Calculating Effect Sizes for Single-Case Research

An Introduction to the SingleCaseES and scdhlm Web Applications and R Packages

James E. Pustejovsky

University of Wisconsin - Madison

April 26, 2023

Acknowledgement

- The work reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grants R324U190002 and R305D190023. The opinions expressed are those of the author and do not represent the views of the Institute or the U.S. Department of Education.

John Ferron

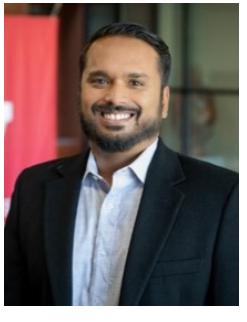
Wendy Machalicek

Man Chen

Paulina Grekov

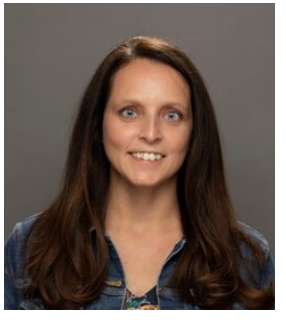
Bethany Hamilton-Bhat

Daniel Swan



David Klingbeil

Ethan van Norman



Jennifer Ledford

Outline

1. Organizing and curating data from single-case designs.
2. Within-study effect sizes.
    - Background
    - The `SingleCaseES` app
3. Between-case standardized mean differences
    - Background
    - The `scdhlm` app

Organizing and curating data from single-case designs

Why organize and curate your data?

1. So that you can do statistical analysis and effect size calculations.
2. So that you can share your data.
    - Make it easily accessible for inclusion in systematic reviews!
3. Because graphing data usually involves loss of information.
4. To fully document your research study.

Tidy SCD data

- One column per variable
- One row per observation session
- Descriptive labels for
  - Case (participant)
  - Phase of design or treatment condition

Case	Phase	Session	Outcome
Deborah's Group	Baseline	1	62.63
Deborah's Group	Baseline	2	40.22
Deborah's Group	Baseline	3	54.26
Deborah's Group	Baseline	4	40.26
Deborah's Group	Baseline	5	46.82
Deborah's Group	Baseline	6	52.45
Deborah's Group	Intervention	7	25.37
Deborah's Group	Intervention	8	26.32
Deborah's Group	Intervention	9	7.65
Deborah's Group	Intervention	10	11.41
Deborah's Group	Intervention	11	13.30
Deborah's Group	Intervention	12	22.66
Deborah's Group	Intervention	13	13.34
Amy's Group	Baseline	1	16.67
Amy's Group	Baseline	2	28.43
Amy's Group	Baseline	3	29.41
Amy's Group	Baseline	4	30.39
Amy's Group	Baseline	5	45.10
Amy's Group	Baseline	6	37.25

Multiple dependent variables

- **Wide format**: Use separate columns for multiple outcome variables

Case	Phase	Session	Problem_Behavior	On_Task_Behavior
Deborah's Group	Baseline	1	16.7	56.7
Deborah's Group	Baseline	2	20.0	70.0
Deborah's Group	Baseline	3	26.7	66.7
Deborah's Group	Baseline	4	20.0	86.7
Deborah's Group	Baseline	5	16.7	56.7
Deborah's Group	Baseline	6	13.3	70.0
Deborah's Group	Intervention	7	16.7	46.7
Deborah's Group	Intervention	8	20.0	73.3
Deborah's Group	Intervention	9	20.0	56.7
Deborah's Group	Intervention	10	30.0	50.0
Deborah's Group	Intervention	11	30.0	63.3
Deborah's Group	Intervention	12	13.3	63.3

Multiple dependent variables

- **Long format**: One row per outcome measure per session

Case	Phase	Session	DV	Outcome
Deborah's Group	Baseline	1	On Task Behavior	56.7
Deborah's Group	Baseline	1	Problem Behavior	16.7
Deborah's Group	Baseline	2	On Task Behavior	70.0
Deborah's Group	Baseline	2	Problem Behavior	20.0
Deborah's Group	Baseline	3	On Task Behavior	66.7
Deborah's Group	Baseline	3	Problem Behavior	26.7
Deborah's Group	Baseline	4	On Task Behavior	86.7
Deborah's Group	Baseline	4	Problem Behavior	20.0
Deborah's Group	Baseline	5	On Task Behavior	56.7
Deborah's Group	Baseline	5	Problem Behavior	16.7
Deborah's Group	Baseline	6	On Task Behavior	70.0
Deborah's Group	Baseline	6	Problem Behavior	13.3

Adding more detail

- Add furthers details about what happened in the study.
- Some ideas:
  - Actual session date + times (YYYY-MM-DD-HH:MM)
  - Observation session lengths
  - Clinician/therapist IDs
  - Notes about events

Case	Phase	Session	Problem Behavior	On-Task Behavior	Date	Session length	Notes
Deborah's Group	Baseline	1	16.7	56.7			
Deborah's Group	Baseline	2	20.0	70.0			
Deborah's Group	Baseline	3	26.7	66.7			

Share Your Data!

Within-case effect size indices

Within-case effect size indices

- Single-number summary of the **direction** and **magnitude** of intervention effect (functional relation) ***for each case*** within a study.
- Use these if you want to:
  - Describe results separately for each participant
  - Examine heterogeneity of effects or associations with individual-level characteristics
  - Compare results across participants and SCED studies that use various outcome measures
- Lots of proposed effect size indices. Today we'll focus on
  - Non-overlap of all pairs
  - Within-case standardized mean difference
  - Log-response ratio

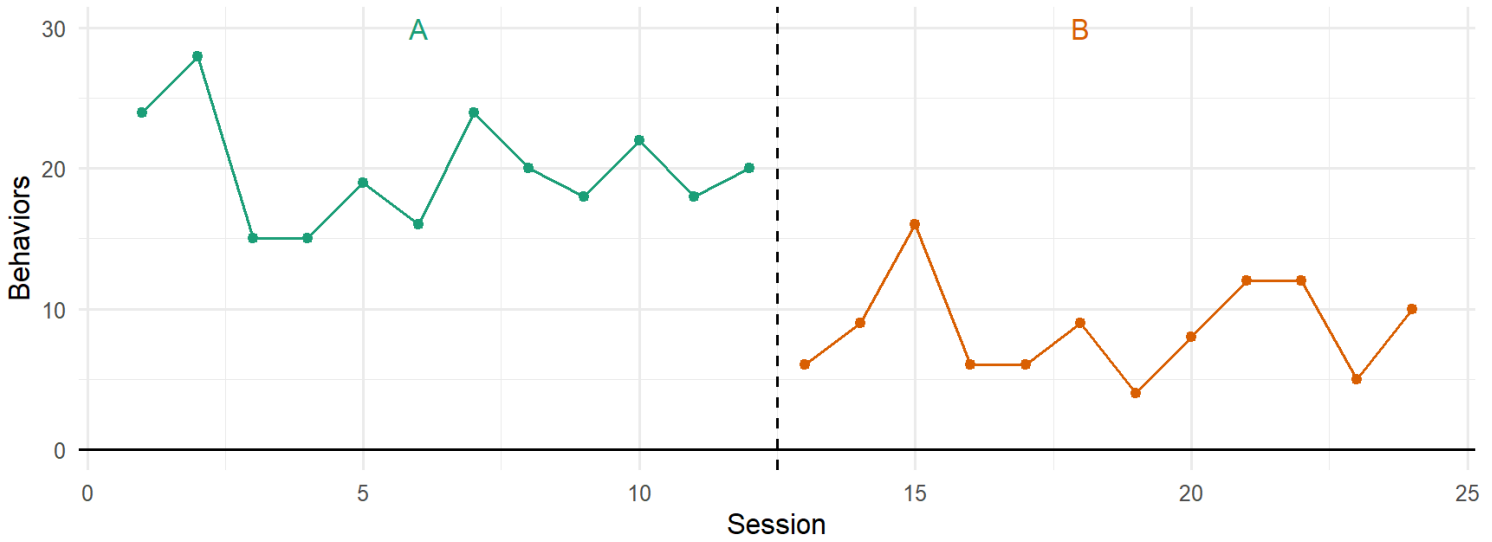
`SingleCaseES`: Single-series calculator

- Access `SingleCaseES` on the web at https://jepusto.shinyapps.io/SCD-effect-sizes/
- Or by opening RStudio and typing

```
library(SingleCaseES)
SCD_effect_sizes()
```

- Two parts to the app:
  - Single-series calculator (direct data entry)
  - Multiple-series calculator (using a data file)

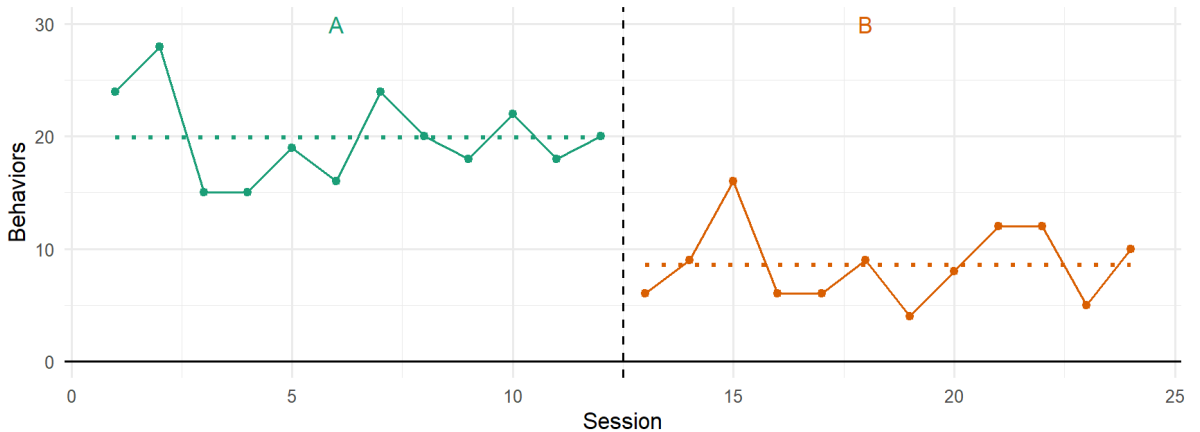
Simplest possible model



- Stable baseline and treatment phases (no time trends)
  - Immediate shift in level due to intervention
- Independence of outcome measurements

Notation

- n_A observations in phase A: $y_1^A, \ldots, y_{n_A}^A$
- n_B observations in phase B: $y_1^B, \ldots, y_{n_B}^B$
- Mean level of the outcome in each phase: μ_A, μ_B
  - Estimated by sample means $\bar{y}_A, \bar{y}_B$
- Standard deviation of the outcome in each phase: σ_A, σ_B
  - Estimated by sample standard deviations S_A, S_B



Non-overlap of all pairs

- Non-overlap measures are defined in terms of ***ordinal comparisons*** of outcomes
- Non-overlap of all pairs (Parker and Vannest, 2009) is defined in terms of all pairs of one observation from phase A and one observation from phase B.
- For every pair $i = 1, \ldots, n_A$ and $j = 1, \ldots, n_B$, take

$$q_{ij} = \begin{cases} 1 & \text{if} \quad y_j^B \text{ better than } y_i^A \\ \frac{1}{2} & \text{if} \quad y_j^B = y_i^A \\ 0 & \text{if} \quad y_j^B \text{ worse than } y_i^A \end{cases}$$

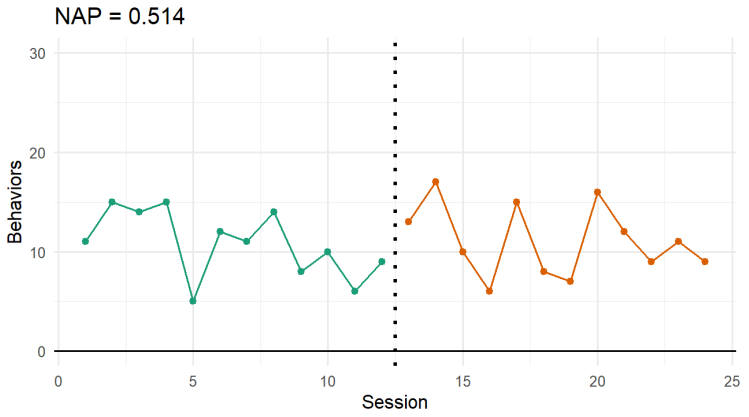
- NAP estimator:

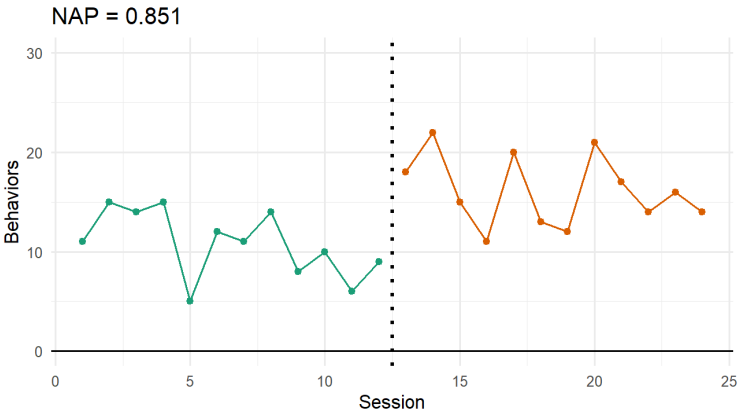
$$NAP = \frac{1}{mn} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} q_{ij}$$

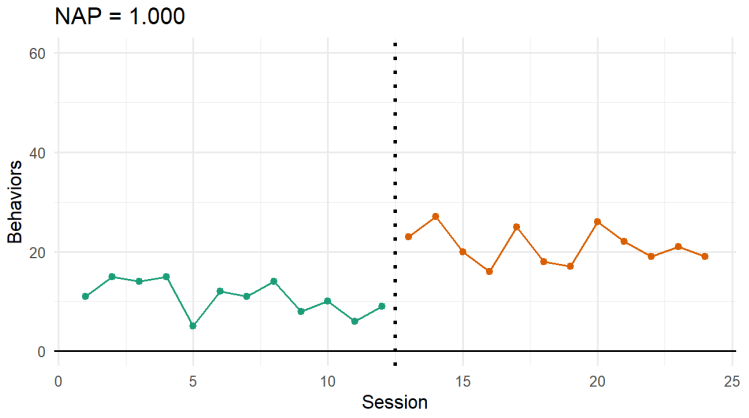
- Standard error based on unbiased estimator (Sen, 1967; Mee, 1990)
    - Methods assume that observations are independent and identically distributed within each phase.

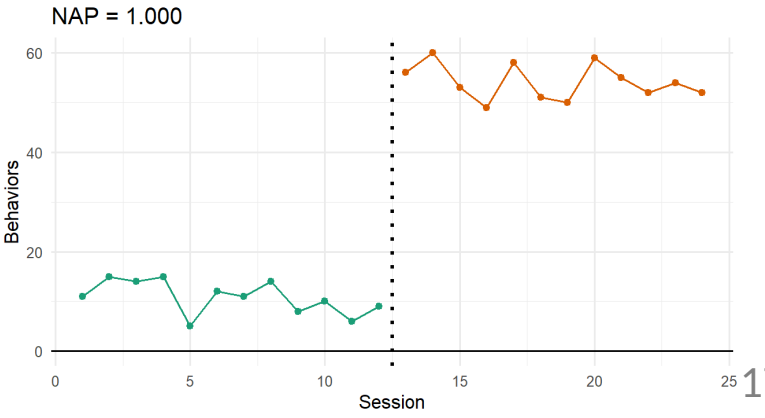
Limited range of sensitivity

- Limited range where NAP (and other non-overlap measures) sensitive to change.









Within-case standardized mean difference

- Proposed by Gingerich (1984) and Busk and Serlin (1992)
- Parameter definition:

$$\delta = \frac{\mu_B - \mu_A}{\sigma_A}$$

  - Difference in means, "standardized" by baseline variation
  - NOT equivalent to between-case SMD because σ_A includes only *within-case variation*.
- Appropriate for interval-scale outcomes
  - Is variability of outcomes approximately constant for different mean levels?
  - Standardizing by within-case variation means this measure will be strongly affected by reliability of measurements
  - Problematic for outcomes with restricted range in baseline

Within-case standardized mean difference: estimation

- Originally proposed estimator:

$$d = \frac{\bar{y}_B - \bar{y}_A}{S_A}$$

- Estimator with small-sample bias correction:

$$g = \left(1 - \frac{3}{4n_A - 5}\right) \times \frac{\bar{y}_B - \bar{y}_A}{S_A}$$

- Approximate standard error, ***assuming independent observations***:

$$SE_g = \left(1 - \frac{3}{4n_A - 5}\right) \sqrt{\frac{1}{n_A} + \frac{S_B^2}{n_B S_A^2} + \frac{d^2}{2(n_A - 1)}}$$

Proportional change in levels

- Percentage (proportional) change from baseline to intervention is an easily interpretable "informal" effect size (Campbell and Herzinger, 2010).
- The **log response ratio** is a formal measure of effect size that describes change in proportional terms (Pustejovsky, 2015; Pustejovsky, 2018).
- Parameter definition:

$$\psi = \log\left(\frac{\mu_B}{\mu_A}\right)$$

  - Appropriate for **ratio-scale** outcomes (frequency counts, percentage duration)
  - Natural logarithm is used to make the range unrestricted.

- Transformation to percentage change:

$$\%\ \text{change} = 100\% \times \left(e^{\psi} - 1\right)$$

Log response ratio: estimation

- Basic estimator (biased if m or n is small):

$$R_1 = \log\left(\frac{\tilde{y}_B}{\tilde{y}_A}\right)$$

- Bias-corrected estimator:

$$R_2 = \log\left(\frac{\tilde{y}_B}{\tilde{y}_A}\right) + \frac{\tilde{S}_B^2}{2n_B\tilde{y}_B^2} - \frac{\tilde{S}_A^2}{2n_A\tilde{y}_A^2}$$

- Approximate standard error for R_2, ***assuming independent observations***:

$$SE_R = \sqrt{\frac{\tilde{S}_A^2}{n_A\tilde{y}_A^2} + \frac{\tilde{S}_B^2}{n_B\tilde{y}_B^2}}$$

Direction of improvement

- Two versions of LRR:
  - **LRRi**: Positive numbers represent **increases** in **desirable outcomes**
  - **LRRd**: Negative numbers represent **decreases** in **undesirable outcomes**
- Use the version that corresponds to ***predominant valence*** of outcomes in your data.
- For count outcomes, LRRi $=$ -LRRd
- For proportion / percentage outcomes, the outcome valence is harmonized before calculation.
  - For proportion / percentage outcomes, LRRi $\neq$ -LRRd

Truncation constants

- If $\bar{y}_A = 0$ or $\bar{y}_B = 0$ then LRR is undefined.
  - If $S_A^2 = 0$ or $S_B^2 = 0$ then SE_R is undefined
- To handle such situations, the app uses **truncated mean** and **truncated SD** estimators:

$$\tilde{y}_A = \max\left\{\bar{y}_A, \frac{1}{2n_A D}\right\}, \qquad \tilde{y}_B = \max\left\{\bar{y}_B, \frac{1}{2n_B D}\right\}$$

and

$$\tilde{S}_A^2 = \max\left\{S_A^2, \frac{1}{n_A^3 D^2}\right\}, \qquad \tilde{S}_B^2 = \max\left\{S_B^2, \frac{1}{n_B^3 D^2}\right\}$$

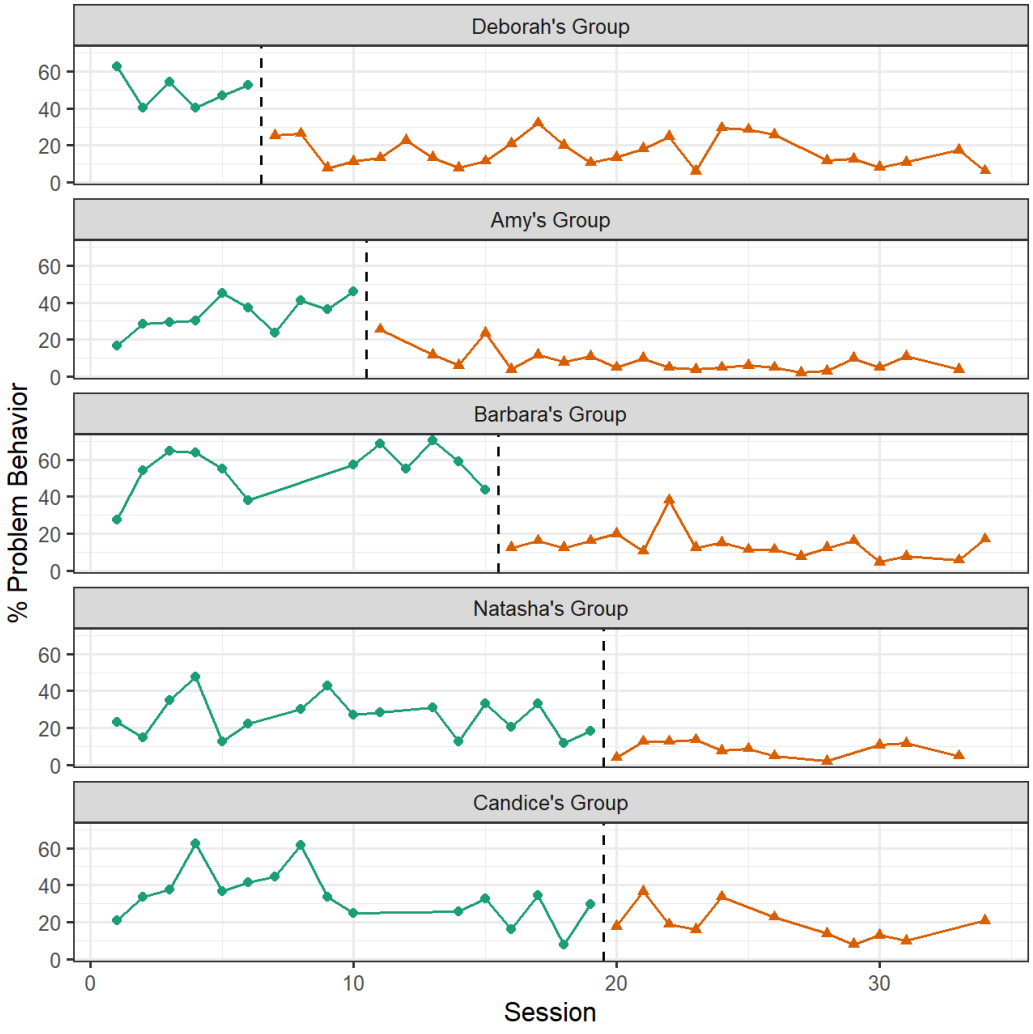
- D is a constant that depends on the outcome scale and measurement procedures
  - Number of intervals / items
  - Session length for direct observation
  - Can also define your own D

`SingleCaseES`: Multiple-series calculator

- Basic walk-through with data from Rodriguez and Anderson (2014)
- Calculating phase-pairs in ABAB designs
- Aggregating effect sizes

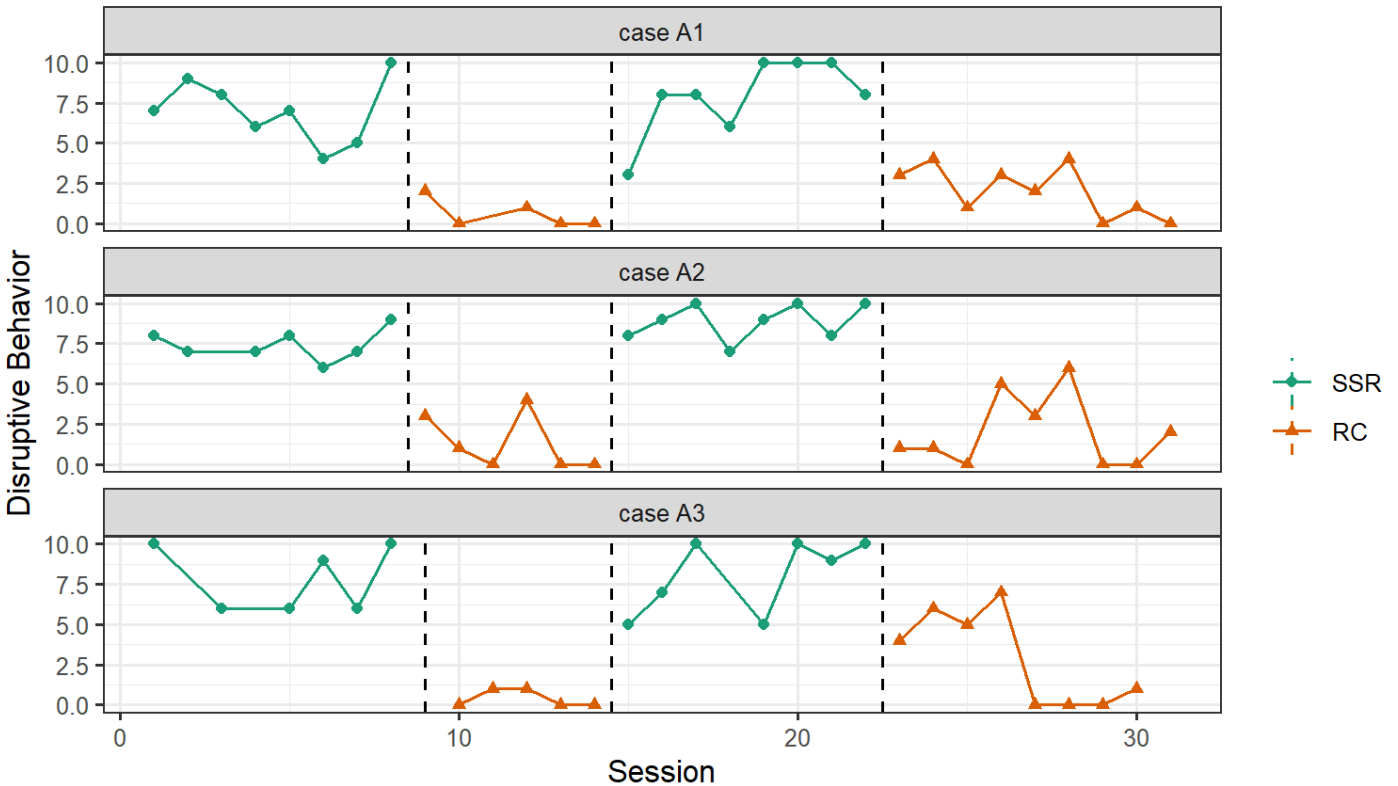
Rodriguez and Anderson (2014)

Integrating a social behavior intervention during small group academic instruction using a total group criterion intervention



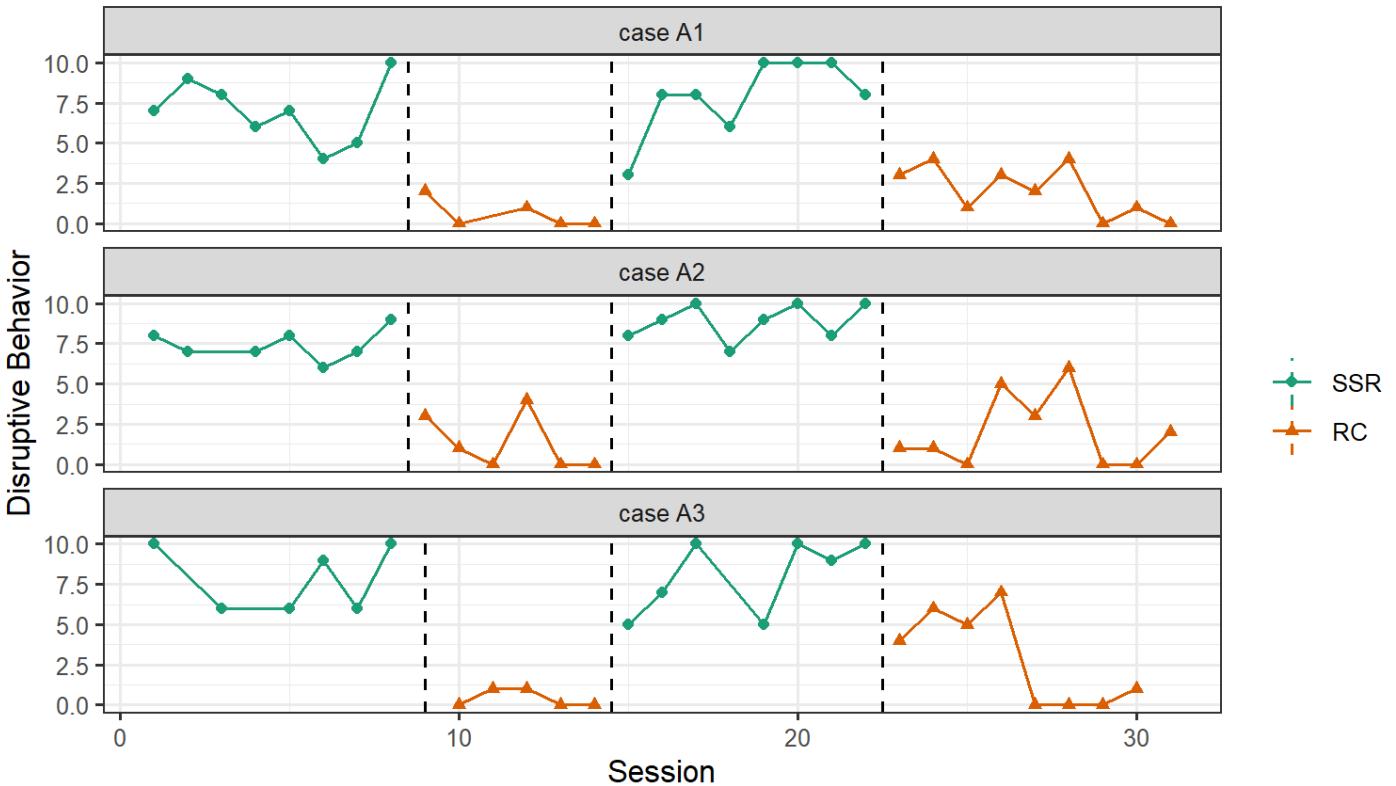
Lambert, Cartledge, Heward et al. (2006)

Effects of response cards on disruptive behavior and academic responding during math lessons by fourth-grade urban students



Calculating phase pairs

- Might want to calculate effect sizes for adjacent pairs of baseline and intervention phases.
- SingleCaseES provides an option to determine phase pairs automatically.



Aggregating effect sizes

- After calculating effect sizes for adjacent pairs of phases, we might want to **average them together** to simplify reporting or further analysis.
  - Average across phase pairs in an ABAB design
  - Average across cases to generate an overall summary effect size estimate
- Several options for taking weighted averages
  - Equal weighting
  - Inverse-variance weighting: $\frac{1}{V}$ (use for LRR)
  - n_A
  - n_B
  - $n_A n_B$ (use for NAP)
  - $\frac{1}{n_A} + \frac{1}{n_B}$ (use for SMD)

Replication code

```r
# Load packages
library(SingleCaseES)

# Load data
library(readxl)
library(janitor)

dat <-
  read_excel(path = "Small-is-Beautiful-effect-size-workshop.xlsx", sheet = "Lambert") %>%
  clean_names(case = "parsed")

# clean data
library(dplyr)

dat <-
  dat %>%
  group_by(case) %>%
  mutate(phase_pair_calculated = calc_phase_pairs(treatment, session = day)) %>%
  ungroup()

# Batch calculation
res <- batch_calc_ES(dat = dat,
                     grouping = c(case),
                     condition = treatment,
                     outcome = outcome,
                     aggregate = c(phase_pair_calculated),
                     weighting = "1/nA + 1/nB",
                     session_number = day,
                     baseline_phase = "SSR",
```

Between-case standardized mean differences

Premises

- **Goal**: Estimate an effect size using data from a single-case design that is *in the same metric* as the standardized mean difference effect size from a between-group experimental design.
- **Why?** (Shadish, Hedges, Horner et al., 2015)
  - **Translation** of single-case research for researchers who work primarily with between-groups designs
  - **Comparison** of results from single-case studies and between-groups studies, for purposes of understanding the utility and limitations of each type of design
  - **Synthesis** involving both single-case and between-groups designs

SMD in between-group experiment

- What is the SMD from a between-group experiment?

$$\delta_{BC} = \frac{\left(\begin{array}{c}\text{Average outcome if}\\ \text{everybody gets intervention}\end{array}\right) - \left(\begin{array}{c}\text{Average outcome if}\\ \text{nobody gets intervention}\end{array}\right)}{\left(\begin{array}{c}\text{SD of outcome if}\\ \text{nobody gets intervention}\end{array}\right)}$$

$$\delta_{BC} = \frac{\left(\begin{array}{c}\text{Average outcome if}\\ \text{everybody gets intervention}\end{array}\right) - \left(\begin{array}{c}\text{Average outcome if}\\ \text{nobody gets intervention}\end{array}\right)}{\sqrt{\left(\begin{array}{c}\text{Between-participant}\\ \text{variance}\end{array}\right) + \left(\begin{array}{c}\text{Within-participant}\\ \text{variance}\end{array}\right)}}$$

- We aim to estimate these component quantities using data from a single-case experimental design.

The broad strategy

(Pustejovsky, Hedges, and Shadish, 2014) described a general strategy for estimating BC-SMD:

1. Develop a hierarchical linear model that describes:
    - The form of time trends and intervention effects
    - How the trends and intervention effects vary across participants
2. Imagine a **hypothetical between-group experiment** with the same population of participants, same intervention, same dependent variable.
    - When is treatment initiated?
    - When are outcomes assessed?
3. Use the hierarchal model to estimate the components of δ_{BC} for the hypothetical experiment.
4. Make a small-sample correction (similar to Hedges' g)

Design translation

A multiple baseline across participants:

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18
X	X	X	X	X	T	X	X	X	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	T	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X	X	T	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X	X	X	X	X	T	X	X	X

A hypothetical between-group design (with pre-test):

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18
				X	T										X		
				X	T										X		
				X											X		
				X											X		

Overview of methods literature

- BC-SMD estimators for a basic hierarchical linear model with no time trends:
  - Hedges, Pustejovsky, and Shadish (2012): Treatment reversal (ABAB) design replicated across 3+ participants
  - Hedges, Pustejovsky, and Shadish (2013): Multiple baseline / multiple probe design with 3+ participants
  - Shadish, Hedges, and Pustejovsky (2014): More worked examples
- Pustejovsky, Hedges, and Shadish (2014) described a more general strategy for multiple baseline / multiple probe designs across participants
  - Valentine, Tanner-Smith, Pustejovsky et al. (2016): Tutorial and practical guidance
- Swaminathan, Rogers, and Horner (2014) proposed Bayesian estimation methods
- Chen, Pustejovsky, Klingbeil et al. (2023) proposed BC-SMD methods for more complex designs:
  - Multiple baseline across behaviors, replicated across 3+ participants
  - Clustered multiple baseline design across participants (3+ clusters)
  - Multivariate multiple baseline design across 3+ participants

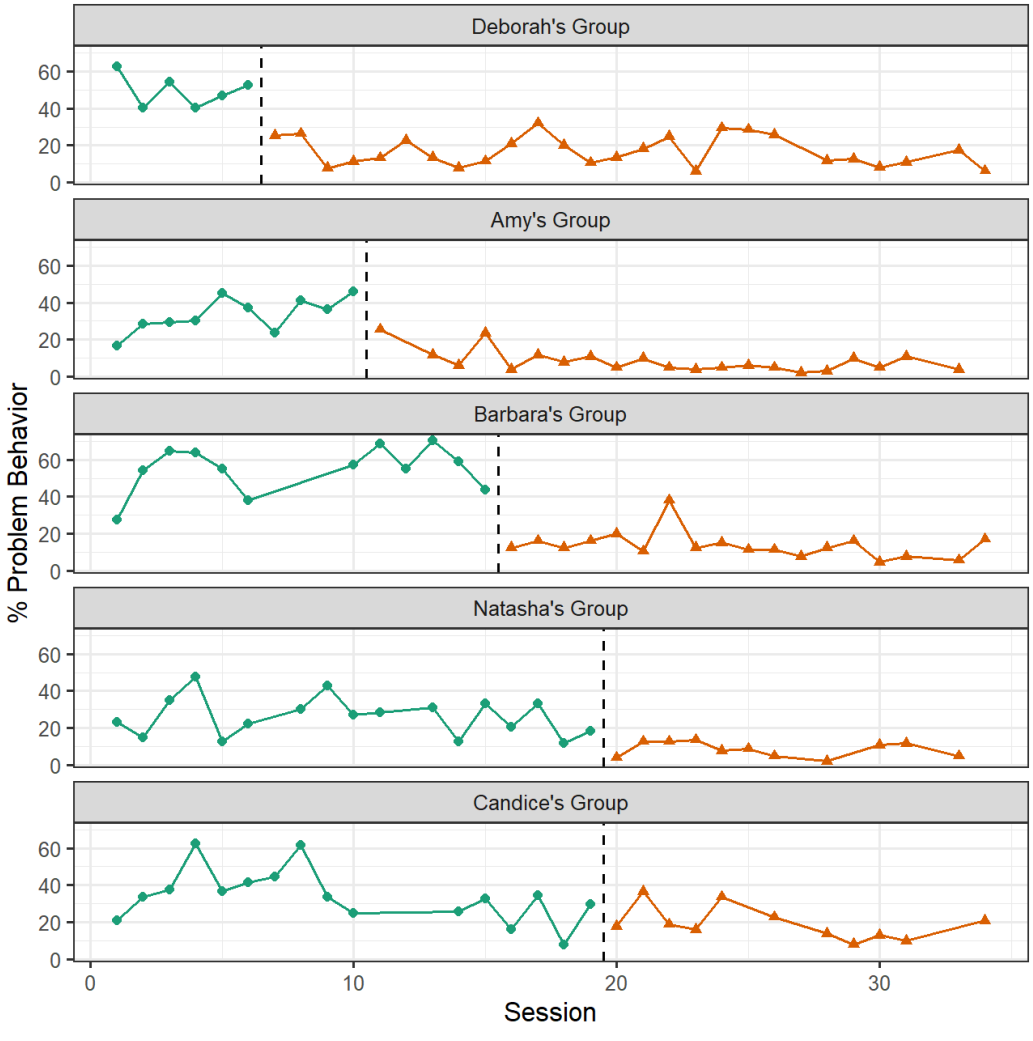
`scdhlm` web app

- Access `scdhlm` on the web at https://jepusto.shinyapps.io/scdhlm/
- Or by opening RStudio and typing

```
library(scdhlm)
shine_scd()
```

Rodriguez and Anderson (2014)

Integrating a social behavior intervention during small group academic instruction using a total group criterion intervention



The most basic HLM

- Level-1 model for each participant:

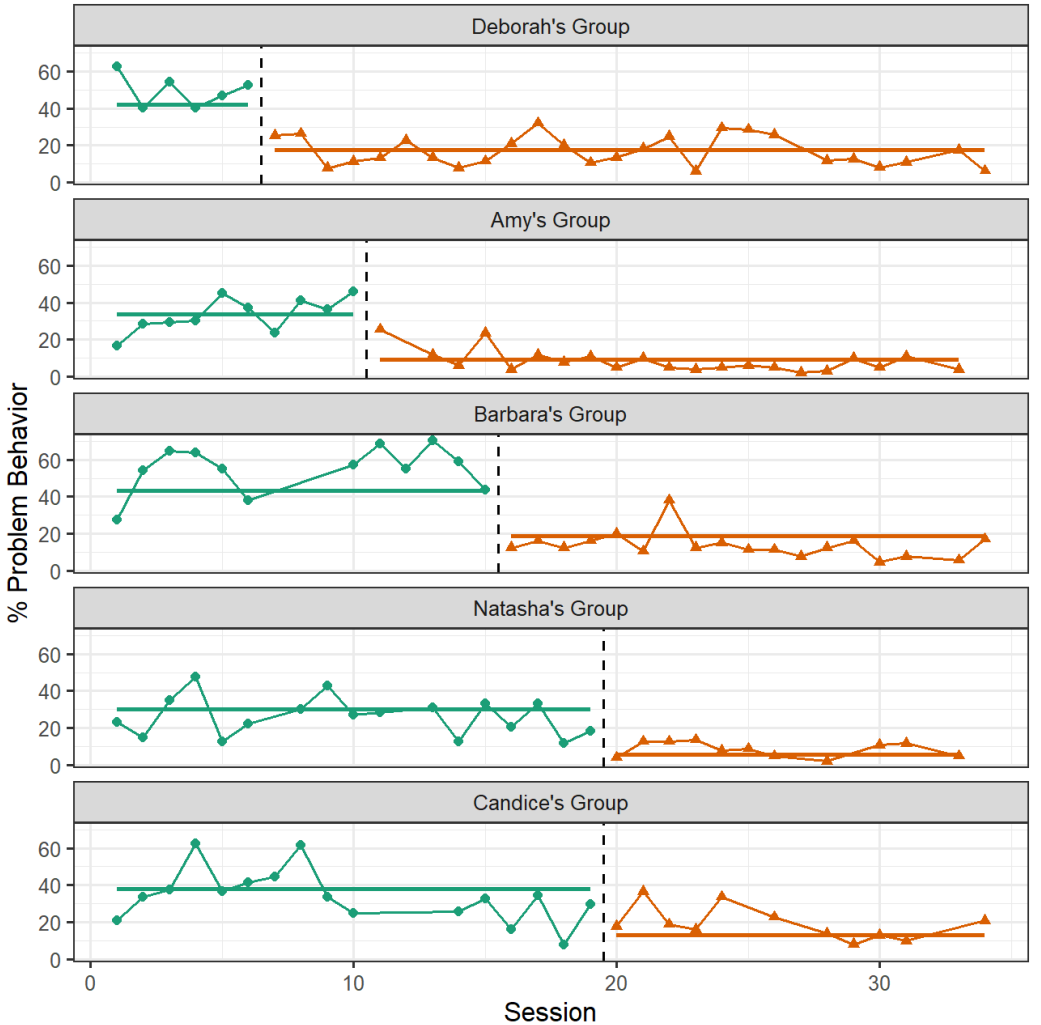
$$Y_{ij} = \beta_{0j} + \beta_{1j}(Tx)_{ij} + e_{ij}$$

  where $\mathrm{Var}(e_{ij}) = \sigma_e^2$ and $e_{1j}, \ldots, e_{Tj} \sim AR_1(\phi)$

- Level-2 model:

$$\begin{aligned} \beta_{0j} &= \theta_{00} + u_{0j}, \qquad u_{0j} \sim N(0, \sigma_{u0}^2) \\ \beta_{1j} &= \theta_{10} \end{aligned}$$

- Under this model:
  - Average outcome if nobody gets intervention: θ_{00}
  - Average outcome if everybody gets intervention: $\theta_{00} + \theta_{10}$
  - SD of outcome if nobody gets intervention: $\sqrt{\sigma_{u0}^2 + \sigma_e^2}$
  - BC-SMD effect size: $\delta_{BC} = \dfrac{\theta_{10}}{\sqrt{\sigma_{u0}^2 + \sigma_e^2}}$



- $\hat{\theta}_{00} = 37.5$
- $\hat{\theta}_{10} = -24.7$
- $\hat{\sigma}^2_e = 112.1$
- $\hat{\sigma}^2_{u0} = 36.7$

$$\hat{\delta}_{BC} = \frac{-24.7}{\sqrt{112.1 + 36.7}} = -2.03$$

$$g_{BC}\ (SE) = -1.99\ (0.31)$$

A more flexible HLM

- Level-1 model for each participant:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Time})_{ij} + \beta_{2j}(Tx)_{ij} + \beta_{3j}(Tx)_{ij} \times \big((\text{Time})_{ij} - k_j\big) + e_{ij}$$

  where $\text{Var}(e_{ij}) = \sigma_e^2$ and $e_{1j}, \ldots, e_{Tj} \sim AR_1(\phi)$ and k_j is last baseline session.

- Level-2 model:

$$\begin{aligned}
\beta_{0j} &= \theta_{00} + u_{0j}, \qquad & u_{0j} \sim N(0, \sigma^2_{u0}) \\
\beta_{1j} &= \theta_{10} + u_{1j}? \\
\beta_{2j} &= \theta_{20} + u_{2j}? \\
\beta_{3j} &= \theta_{30} + u_{3j}?
\end{aligned}$$

  - Adding a random effect $\rightarrow$ allowing slope / Tx effect to ***vary across cases***
  - Omitting a random effect $\rightarrow$ assuming slope / Tx effect is ***constant***
  - Models with more random effects require more cases

A more flexible HLM

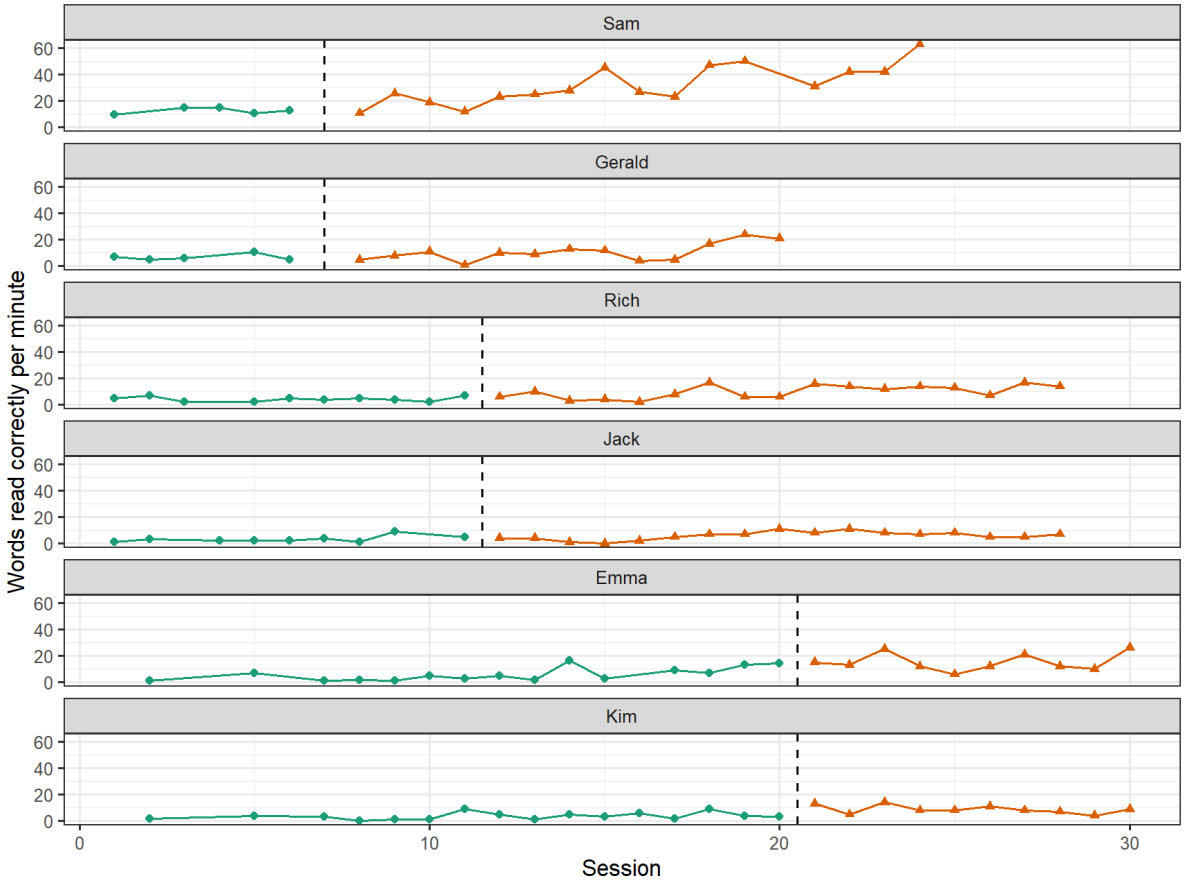
- Can also modify assumptions about level-1 errors
- Different variance by treatment phase:

$$\text{Var}(e_{ij}) = \begin{cases} \sigma^2_{eC} & \text{if} \quad (Tx)_{ij} = 0 \\ \sigma^2_{eT} & \text{if} \quad (Tx)_{ij} = 1 \end{cases}$$

- Correlation structure of level-1 errors:
  - First order auto-regression $(AR_1(\phi))$
  - First order moving average $(MA_1(\phi))$
  - Independent

Model should be informed by **theoretical expectations** and **visual inspection**

Barton-Arwood, Wehby, and Falk (2005) Reading instruction for elementary-age students with emotional and behavioral disorders: Academic and behavioral outcomes



Models with time trends

- For models with time trends, we need to specify ***timing*** of pre-test and post-test for the hypothetical between-group design.
- **Initial treatment time**: Last session of baseline phase before being assigned to intervention or comparison condition.
  - Default: Length of shortest baseline phase
- **Focal follow-up time**: Session during which outcomes would be assessed in hypothetical experiment.
  - Default: Last measurement occasion for first case to enter intervention
  - This is not a particularly good default
  - Ideally, pick a focal follow-up time based on a meaningful or typical treatment duration

Barton-Arwood, Wehby, and Falk (2005) effect size calculations

- Model specification
  - Baseline level (random)
  - Baseline time trends (constant)
  - Treatment level change (constant)
  - Treatment trend change (random)
  - Level-1 variance differs by phase
- Initial treatment time: After 6 sessions
- Focal follow-up time of session 16 (10 sessions of treatment).
- BC-SMD estimate:

$$g_{BC}\ (SE) = 0.82\ (0.75)$$

Illustrative application of BC-SMDs

- Calder and colleagues (2020, 2021) studied an explicit grammar instruction intervention for children with developmental language disorder.
- Calder, Claessen, Ebbels et al. (2020): multiple baseline across nine participants
  - Data available in the Excel workbook
  - Try calculating a BC-SMD estimate after 10 weeks of intervention
- Calder, Claessen, Ebbels et al. (2021): crossover randomized trial with $N = 21$ participants
  - 10 weekly intervention sessions
  - $g = 1.97, SE = 0.11$ for expressive morphosyntax
  - $g = 0.06, SE = 0.06$ for grammaticality judgements

Limitations of between-case SMD

- Tool for translating from single-case logic to group-design logic.
    - Premised on the idea that a hypothetical group design is theoretically plausible
- Describes a **summary**, **average effect** across a set of cases
    - Potentially concealing individual-level heterogeneity
- For some models, magnitude depends on the features (timing) of hypothetical between-group design
- Technical limitations
    - Only available for some designs
    - Requires at least 3 participants (preferably more!)
    - Models assume normal (Gaussian) errors
    - Care needed for model selection

References

Barton-Arwood, S. M., J. H. Wehby, and K. B. Falk (2005). "Reading Instruction for Elementary-Age Students with Emotional and Behavioral Disorders: Academic and Behavioral Outcomes". In: *Exceptional Children* 72.1, pp. 7-27. ISSN: 0014-4029, 2163-5560. DOI: 10.1177/001440290507200101. (Visited on Apr. 26, 2023).

Busk, P. L. and R. C. Serlin (1992). "Meta-Analysis for Single-Case Research". In: *Single-Case Research Design and Analysis: New Directions for Psychology and Education*. Ed. by T. R. Kratochwill and J. R. Levin. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc, pp. 187-212.

Calder, S. D., M. Claessen, S. Ebbels, et al. (2020). "Explicit Grammar Intervention in Young School-Aged Children With Developmental Language Disorder: An Efficacy Study Using Single-Case Experimental Design". In: *Language, Speech, and Hearing Services in Schools* 51.2, pp. 298-316. ISSN: 0161-1461, 1558-9129. DOI: 10.1044/2019_LSHSS-19-00060.

Calder, S. D., M. Claessen, S. Ebbels, et al. (2021). "The Efficacy of an Explicit Intervention Approach to Improve Past Tense Marking for Early School-Age Children with Developmental Language Disorder". In: *Journal of Speech, Language, and Hearing Research* 64.1, pp. 91-104. ISSN: 1092-4388, 1558-9102. DOI: 10.1044/2020_JSLHR-20-00132. (Visited on Mar. 17, 2022).

Campbell, J. M. and C. V. Herzinger (2010). "Statistics and Single Subject Research Methodology". In: *Single Subject Research Methodology in Behavioral Sciences*. Ed. by D. L. Gast. New York, NY: Routledge. Chap. 14, pp. 417-450.

Chen, M., J. E. Pustejovsky, D. A. Klingbeil, et al. (2023). "Between-Case Standardized Mean Differences: Flexible Methods for Single-Case Designs". In: *Journal of School Psychology* 98, pp. 16-38. ISSN: 00224405. DOI: 10.1016/j.jsp.2023.02.002. (Visited on Apr. 26, 2023).

Gingerich, W. J. (1984). "Meta-Analysis of Applied Time-Series Data". In: *The Journal of Applied Behavioral Science* 20.1, pp. 71-79. ISSN: 0021-8863. DOI: 10.1177/002188638402000113.

Hedges, L. V., J. E. Pustejovsky, and W. R. Shadish (2012). "A Standardized Mean Difference Effect Size for Single Case Designs". In: *Research Synthesis Methods* 3, pp. 224-239. ISSN: 17592879. DOI: 10.1002/jrsm.1052.

Hedges, L. V., J. E. Pustejovsky, and W. R. Shadish (2013). "A Standardized Mean Difference Effect Size for Multiple Baseline Designs across Individuals". In: *Research Synthesis Methods*. ISSN: 17592879. DOI: 10.1002/jrsm.1086.

References

Lambert, M. C., G. Cartledge, W. L. Heward, et al. (2006). "Effects of Response Cards on Disruptive Behavior and Academic Responding during Math Lessons by Fourth-Grade Urban Students". In: *Journal of Positive Behavior Interventions* 8.2, p. 88.

Mee, R. W. (1990). "Confidence intervals for probabilities and tolerance Regions based on a generalization of the Mann-Whitney statistic". In: *Journal of the American Statistical Association* 85.411, p.

1. DOI: 10.2307/2290017.

Parker, R. I. and K. Vannest (2009). "An improved Effect Size for Single-Case Research: Nonoverlap of All Pairs". In: *Behavior Therapy* 40.4, pp. 357-367. DOI: 10.1016/j.beth.2008.10.006.

Pustejovsky, J. E. (2015). "Measurement-Comparable Effect Sizes for Single-Case Studies of Free-Operant Behavior." In: *Psychological Methods* 20.3, pp. 342-359. ISSN: 1939-1463, 1082-989X. DOI: 10.1037/met0000019.

Pustejovsky, J. E. (2018). "Using Response Ratios for Meta-Analyzing Single-Case Designs with Behavioral Outcomes". In: *Journal of School Psychology* 68, pp. 99-112. ISSN: 00224405. DOI: 10.1016/j.jsp.2018.02.003.

Pustejovsky, J. E., L. V. Hedges, and W. R. Shadish (2014). "Design-Comparable Effect Sizes in Multiple Baseline Designs: A General Modeling Framework". In: *Journal of Educational and Behavioral Statistics* 39.5, pp. 368-393. ISSN: 1076-9986. DOI: 10.3102/1076998614547577.

Rodriguez, B. J. and C. M. Anderson (2014). "Integrating a Social Behavior Intervention During Small Group Academic Instruction Using a Total Group Criterion Intervention". In: *Journal of Positive Behavior Interventions* 16.4, pp. 234-245. ISSN: 1098-3007, 1538-4772. DOI: 10.1177/1098300713492858. (Visited on Apr. 21, 2023).

Sen, P. K. (1967). "A note on asymptotically Distribution-Free Confidence Bounds for P{X $<$ Y}, Based on Two Independent Samples". In: *Sankhy\=a: The Indian Journal of Statistics, Series A (1961-2002)* 29.1, pp. 95-102.

Shadish, W. R., L. V. Hedges, R. H. Horner, et al. (2015). *The Role of Between-Case Effect Size in Conducting, Interpreting, and Summarizing Single-Case Research*. NCER 2015-002. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

References

Shadish, W. R., L. V. Hedges, and J. E. Pustejovsky (2014). "Analysis and Meta-Analysis of Single-Case Designs with a Standardized Mean Difference Statistic: A Primer and Applications". In: *Journal of School Psychology* 52.2, pp. 123-147. ISSN: 0022-4405. DOI: 10.1016/j.jsp.2013.11.005.

Swaminathan, H., H. J. Rogers, and R. H. Horner (2014). "An Effect Size Measure and Bayesian Analysis of Single-Case Designs". In: *Journal of School Psychology*. ISSN: 00224405. DOI: 10.1016/j.jsp.2013.12.002.

Valentine, J. C., E. E. Tanner-Smith, J. E. Pustejovsky, et al. (2016). "Between-Case Standardized Mean Difference Effect Sizes for Single-Case Designs: A Primer and Tutorial Using the Scdhlm Web Application". In: *Campbell Systematic Reviews* 12.1, pp. 1-31. ISSN: 18911803. DOI: 10.4073/cmdp.2016.1.