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

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

- Mudford et al. (2009) reviewed articles published 1995-2005 in JABA.
- 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:
  \[
  \text{PND} = \% \text{ of observations in treatment condition that are larger than the maximum observation in baseline}
  \]

- For “negative” behaviors:
  \[
  \text{PND} = \% \text{ of observations in treatment condition that are smaller than the minimum observation in baseline}
  \]
Problems with PND
1. Does not capture direction of effect
2. No standard error
3. Magnitude depends on length of baseline

<table>
<thead>
<tr>
<th>Case</th>
<th>Function</th>
<th>PND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brooke</td>
<td>Escape</td>
<td>100%</td>
</tr>
<tr>
<td>Gary</td>
<td>Escape</td>
<td>50%</td>
</tr>
<tr>
<td>Maggie</td>
<td>Escape</td>
<td>38%</td>
</tr>
<tr>
<td>Christy</td>
<td>Attention</td>
<td>0%</td>
</tr>
<tr>
<td>Rick</td>
<td>Attention</td>
<td>0%</td>
</tr>
<tr>
<td>Riley</td>
<td>Attention</td>
<td>0%</td>
</tr>
</tbody>
</table>
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

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

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:

\[
lRR = \log(\text{Response Ratio}) = \log(\bar{y}_T) - \log(\bar{y}_B)
\]

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

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

\[
[lRR - 1.96 \times s.e., \ lRR + 1.96 \times s.e.]
\]

- A 95% confidence interval for % change:

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

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

- **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.
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.

<table>
<thead>
<tr>
<th>Design</th>
<th>Analytic model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Random</td>
<td>Right</td>
</tr>
<tr>
<td>Triage - known</td>
<td>Wrong</td>
</tr>
<tr>
<td>Triage - measured</td>
<td>Wrong</td>
</tr>
</tbody>
</table>

- 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