Design-Comparable Effect Size for Single-Case Design Studies

Mariola Moeyaert, Ph.D.
Associate Professor
The State University of New York at Albany

Daniel M. Swan, Ph.D.
Postdoctoral Scholar
University of Oregon

Emily Tanner-Smith, Ph.D.
Thomson Professor, University of Oregon
What Works Clearinghouse
Presenters

Mariola Moeyaert, Ph.D.
The State University of New York at Albany

Daniel M. Swan, Ph.D.
University of Oregon
Learning goals for this webinar

• After this webinar, you will be able to
  – Understand the difference in quantifying intervention effectiveness between Version 4.0 and Version 4.1 of the What Works Clearinghouse (WWC) Standards.
  – Quantify intervention effectiveness in single-case design (SCD) studies using the design-comparable effect size (D-CES).
  – Understand the logic behind the D-CES and its underlying assumptions.
  – Understand the advantages and disadvantages of the D-CES.
  – Use a point-and-click shiny application to calculate the D-CES.
Part 1: Introduction to SCDs and D-CES

1. Introduction to SCD types
2. Approaches to quantifying intervention effects
3. Examining evidence for intervention effects:
   What Works Clearinghouse Version 4.0 versus Version 4.1
4. Introduction to D-CES
1. Introduction to SCD types

- SCDs are experiments in which one unit is observed repeatedly during a certain period of time under different levels of at least one manipulated variable.
- SCDs can demonstrate a causal effect.
- SCDs involve repeated, systematic measurement of a dependent variable before, during, and after the active manipulation of an independent variable.
1. Introduction to SCD types

D-CES can be calculated

- Reversal/withdrawal: ≥ 4 phases with ≥ 5 points
- Multiple baseline: ≥ 6 phases with ≥ 5 points
- Alternating treatment: ≥ 5 points per condition with ≤ 2 points per phase

- Reversal/withdrawal: ≥ 4 phases with ≥ 3 points
- Multiple baseline: ≥ 6 phases with ≥ 3 points
- Alternating treatment: ≥ 4 points per condition with ≤ 2 points per phase

- Reversal/withdrawal: ≤ 3 phases or ≤ 2 points
- Multiple baseline: ≤ 5 phases or ≤ 2 points
- Alternating treatment: ≤ 3 points per condition or > 2 points per phase

Meets WWC SCD Standards Without Reservations
Meets WWC SCD Standards With Reservations
Does Not Meet WWC SCD Standards

1. Introduction to SCD types: Treatment reversal (AB\textsuperscript{k})

- Outcome: Percentage of appropriate social skills
- Intervention: Peer tutor training program
- Sample: Children with autism
- Three potential demonstrations for intervention effectiveness
- Data points per phase:
  - Baseline 1 and Intervention 1 $\rightarrow$ five data points
  - Baseline 2 and Intervention 2 $\rightarrow$ four data points

1. Introduction to SCD types: Multiple baseline design (MB)

- Outcome: Quality of writing points
- Intervention: Self-regulated strategy development
- Sample: Three students with learning disorders
- Data points per phase:
  - Baseline → three/five data points
  - Intervention → three data points
- A least three potential demonstrations for intervention effectiveness

2. Approaches to quantifying intervention effects

2.1 Within-case quantification
- Non-overlap measures.
- Log response ratio.
- Within-case standardized mean difference.
- Regression-based measures.

2.2 Across-case quantification
- Mean/median/range of the within-case quantification.
- Hierarchical linear modeling.
- Design-comparable effect size.
3. Examining evidence for intervention effects

WWC Version 4.0

- Intervention effects were synthesized if studies met the 5-3-20 rule:
  - At least five studies met WWC SCD standards with or without reservations, the studies were conducted by at least three different research teams with no overlapping authorship at three different institutions, and the combined number of cases was at least 20.
- Review teams tallied and computed the proportion of SCD intervention effects that were positive or negative.
- SCD effects were not synthesized with effects from group design studies.
3. Examining evidence for intervention effects

**WWC Version 4.1**

- Intervention effects are synthesized using the D-CES.
- The D-CES can be synthesized with effects from group-design studies.
- The D-CES are combined across studies using a fixed-effects meta-analysis. Fixed-effects meta-analysis involves computing a weighted average effect size. Studies are weighted by the inverse of the sampling variance of their effect sizes.
4. Introduction to D-CES

- D-CES estimates the same parameter from what may be different designs (Pustejovsky, Hedges, & Shadish, 2014).
- Study requirements for D-CES:
  - Design: Treatment reversal ($AB^k$), and multiple baseline (MB)/multiple probe
- The outcome is measured on a continuous scale that is common across cases.
4. Introduction to D-CES

- Sample size: Three or more cases
- Use of hierarchical linear modeling
- Estimation procedure: Moment estimation techniques (restrictive assumptions) or restricted maximum likelihood estimation (more general and flexible)
4. D-CES general framework

D-CES can be estimated using two-level hierarchical modeling:

\[ Y_{it} = \beta_{0i} + \beta_{1i} \text{Intervention}_{it} + e_{it} \]

\( \beta_{1i} \) indicates the individual-specific intervention effect.

\[ \beta_{0i} = \theta_0 + u_i \]

\( \beta_{1i} = \theta_1 \)

\( \theta_1 \) indicates the unstandardized intervention effect across the \( i \) cases.

\[ \delta = \frac{\theta_1}{\sqrt{\sigma^2 + \tau^2}} \]

Numerator: Unstandardized intervention effect
Denominator: The standard deviation of the outcome, including both within- and between-case variation

- \( i \) indicates the case (\( i = 1 \) to \( I \)), and case i is measured for a total of \( n_m \) measurement occasions (\( t = 1, \ldots, n_m \)).
- \( Y_{it} \) indicates the outcome for case i at measurement occasion t.
- \( \text{Intervention}_{it} \) is a dummy variable indicating whether \( Y_{it} \) is obtained during the baseline or the intervention phase.
- \( e_{it} \sim N (0, \sigma^2) \), and the errors for case i follow an AR(1) process; \( u_i \sim N (0, \tau^2) \).
4. D-CES general framework

- Basic D-CES assume changes in level and constant intervention effects across cases.

- More general/complex D-CES
  - Allow intervention effect to vary across cases.
  - Can include linear/polynomial time trends.
4. D-CES general framework

- If the individual-level model includes time trends for the baseline or intervention phase, one must make assumptions about how the time trends vary across cases:

  \[ Y_{it} = \beta_0i + \beta_1i \times Time_{it} + \beta_2i Intervention_{it} + (Time'_{it} \times \beta_3i Intervention_{it}) + e_{it} \]

- If time trends in either phase vary across cases, then the total variation in the outcome is not constant \( \rightarrow \) consequences for the denominator of the D-CES.
  - See Pustejovsky and colleagues (2014).
4. D-CES general framework

- D-CES for both $AB^k$ and MB can be corrected for small-sample-size bias,

$$J(v) = 1 - \frac{3}{4v - 1}$$

Bias-corrected effect size:

$$g = J(v) \times \hat{\delta}$$

where $v$ is an estimated degrees of freedom (this will be somewhere between the number of cases and the total number of time points; computation of $v$ is different for $AB^k$ and MB).
Part 2: Appropriate use and application of D-CES

1. Appropriately using D-CES
2. Using scdhlm
Appropriately using D-CES

• The D-CES is appropriate only for designs in which comparisons can be made across participants.
  – It is inappropriate for designs with multiple baselines across contexts or behaviors.
• The outcomes used to estimate a single D-CES should all be the same measure.
Appropriately using D-CES

- Expanding the hierarchical model
  - When there are only a few observations per phase, apparent trends might be variability.
  - When there are only a few cases, additional model parameters are difficult to estimate.
Using scdhlm

• The R package scdhlm was written by James E. Pustejovsky (Pustejovsky, 2016) for estimating the D-CES, which he also calls the “between-case standardized mean difference.”
  – App website: https://jepusto.shinyapps.io/scdhlm/

• The app can be run locally on your computer by installing R and installing the scdhlm package.

• It is also possible to estimate the effect size using R code instead of the app, if you are familiar with using R.
Data organization

For the scdhlm app, data need to be organized in a “long” format, with columns including the following, at a minimum:

- The case identifier.
- The phase identifier.
- The observation session number (whole numbers are best).
- The outcome.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Participant Name or ID</td>
<td></td>
<td>Phase Name</td>
<td>Outcome Value (y-axis)</td>
</tr>
<tr>
<td>2</td>
<td>Gizelle</td>
<td></td>
<td>Baseline</td>
<td>11.9920291</td>
</tr>
<tr>
<td>3</td>
<td>Gizelle</td>
<td></td>
<td>Baseline</td>
<td>20.00336655</td>
</tr>
<tr>
<td>4</td>
<td>Gizelle</td>
<td></td>
<td>Baseline</td>
<td>9.936649906</td>
</tr>
<tr>
<td>5</td>
<td>Gizelle</td>
<td></td>
<td>Baseline</td>
<td>21.98753389</td>
</tr>
<tr>
<td>6</td>
<td>Gizelle</td>
<td></td>
<td>Intervention</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>Gizelle</td>
<td></td>
<td>Intervention</td>
<td>100.1008411</td>
</tr>
<tr>
<td>8</td>
<td>Gizelle</td>
<td></td>
<td>Intervention</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>Gizelle</td>
<td></td>
<td>Intervention</td>
<td>100.1457006</td>
</tr>
<tr>
<td>10</td>
<td>Gizelle</td>
<td></td>
<td>Intervention</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>Gizelle</td>
<td></td>
<td>Intervention</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Task Engagement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
About scdhlm

Between-case standardized mean difference estimator

scdhlm
Version 0.4.2.9000
Designed and maintained by James E. Pustejovsky
- pustejovsky@wisc.edu
- https://jepusto.com
Contributions from
- Bethany Hamilton
- Man Chen

Source code available on Github
Your comments, suggestions, and feedback are welcome.

Suggested citation

Tutorial paper
The Campbell Collaboration. DOI: 10.4073/cmpn.2016.3
Load data

• Select the “Load” tab along the top.

• If you are going to estimate the effect size from the Study Review Guide, select the “Upload data from a .xlsx file” radio button.

• Browse to the Study Review Guide.

• Leave “File has header?” selected.

• Select the “Data” sheet.
Specify variables

1. Please specify the study design.
   - Treatment Reversal

2. Please select the variable containing each type of information.
   - **Case identifier**
     - Participant Name or ID
   - **Phase identifier**
     - Phase Name
   - **Session number**
     - Session Number (x-axis)
   - **Outcome variable**
     - Outcome Value (y-axis)

3. Please specify the baseline and treatment levels.
   - **Baseline level**
     - Baseline
   - **Treatment level**
     - Intervention

- Select “Treatment Reversal” under study design. If using a multiple baseline or multiple probe design, select “Multiple Baseline.”

- Select the variables that correspond to the case (“Participant Name or ID”), the phase identifier (“Phase Name”), the session number (“Session Number (x-axis)”), and the outcome variable (“Outcome Value (y-axis)”).

- Select the values from the “Phase Name” variable that correspond to the baseline and treatment levels.
Inspect the data

• The app will replot the data on the “Inspect” tab.

• Take the time to ensure that the replotted data look similar to the plots from which you extracted them.
Inspect the data

- The four values you specified for the case, session number, phase identifier, and outcome variable are found in the first four columns.

- The last two columns are values that the app will use internally to estimate the multilevel model from which the app extracts parameters to estimate the D-CES.

<table>
<thead>
<tr>
<th>case</th>
<th>session</th>
<th>phase</th>
<th>outcome</th>
<th>trt</th>
<th>phase_pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gizelle</td>
<td>1</td>
<td>Baseline</td>
<td>11.99</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Gizelle</td>
<td>2</td>
<td>Baseline</td>
<td>20.03</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Gizelle</td>
<td>3</td>
<td>Baseline</td>
<td>9.94</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Gizelle</td>
<td>4</td>
<td>Baseline</td>
<td>21.99</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Gizelle</td>
<td>5</td>
<td>Intervention</td>
<td>100.10</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Gizelle</td>
<td>6</td>
<td>Intervention</td>
<td>100.12</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Gizelle</td>
<td>7</td>
<td>Intervention</td>
<td>100.15</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Gizelle</td>
<td>8</td>
<td>Intervention</td>
<td>100.17</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Gizelle</td>
<td>9</td>
<td>Intervention</td>
<td>100.19</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Gizelle</td>
<td>10</td>
<td>Baseline</td>
<td>7.04</td>
<td>0.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Gizelle</td>
<td>11</td>
<td>Baseline</td>
<td>10.12</td>
<td>0.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Gizelle</td>
<td>12</td>
<td>Baseline</td>
<td>39.92</td>
<td>0.00</td>
<td>2.00</td>
</tr>
</tbody>
</table>
Estimate the model

The WWC’s default specification is to estimate a fixed effect and a random effect for the baseline phase, and a fixed effect for the treatment phase.
Record the design-comparable effect size

The “Effect size” tab contains the important information you need to record.
Record the design-comparable effect size

The “Effect size” tab contains the important information you need to record.

<table>
<thead>
<tr>
<th>Design Comparable Effect Size Estimate</th>
<th>Change Sign of Effect for Meta-Analysis?</th>
<th>Standard Error</th>
<th>Autocorrelation Estimate</th>
<th>Note any deviations from default modeling guidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.99</td>
<td>No</td>
<td>1.22</td>
<td>-0.45</td>
<td>None</td>
</tr>
</tbody>
</table>
The design-comparable effect size infographic

Individual Review

Is the experiment a design for which there is an effect size*, AND that includes three or more individuals? Yes  No

Stop! Recording or extracting study data is not necessary. Do not estimate an effect size.

The study provides data in a tabular format? Yes  No

Send an author query.

Record data in Data tab of the study review guide.

Study author provides data in a tabular format? Yes  No

Estimate an effect size for each eligible experiment and record the estimates in the study review guide.

Save a copy of each plot for which an effect size will be estimated to the Plots tab of the study review guide.

Carefully extract data from plots using software.†

*At the release of the What Works Clearinghouse Standards and Procedures Handbooks, version 4.1, the only designs for which there was an effect size were multiple baseline across participants, multiple probe across participants, and treatment reversal designs. Future research may allow effect sizes to be estimated for a wider range of designs.

†Check the extracted data to ensure that the values are close to what is shown on the plot. Some measurement error is to be expected. However, values that appear very different** from what is observed on the plot or that are outside the minimum or maximum of the y-axis may indicate a need to recalibrate the axis reference points.
The design-comparable effect size infographic

Reconciling Reviews

1. Were the data extracted from plots?
   - Yes: Check reliability of extracted outcomes across reviewers.
   - No: Are the effect size estimates identical?
     - Yes: Direct the reviewer with obvious errors to independently extract the outcomes again and correct the study review guide.
     - No: Direct both reviewers to independently extract the outcomes again and correct the study review guide.

2. Were outcomes reliably extracted?
   - Yes: Comparing the extracted values from both reviewers with the plots, is one obviously wrong?**
     - Yes: Record the reliability estimate, type or reliability estimate used, and other data notes in the appropriate columns on the Review tab.
     - No: Direct both reviewers to independently extract the outcomes again and correct the study review guide.
   - No: Check for transcription errors or deviations from modeling assumptions. Instruct reviewers to correct errors.

3. Ensure that a copy of the Data and Plots tabs for both reviewers is saved and labeled in the final study review guide.

**Making this judgement includes considering the scale of measurement. Obviously, wrong values might include (but are not limited to) values outside the scale of measurement, such as proportions/percentages well above 100% or proportions/percentages/counts well below 0.
Questions?
Have questions? Contact us: https://ies.ed.gov/ncee/wwc/help

Mariola Moeyaert, Ph.D.
The State University of New York at Albany

Daniel M. Swan, Ph.D.
University of Oregon

Emily Tanner-Smith, Ph.D.
University of Oregon and What Works Clearinghouse
References and further reading


References and further reading

