More Power to You: Calculating Your Effect Size

Kelly Pelzel & Beth Troutman, University of Iowa
with assistance from
Allison Momany, University of Iowa
& John Paul Abner, Milligan College

UC Davis 18th Annual Conference on PCIT
for Traumatized Children
September 17, 2018
• What is effect size and how does it relate to statistical significance?
• Why calculate effect size with your own PCIT data?
• How do you handle data from families who drop out?
• Step-by-step instructions on how to calculate effect size (with example).
• Why interpret PCIT effect sizes with caution?
• Time to calculate your effect size (if you brought deidentified data).
What is Effect Size (Hartmann et al., 2015)

• Effect size is an estimate of the magnitude of a difference

• It is used to convey the importance or strength of result.

• Whether or not a finding is statistically significant depends on both the effect size and the sample size.

• If you have an estimate of effect size, you can estimate how many participants you would likely need to get a statistically significant result. This is called a power analysis.
Effect Size Examples (Cohen’s $d$)

- Effect sizes are generally categorized as trivial ($d < .2$), small ($d = .2$ to $.5$), medium ($d = .5$ to $.8$), or large ($d > .8$) (Cohen, 1992)
- Even trivial effect sizes can have an impact at pop. level (e.g., baby aspirin)
• Gardner’s Effect Size Illustrator:

http://esi.medicine.dal.ca/effect-size-illustrator.html#app=fb33&1a02-selectedIndex=0
• 4 multiple choice items

• Illustration displayed represents the effect size of one of the three studies listed (ECBI intensity scores)

• Studies were included in the Thomas et al., 2016 meta-analysis

• Blue curve received PCIT

• Grading on the honor system

• Candy for winners!
a. Eyberg et al., 1995
(-1.23)

b. McCabe et al., 2009
(-0.67)

c. Mersky et al., 2016
(-0.27)
a. Danko, 2015  
(-0.25)

b. Webb et al., 2016  
(-0.13)

c. Solomon et al., 2008  
(-0.32)
a. Thomas & Zimmer-Gembeck, 2011 (-0.28)

b. Chaffin et al., 2004 (0.22)

c. Thomas & Zimmer-Gembeck, 2012 (-0.25)
a. Bagner et al., 2010 (-2.72)
b. McNeil et al., 1999 (-2.65)
c. Bagner & Eyberg, 2007 (-1.43)
### Table: Study or Subgroup Analysis

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>PCIT Mean</th>
<th>PCIT SD</th>
<th>PCIT Total</th>
<th>Control Mean</th>
<th>Control SD</th>
<th>Control Total</th>
<th>Weight</th>
<th>IV, Random, 95% CI</th>
<th>SMD</th>
<th>IV, Random, 95% CI</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abraham et al, 2015</td>
<td>104.65</td>
<td>39.92</td>
<td>17</td>
<td>124.8</td>
<td>34.46</td>
<td>25</td>
<td>4.7</td>
<td>-0.54 (-1.17 to 0.09)</td>
<td>-0.54</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Bagner and Eyberg, 2007</td>
<td>100.63</td>
<td>26.22</td>
<td>10</td>
<td>143.14</td>
<td>30.33</td>
<td>12</td>
<td>3.7</td>
<td>-1.43 (-2.39 to -0.47)</td>
<td>-1.43</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Bagner et al, 2010</td>
<td>43</td>
<td>4.3</td>
<td>11</td>
<td>64.6</td>
<td>9.5</td>
<td>14</td>
<td>3.2</td>
<td>-2.72 (-3.86 to -1.58)</td>
<td>-2.72</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Bretan et al, 1997</td>
<td>133</td>
<td>37.7</td>
<td>16</td>
<td>170</td>
<td>36</td>
<td>13</td>
<td>4.2</td>
<td>-0.97 (-1.75 to -0.19)</td>
<td>-0.97</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Chaffin et al, 2004</td>
<td>57.14</td>
<td>3.0948</td>
<td>75</td>
<td>56.4</td>
<td>4</td>
<td>35</td>
<td>5.4</td>
<td>0.22 (-0.18 to 0.62)</td>
<td>0.22</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Danko, 2015</td>
<td>117</td>
<td>27.86</td>
<td>7</td>
<td>125.71</td>
<td>36.8</td>
<td>7</td>
<td>3.4</td>
<td>-0.25 (-1.30 to 0.80)</td>
<td>-0.25</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Eyerberg et al, 1995</td>
<td>120.4</td>
<td>18.8</td>
<td>10</td>
<td>172.4</td>
<td>62</td>
<td>6</td>
<td>3.2</td>
<td>-1.23 (-2.35 to -0.10)</td>
<td>-1.23</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Foley, 2011</td>
<td>89.65</td>
<td>46.04</td>
<td>19</td>
<td>92.08</td>
<td>42.73</td>
<td>24</td>
<td>4.8</td>
<td>-0.05 (-0.66 to 0.55)</td>
<td>0.05</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Leung et al, 2009</td>
<td>102.21</td>
<td>26</td>
<td>48</td>
<td>140.19</td>
<td>22.17</td>
<td>62</td>
<td>5.3</td>
<td>-1.58 (-2.01 to -1.14)</td>
<td>-1.58</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Leung et al, 2015</td>
<td>111.65</td>
<td>26.37</td>
<td>54</td>
<td>152.99</td>
<td>32.26</td>
<td>57</td>
<td>5.3</td>
<td>-1.39 (-1.81 to -0.97)</td>
<td>-1.39</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Matos et al, 2009</td>
<td>51.52</td>
<td>10.51</td>
<td>20</td>
<td>68.36</td>
<td>9.74</td>
<td>12</td>
<td>4.1</td>
<td>-1.60 (-2.43 to -0.77)</td>
<td>-1.60</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>McCabe et al, 2009</td>
<td>89.5915</td>
<td>39.7685</td>
<td>40</td>
<td>118.5</td>
<td>48.34</td>
<td>18</td>
<td>4.9</td>
<td>-0.67 (-1.24 to -0.10)</td>
<td>-0.67</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>McNeil et al, 1999</td>
<td>105.5</td>
<td>26.55</td>
<td>18</td>
<td>176.79</td>
<td>25.9</td>
<td>14</td>
<td>3.6</td>
<td>-2.65 (-3.63 to -1.66)</td>
<td>-2.65</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Mersky et al, 2016</td>
<td>123.8483</td>
<td>37.4731</td>
<td>58</td>
<td>134</td>
<td>36.65</td>
<td>33</td>
<td>5.3</td>
<td>-0.27 (-0.70 to 0.16)</td>
<td>0.27</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Nixon et al, 2003</td>
<td>125.9751</td>
<td>19.6893</td>
<td>37</td>
<td>148.35</td>
<td>19.05</td>
<td>17</td>
<td>4.7</td>
<td>-1.13 (-1.75 to -0.52)</td>
<td>-1.13</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Querido, 2004</td>
<td>10.2</td>
<td>3.27</td>
<td>5</td>
<td>21</td>
<td>10.03</td>
<td>6</td>
<td>2.7</td>
<td>-1.27 (-2.63 to 0.09)</td>
<td>-1.27</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Schuhmann et al, 1998</td>
<td>117.6</td>
<td>40.4</td>
<td>22</td>
<td>169.7</td>
<td>34.1</td>
<td>20</td>
<td>4.6</td>
<td>-1.36 (-2.04 to -0.68)</td>
<td>-1.36</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Solomon et al, 2008</td>
<td>59.7</td>
<td>4.95</td>
<td>10</td>
<td>62.22</td>
<td>9.77</td>
<td>9</td>
<td>3.8</td>
<td>-0.32 (-1.22 to 0.59)</td>
<td>0.32</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Stokes, 2015</td>
<td>98.67</td>
<td>25.86</td>
<td>6</td>
<td>150.5</td>
<td>53.4</td>
<td>10</td>
<td>3.3</td>
<td>-1.08 (-2.18 to 0.02)</td>
<td>1.08</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Terao, 1999</td>
<td>100.41</td>
<td>36.16</td>
<td>17</td>
<td>127.65</td>
<td>37.87</td>
<td>17</td>
<td>4.5</td>
<td>-0.72 (-1.41 to -0.02)</td>
<td>0.72</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Thomas and Zimmer-Gembeck, 2011</td>
<td>139.1</td>
<td>35.4</td>
<td>42</td>
<td>148.9</td>
<td>33.4</td>
<td>34</td>
<td>5.2</td>
<td>-0.28 (-0.74 to 0.17)</td>
<td>0.28</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Thomas and Zimmer-Gembeck, 2012</td>
<td>133.7</td>
<td>38.1</td>
<td>60</td>
<td>143.1</td>
<td>36.7</td>
<td>40</td>
<td>5.4</td>
<td>-0.25 (-0.65 to 0.15)</td>
<td>0.25</td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

**Total (95% CI):**

- **SMD:** 0.87 (-1.17 to 0.58)

**Risk of Bias:**

- **A:** ✦
- **B:** ✦
- **C:** ✦
Why Calculate Effect Size?

• Compare your effectiveness to the benchmark.

• Compare the effectiveness of a modification to the effectiveness of the standard approach.

• Compare effectiveness with a specific population.
What about the Non-Completers?

How do you handle data from families who drop out?

- Leave them out
- Include them ("intent to treat with last observation carried forward")
- Include some of them (e.g., everyone who had at least two sessions, everyone who had at least two ECBIs, everyone who at least made it through CDI)
One sample, all given same outcome measure

1. To calculate effect size, you will need:
   a. Pre-treatment outcome measure mean
   b. Pre-treatment outcome measure SD
   c. “Post”-treatment outcome measure mean
   d. “Post”-treatment outcome measure SD
   e. Correlation between pre-tx & “post”-tx scores

2. Normally distributed scores

*all of the above can be calculated in Excel*

3. A effect size calculator that can handle repeated measures data:

https://www.psychometrica.de/effect_size.html
Why interpret PCIT effect sizes with caution?

- ECBI scores are part of graduation criteria and are also an outcome measure.
  (You might also tend to exclude individuals with low pre-treatment ECBI scores from PCIT.)

- Small samples can have non-normal distributions.

- Missing data

- Pre/post EBCI scores don’t tell the whole story.
IoWA-PCIT
Integration of Working Models of Attachment into Parent Child Interaction Therapy.

Kelly Pelzel, Ph.D.
Clinical Assistant Professor
Department of Psychiatry
Carver College of Medicine
University of Iowa
Iowa City, Iowa
kelly-pelzel@uiowa.edu

Beth Troutman, Ph.D.
Clinical Professor
Department of Psychiatry
Carver College of Medicine
University of Iowa
Iowa City, Iowa
beth-troutman@uiowa.edu