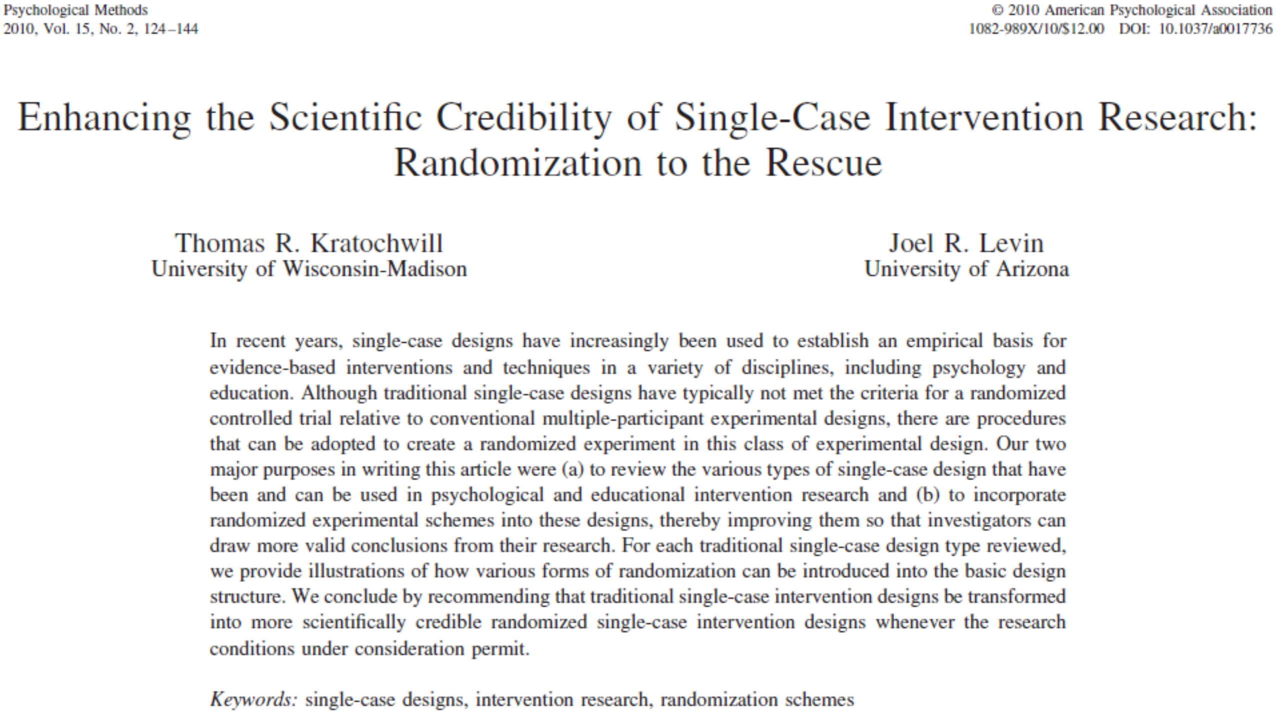
JAMES E. PUSTEJOVSKY
UNIVERSITY OF TEXAS AT AUSTIN

Randomization Inference for Single-Case Experimental Designs

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D160002. The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education.

USING RANDOMIZATION ENHANCES INTERNAL VALIDITY



Psychological Methods
2010, Vol. 15, No. 2, 124–144

© 2010 American Psychological Association
1082-989X/10/$12.00 DOI: 10.1037/a0017736

Enhancing the Scientific Credibility of Single-Case Intervention Research: Randomization to the Rescue

Thomas R. Kratochwill
University of Wisconsin-Madison

Joel R. Levin
University of Arizona

In recent years, single-case designs have increasingly been used to establish an empirical basis for evidence-based interventions and techniques in a variety of disciplines, including psychology and education. Although traditional single-case designs have typically not met the criteria for a randomized controlled trial relative to conventional multiple-participant experimental designs, there are procedures that can be adopted to create a randomized experiment in this class of experimental design. Our two major purposes in writing this article were (a) to review the various types of single-case design that have been and can be used in psychological and educational intervention research and (b) to incorporate randomized experimental schemes into these designs, thereby improving them so that investigators can draw more valid conclusions from their research. For each traditional single-case design type reviewed, we provide illustrations of how various forms of randomization can be introduced into the basic design structure. We conclude by recommending that traditional single-case intervention designs be transformed into more scientifically credible randomized single-case intervention designs whenever the research conditions under consideration permit.

Keywords: single-case designs, intervention research, randomization schemes

RANDOMIZATION IS OF LIMITED UTILITY OR IRRELEVANT

- "Randomization, in some situations, is used in SCR…. In general, however, randomization of when to change conditions and other recommended instances by Kratochwill and Levin are not accepted by most SCR investigators, and many, as do I, find their arguments unconvincing" (Wolery, 2012).

- "Randomization is neither necessary nor sufficient for establishing causal relations in SCDs" (Ledford, 2017).

OVERVIEW

- Logic of randomization and randomization tests
- Examples
- Pros & Cons of randomization tests

RANDOMIZATION IN SINGLE CASE DESIGNS

- SCDs involve
  - One or more cases
  - Repeated measurement of an outcome on each case
  - Assignment of treatment conditions within each case.

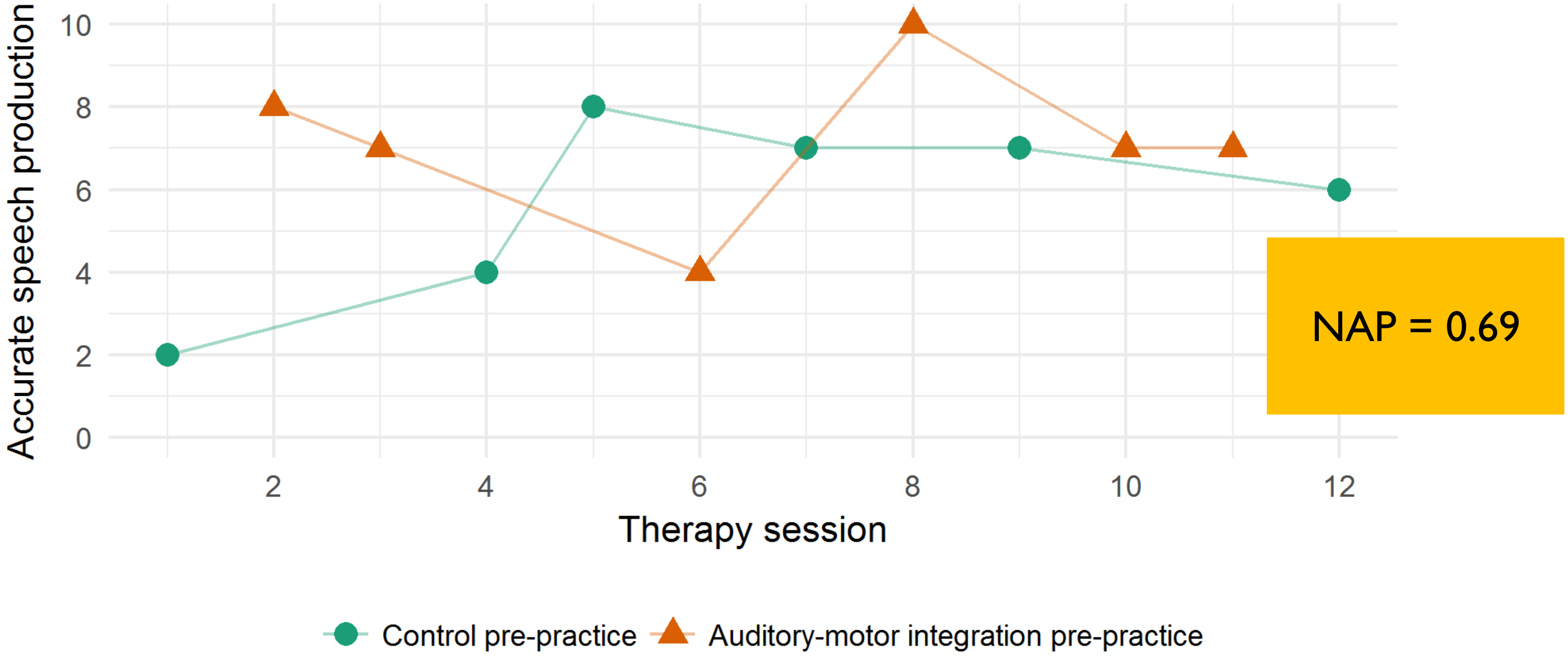
- Randomization of measurement occasions to treatment conditions ***within cases*** (possibly also ***across*** cases).

RANDOMIZATION IN SINGLE CASE DESIGNS

- Alternating treatment designs
  - Randomize conditions across time-points.
  - Blocking or other constraints to prevent repetition of conditions.
- AB designs
  - Randomize timing of phase change
  - Constraints on minimum phase length, baseline stability ("range bound randomization")
- Treatment reversal (ABAB) designs
  - Randomize timing of phase changes
- Multiple baseline designs
  - Randomize timing of phase change for each case/tier.
  - Constraints on minimum phase length, spacing of phase changes.

RVACHEW & MATHEWS (2017)

- 8 year old boy with childhood apraxia of speech.
- Randomized alternating treatment design, blocked by week (2 sessions per week)



NAP = 0.69

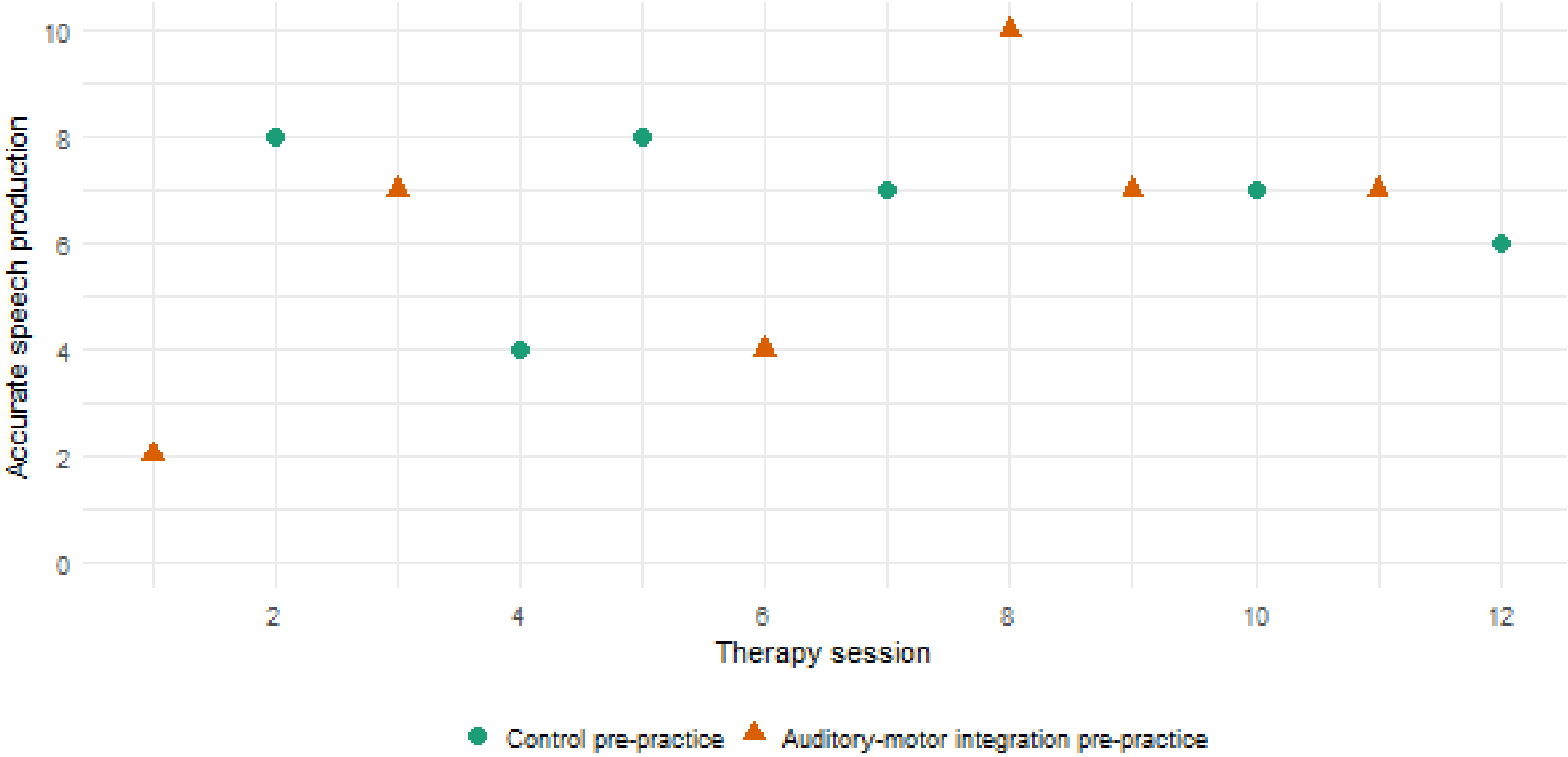
LOGIC OF RANDOMIZATION INFERENCE

H_0: ***Intervention has no effect on outcomes whatsoever (no functional relation)***.

1. Using the collected data, calculate a summary statistic T* describing the functional relation of interest.
2. Compare T* to the distribution of values that ***would have be observed*** if the null hypothesis is true.
3. If T* is very unlikely under the null distribution, reject H_0 and conclude that some functional relation exists.

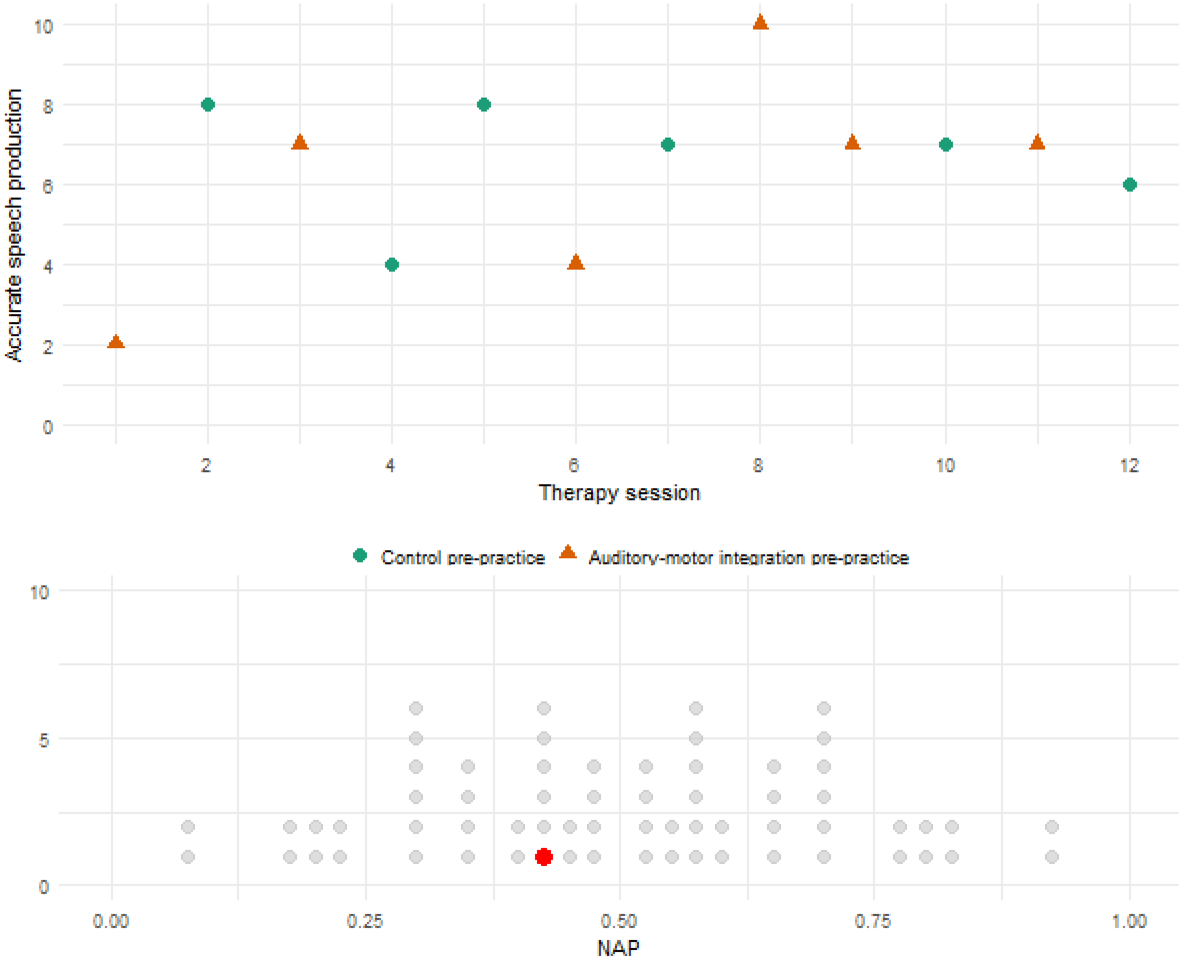
FINDING THE NULL DISTRIBUTION

If H_0 is true, then using a different schedule of treatment sessions would result in an ***identical data series***.

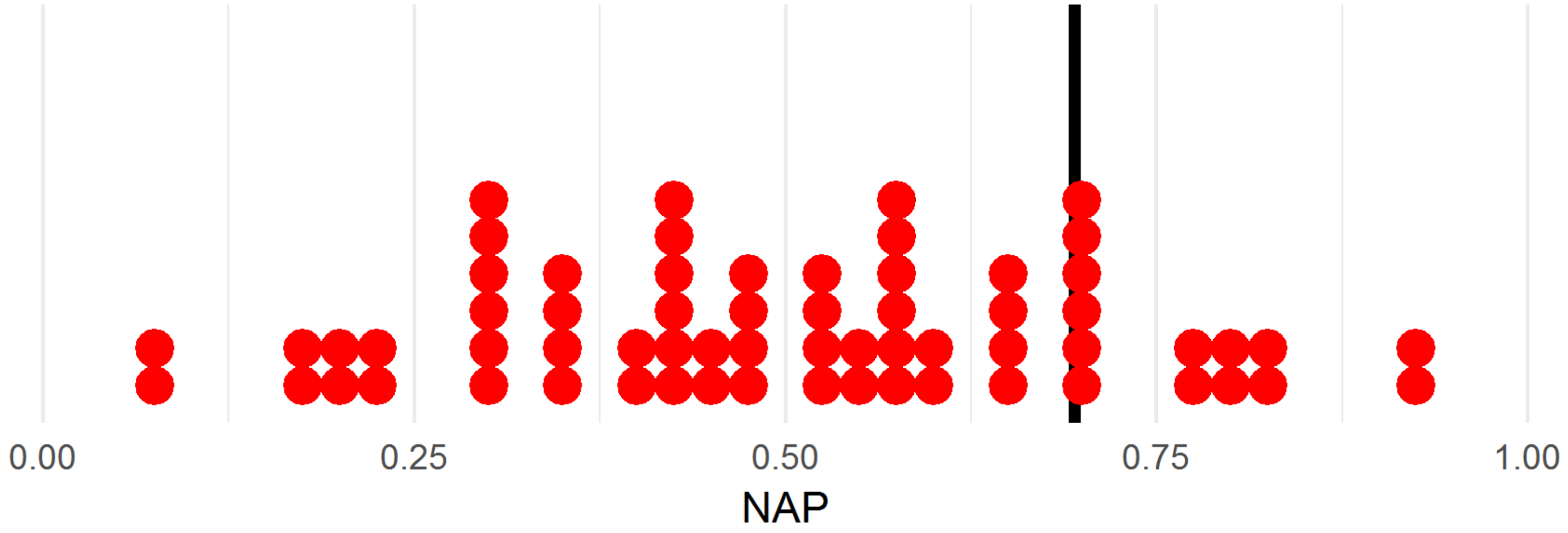


FINDING THE NULL DISTRIBUTION

Calculate NAP for every possible schedule of treatment assignments.



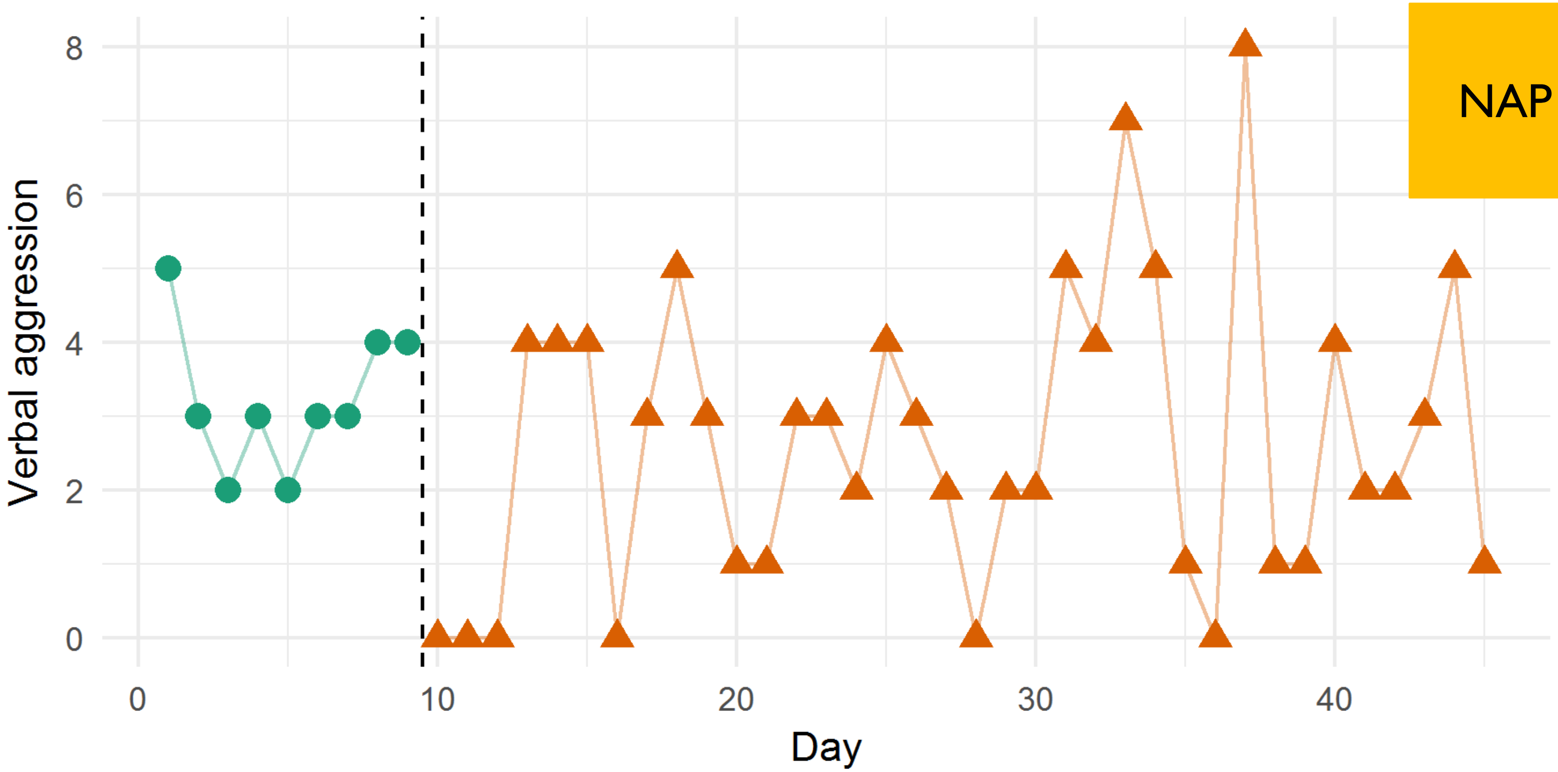
COMPARE OBSERVED T* TO NULL DISTRIBUTION



- Actual data NAP: 0.69
- Proportion of null distribution ≥ actual NAP: 14 / 64 = .22
- Cannot rule out the possibility that treatment has no effect.

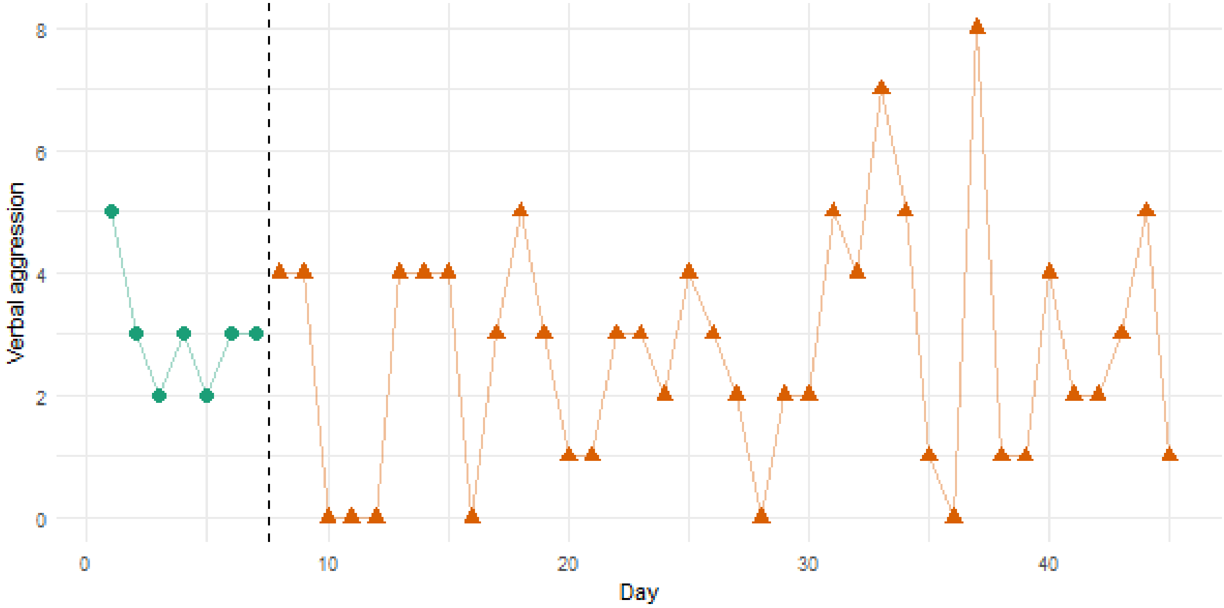
WINKENS ET AL. (2014)

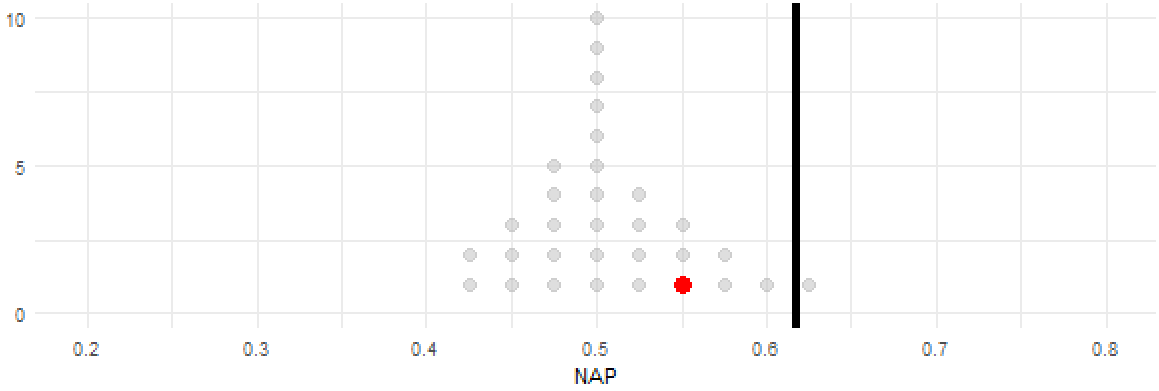
- Evaluated effects of a simplified behavior modification therapy for use by nurses.



NAP = 0.61

WINKENS ET AL. (2014)





- Actual data NAP: 0.61
- Proportion of null distribution ≥ actual NAP: 1 / 31 = .03
- Conclude: treatment has some functional relation with verbal aggression.

SOFTWARE FOR RANDOMIZATION INFERENCE

- Excel Package of Randomization Tests (Levin, Evmenova, & Gafurov, 2014)
  - Freely available at https://ex-prt.weebly.com/
- R package SCRT (Bulté & Onghena, 2008)
  - Freely available on CRAN

E IS

Look, Ma! No modeling assumptions!

- Does NOT require making modeling assumptions about:
  - Baseline time trends
  - Auto-correlation of outcomes
  - Long time series

- ***But it only works if you actually randomize*.**

ADVANTAGE: STUDY DESIGN PROCEDURES ARE PRECISE AND OPERATIONAL

- Set of possible study designs is fully enumerated.
- Replicable on a procedural level.
- Is this true of response-guided experimentation?

LIMITATION: CEDING CONTROL

- Randomization requires more advanced planning than response-guided methods.
- Risk that results will be more ambiguous, less interpretable by visual analysis.

LIMITATION: STRONG NULL HYPOTHESIS

- Randomization tests address the null hypothesis that there is ***no functional relation at all***.
- In designs with multiple cases, no functional relation ***for any case***.

LIMITATION: BEYOND STATISTICAL SIGNIFICANCE

- APA guidelines emphasize reporting of ***effect sizes*** and ***confidence intervals*** (Wilkinson, 1999)
- American Statistical Association (2016):
  - *Statistical significance does not measure the size of an effect or the importance of a result.*
  - *By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.*
  - *Scientific conclusions…should not be based only on whether a p-value passes a specific threshold.*
- McShane, Gal, Gelman, & Tackett (2017): *"Abandon statistical significance."*

LIMITATION: BEYOND STATISTICAL SIGNIFICANCE

More than a p-val

BEYOND STATISTICAL SIGNIFICANCE

- Confidence intervals by randomization test inversion (Michiels et al., 2017).
- Methods require modeling assumptions about:
  - ***exact form*** of functional relation
  - ***effect homogeneity*** across cases.

SUMMARY

- Using randomization and randomization tests provides a way to test for ***presence*** of functional relation.
    - Procedural replicability.
    - Avoids strong statistical modeling assumptions.
- But statistical hypothesis tests are only one part of inference, not sufficient alone.
- More precise description of response-guided design procedures would help bridge gap with randomization approaches.

REFERENCES

Bulté, I., & Onghena, P. (2008). An R package for single-case randomization tests. Behavior Research Methods, 40,467–478, http://dx.doi.org/10.3758/BRM.40.2.467.

Gast, D. L., & Hammond, D. (2010). Withdrawal and reversal designs. In D. L. Gast (Ed.), *Single Subject Research Methodology in Behavioral Sciences*. New York, NY: Routledge.

Gast, D. L., & Ledford, J. R. (2010). Multiple baseline and multiple probe designs. In D. L. Gast (Ed.), *Single Subject Research Methodology in Behavioral Sciences*. New York, NY: Routledge.

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. *Psychological Methods*, *15*(2), 124–44. https://doi.org/10.1037/a0017736

Ledford, J. R. (2017). No randomization? No problem: Experimental control and random assignment in single case research. *American Journal of Evaluation*. https://doi.org/10.1177/1098214017723110

Levin, J. R., Evmenova, A. S., & Gafurov, B. S. (2014). The single-case data-analysis ExPRT (Excel Package of Randomization Tests).

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *The Annals of Applied Statistics*, 7(1), 295–318. http://doi.org/10.1214/12-AOAS583

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2017). Abandon statistical significance. *arXiv preprint arXiv:1709.07588*

Michiels, B., Heyvaert, M., Meulders, A., & Onghena, P. (2017). Confidence intervals for single-case effect size measures based on randomization test inversion. *Behavior Research Methods*, *49*(1), 363–381. http://doi.org/10.3758/s13428-016-0714-4

Rvachew, S., & Matthews, T. (2017). Demonstrating treatment efficacy using the single subject randomization design: A tutorial and demonstration. *Journal of Communication Disorders*, *67*(August 2016), 1–13. http://doi.org/10.1016/j.jcomdis.2017.04.003

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's Statement on *p*-Values: Context, Process, and Purpose. *The American Statistician*, *70*(2), 129–133. http://doi.org/10.1080/00031305.2016.1154108

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–604.

Winkens, I., Ponds, R., Pouwels, C., Eilander, H., & Van Heugten Ae, C. (2014). Using single-case experimental design methodology to evaluate the effects of the ABC method for nursing staff on verbal aggressive behaviour after acquired brain injury. *Neuropsychological Rehabilitation*, *24*, 349–364. http://doi.org/10.1080/09602011.2014.901229

RESPONSE-GUIDED EXPERIMENTATION

- Traditional single case research emphasizes continual monitoring of outcomes, using data to determine when to change treatment conditions.
- For ABAB design:
  - *Introduce the intervention (B_1) only after a stable contratherapeutic or zero-celebration trend has been established in the initial baseline condition (A_1).Withdraw (or reverse) the intervention and return to the baseline condition (A_2) only after acceptable stability in both trend and level has been established in the first intervention condition.* (Gast & Hammond, 2010, p. 250)

RESPONSE-GUIDED EXPERIMENTATION

- For multiple baseline across behaviors:
  - *After performance is stable for all of the behaviors, the intervention is applied to the first behavior….After performance stabilizes across all behaviors, the intervention is applied to the second behavior.* (Kazdin, 2011, p. 145)
  - *Introduce the intervention when the data path of at least one behavior (ideally all behaviors) show acceptable stability in level and trend while maintaining other behaviors in the pre-intervention condition. Introduce the intervention to a new behavior only after criterion-level responding is demonstrated with the preceding behavior.* (Gast & Ledford, 2010, p. 285)