Operational sensitivities of non-overlap effect sizes for single-case experimental designs

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

TEXAS Education
The University of Texas at Austin
College of Education

Introduction

One of the basic decisions that must be made in any quantitative synthesis of intervention research is what effect size to use to quantify the magnitude of treatment effects. Ideally, an effect size measure should be on a scale that is easy to interpret and that also allows for comparison with other studies in the same area (Hedges, 2008; Lipsey & Wilson, 2001). Good effect size measures should therefore be relatively insensitive to incidental features of a study's design, such as the sample size or the choice of measurement procedures. Using effect sizes metrics that are sensitive to such operational details will tend to obscure substantive variation in study results and reduce the interpretability of synthesis results.

A wide array of effect sizes have been proposed for use with single-case experimental designs, but there remains considerable disagreement regarding their merits (Shadish, Rindskopf, & Hedges, 2008). The most widely used effect size measures in single-case research are in the family of non-overlap statistics. The non-overlap measures are sometimes described as non-parametric effect sizes, in the sense that they are not premised on assumptions regarding the normality of outcome measures and are insensitive to outliers (Parker, Vannest, & Davis, 2011). These properties are seen as advantageous because many outcome measures used in single-case research are not well-modeled by normal distributions. However, little previous research has examined the characteristics of non-overlap effect sizes under data-generating models that are more plausible for the types of outcome measures used in single-case research. The present study aims to fill that gap, by studying the behavior of non-overlap effect sizes using data simulated from a realistic model for direct observation of behavior. The simulation results demonstrate that all of the non-overlap measures are in fact sensitive to operational details of the study design.

Notation

$Y_1^A, \ldots, Y_m^A$ outcome data in baseline phase

$Y_1^B, \ldots, Y_n^B$ outcome data in treatment phase

Simulation design

- Simple AB design with m observation sessions in baseline phase and n observation sessions in treatment phase.
- Each observation lasted L min.
- For each observation session, a behavior stream was generated based on an Alternating Poisson Process model for state behavior with specified prevalence and incidence.
- Treatment leads to a percentage decrease in prevalence and incidence.
- Various observation recording procedures applied to the simulated behavior stream.
- Simulations carried out with the R package ARPobservation (Pustejovsky, 2014; available on CRAN).
- 10,000 replications of each combination of factor levels.

Type	Factor	Levels	Citation
Behavioral	Prevalence (%)	20, 50, 80	None available
	Incidence (per min)	1, 2	Mudford, Taylor, & Martin (2009)
	Treatment phase decrease (%)	0, 50, 80	None available
Procedural	Recording procedure	Continuous recording momentary time sampling (10, 20, 30 s) partial interval recording (10, 20, 30 s)	Mudford, Taylor, & Martin (2009) Adamson & Wachsmuth (2014)
	Session length (L min)	5, 10, 15, 20	Gage et al. (2012)
	Baseline sessions (m)	5, 10, 15, 20	Smith (2012) Shadish & Sullivan (2011)
	Treatment sessions (n)	5, 10, 15, 20	

Percentage of non-overlapping data (PND)

Def: Percentage of measurements in the treatment phase that exceed the highest measurement from the baseline phase (Scruggs et al., 1987).

$$\text{PND} = 100\% \times \frac{1}{n}\sum_{i=1}^{n} I\left(Y_i^B > \max\left\{Y_1^A, \ldots, Y_m^A\right\}\right)$$

Figure 1. Expected magnitude of PND when the intervention has no effect, for L = 5 and varying session lengths.

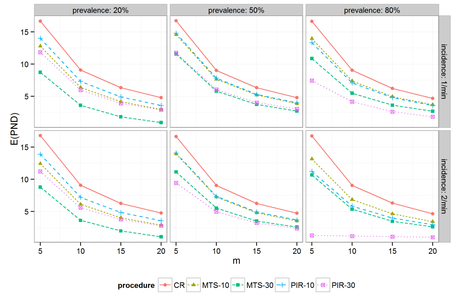
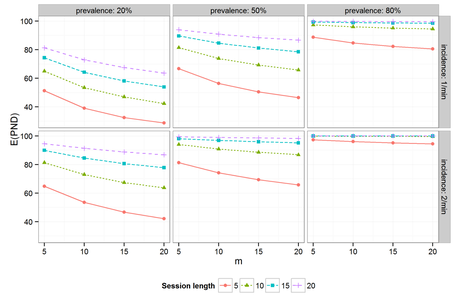


Figure 2. Expected magnitude of PND based on continuous recording data when the intervention leads to a 50% change, for varying session lengths.

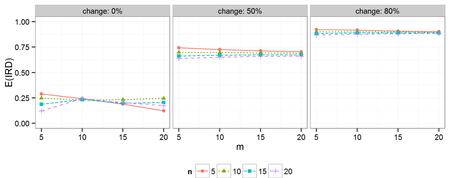


Improvement Rate Difference (IRD)

Def: Pearson's phi coefficient corresponding to a 2 × 2 table arrangement of the numbers obtained in calculating PAND, where overlapping observations are evenly divided between the lower left and upper right cells of the table (Parker, Vannest, & Brown, 2009).

$$\text{IRD} = \frac{1}{2mn}\left[(m+n)^2 \frac{\text{PAND}}{100\%} - m^2 - n^2\right]$$

Figure 5. Expected magnitude of IRD based on continuous recording data with L = 5, when incidence is once per min., for varying baseline and treatment phase lengths.



Percentage of all non-overlapping data (PAND)

Def: Percentage of the total number of observations remaining after removing the minimum number of observations from either phase such that the highest remaining measurement from the baseline phase is less than the lowest remaining measurement from the treatment phase (Parker, Vannest, & Davis, 2011, 2014).

$$\text{PAND} = 100\% \times \frac{1}{m+n}\max_{i,j}\left\{(i+j)\, I\left(Y_{(i)}^A < Y_{(n+1-j)}^B\right)\right\}$$

for baseline phase order statistics $Y_{(1)}^A, Y_{(2)}^A, \ldots, Y_{(m)}^A$

and treatment phase order statistics $Y_{(1)}^B, Y_{(2)}^B, \ldots, Y_{(n)}^B$

Figure 3. Expected magnitude of PAND based on continuous recording data with L = 5, when incidence is once per min., for varying baseline and treatment phase lengths.

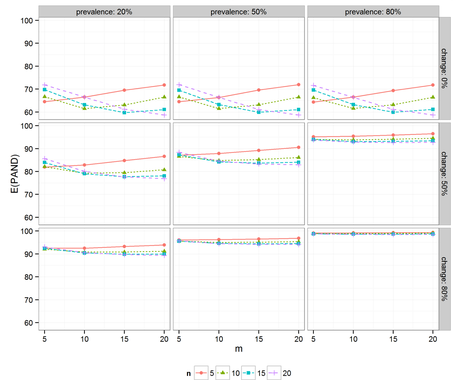
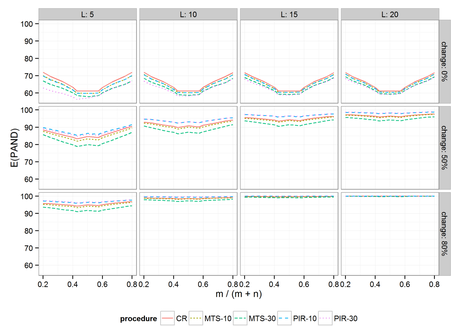


Figure 4. Expected magnitude of PAND based on continuous recording data when prevalence is 50% and incidence is once per min., for varying session lengths and recording procedures.



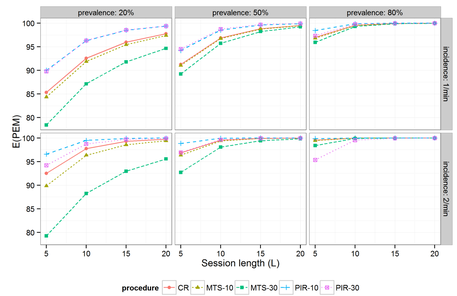
Percentage exceeding the median (PEM)

Def: Percentage of measurements in the treatment phase that exceed the median of the baseline phase measurements; to account for the possibility of ties in the data, measurements in the treatment phase that are exactly equal to the median of the baseline phase are counted as half an observation (Ma, 2006).

$$\text{PEM} = 100\% \times \frac{1}{n}\sum_{i=1}^{n}\left[I\left(Y_i^B > M^A\right) + 0.5 I\left(Y_i^B = M^A\right)\right]$$

where $M^A = \text{median}\left\{Y_1^A, \ldots, Y_m^A\right\}$

Figure 6. Expected magnitude of PEM.

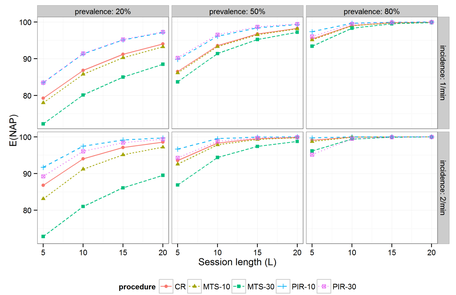


Non-overlap of all pairs (NAP)

Def: Percentage of all pairwise comparisons where the measurement from the treatment phase exceeds the measurement from the baseline phase (Parker & Vannest, 2009).

$$\text{NAP} = 100\% \times \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[I\left(Y_j^B > Y_i^A\right) + 0.5 I\left(Y_j^B = Y_i^A\right)\right]$$

Figure 7. Expected magnitude of NAP.



Conclusions

- Magnitude of non-overlap measures is affected by arbitrary procedural factors, not solely by the magnitude of a treatment effect.
- The non-overlap measures are unsuitable for use as effect sizes because they do not provide a fair basis for comparison across studies that use different procedures.
- Syntheses of single-case experimental designs need to pay greater attention to the outcome measurement procedures and details of study design.

Extent to which non-overlap measures are sensitive to procedural factors

Non-overlap measure	Session length	Recording procedure	Baseline sessions	Treatment sessions
PND	High	Moderate	High	-
PAND	Moderate	Slight	High	High
IRD	Moderate	Slight	High	High
PEM	High	High	-	-
NAP	High	High	-	-