Using response ratios for meta-analyzing single-case designs with behavioral outcomes

James E. Pustejovsky – pusto@austin.utexas.edu

TEXAS Education
The University of Texas at Austin
College of Education

Paper and materials available at: https://osf.io/4fe6u/

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

Introduction

Methods for meta-analyzing single-case designs (SCDs) are needed in order to inform evidence-based practice in special education and to draw broader and more defensible generalizations in areas where SCDs comprise a large part of the research base. The most widely used outcomes in single-case research are measures of behavior collected using systematic direct observation, which are typically quantified as natural rates or proportions. For studies that use such measures, a simple and intuitive way to quantify effect sizes is in terms of proportionate change from baseline, using an effect size known as the log response ratio. This paper describes methods for estimating log response ratios and for combining the estimates using meta-analysis. The methods are based on a simple model for comparing two phases, where the level of the outcome is stable within each phase and the repeated outcome measurements are independent. Auto-correlation in the outcome measures will lead to inconsistent estimates of the sampling variance of the effect size. However, meta-analysis of response ratios can be conducted with robust variance estimation procedures that remain valid even when sampling variance estimates are inconsistent. The methods are demonstrated using data from a recent meta-analysis on the effects of group contingency interventions for student problem behavior.

Assumptions and parameter definition

- Average level of the outcome is constant within each phase
- Outcome measures are mutually independent (working assumption)
- The log-response ratio parameter is defined as

$$\psi = \ln\left(\mu_B / \mu_A\right) = \ln(\mu_B) - \ln(\mu_A)$$

where μ_A is level during baseline and μ_B is level during treatment
- Transformation to percentage change:

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

Estimation

$Y_1^A, ..., Y_m^A$ outcome data in baseline phase

$Y_1^B, ..., Y_n^B$ outcome data in treatment phase

$$\tilde{y}_A = \max\left\{\frac{1}{2mD}, \frac{1}{m}\sum_{i=1}^{m} Y_i^A\right\} \quad \text{truncated baseline phase mean}$$

$$\tilde{y}_B = \max\left\{\frac{1}{2nD}, \frac{1}{n}\sum_{i=1}^{n} Y_i^B\right\} \quad \text{truncated treatment phase mean}$$

s_A baseline phase standard deviation

s_B treatment phase standard deviation

- LRR estimator (with small-sample bias correction):

$$R_2 = \ln\left(\tilde{y}_B\right) + \frac{s_B^2}{2n\tilde{y}_B^2} - \ln\left(\tilde{y}_A\right) - \frac{s_A^2}{2m\tilde{y}_A^2}$$

- Variance estimator (assuming mutual independence):

$$V^R = \frac{s_A^2}{m\tilde{y}_A^2} + \frac{s_B^2}{n\tilde{y}_B^2}$$

Advantages of log-response ratio

- Direct relationship to ***percentage change from baseline***, which is a familiar and readily interpretable conceptualization of effect size (Marquis et al., 2000; Campbell & Herzinger, 2010).
- Relatively insensitive to outcome measurement procedures (e.g., observation session length, recording system) because it is a function of mean levels alone (Pustejovsky, 2015, 2016).
- Under certain conditions, LRRs are comparable across different dimensional characteristics, such as frequency and percentage duration (Pustejovsky, 2015).

Limitations of log-response ratio

- Currently available methods assume outcome levels are stable (without time trends) within each phase.
- Variance estimator is based on assumption that outcome measures are mutually independent. However, this limitation can be addressed using robust variance estimation.
- Use requires attention to operational definition of behavioral outcomes, which makes calculation more complicated.
- May not be appropriate for behaviors with zero or near-zero baselines.

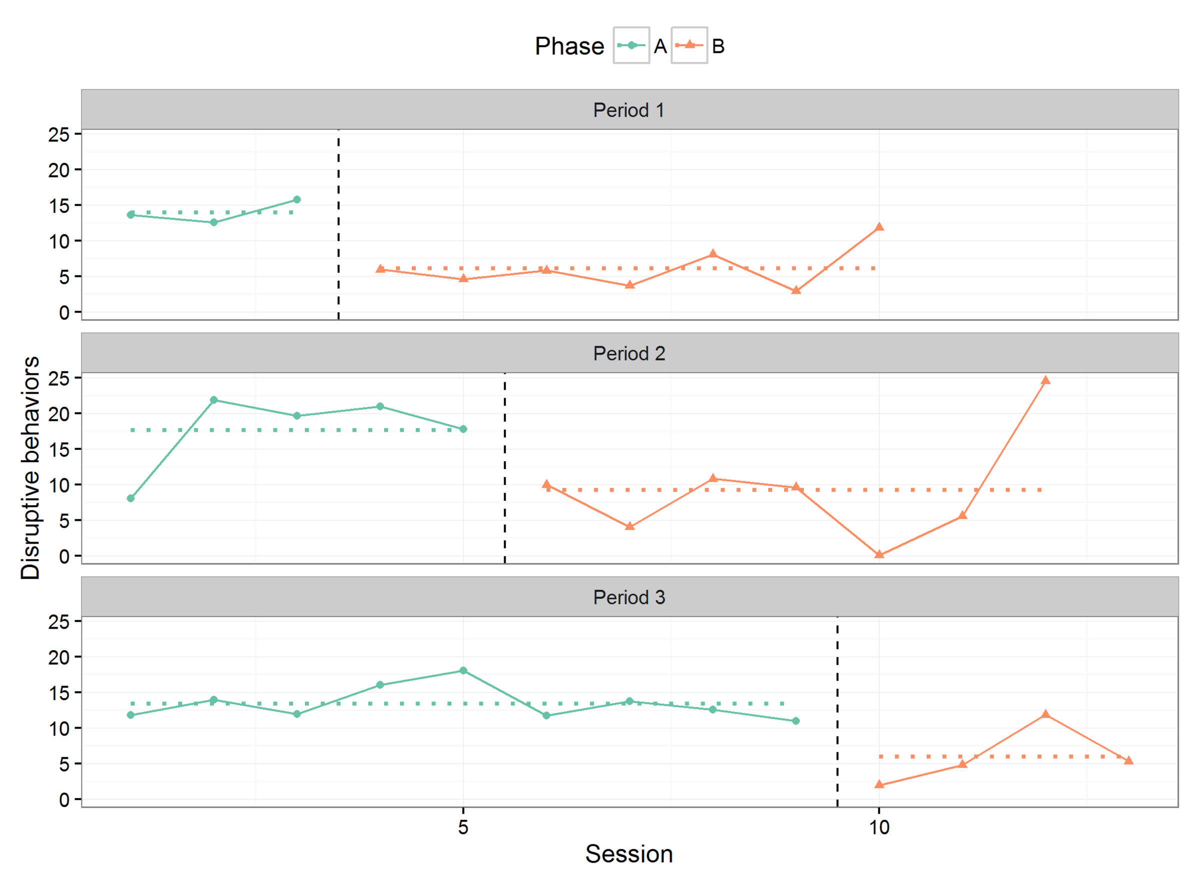
Example 1: McKissick et al. (2010)

- McKissick et al. (2010) examined effects of interdependent group contingency involving randomization of component contingencies in a second-grade, general education classroom.
- Multiple baseline design across class periods.
- Disruptive behaviors measured using frequency counting in 20 min sessions across the entire sample of 26 students.

Table 1. Summary statistics and LRR effect size estimates for frequency of disruptive behavior data from McKissick et al. (2010).

	Baseline phase			Treatment phase				
Case	$\tilde{y}_A$	s_A	m	$\tilde{y}_B$	s_B	n	R_2	SE^R
Period 1	13.983	1.626	3	6.146	3.025	7	-0.807	0.198
Period 2	17.652	5.577	5	9.211	7.766	7	-0.610	0.349
Period 3	13.441	2.330	9	5.997	4.183	4	-0.748	0.354

Figure 1. Disruptive behaviors during multiple baseline design from McKissick et al. (2010)



Valence transformation

- If studies with both positive- and negative-valence outcomes are to be included in a meta-analysis, then the effect sizes should first be transformed so that the sign of the estimate is consistent with the direction of therapeutic improvement across all of the outcomes.
- For LRR, the transformation method depends on whether the metric of the outcome variable is quantified as a natural rate or as a proportion.
- For studies that measure behavior on a natural rate metric, reverse the sign of effect size indices (i.e., multiplying by -1) to be consistent with the direction of therapeutic improvement.
- For studies that measure behavior on a proportion or percentage metric, redefine the outcome variable so that the direction of therapeutic improvement is consistent.
  - For example, suppose that most studies examine behavior where decrease is desirable, but one study measures percentage of time on-task (positive valence). Before calculating the LRR, the data from this study need to first be transformed by subtracting the original scores from 100%, yielding percentage of time off-task.
- Two distinct ways that the LRR can be applied to proportion outcomes, depending on whether therapeutic improvement corresponds to negative or positive values of the LRR:
  - LRR-d when negative values correspond to therapeutic improvement
  - LRR-i when positive values correspond to therapeutic improvement

Handling multiple pairs of phases

- Some common types of SCDs involve multiple replications of the baseline-treatment contrast for each case (e.g., treatment reversal designs).
- Three possible approaches to estimate LRR:
  - ***select a single pair of phases*** that best represents the functional relationship of interest (e.g., intial baseline and initial treatment phase);
  - ***pool data across multiple phases*** and calculate a single LRR estimate comparing all baseline phases to all treatment phases; or
  - ***calculate estimates for each pair*** of adjacent phases, then average those estimates together to obtain a single summary effect size for each case.

Meta-analysis with robust variance estimation

- Purposes of meta-analysis:
  - Estimate overall average effects across cases and studies
  - Characterize heterogeneity of effects across cases and studies
  - Identify moderators of effect magnitude
- Multi-level meta-analysis model for case-level effect size estimates (Van den Noortgate & Onghena, 2003, 2008):

$$R_{jk} = \gamma + v_k + u_{jk} + e_{jk}$$

where
- R_{jk} is LRR effect size estimate for case j in study k,
- e_{jk} is sampling error with mean 0 and variance V_{jk},
- u_{jk} is case-level random deviation with mean 0 and variance ω^2
- v_k is study-level random deviation with mean 0 and variance τ^2
- γ is overall average effect size estimate across

- Restricted maximum likelihood estimation of variance components ω^2, τ^2
- Estimate γ using inverse-variance weighted average
- Standard errors and confidence intervals using robust variance estimation (Hedges, Tipton, & Johnson, 2010) to account for possible mis-estimation of sampling variances (V_{jk}) due to auto-correlation in outcome measures.
- Estimation in R using **metafor** (Viechtbauer, 2010) and **clubSandwich** (Pustejovsky, 2017) packages.

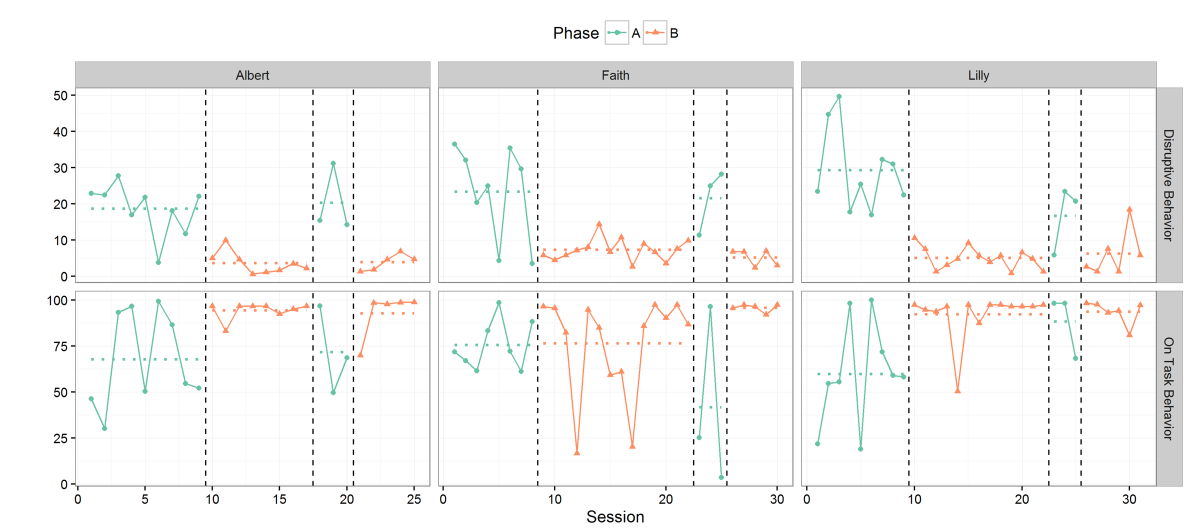
Example 2: Schmidt (2007)

- Schmidt (2007) evaluated the effects of Class-wide Function-Based Intervention Teams (CW-FIT; a group-contingency intervention), on on-task and disruptive behavior of three students in a first-grade class.
- Daily 10 minute observation sessions
- On task behavior measured using duration recording
- Disruptive behavior measured using frequency counting

Table 2. LRR-d and LRR-i effect size estimates and variances for disruptive behavior and on-task behavior data from Schmidt (2007).

	LRR-d						LRR-i	
	A1B1		A2B2		Combined		Combined	
Case	R_2^1	V^{R1}	R_2^2	V^{R2}	R_2	V^R	R_2	V^R
Disruptive behaviors (frequency count)								
Albert	-1.605	0.104	-1.651	0.141	-1.628	0.061	1.628	0.061
Faith	-1.168	0.051	-1.428	0.096	-1.298	0.037	1.298	0.037
Lilly	-1.749	0.044	-0.947	0.289	-1.348	0.083	1.348	0.083
On-task behavior (% duration)								
Albert	-1.716	0.155	-1.165	0.842	-1.440	0.249	0.282	0.014
Faith	-0.009	0.132	-2.698	0.285	-1.353	0.104	0.310	0.116
Lilly	-1.560	0.261	-0.870	0.884	-1.215	0.286	0.237	0.010

Figure 2. Disruptive and on-task behaviors during a replicated ABAB design from Schmidt (2007)



Example 3: Maggin et al. (2017)

- Maggin and colleagues (2017) conducted a systematic review and synthesis of single-case studies on group-contingency interventions.
- Original meta-analysis was based on between-case standardized mean difference effect size indices (Shadish, Hedges, & Pustejovsky, 2013), which could be estimated for only 27 (68%) of included studies.
- Re-analysis using the LRR effect size
  - Can be applied to data from all studies that met inclusion criteria
  - Case-level index that allows for examination of heterogeneity within and across studies
  - Illustration used 33 studies and 110 cases that assessed effects of group contingency interventions on problem behavior.
  - Meta-regression controlling for setting (general education or special education) and unit of analysis (group or individual)

Table 3. Meta-analysis results based on LRR-d effect sizes for problem behavior outcomes from Maggin et al. (2017).

	Studies	Effects	Est.	SE	d.f.	95% CI	$\hat{\omega}$	$\hat{\tau}$
Model 1							0.21	0.43
Overall average	33	110	-1.18	0.08	31.1	[-1.35, -1.01]		
Model 2							0.21	0.32
General Ed., group	19	46	-0.95	0.07	17.1	[-1.11, -0.80]		
General Ed., individual	5	22	-1.65	0.25	3.9	[-2.36, -0.95]		
Special Ed., group	5	15	-1.53	0.15	3.6	[-1.98, -1.09]		
Special Ed., individual	4	27	-1.20	0.24	2.8	[-2.01, -0.39]		