

A Systematic Appraisal of Individual Effect Size Estimates in Aphasia Single-Case Designs via Simulation

^{1,2}Robert Cavanaugh, ³Lauren Terhorst, ^{1,2}Alexander M. Swiderski,
²William D. Hula, ^{1,2}William S. Evans

¹ Communication Sciences & Disorders, University of Pittsburgh

² VA Pittsburgh Healthcare System

³ Occupational Therapy, University of Pittsburgh



University of
Pittsburgh

Introduction

Single-case designs are common in aphasia research (SCD; small-N; single-subject)

Effect sizes are central to evaluating treatment response in SCDs

Testing behavioral & neuro predictors requires accurate and precise effect sizes

A Range of effect sizes have been employed in aphasia single-case design studies

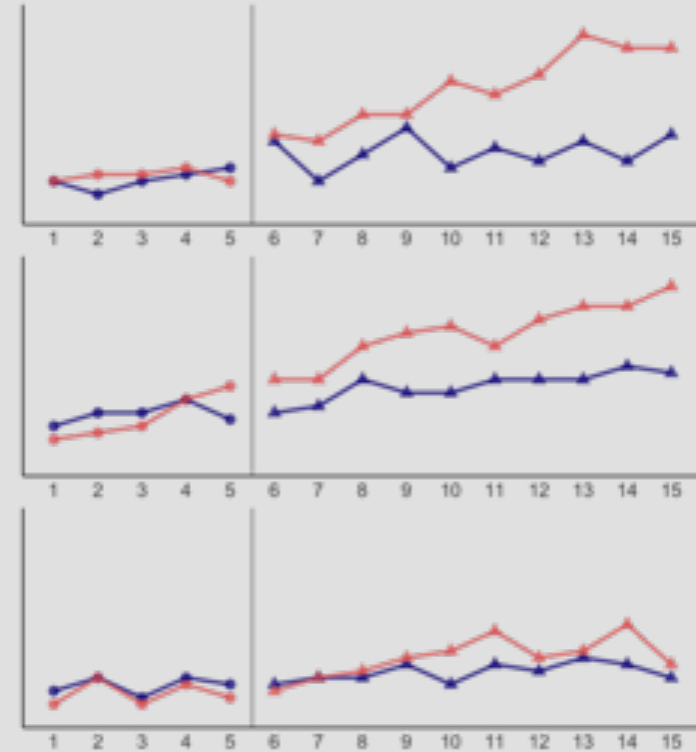
Not all are equivalent nor equally rigorous

Few empirical comparisons & limited guidance for aphasia researchers

Purpose: systematically compare effect sizes used in contemporary aphasia single case design research

Methods

1. Simulate (Manolov & Solonas, 2008)
 - $N = 100$, Multiple-baseline (AB) design
 - 5 baseline and 10 treatment probes
 - 30 treated & 30 untreated words
- intercept: $N(\mu = -1.75, \sigma^2 = .25)$
- participant: $unif(0, 2)$
- item: $N(\mu = 0, \sigma^2 = 0.6)$
- condition: treated = 1; untreated = 0
- logistic link function probability of correct response
- probabilistic simulation: autocorrelated data ($\rho = 0.5$)



$$\begin{aligned} response = & intercept + 0.06 * baselineSlope + 0.3 * levelChange * participant * condition \\ & + 0.15 * slopeChange * participant * condition + itemDifficulty \end{aligned}$$

Visual Inspection; Recover parameters Beeson & Robey quantiles: 3.2 (4.0), 6.8 (7.0), 10.6 (10.1)

Methods

2. Calculate Effect Sizes:*

- Standardized mean difference (**SMD**; Busk & Serlin 1992, Beeson & Robey, 2006)
- Non-overlap of all pairs (**NAP**; Parker & Vannest, 2009)
- **Tau-U** (Parker et al., 2011; Lee & Cherney, 2018)
- Proportion of potential maximal gain (**PMG**; Lambon Ralph et al., 2010)
- Generalized linear mixed-effects models (**GLMM**; Wiley & Rapp, 2018; Meier et al., 2019)
- Bayesian mixed-effects models (**BMEM** Hutema & McKean, 2000; Evans et al., 2020)

3. Compare agreement between effect sizes:

- Standardized (z-scored) effect sizes within method & condition
- Agreement estimated using concordance correlation coefficients (Lin, 1989)
 - Estimated via non-parametric U-statistics & bootstrapped 95% CIs (King, 2001)

[*] for a primer on each measure, see link at end of slides

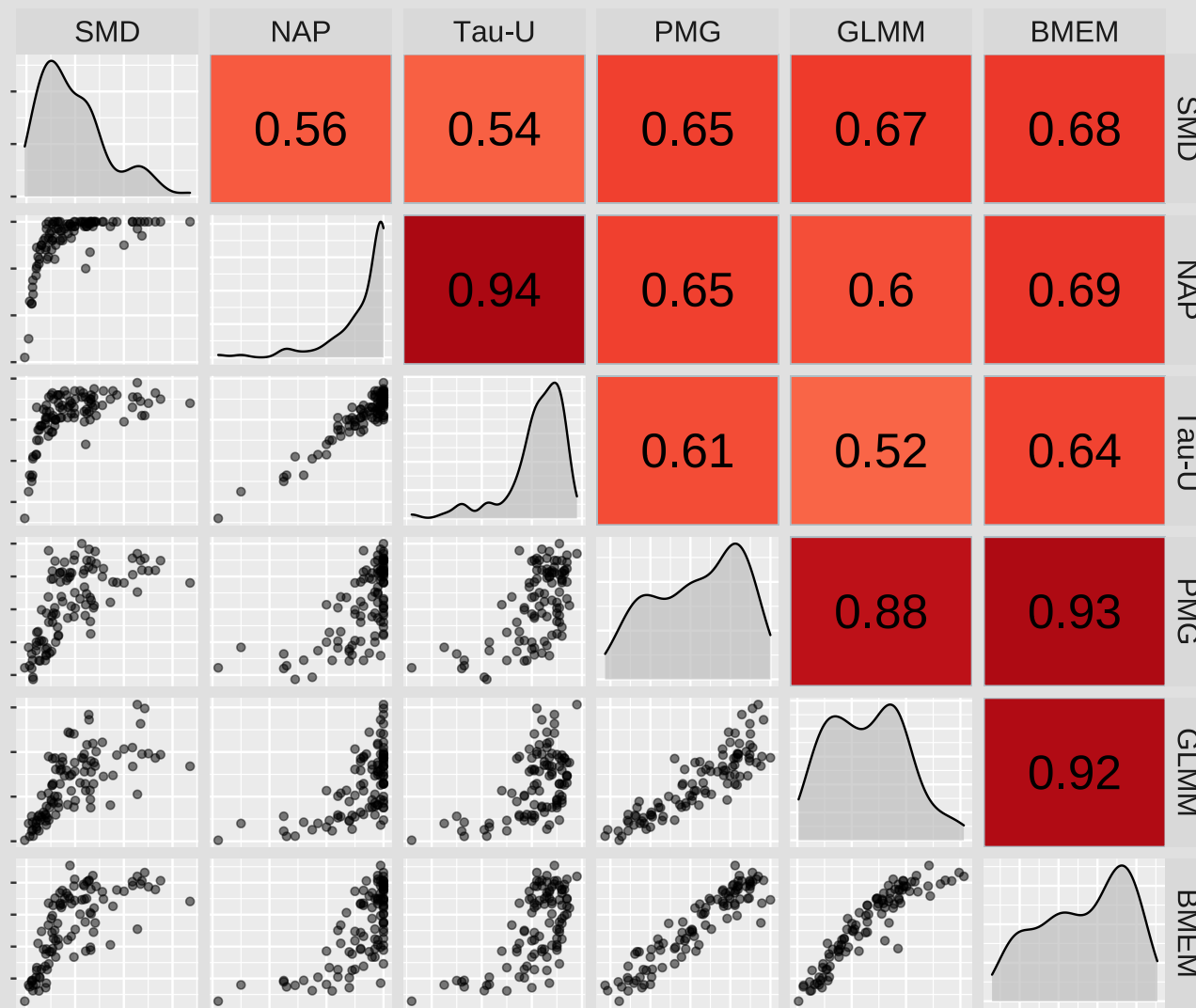
Results

Summary - Treated

1. Excellent agreement: mixed effect measures & PMG
2. Excellent agreement: Tau-U & NAP
3. Ceiling effects: NAP; Tau-U
4. SMD: disagreement increasing with effect size magnitude

Coefficients: < 0.40 = poor; 0.40 - 0.75 = fair to good; > 0.75 = good to excellent;

Concordance Correlation Coefficients: Treated Condition



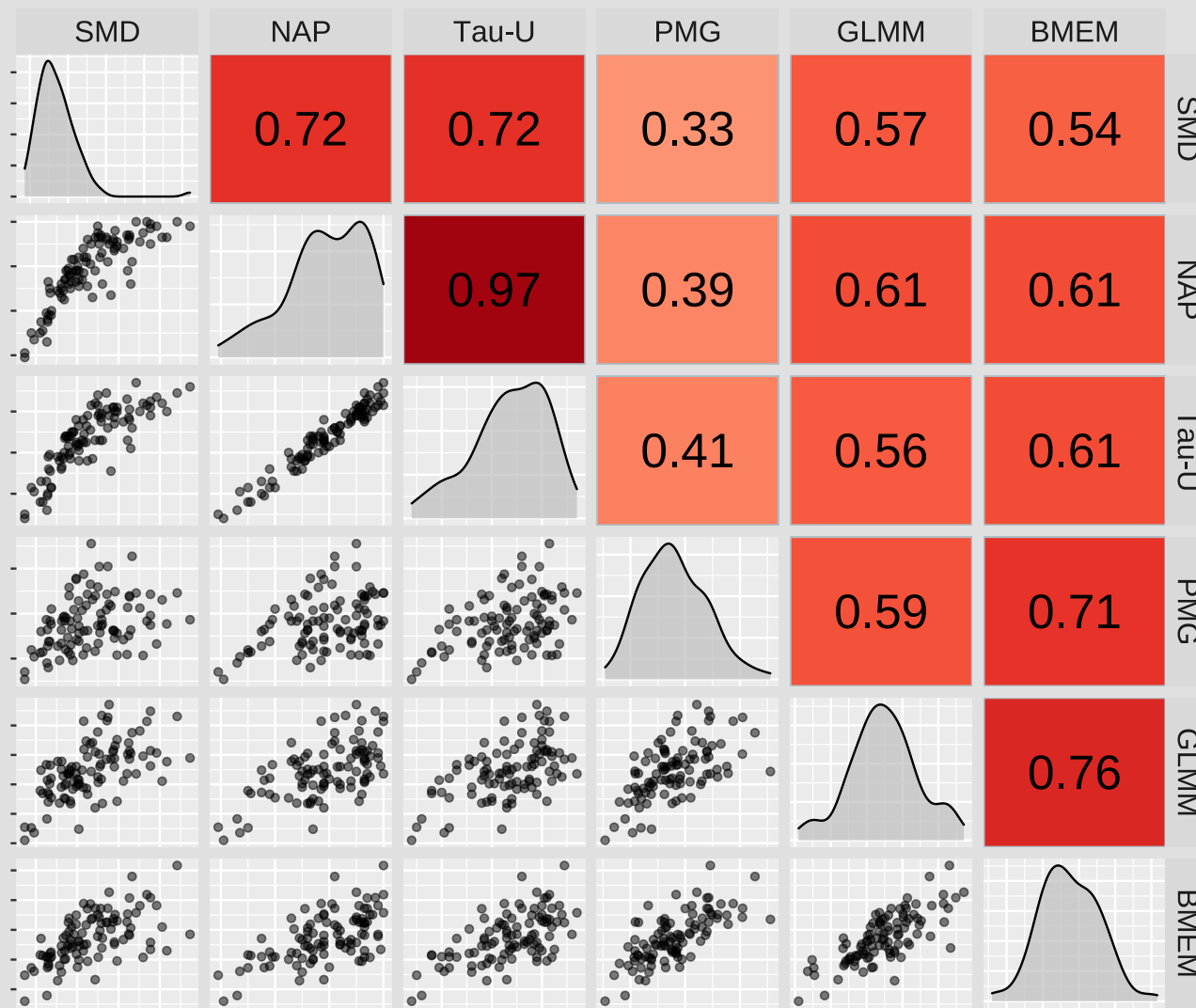
Results

Summary - Untreated

1. Excellent agreement: Tau-U & NAP
2. No ceiling effects
3. Poor agreement between PMG & SMD/NAP/Tau-U
4. SMD: heteroscedasticity remains

Coefficients: < 0.40 = poor; $0.40 - 0.75$ = fair to good; > 0.75 = good to excellent;

Concordance Correlation Coefficients: Untreated Condition



Findings

1. Careful (a priori) selection & use of effect size measures matters
 - measures are far from interchangeable
 - Effect size selection likely to affect study findings
2. Ceiling effects and rising baselines remain core unresolved issues
 - should be addressed with careful experimental design.
3. Continue to evaluate new methods:
 - Revised Baseline-corrected Tau (Tarlow, 2017:
<https://doi.org/10.1177/0145445516676750>)
 - Log Response Ratio or Log Odds Ratio (Pustejovsky, 2015:
<https://doi.org/10.1037/met0000019>)

Selection = dependent variable + experimental design + expected treatment pattern

Specific Recommendations

1. Given criticisms + available alternatives, recommend against future use of SMD.
 - no meaningful changes substituting Beeson & Robey's d for SMD
2. GLMMs and their Bayesian extensions satisfy most criteria for effect sizes:
 - accounting for baseline trends (within reason) & minimizing ceiling effects
 - providing relatively clear estimates of effect *magnitude and uncertainty*.
3. PMG approximates mixed effects approaches but lacks certainty/significance.
 - need to combine with an approach such as weighted statistics
4. NAP & Tau-U provide similar results when treatments effects are gradual and/or small but lose sensitivity when effect sizes are medium to large.

Selected References

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review*, 16(4), 161–169. <https://doi.org/10.1007/s11065-006-9013-7>

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (p. 187–212). Lawrence Erlbaum Associates, Inc.

Evans, W. S., Cavanaugh, R. B., Quique, Y. L., Boss, E., Starns, J. J., Hula, W. D. (manuscript under review). Playing with BEARS: Balancing Effort, Accuracy, and Response Speed in a Semantic Feature Verification Anomia Treatment Game.

Goldfeld, K. (2019). *simstudy: Simulation of Study Data* (R package). <https://cran.r-project.org/package=simstudy>

Huitema, B. E., & McKean, J. W. (2000). Design specification issues in time-series intervention models. *Educational and Psychological Measurement*, 60(1), 38–58. <https://doi.org/10.1177/00131640021970358>

King, T. S., & Chinchilli, V. M. (2001). A generalized concordance correlation coefficient for continuous and categorical data. *Statistics in Medicine*, 20(14), 2131–2147. <https://doi.org/10.1002/sim.845>

Lin, L. I.-K. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*, 45(1), 255–268. JSTOR. <https://doi.org/10.2307/2532051>

Lambon Ralph, M. A., Snell, C., Fillingham, J. K., Conroy, P., & Sage, K. (2010). Predicting the outcome of anomia therapy for people with aphasia post CVA: both language and cognitive status are key predictors. *Neuropsychological Rehabilitation*, 20(2), 289–305. <https://doi.org/10.1080/09602010903237875>

Manolov, R., & Solanas, A. (2008). Comparing N = 1 Effect Size Indices in Presence of Autocorrelation. *Behavior Modification*, 32(6), 860–875. <https://doi.org/10.1177/0145445508318866>

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, 40(4), 357–367. <https://doi.org/10.1016/j.beth.2008.10.006>

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, 42(2), 284–299. <https://doi.org/10.1016/j.beth.2010.08.006>

Wiley, R. W., & Rapp, B. (2018). Statistical analysis in Small-N Designs: Using linear mixed-effects modeling for evaluating intervention effectiveness. *Aphasiology*, 33(1), 1–30. <https://doi.org/10.1080/02687038.2018.1454884>

Thank You -

More Resources:

- Effect Size Primer:
rb-cavanaugh.shinyapps.io/aphasia-effect-sizes/
- R code & data: <https://osf.io/6x5pd/>

Acknowledgements

- Natalie Gilmore & Robery Wiley
- Pitt ICRE TL1 Predoctoral Fellowship Program
(1TL1TR001858, PI: Kraemer)

Contact: rob.cavanaugh@pitt.edu

