Discussion: Experimental Methods

Yuki Atsusaka
atsusaka.org

Hobby School of Public Affairs
University of Houston

TexMeth 2025
University of Houston, TX
2/22/2025

Diaz & Rossiter

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)
- Motivation

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)
- Motivation
  - Section 4 (literature review) can be the intro?

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)
- Motivation
  - Section 4 (literature review) can be the intro?
  - Any estimates of sample loss in existing studies ($<$ 1%, p.19!?)?

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)
- Motivation
  - Section 4 (literature review) can be the intro?
  - Any estimates of sample loss in existing studies ($< 1\%$, p.19!?)?
  - Maybe some of the 184 studies did not block/repeat due to the sample loss concern?

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)
- Motivation
  - Section 4 (literature review) can be the intro?
  - Any estimates of sample loss in existing studies ($<$ 1%, p.19!?)?
  - Maybe some of the 184 studies did not block/repeat due to the sample loss concern?
    - $\rightarrow$ Could you show the prevalence of this practical trade-off?

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)
- Motivation
  - Section 4 (literature review) can be the intro?
  - Any estimates of sample loss in existing studies ($<$ 1%, p.19!?)?
  - Maybe some of the 184 studies did not block/repeat due to the sample loss concern?
    - $\rightarrow$ Could you show the prevalence of this practical trade-off?
- Methods and analysis

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)
- Motivation
  - Section 4 (literature review) can be the intro?
  - Any estimates of sample loss in existing studies ($<$ 1%, p.19!?)?
  - Maybe some of the 184 studies did not block/repeat due to the sample loss concern?
    - $\rightarrow$ Could you show the prevalence of this practical trade-off?
- Methods and analysis
  - How *ignorable* is (esp. explicit) sample loss? (cf. Section 5.2)

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)
- Motivation
  - Section 4 (literature review) can be the intro?
  - Any estimates of sample loss in existing studies ($<$ 1%, p.19!?)?
  - Maybe some of the 184 studies did not block/repeat due to the sample loss concern?
    - $\rightarrow$ Could you show the prevalence of this practical trade-off?
- Methods and analysis
  - How *ignorable* is (esp. explicit) sample loss? (cf. Section 5.2)
  - Formal analysis of *selection* may be informative in Section 3

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)
- Motivation
  - Section 4 (literature review) can be the intro?
  - Any estimates of sample loss in existing studies ($<$ 1%, p.19!?)?
  - Maybe some of the 184 studies did not block/repeat due to the sample loss concern?
    - $\rightarrow$ Could you show the prevalence of this practical trade-off?
- Methods and analysis
  - How *ignorable* is (esp. explicit) sample loss? (cf. Section 5.2)
  - Formal analysis of *selection* may be informative in Section 3
  - Any guideline on # treated and control?

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)
- Motivation
  - Section 4 (literature review) can be the intro?
  - Any estimates of sample loss in existing studies ($<$ 1%, p.19!?)?
  - Maybe some of the 184 studies did not block/repeat due to the sample loss concern?
    - $\rightarrow$ Could you show the prevalence of this practical trade-off?
- Methods and analysis
  - How *ignorable* is (esp. explicit) sample loss? (cf. Section 5.2)
  - Formal analysis of *selection* may be informative in Section 3
  - Any guideline on # treated and control?
  - How much do **substantive conclusions** change across designs and degress of sample loss?

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)
- Motivation
  - Section 4 (literature review) can be the intro?
  - Any estimates of sample loss in existing studies ($<$ 1%, p.19!?)?
  - Maybe some of the 184 studies did not block/repeat due to the sample loss concern?
    - $\rightarrow$ Could you show the prevalence of this practical trade-off?
- Methods and analysis
  - How *ignorable* is (esp. explicit) sample loss? (cf. Section 5.2)
  - Formal analysis of *selection* may be informative in Section 3
  - Any guideline on # treated and control?
  - How much do **substantive conclusions** change across designs and degress of sample loss?
    - Does "significance" change?

Diaz & Rossiter

- Empirical analysis of the impact of sample loss on precision in block randomization and repeated designs
  - Important & extensive work connecting practice and statistics
  - Develop a *guideline* for researchers & R package (must be in the intro?)
- Motivation
  - Section 4 (literature review) can be the intro?
  - Any estimates of sample loss in existing studies ($<$ 1%, p.19!?)?
  - Maybe some of the 184 studies did not block/repeat due to the sample loss concern?
    - $\rightarrow$ Could you show the prevalence of this practical trade-off?
- Methods and analysis
  - How *ignorable* is (esp. explicit) sample loss? (cf. Section 5.2)
  - Formal analysis of *selection* may be informative in Section 3
  - Any guideline on # treated and control?
  - How much do **substantive conclusions** change across designs and degress of sample loss?
    - Does "significance" change?
    - Do confidence intervals overlap across designs?

Jordan, Ollerenshaw, & Trexler

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?
    - → What do existing **applications** do wrong?

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?
    - → What do existing **applications** do wrong?
  - One more contribution (beyond "just" replication)?

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?
    - → What do existing **applications** do wrong?
  - One more contribution (beyond "just" replication)?
    - Sensitivity analysis/practical tests for repeated measure designs

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?
    - → What do existing **applications** do wrong?
  - One more contribution (beyond "just" replication)?
    - Sensitivity analysis/practical tests for repeated measure designs
    - Formal analysis of precision

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?
    - → What do existing **applications** do wrong?
  - One more contribution (beyond "just" replication)?
    - Sensitivity analysis/practical tests for repeated measure designs
    - Formal analysis of precision
- Methods and analysis

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?
    - → What do existing **applications** do wrong?
  - One more contribution (beyond "just" replication)?
    - Sensitivity analysis/practical tests for repeated measure designs
    - Formal analysis of precision
- Methods and analysis
  - A bit more explanation would be helpful for

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?
    - → What do existing **applications** do wrong?
  - One more contribution (beyond "just" replication)?
    - Sensitivity analysis/practical tests for repeated measure designs
    - Formal analysis of precision
- Methods and analysis
  - A bit more explanation would be helpful for
    - Repeated measure designs (how do they improve precision?)

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?
    - → What do existing **applications** do wrong?
  - One more contribution (beyond "just" replication)?
    - Sensitivity analysis/practical tests for repeated measure designs
    - Formal analysis of precision
- Methods and analysis
  - A bit more explanation would be helpful for
    - Repeated measure designs (how do they improve precision?)
    - Seemingly unrelated regression/random effects (why?)

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?
    - → What do existing **applications** do wrong?
  - One more contribution (beyond "just" replication)?
    - Sensitivity analysis/practical tests for repeated measure designs
    - Formal analysis of precision
- Methods and analysis
  - A bit more explanation would be helpful for
    - Repeated measure designs (how do they improve precision?)
    - Seemingly unrelated regression/random effects (why?)
  - What are hidden assumptions?

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?
    - → What do existing **applications** do wrong?
  - One more contribution (beyond "just" replication)?
    - Sensitivity analysis/practical tests for repeated measure designs
    - Formal analysis of precision
- Methods and analysis
  - A bit more explanation would be helpful for
    - Repeated measure designs (how do they improve precision?)
    - Seemingly unrelated regression/random effects (why?)
  - What are hidden assumptions?
    - p(random) is the same across repeated and post-designs

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?
    - → What do existing **applications** do wrong?
  - One more contribution (beyond "just" replication)?
    - Sensitivity analysis/practical tests for repeated measure designs
    - Formal analysis of precision
- Methods and analysis
  - A bit more explanation would be helpful for
    - Repeated measure designs (how do they improve precision?)
    - Seemingly unrelated regression/random effects (why?)
  - What are hidden assumptions?
    - p(random) is the same across repeated and post-designs
    - Post-design provides the ground truth (unbiased estimated of the parameter)

Jordan, Ollerenshaw, & Trexler

- Empirical validation of *repeated measure designs* for improving precision in treatment effect estimation
  - Great paper, relevant Q, extensive (seriously) large-scale replication
  - Evidence that repeated designs work well in survey experiments
- Motivation
  - Many caveats and unanswered questions in CSP (2021)
    - → Let existing studies highlight these **theoretical** limitations?
    - → What do existing **applications** do wrong?
  - One more contribution (beyond "just" replication)?
    - Sensitivity analysis/practical tests for repeated measure designs
    - Formal analysis of precision
- Methods and analysis
  - A bit more explanation would be helpful for
    - Repeated measure designs (how do they improve precision?)
    - Seemingly unrelated regression/random effects (why?)
  - What are hidden assumptions?
    - p(random) is the same across repeated and post-designs
    - Post-design provides the ground truth (unbiased estimated of the parameter) → Is "attenuation bias" the right expression here?

Jordan, Ollerenshaw, & Trexler (Minor Commnets)

- Visualization of the design may be helpful (p. 11)
- Notation for N_i is not standard? What's i?
- Is the main concern about external validity of CSP?
  - "CSP's definitive conclusion rests primarily on just six experiments" (p. 2)
- What's statistical noise? (p. 4)
  - Recent studies (such as mine and Matt Tyler's) show that these are not really "noise" (that only inflates SE while not affecting point estimate)
- Minor notes
  - P. 5 says researchers have "historically avoided" the design, but it's very new?
  - P. 7 notes "Further, close proximity may even be advantageous if it reduces random noise and strengthens the correlation between pre- and post-treatment measures, which increases precision." → This may require more explanations as it seems very important
  - Any reasons why repeated designs may fail for probability-based samples?
  - Is consistency in response bad or good? (not very clear)
- P. 32 really changes the story (**people don't really care about the exact value of ATE**)? What is your quantity of interest? p-value? Effect direction?

Overall Theme and Questions

Precision in ATE Estimation

Overall Theme and Questions

Precision in ATE Estimation

- **What's precision** as a statistical terminology? (both papers do not define it?)

Overall Theme and Questions

Precision in ATE Estimation

- **What's precision** as a statistical terminology? (both papers do not define it?)
- Why do we need precision?

Overall Theme and Questions

Precision in ATE Estimation

- **What's precision** as a statistical terminology? (both papers do not define it?)
- Why do we need precision?
  - Accurate inference on the true/exact value of ATE?

Overall Theme and Questions

Precision in ATE Estimation

- **What's precision** as a statistical terminology? (both papers do not define it?)
- Why do we need precision?
  - Accurate inference on the true/exact value of ATE?
  - Hypothesis testing (with null = 0)?

Overall Theme and Questions

Precision in ATE Estimation

- **What's precision** as a statistical terminology? (both papers do not define it?)
- Why do we need precision?
  - Accurate inference on the true/exact value of ATE?
  - Hypothesis testing (with null = 0)?
  - Directional hypothesis (with null ≤ 0)?

Overall Theme and Questions

Precision in ATE Estimation

- **What's precision** as a statistical terminology? (both papers do not define it?)
- Why do we need precision?
  - Accurate inference on the true/exact value of ATE?
  - Hypothesis testing (with null = 0)?
  - Directional hypothesis (with null ≤ 0)?
- Precision is something related to **statistical power**:

Overall Theme and Questions

Precision in ATE Estimation

- **What's precision** as a statistical terminology? (both papers do not define it?)
- Why do we need precision?
  - Accurate inference on the true/exact value of ATE?
  - Hypothesis testing (with null = 0)?
  - Directional hypothesis (with null ≤ 0)?
- Precision is something related to **statistical power**: the chance of rejecting the null for a given treatment effect τ, as a function of

Overall Theme and Questions

Precision in ATE Estimation

- **What's precision** as a statistical terminology? (both papers do not define it?)
- Why do we need precision?
  - Accurate inference on the true/exact value of ATE?
  - Hypothesis testing (with null = 0)?
  - Directional hypothesis (with null ≤ 0)?
- Precision is something related to **statistical power**: the chance of rejecting the null for a given treatment effect τ, as a function of
  - Null hypothesis (zero, negative, etc)

Overall Theme and Questions

Precision in ATE Estimation

- **What's precision** as a statistical terminology? (both papers do not define it?)
- Why do we need precision?
  - Accurate inference on the true/exact value of ATE?
  - Hypothesis testing (with null = 0)?
  - Directional hypothesis (with null ≤ 0)?
- Precision is something related to **statistical power**: the chance of rejecting the null for a given treatment effect τ, as a function of
  - Null hypothesis (zero, negative, etc)
  - Expected treatment effect (effect size)

Overall Theme and Questions

Precision in ATE Estimation

- **What's precision** as a statistical terminology? (both papers do not define it?)
- Why do we need precision?
  - Accurate inference on the true/exact value of ATE?
  - Hypothesis testing (with null = 0)?
  - Directional hypothesis (with null ≤ 0)?
- Precision is something related to **statistical power**: the chance of rejecting the null for a given treatment effect τ, as a function of
  - Null hypothesis (zero, negative, etc)
  - Expected treatment effect (effect size)
  - Standard error of the estimated effect (sample size, effect heterogeneity, other noise)

Overall Theme and Questions

Precision in ATE Estimation

- **What's precision** as a statistical terminology? (both papers do not define it?)
- Why do we need precision?
  - Accurate inference on the true/exact value of ATE?
  - Hypothesis testing (with null = 0)?
  - Directional hypothesis (with null ≤ 0)?
- Precision is something related to **statistical power**: the chance of rejecting the null for a given treatment effect τ, as a function of
  - Null hypothesis (zero, negative, etc)
  - Expected treatment effect (effect size)
  - Standard error of the estimated effect (sample size, effect heterogeneity, other noise)

  $\rightarrow$ What roles do block randomization and repeated designs (and sample loss) play in this big picture?