Determining the Timing of Phase Changes

Some Statistical Perspective

James E. Pustejovsky

University of Wisconsin - Madison

May 18, 2023

A birthday party analogy

Randomized phase changes =
peanut butter chocolate cupcake

Response-guided phase changes =
carrot cake cupcake

Fixed phase changes =
yellow cake cupcake

Design Criteria

How do these approaches compare in terms of

- Theoretical support (statistical theory)
- Procedural reproducibility
- Feasibility & pragmatic considerations

> The best rule of thumb for evaluating a procedure description as technological is probably to ask whether a typically trained reader could replicate that procedure well enough to produce the same results, given only a reading of the description
>
> — Baer, Wolf, and Risley (1968)

Randomized phase changes

Randomized between-group designs: Theory

- Clear target of estimation and inference: the **Sample Average Treatment Effect**:

$$SATE = \left(\begin{array}{c}\text{Average outcome if}\\ \text{everybody gets B}\end{array}\right) - \left(\begin{array}{c}\text{Average outcome if}\\ \text{everybody gets A}\end{array}\right)$$

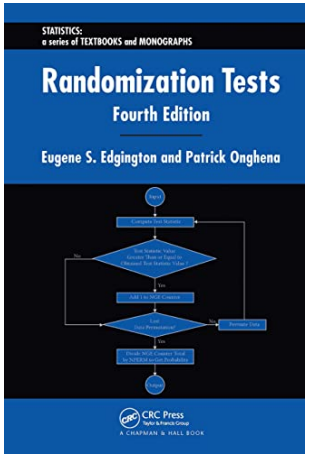
- Deep formal statistical theory based on randomization

  😎 Hypothesis testing

  😎 Effect estimation

- Extensive methodology for handling common problems
  - Interference between units
  - Attrition
  - Non-compliance

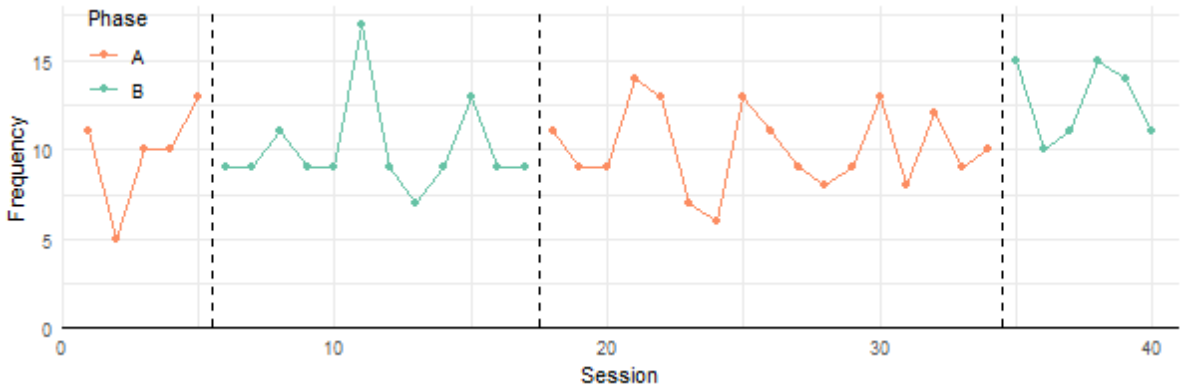
Randomized single-case designs: Theory



- Formal randomization-based theory for hypothesis testing (Edgington and Onghena, 2007)
- But constrained to tests of **sharp null hypothesis**:

H_0 : Intervention has no effect on outcomes whatsoever

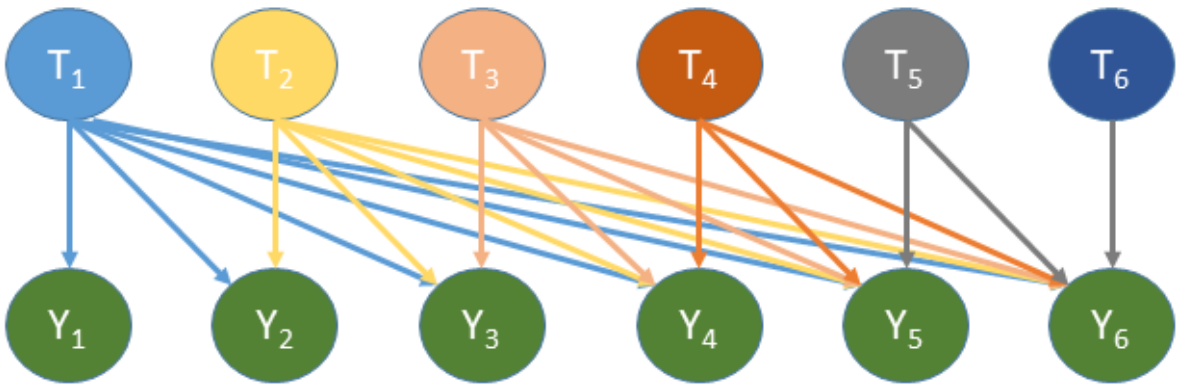
H_A : Intervention alters at least one outcome at some time-point



Randomized single-case designs: Theory

🤔 Effect estimation

- Confidence intervals based on randomization distribution (Michiels, Heyvaert, Meulders, and Onghena, 2017)
- Estimation based on randomization distribution requires a ***completely specified response function***
  - How do *all past treatment assignments* affect current outcome?



Randomized phase changes

- 😎 Hypothesis testing
- 🤔 Effect estimation
- 🤔 Reproducibility
  - Reproducible in theory
  - R package SCRT, web app SCDA
  - How well do researchers implement randomized designs?
- 🤷 Pragmatic considerations
  - Sometimes fairly feasible, sometimes not

Response-guided phase changes

Response-guided designs: Theory

😟 Response-guided design can distort conventional hypothesis tests (Allison, Franklin, and Heshka, 1992; Ferron, Foster-Johnson, and Kromrey, 2003)

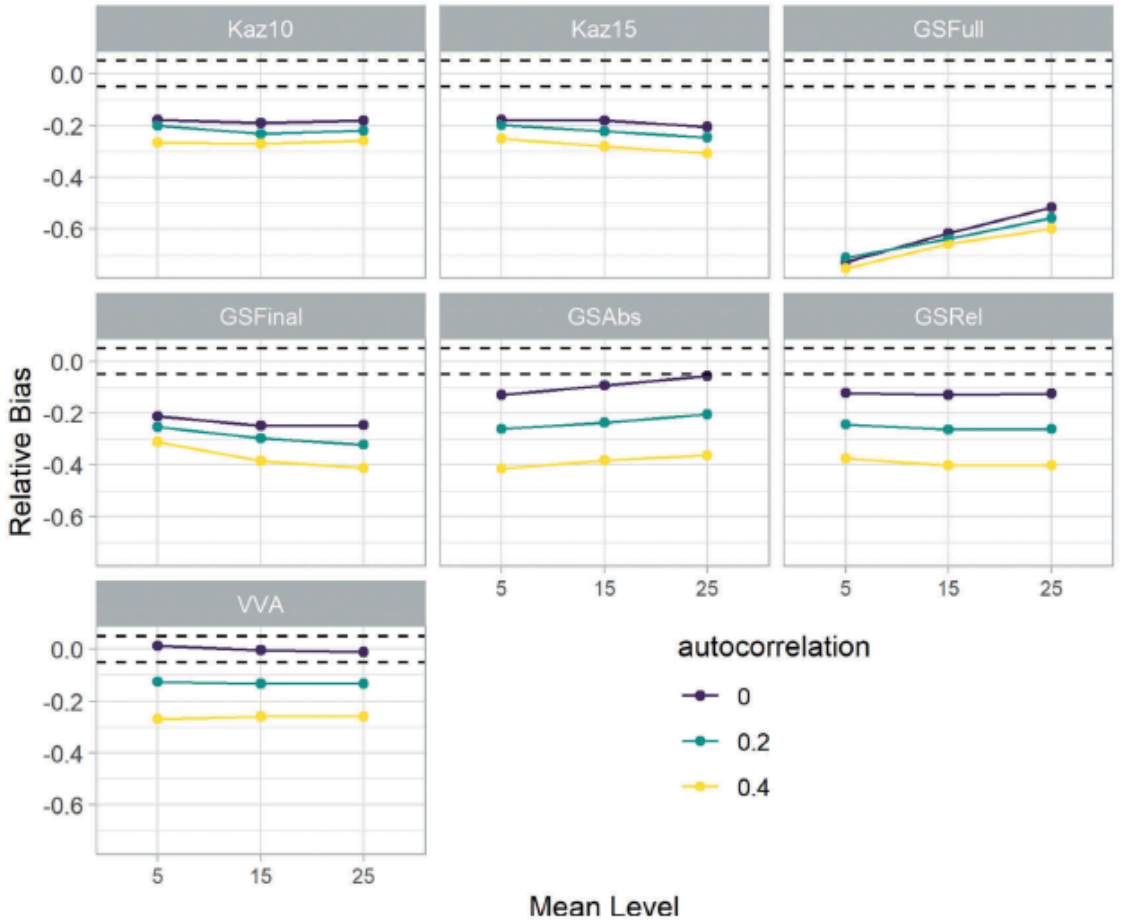
- 🤓 Masked Visual Analysis provides a theory for hypothesis testing (Ferron and Jones, 2006)

😟 Estimation

- Joo, Ferron, Beretvas, Moeyaert, and Van den Noortgate (2018) found little impact of response-guided decisions on estimates from multi-level models
- Swan, Pustejovsky, and Beretvas (2020) found that response-guided decisions lead to under-estimation of baseline variance

Swan, Pustejovsky, and Beretvas (2020)

Relative bias of baseline variance for stable, Poisson-distributed baselines



Response-guided phase changes

- 😟 / 🤓 Hypothesis testing
- 😟 Effect estimation
- 😨 Reproducibility
  - Various operationally defined criteria have been described (Gast and Spriggs, 2010; Kazdin, 2011; Ferron, Joo, and Levin, 2017)
  - But not clear what researchers actually do in practice
- 😁 Feasibility and pragmatic considerations
  - Although stringent criteria might create scheduling challenges

Fixed phase changes

Fixed-phase length designs: Theory

🤷 No particular supporting statistical theory

- Conventional (parametric) statistical inference / estimation approaches **assume that the phase lengths are fixed (constant)**.
    - Standard error of a sample mean is holding n fixed:

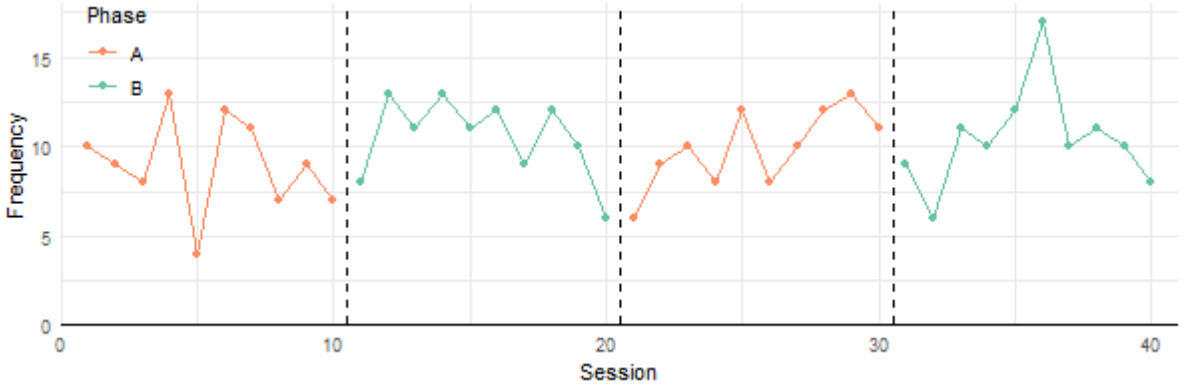
$$SE(\bar{y}) = \frac{S}{\sqrt{n}}$$

    - Standard error of regression coefficient is estimated while holding the predictors fixed:

$$SE\left(\hat{\beta}\right) = \sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta} \mid \mathbf{X}\right)}$$

Fixed-phase length designs: Theory

Conventional statistical analysis assumes phases are fixed and imagines that the *outcomes* would change across replications of the study.



- Using fixed phase lengths therefore aligns with conventional statistical analysis.
    - Less likely to mess up a statistical analysis than with response-guided phase changes.
    - Power analysis can be used to determine phase lengths needed to achieve desired precision.

Fixed-phase length designs

- 🤷 Hypothesis testing
- 🤷 Effect estimation

- 😀 Reproducibility
  - Easy to describe procedures

- 😀 Feasibility / pragmatic considerations
  - Phase changes can be planned under scheduling constraints

My ratings

Criterion	Randomized	Response-guided	Fixed phase length
Statistical theory: Hypothesis testing	😎	😟	🤷
Statistical theory: Effect estimation	🤔	😟	🤷
Reproducibility	🤔	😱	😀
Feasibility & pragmatic considerations	🤷	😀	😀

Research priorities

- Procedural reproducibility and transparency are core principles. They need to be prioritized with any approach to phase change decisions.
- Trade-offs between:
  - Using **statistical analysis**, gathering evidence about **magnitude** of effects
  - Using **visual analysis**, making **binary decisions** (presence/absence) about functional relations

References

Allison, D. B., R. D. Franklin, and S. Heshka (1992). "Reflections on Visual Inspection, Response Guided Experimentation, and Type I Error Rate in Single-Case Designs". In: *The Journal of Experimental Education* 61.1, pp. 45-51. DOI: 10.1080/00220973.1992.9943848.

Baer, D. M., M. M. Wolf, and T. R. Risley (1968). "Some Current Dimensions of Applied Behavior Analysis." In: *Journal of applied behavior analysis* 1.1, pp. 91-7.

Edgington, E. and P. Onghena (2007). *Randomization Tests*. Boca Raton, FL: Chapman & Hall.

Ferron, J. M. and P. K. Jones (2006). "Tests for the Visual Analysis of Response-Guided Multiple-Baseline Data". In: *The Journal of Experimental Education* 75.1, pp. 66-81. DOI: 10.3200/JEXE.75.1.66-81.

Ferron, J. M., S. Joo, and J. R. Levin (2017). "A Monte Carlo Evaluation of Masked Visual Analysis in Response-Guided versus Fixed-Criteria Multiple-Baseline Designs". In: *Journal of Applied Behavior Analysis* 50.4, pp. 701-716. DOI: 10.1002/jaba.410.

Ferron, J., L. Foster-Johnson, and J. D. Kromrey (2003). "The Functioning of Single-Case Randomization Tests With and Without Random Assignment". In: *The Journal of Experimental Education* 71.3, pp. 267-288. DOI: 10.1080/00220970309602066.

Gast, D. L. and A. D. Spriggs (2010). "Visual Analysis of Graphic Data". In: *Single Subject Research Methodology in Behavioral Sciences*. Ed. by D. L. Gast. New York, NY: Routledge. Chap. 9, pp. 199-233.

Joo, S., J. M. Ferron, S. N. Beretvas, et al. (2018). "The Impact of Response-Guided Baseline Phase Extensions on Treatment Effect Estimates". In: *Research in Developmental Disabilities* 79, pp. 77-87. ISSN: 08914222. DOI: 10.1016/j.ridd.2017.12.018.

Kazdin, A. E. (2011). *Single-Case Research Designs: Methods for Clinical and Applied Settings*. New York, NY: Oxford University Press.

Michiels, B., M. Heyvaert, A. Meulders, et al. (2017). "Confidence Intervals for Single-Case Effect Size Measures Based on Randomization Test Inversion". In: *Behavior Research Methods* 49.1, pp. 363-381. DOI: 10.3758/s13428-016-0714-4.

References

Swan, D. M., J. E. Pustejovsky, and S. N. Beretvas (2020). "The Impact of Response-Guided Designs on Count Outcomes in Single-Case Experimental Design Baselines". In: *Evidence-Based Communication Assessment and Intervention*, pp. 1-26. DOI: 10.1080/17489539.2020.1739048.

Randomized between-group designs

R.A. Fisher

> The purpose of randomisation ... is to guarantee the validity of the test of significance, this test being based on an estimate of error made possible by replication (Fisher, 1935).
>
> The theory of estimation presupposes a process of random sampling. All our conclusions within that theory rest on this basis; without it our tests of significance would be worthless (Fisher, 1947).

> *Sure, but you don't have to be a such a jerk about it!*
>
> — Neyman (probably at some point?)

Jerzy Neyman

Randomized between-group designs

A. Bradford Hill

> [Randomized allocation to treatment] ensures that neither our personal idiosyncrasies (our likes or dislikes consciously or unwittingly applied) nor our lack of balanced judgement has entered into the construction of the different treatment groups—the allocation has been outside our control and the groups are therefore unbiased (Hill, 1952).