Synthesis of single-case research: Meta-analytic methods & challenges

James E. Pustejovsky

pusto@austin.utexas.edu

April 10, 2014

Outline

1. Research synthesis & meta-analysis
2. Effect sizes for single-case research
   - Desiderata
   - Examples using direct observation of behavior
3. Challenges & data quality issues

Research synthesis & meta-analysis

- **Research synthesis**: the systematic integration of empirical research for purposes of drawing generalizations (Cooper & Hedges, 2009).

- **Meta-analysis**: statistical methods that support research synthesis, especially methods for combining results from a collection of studies.

Disciplines that rely on research synthesis

- Medicine (cf. the Cochrane Collaboration)
- Education
- Psychology
- Social policy (justice, welfare, etc.)
- Physical sciences
- Economics, international development

Synthesis of single-case research (SCR)

- Goals of synthesis:
  - Improve generalizability of findings from small studies
  - Understanding moderators of effectiveness
  - Establish evidence-based practices

- Synthesis across two levels
  - multiple cases in a single study
  - multiple studies

Quantitative Syntheses of SCR for students with disabilities: 1985-2009

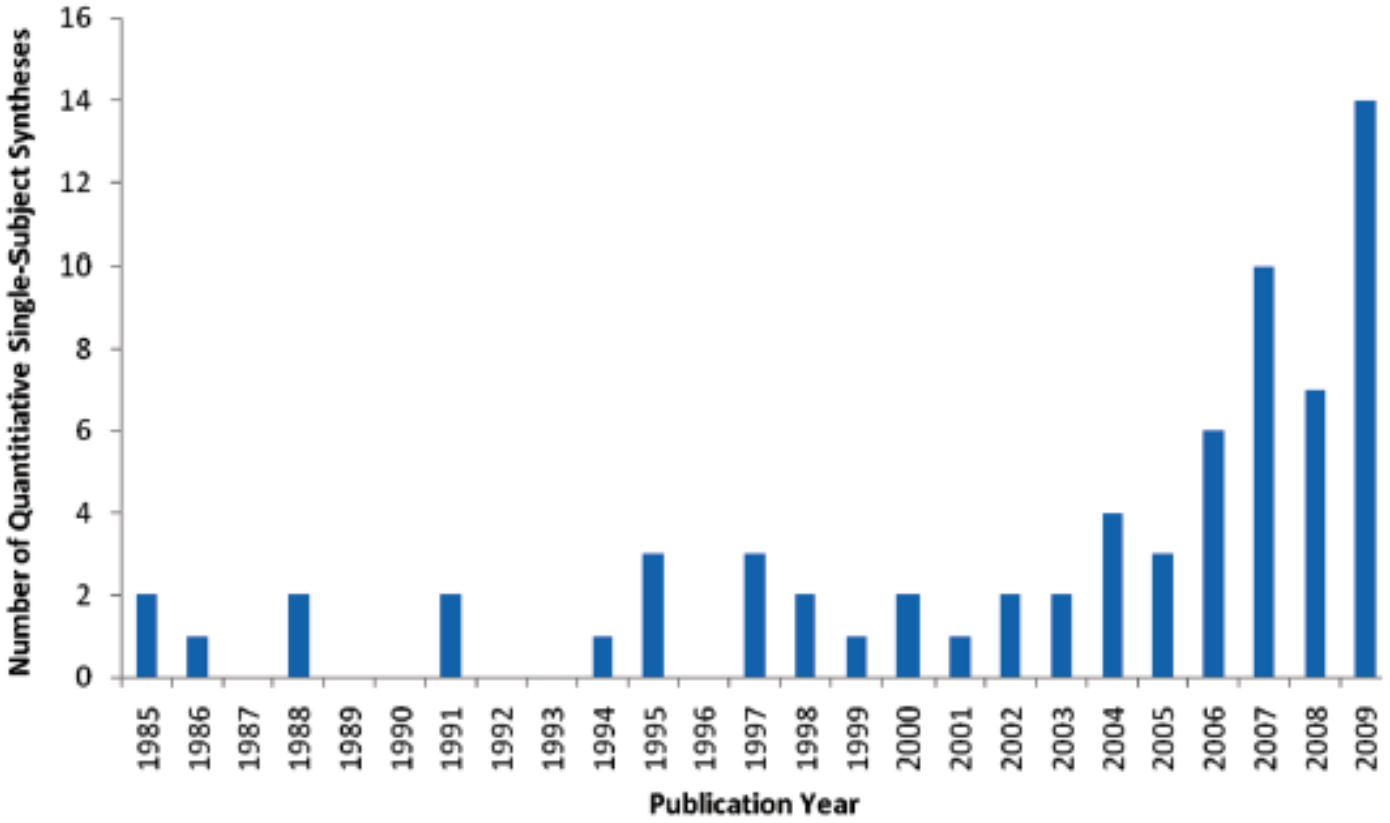


FIGURE 2 Annual frequency of quantitative syntheses using single-subject research that included students with disabilities published between 1985 and 2009 (color figure available online).

Source: Maggin, O'Keeffe, & Johnson (2011)

Effect sizes

- Quantitative measure of treatment effect magnitude & direction
- Basic inputs in a meta-analysis
- Many different families of effect sizes
- Should allow for fair comparisons across a set of studies to be synthesized.

Effect size desiderata (Lipsey & Wilson, 2001)

1. Interpretable measure of *magnitude* & *direction* of treatment effect
2. Comparable across cases & studies
3. Not influenced by arbitrary study design characteristics:
   - sample size
   - outcome measurement procedures
   - other study design features
4. Computable from available data
5. Accompanied by a measure of uncertainty (i.e., a standard error)

Procedures for direct observation of behavior

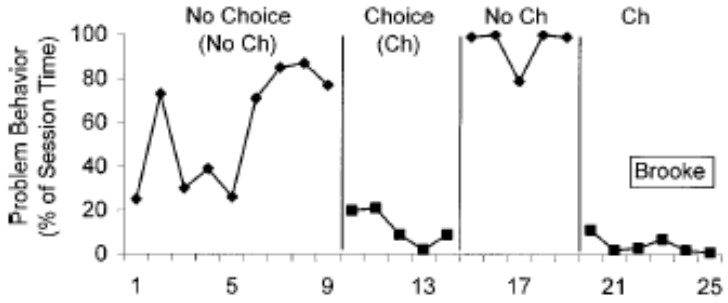
Recording procedure	% of Studies			
	Mudford et al. (n=168)	Laine & Ledford (n = 100)	Shadish & Sullivan (n = 68)	Shogren et al. (n = 32)
Event counting	52	55	60	9
Continuous recording	20		10	16
Interval recording	45	34	19	59
Momentary time sampling		11	7	3
Other			16	16

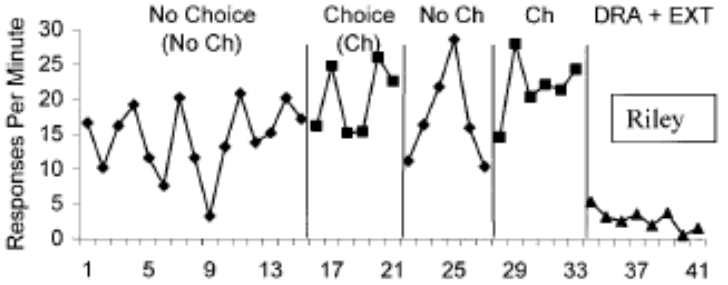
- Mudford et al. (2009) reviewed articles published 1995-2005 in JABA.
- Laine & Ledford (2014) reviewed articles published 2008-2012 in 4 journals that publish SCR on interventions for young children with disabilities.
- Shadish & Sullivan (2011) reviewed articles published in 2008 and reporting SCR.
- Shogren et al. (2008) is a research synthesis on the effects of providing choice-making opportunities on problem behavior of children with disabilities.

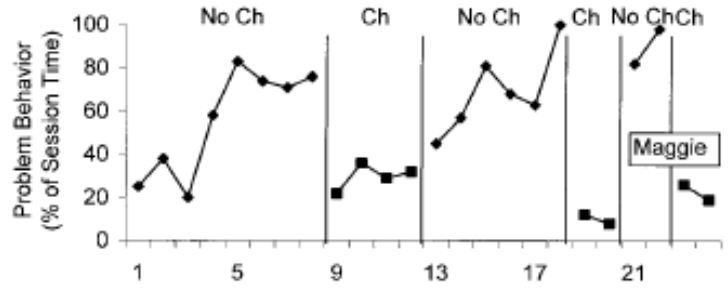
Effect sizes for single-case research

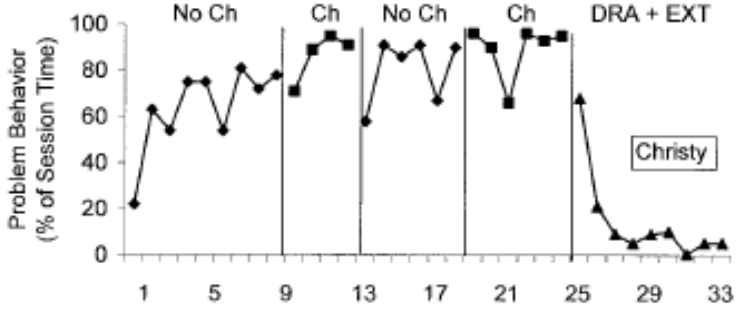
- Non-overlap measures
  - Percentage of non-overlapping data (PND; Scruggs et al., 1987)
  - Percentage exceeding the median (PEM; Ma, 2006)
  - Non-overlap of all pairs (Parker & Vannest, 2009)
- Within-case standardized mean differences (Busk & Serlin, 1992)
- Ratio/log-ratio measures (Pustejovsky, 2014)
- Design-comparable standardized mean differences (Hedges, Pustejovsky, & Shadish, 2012, 2013)

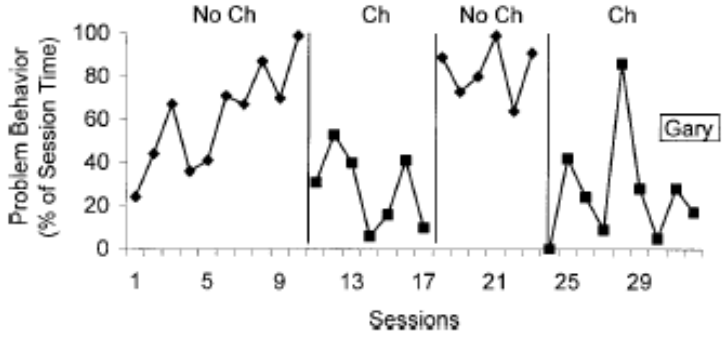
Romaniuk et al. (2002). The influence of activity choice on problem behaviors maintained by escape versus attention.

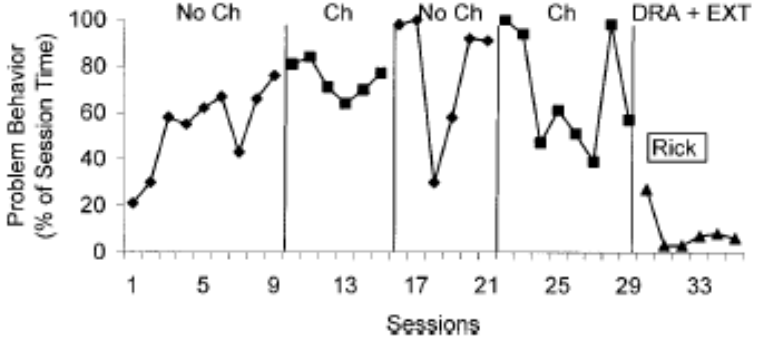












Percentage of Non-overlapping Data

- Most commonly applied effect size measure in synthesis of SCR for students with disabilities (Maggin et al., 2011)

- For "positive" behaviors:
  PND = % of observations in treatment condition that are larger than the maximum observation in baseline

- For "negative" behaviors:
  PND = % of observations in treatment condition that are smaller than the minimum observation in baseline

Romaniuk example

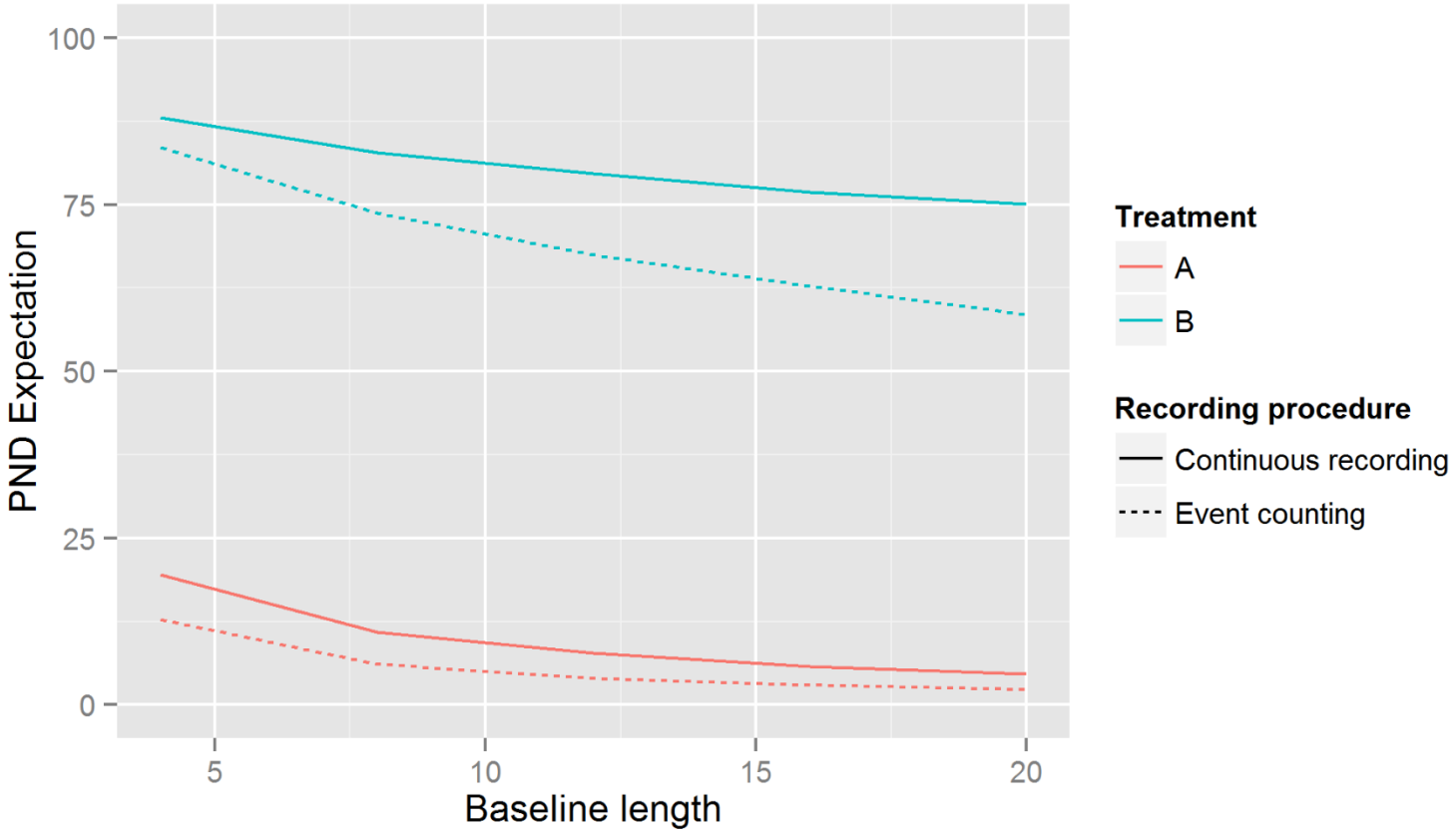
Case	Function	PND
Brooke	Escape	100%
Gary	Escape	50%
Maggie	Escape	38%
Christy	Attention	0%
Rick	Attention	0%
Riley	Attention	0%

Problems with PND

1. Does not capture direction of effect
2. No standard error
3. Magnitude depends on length of baseline

A simulated example

- Baseline behavior duration = 50%, frequency = 0.75/min
- Treatment A does nothing
- Treatment B reduces behavior to duration = 10%, frequency = 0.15/min



Within-case standardized mean difference

SMD measures differences in standard-deviation units:

$$d = \frac{\bar{y}_T - \bar{y}_B}{s_p}$$

where s_p is the pooled standard deviation, i.e., the square-root of the pooled variance

$$s_p^2 = \frac{(n_T - 1)s_T^2 + (n_B - 1)s_B^2}{n_T + n_B - 2}$$

Romaniuk example

Case	Function	PND	SMD (s.e.)
Brooke	Escape	100%	-2.95 (0.59)
Gary	Escape	50%	-1.95 (0.43)
Maggie	Escape	38%	-2.16 (0.54)
Christy	Attention	0%	1.12 (0.44)
Rick	Attention	0%	0.36 (0.37)
Riley	Attention	0%	1.03 (0.38)
Meta-analysis	Escape		-2.26 (0.29)
	Attention		0.81 (0.23)

Problems with SMD

- What if duration during baseline has mean = 0.5, SD = 0.3?

Response ratio

- Ratio measures are closely connected to % changes:

$$\text{Response Ratio} = \frac{\bar{y}_T}{\bar{y}_B}$$

- Log-transformation is used to make sampling distribution closer to normal:

$$lRR = \log\left(\text{Response Ratio}\right) = \log\left(\bar{y}_T\right) - \log\left(\bar{y}_B\right)$$

$$s.e. \approx \sqrt{\frac{s_T^2}{n_T \bar{y}_T^2} + \frac{s_B^2}{n_B \bar{y}_B^2}}$$

Response ratio

- A 95% confidence interval for the log-response ratio:

$$\left[lRR - 1.96 \times s.e., \quad lRR + 1.96 \times s.e.\right]$$

- A 95% confidence interval for % change:

$$100\% \times \left[\exp\left(lRR - 1.96 \times s.e.\right) - 1, \quad \exp\left(lRR + 1.96 \times s.e.\right) - 1\right]$$

Romaniuk example

Case	Function	PND	SMD (s.e.)	IRR (s.e.)
Brooke	Escape	100%	-2.95 (0.59)	-2.39 (0.37)
Gary	Escape	50%	-1.95 (0.43)	-0.96 (0.23)
Maggie	Escape	38%	-2.16 (0.54)	-1.09 (0.19)
Christy	Attention	0%	1.12 (0.44)	0.22 (0.08)
Rick	Attention	0%	0.36 (0.37)	0.12 (0.13)
Riley	Attention	0%	1.03 (0.38)	0.31 (0.10)
Meta-analysis	Escape		-2.26 (0.29)	-1.22 (0.13)
	Attention		0.81 (0.23)	0.23 (0.06)

- Escape: 66-77% reduction in problem behavior
- Attention: 13-40% increase in problem behavior

Response Ratio

- Can sometimes be used to make comparisons across recording procedures (Pustejovsky, 2014)

- Current methods don't handle
  - Serial dependence
  - Time trends
  - Floors/ceilings in the measurements
- …but PND and SMD have problems with these too.

- Interval recording procedures need special treatment

Challenges & data-quality issues

- Construct validity of interval recording data
- Study design procedures & internal validity
- Selective reporting

Interval recording

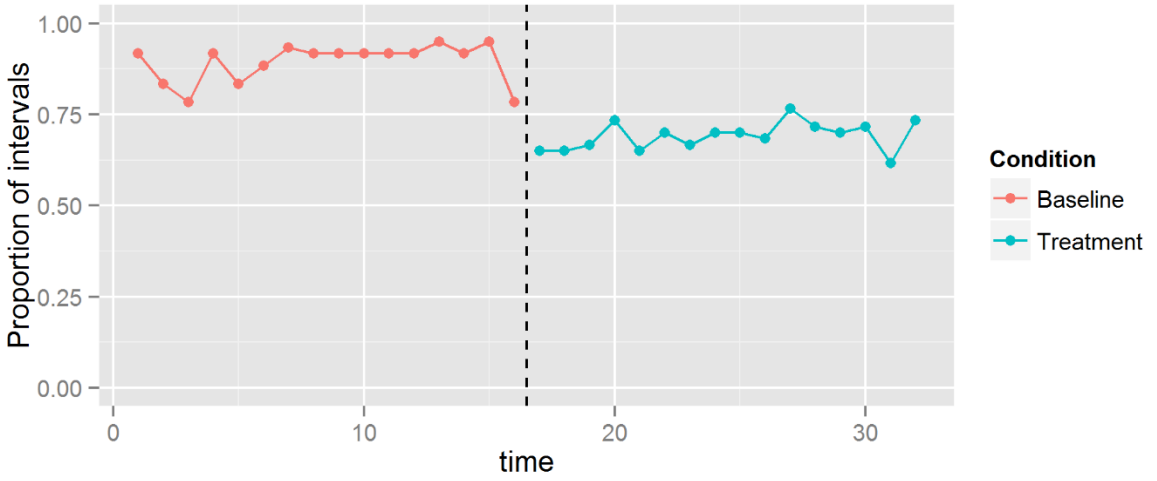
- Partial interval recording over-estimates % duration.
- Whole interval recording under-estimates % duration.

- Extent of systematic bias depends on
  - % duration
  - Frequency of the behavior
  - Length of intervals
  - Distribution of inter-event times

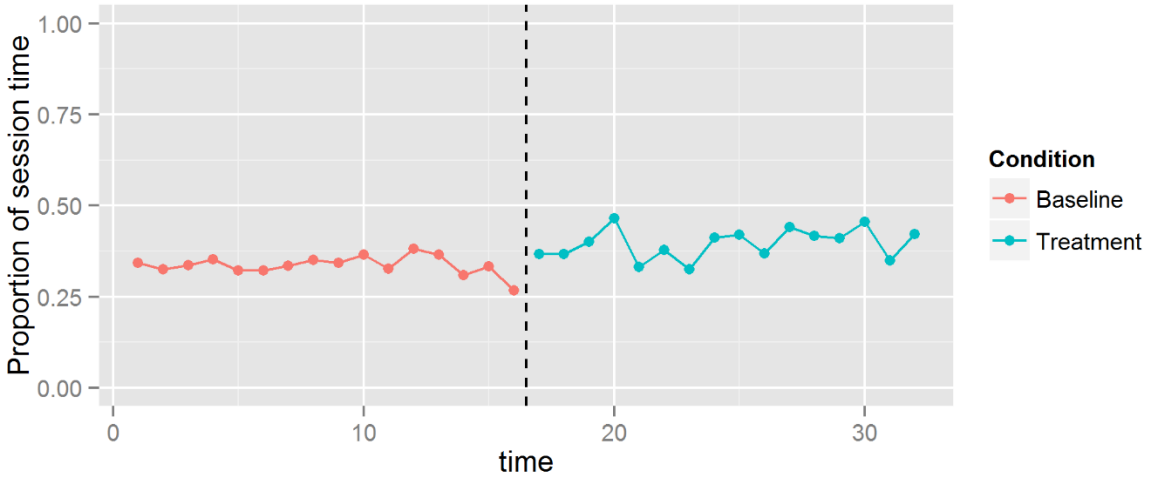
- Systematic bias can lead to systematically wrong inferences.

A simulated example of partial interval recording

Using PIR, it appears that prevalence decreases…



…when sample prevalence has instead increased slightly.

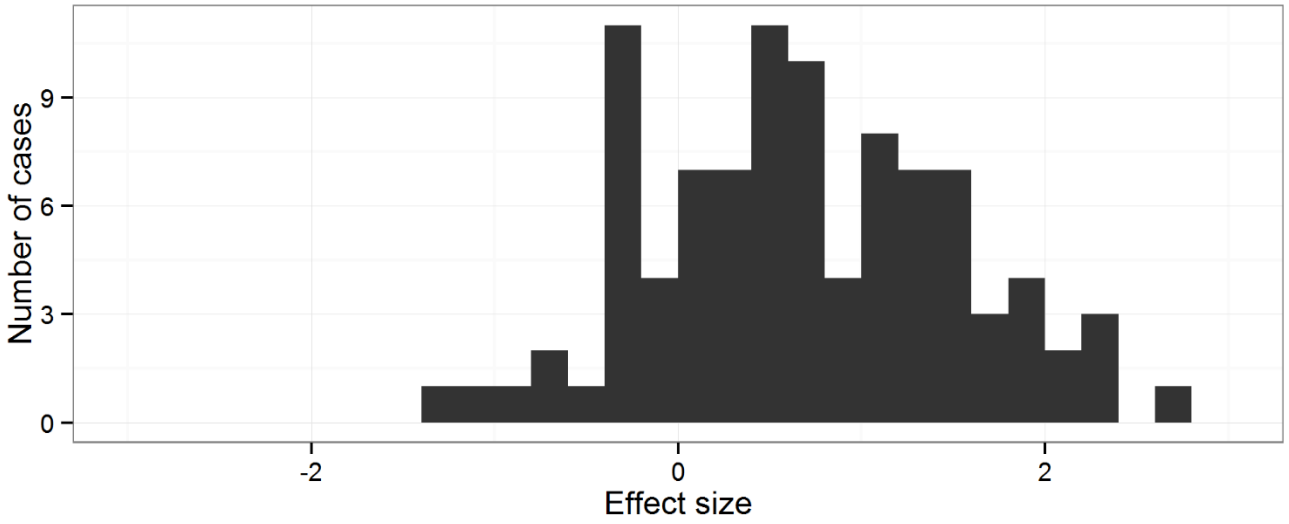


Interval recording

- If you are conducting a study…
  - DON'T USE INTERVAL RECORDING TO MEASURE BEHAVIOR.
  - Unless you already know a lot about the behavior.

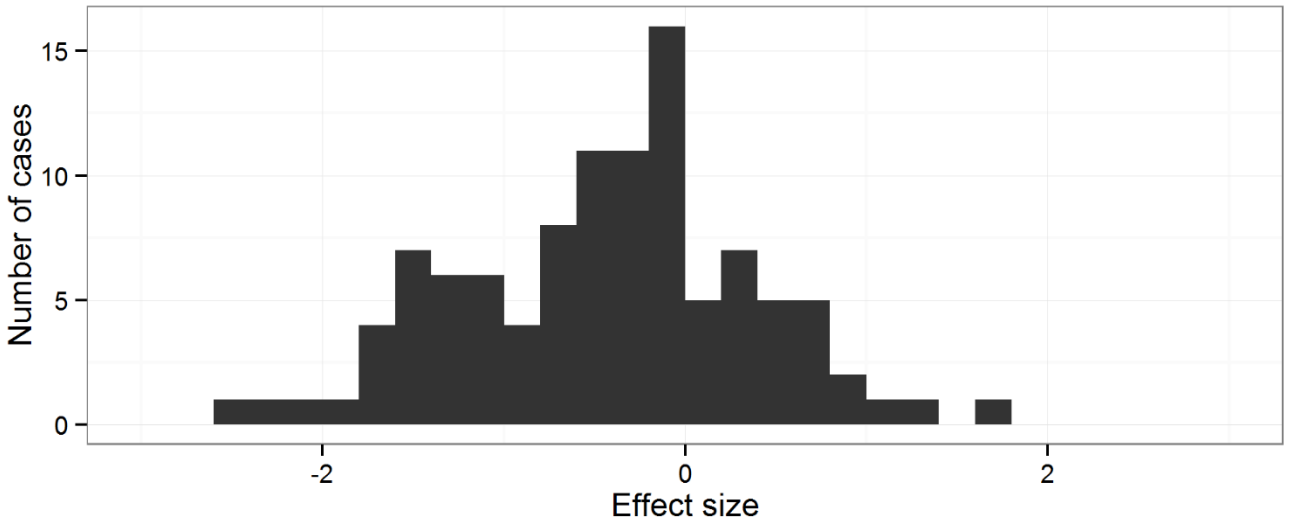
- If studies to be synthesized use interval recording…
  - Need specialized methods for estimating valid effect sizes
  - These require prior knowledge about the behavior.
  - More details: Pustejovsky & Swan (2014).

Selective reporting



Published studies
Mean ES: 0.68
78% of effects are > 0

All studies
Mean ES: 0.08
51% of effects are > 0



Unpublished studies
Mean ES: -0.46
25% of effects are > 0

Study design procedures & internal validity

Three procedures for conducting a multiple baseline study:

1. Randomly assign cases to treatment times.
2. Triage on known characteristics.
   - Suppose that the investigator knows how severe each case is before the study starts.
   - Assign worst case to first treatment time.
   - Assign best case to last treatment time.
3. Triage on measured baselines
   - Measure baseline outcomes on all cases until first treatment time.
   - Assign case with worst baseline outcomes to first treatment time.
   - Continue to measure outcomes.
   - Assign case with next-worst outcomes to second treatment time.

Study design procedures & internal validity

- Proper analysis depends on which procedure was used.

Design	Analytic model		
	A	B	C
Random	Right	Wrong	Wrong
Triage – known	Wrong	Right	Wrong
Triage - measured	Wrong	Wrong	Right

- Using the wrong method will lead to biased estimates of treatment effects.
- Descriptions of methods need to include details about how cases were assigned to treatment times.

References

- 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* (pp. 187–212). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Cooper, H. M., Hedges, L. V, & Valentine, J. C. (2009). *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation Publications.
- Hedges, L. V, Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3*, 224–239.
- Hedges, L. V, Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, *4*(4), 324–341. doi:10.1002/jrsm.1086
- Lane, J. D., & Ledford, J. R. (2014). Using Interval-Based Systems to Measure Behavior in Early Childhood Special Education and Early Intervention. *Topics in Early Childhood Special Education*.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49). Sage Publications, Inc.
- Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification*, *30*(5), 598.
- Maggin, D. M., O'Keeffe, B. V, & Johnson, A. H. (2011). A quantitative synthesis of methodology in the meta-analysis of single-subject research for students with disabilities: 1985-2009. *Exceptionality*, *19*(2), 109–135.
- Mudford, O. C., Taylor, S. A., & Martin, N. T. (2009). Continuous recording and interobserver agreement algorithms reported in the Journal of Applied Behavior Analysis (1995-2005). *Journal of Applied Behavior Analysis*, *42*(1), 165–169.
- Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, *40*(4), 357–67.
- Pustejovsky, J. E. (2014). Measurement-comparable effect sizes for single-case studies of free operant behavior. *Psychological Methods*. doi:10.1037/met0000019
- Pustejovsky, J. E., & Swan, D. M. (2014). *Four methods for analyzing partial interval recording data, with application to single-case research*. Paper presented at the annual meeting of the American Educational Research Association, Philadelphia, PA.
- Romaniuk, C., Miltenberger, R., Conyers, C., Jenner, N., Jurgens, M., & Ringenberg, C. (2002). The influence of activity choice on problem behaviors maintained by escape versus attention. *Journal of Applied Behavior Analysis*, *35*(4), 349–62.
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research. *Remedial and Special Education*, *8*(2), 24–43.
- Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, *43*(4), 971–980.
- Shogren, K. A., Faggella-Luby, M. N., Bae, S. J., & Wehmeyer, M. L. (2004). The effect of choice-making as an intervention for problem behavior. *Journal of Positive Behavior Interventions*, *6*(4), 228–237.