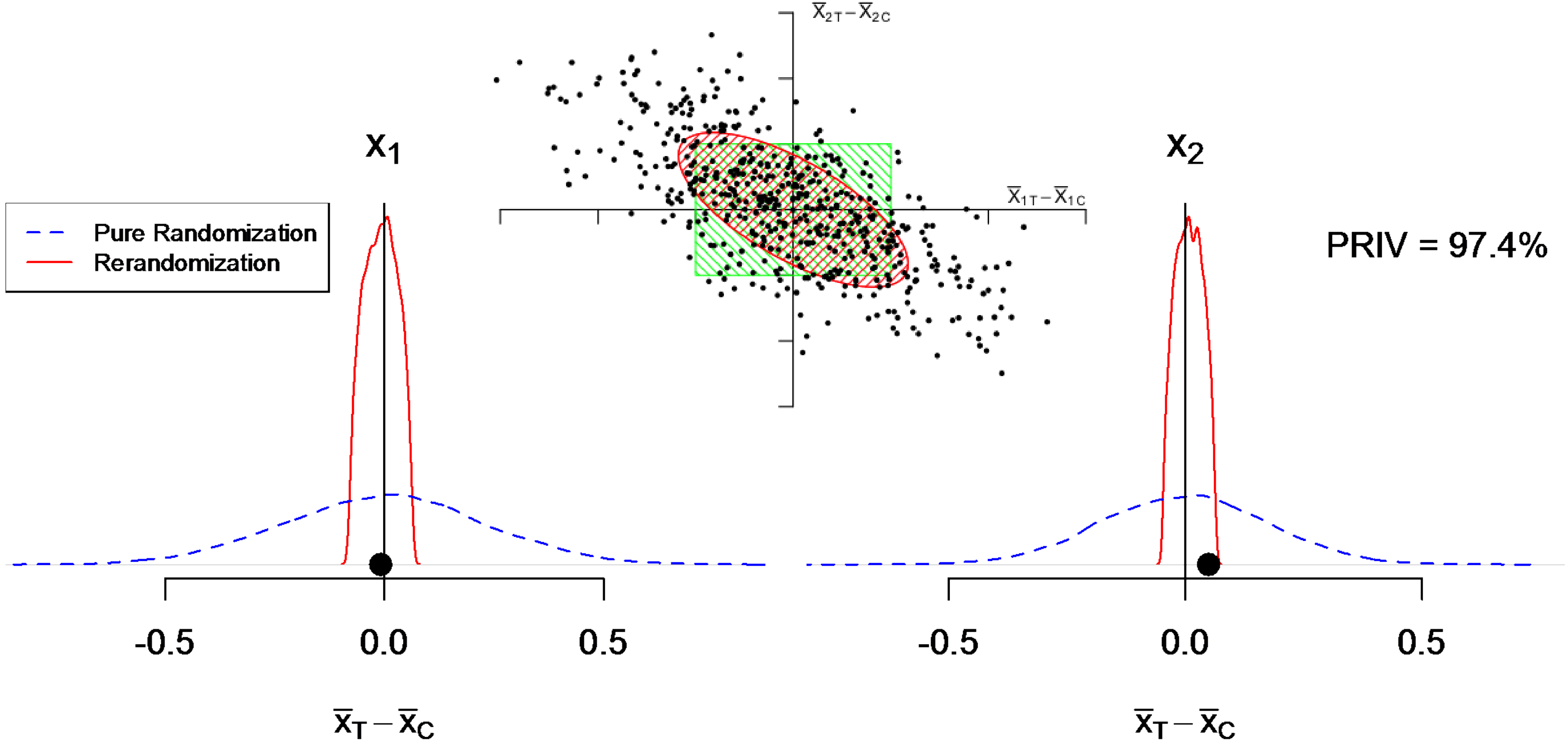
$\Rightarrow$ Preserves the correlations of $\bar{\boldsymbol{X}}_T - \bar{\boldsymbol{X}}_c$

$\Rightarrow$ Balance improvement equal for each x_j (equal percent variance reducing)

(Morgan and Rubin, Annals of Statistics, 2012)

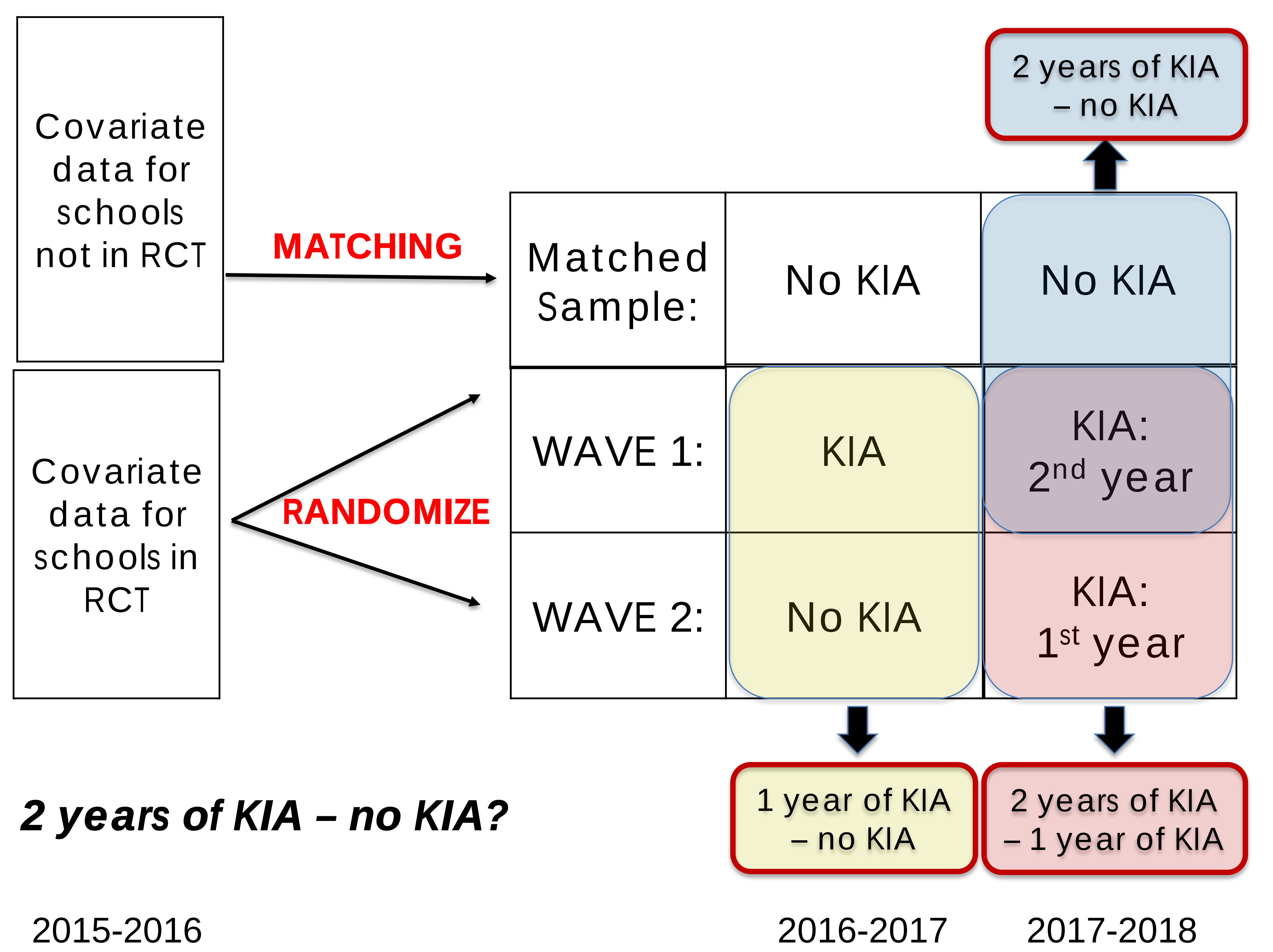
Mahalanobis

Mahalanobis: $(\bar{\boldsymbol{X}}_T - \bar{\boldsymbol{X}}_c)'[\text{cov}(\boldsymbol{x})]^{-1}(\bar{\boldsymbol{X}}_T - \bar{\boldsymbol{X}}_c)$

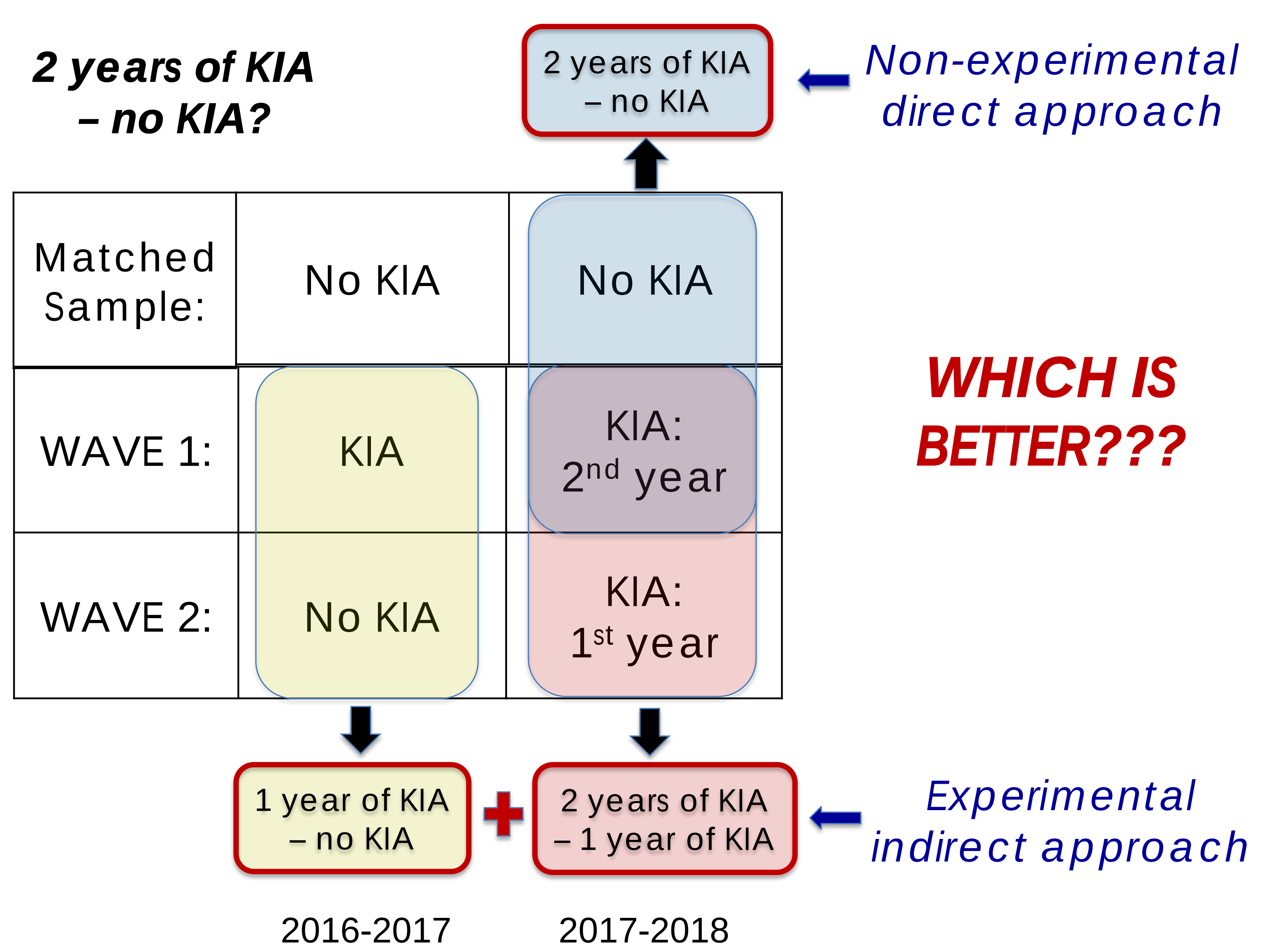


Knowledge in Action

- Part 1 (“Efficacy Study”): randomize schools to KIA or control; compare outcomes after 1 year
- Part 2 (“Maturation Study”): continue to follow schools another year (experimental & observational)



Covariate data for schools not in RCT
MATCHING
Matched Sample:
No KIA
No KIA
2 years of KIA – no KIA
Covariate data for schools in RCT
RANDOMIZE
WAVE 1:
KIA
KIA: 2nd year
WAVE 2:
No KIA
KIA: 1st year
1 year of KIA – no KIA
2 years of KIA – 1 year of KIA
2 years of KIA – no KIA?
2015-2016
2016-2017
2017-2018


2 years of KIA
– no KIA?
2 years of KIA
– no KIA
Non-experimental
direct approach
Matched Sample:
No KIA
No KIA
WAVE 1:
KIA
KIA:
2nd year
WHICH IS
BETTER???
WAVE 2:
No KIA
KIA:
1st year
1 year of KIA
– no KIA
2 years of KIA
– 1 year of KIA
Experimental
indirect approach
2016-2017
2017-2018

Potential Outcomes & Estimands

- $\bar{Y}_j(W_j, t)$= potential outcome for school *j* under treatment W_j in year *t*
- ***Causal effect: compare potential outcomes under different treatments***

$$\tau_{1,t} \equiv \overline{Y(1,t)} - \overline{Y(0,t)} = \frac{\sum_{j=1}^{n} \bar{Y}_j(1,t)}{n} - \frac{\sum_{j=1}^{n} \bar{Y}_j(0,t)}{n}$$

$$\tau_{2-1,t} \equiv \overline{Y(2,t)} - \overline{Y(1,t)} = \frac{\sum_{j=1}^{n} \bar{Y}_j(2,t)}{n} - \frac{\sum_{j=1}^{n} \bar{Y}_j(1,t)}{n}$$

$$\tau_{2,t} \equiv \overline{Y(2,t)} - \overline{Y(0,t)} = \frac{\sum_{j=1}^{n} \bar{Y}_j(2,t)}{n} - \frac{\sum_{j=1}^{n} \bar{Y}_j(1,t)}{n}$$

Estimators

$$\hat{\tau}_{1,2017} \equiv \frac{\sum_{j=1}^{n} W_j \bar{Y}_j(1,2017)}{\sum_{j=1}^{n} W_j} - \frac{\sum_{j=1}^{n}(1-W_j)\bar{Y}_j(0,2017)}{\sum_{j=1}^{n}(1-W_j)}$$

$$\hat{\tau}_{2-1,2018} \equiv \frac{\sum_{j=1}^{n} I_{W_j=2}\bar{Y}_j(2,2018)}{\sum_{j=1}^{n} I_{W_j=2}} - \frac{\sum_{j=1}^{n} I_{W_j=1}\bar{Y}_j(1,2018)}{\sum_{j=1}^{n} I_{W_j=1}}$$

$$\hat{\tau}_{2,2018} \equiv \frac{\sum_{j=1}^{n} I_{W_j=2}\bar{Y}_j(2,2018)}{\sum_{j=1}^{n} I_{W_j=2}} - \frac{\sum_{j=1}^{n} I_{W_j=1}\bar{Y}_j(0,2018)}{\sum_{j=1}^{n} I_{W_j=1}}$$

2 years of KIA – no KIA?

2 years of KIA – no KIA

Non-experimental direct approach

Matched Sample:	No KIA	No KIA
WAVE 1:	KIA	KIA: 2^{nd} year
WAVE 2:	No KIA	KIA: 1^{st} year
	2016-2017	2017-2018

WHICH IS BETTER???

1 year of KIA – no KIA + 2 years of KIA – 1 year of KIA

Experimental indirect approach

Propensity Score Matching

- $W_j = \begin{cases} 1 \text{ if in Wave 1 of experiment} \\ 0 \text{ if not in experiment} \end{cases}$
- Propensity score: $e_j = P\left(W_j = 1 \,\middle|\, \boldsymbol{x}_j\right)$
- Match each Wave 1 teacher with a control with a similar propensity score

Criteria for success:

- Quality of observed covariate data
  - can only balance *observed* data
- Good matches available
  - adequate overlap between groups
  - large enough pool of potential controls

Propensity Score Matching

- If we have good matches, we can balance observed covariates
- <u>Key point</u>: unless we have data on all relevant covariates (which we won't), there will still be ***bias*** (baseline differences)
- Usually hard to quantify this bias…
- … BUT we have a very rare feature!!

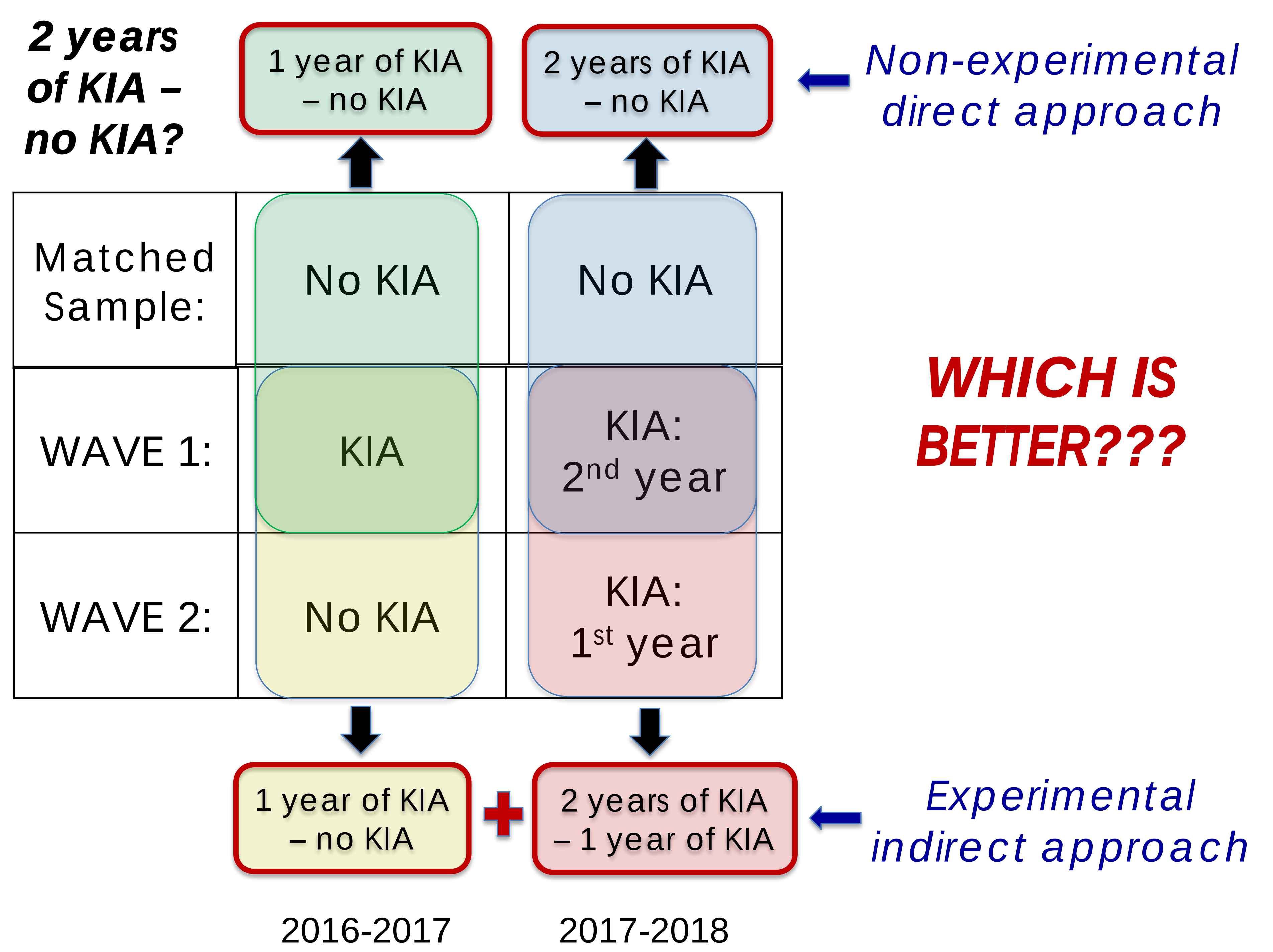
1 year of KIA – no KIA

2 years of KIA – no KIA

	2016-2017	2017-2018
Matched Sample:	No KIA	No KIA
WAVE 1:	KIA	KIA: 2nd year
WAVE 2:	No KIA	KIA: 1st year

1 year of KIA – no KIA

We can validate the non-experimental approach by comparing 1 year impact estimates!


2 years of KIA – no KIA?
1 year of KIA – no KIA
2 years of KIA – no KIA
Non-experimental direct approach
Matched Sample:
No KIA
No KIA
WAVE 1:
KIA
KIA: 2nd year
WAVE 2:
No KIA
KIA: 1st year
WHICH IS BETTER???
1 year of KIA – no KIA
2 years of KIA – 1 year of KIA
Experimental indirect approach
2016-2017
2017-2018

Experimental Indirect Approach

$$\tau_{2-1,2018} + \tau_{1,2017} = \overline{Y(2,2018)} - \overline{Y(1,2018)} + \overline{Y(1,2017)} - \overline{Y(0,2017)}$$

- Critical assumption: potential outcomes may depend on year, but *treatment effects* do not
- That is, $\overline{Y(1,2017)} \neq \overline{Y(1,2018)}$, but

$$\tau_{1,2017} = \tau_{1,2018} \equiv \tau_1$$

- This implies $\tau_1 + \tau_{2-1} = \tau_2$

Unbiased

Define $\hat{\tau}_2 \equiv \hat{\tau}_1 + \hat{\tau}_{2-1}$

Theorem: Assuming treatment effects do not vary by year, $\mathrm{E}(\hat{\tau}_2) = \tau_2$.

Proof: $\mathrm{E}(\hat{\tau}_2) = \mathrm{E}(\hat{\tau}_1 + \hat{\tau}_{2-1}) = \tau_1 + \tau_{2-1} = \tau_2$.

Variance

$$var(\hat{\tau}_2) = var(\hat{\tau}_1 + \hat{\tau}_{2-1})$$

$$= var(\hat{\tau}_1) + var(\hat{\tau}_{2-1}) + 2\mathrm{cov}(\hat{\tau}_1, \hat{\tau}_{2-1})$$

- Both estimates are comparisons of the *same* teachers; likely to be highly positively correlated
- More than double the variance of each individual estimate

Constant Treatment Effect?

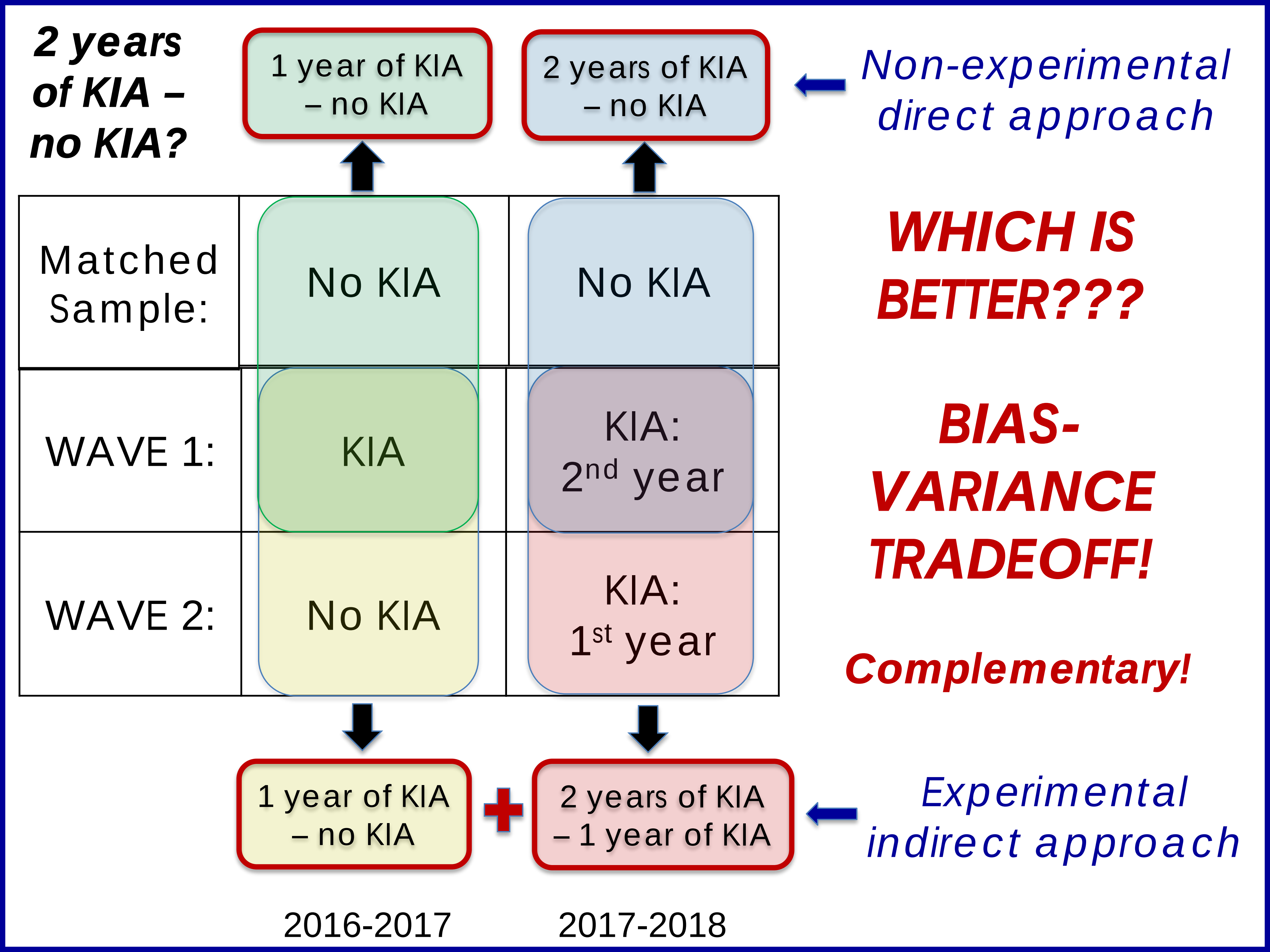
- Suppose constant treatment effect, so $\bar{Y}_j(1,t) = \bar{Y}_j(0,t) + \tau_1$ and $\bar{Y}_j(2,t) = \bar{Y}_j(1,t) + \tau_{2-1}$ $\forall j$.

- Then:
  - o $\hat{\tau}_1 = \tau_1 + \left[\overline{Y_{Wave1}(0,2017)} - \overline{Y_{Wave2}(0,2017)}\right]$
  - o $\hat{\tau}_{2-1} = \tau_{2-1} + \left[\overline{Y_{Wave1}(0,2018)} - \overline{Y_{Wave2}(0,2018)}\right]$

- Under additivity, and if we again assume differences in time cancel with comparisons within the same year, then $\hat{\tau}_1$ and $\hat{\tau}_{2-1}$ are perfectly correlated!

- $var(\hat{\tau}_2) = var(\hat{\tau}_1)+var(\hat{\tau}_{2-1}) + 2\sqrt{var(\hat{\tau}_1)var(\hat{\tau}_{2-1})}$

- If $var(\hat{\tau}_1) \approx var(\hat{\tau}_{2-1})$, then $var(\hat{\tau}_2) \approx 4var(\hat{\tau}_1)$



2 years of KIA – no KIA?
1 year of KIA – no KIA
2 years of KIA – no KIA
Non-experimental direct approach
Matched Sample:
No KIA
No KIA
WHICH IS BETTER???
WAVE 1:
KIA
KIA: 2nd year
BIAS-VARIANCE TRADEOFF!
WAVE 2:
No KIA
KIA: 1st year
Complementary!
1 year of KIA – no KIA
2 years of KIA – 1 year of KIA
Experimental indirect approach
2016-2017
2017-2018

Other Interesting Tidbits

- Student-level versus school level analysis
- Combined analyses?
- Student/parental consent => missing data
- “Joiners”
- Non-compliance
- Teachers switching schools/courses
- Anticipation bias
- … and more!

Conclusion

- Rerandomization can improve experimental design
- Propensity score matching can improve observational studies
- Bias-variance tradeoff for 2 year impact
- Lots of fun statistics in rich applied problems!

klm47@psu.edu

- Funded by George Lucas Educational Foundation
- Joint work with Anna Saavedra, Amie Rappaport, Ying Liu, and Juan Saavedra

Weighting

- Option 1: Weight schools equally

$$\hat{\tau}_1 = \frac{\sum_{j=1}^{n} W_j \bar{Y}_j(1)}{\sum_{j=1}^{n} W_j} - \frac{\sum_{j=1}^{n} (1 - W_j) \bar{Y}_j(0)}{\sum_{j=1}^{n} (1 - W_j)}$$

Differing number of students (3-127)

- Option 2: Weight students equally

$$\hat{\tau}_1 = \frac{\sum_{j=1}^{n} W_j \bar{Y}_j(1)\, n_j}{\sum_{j=1}^{n} W_j\, n_j} - \frac{\sum_{j=1}^{n} (1 - W_j) \bar{Y}_j(0) n_j}{\sum_{j=1}^{n} (1 - W_j) n_j}$$

τ may vary with class size

$$= \frac{\sum_{j=1}^{n} W_j \sum_{i=1}^{n_j} Y_i(1)}{\sum_{j=1}^{n} W_j n_j} - \frac{\sum_{j=1}^{n} (1 - W_j) \sum_{i=1}^{n_j} Y_i(0)}{\sum_{j=1}^{n} (1 - W_j) n_j}$$

Multilevel Model

Student-level: $Y_i(W_{j[i]}) \sim N(\mu_j(W_j) + \boldsymbol{\beta}_1 \boldsymbol{x}_1, \sigma_Y^2)$

School-level: $\mu_j(W_j) \sim N(\alpha_k + \tau W_j + \boldsymbol{\beta}_2 \boldsymbol{x}_2, \sigma_\mu^2)$

District-level: $\alpha_k \sim N(\alpha + \boldsymbol{\beta}_3 \boldsymbol{x}_3, \sigma_\alpha^2)$

- Smaller schools “shrink” more; in between the two weighting extremes