Responsiveness and Validity of 3 Outcome Measures of Motor Function After Stroke Rehabilitation

To the Editor:

With interest we read the recently published letter of Dr Sivan, “Interpreting Effect Size to Estimate Responsiveness of Outcome Measures,”1 as a response to a paper by Hsieh et al2 in which they provided indices of the magnitude of treatment-related intraindividual change assessed with the Fugl-Meyer Assessment (FMA), Action Research Arm Test (ARAT), and Wolf Motor Function Test performance time (WMFT-TIME) and functional ability scores (WMFT-FAS). As an effect size index, Hsieh et al used the method of the so-called standardized response mean (SRM) by which mean change in scores over time is divided by the SD of these change scores (see Formula A).

\[ \text{SRM} = \frac{\bar{X}_{\text{change}}}{\text{SD}_{\text{Xchange}}} \]

As Sivan argued in his letter,1 the interpretation of the magnitude of intraindividual change estimated with a SRM may lead to overestimation or underestimation of treatment-related effects when the widely used thresholds of Cohen3 are used. These thresholds for classification of the magnitude of mean differences were developed with an effect size index based on standardizing these mean differences using the pooled SD (see Formula B). Dunlap et al convincingly argued that only the pooled SD should be used to compute effect size (ES) for correlated designs and concluded that if the SDpooled is correct for the amount of correlation between the measures, then the ES estimate will be an overestimate of the actual ES.4 It is essential for clinical investigators to understand the differences between the SRM and ES in classifying treatment-related change in terms of Cohen’s thresholds (ES < 0.20 indicating “trivial” change, ES between ≥0.20 and <0.50 “small,” ES of ≥0.50 to <0.80 a moderate, and ES ≥0.80 a large change).3

\[ \text{ES}_{\text{pooled}} = \frac{\bar{X}_{\text{change}}}{\text{SD}_{\text{Xchange}}(\text{pooled})} \]

Hsieh et al refer to our earlier work concerning the risk of misclassification of an SRM when using Cohen’s thresholds in their response to Sivan’s critic. However, their calculation of adjusted ES estimates is based on a false assumption. Consequently, adjusting ES for the size of the correlation between baseline and follow-up as computed by Hsieh et al5 in the Table leads to an ES estimate more than twice the magnitude of the ES computed using the SDpooled when the correlation between the baseline and follow-up scores is at least 0.8.6

Adjustment of a SRM to ES comprises 2 components. First, Cohen introduced a \( \sqrt{2} \) correction as necessary for an appropriate use of his tables for sample size calculation. This correction for looking in Cohen’s power tables is necessary because these assume 2(N-1) degrees of freedom (2 independent samples), whereas in, for example, pre-/posttest evaluation, only n-1 are actually available3 (pp 46 to 48). Thus, following Cohen’s theory, “multiplying SRM by \( \sqrt{2} \) (approximately 1.41) compensates for the sample size tables’ assumption of double the error variance”7 (p 46).

Second, because the \( t \) test prescribed in “own control” study designs (baseline to follow-up) is based on correlated means (p 48), we also have to compensate for the correlation \( r \) between paired observations. Therefore, according to Cohen, the relative size of the standardizing unit for the SRM to the ESpooled is not

\[ d = \frac{d'}{\sqrt{1-r}} \]

but

\[ \sqrt{2(1-r)} \]

Thus, the difference between means for paired (dependent) samples needs to be standardized by a value “which is \( \sqrt{2(1-r)} \) as large as would be the case were they independent”8 (p 49).

As was shown in an earlier publication, \( (d'/\sqrt{2})/\sqrt{(1-r)} \) is equivalent to the SRM and alternatively SRM * \( \sqrt{2} \) * \( \sqrt{(1-r)} \)

Table. Effect Size Estimates SRM, ESP, and ES as Adjusted by Hsieh et al

<table>
<thead>
<tr>
<th>Scale</th>
<th>( r )</th>
<th>( r^2 )</th>
<th>( [A] ) Mean Change</th>
<th>( [B] ) ( \text{SD}_{\text{Xchange}} )</th>
<th>( [C] ) ( \text{SD}_{\text{Xchange}}(\text{pooled}) )</th>
<th>A/B Effect Size d (SRM)</th>
<th>A/C Effect Size ( \text{ES}_{\text{pooled}} ) (ESP)</th>
<th>Adjusted Effect Size ( \text{ES}_{\text{pooled}}(\text{ESP})/\sqrt{(1-r)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMA</td>
<td>0.901</td>
<td>0.81</td>
<td>5.75</td>
<td>4.06</td>
<td>8.67</td>
<td>1.42</td>
<td>0.66</td>
<td>2.06</td>
</tr>
<tr>
<td>ARAT</td>
<td>0.915</td>
<td>0.84</td>
<td>4.68</td>
<td>4.95</td>
<td>11.24</td>
<td>0.95</td>
<td>0.41</td>
<td>1.40</td>
</tr>
<tr>
<td>WMFT-TIME</td>
<td>0.594</td>
<td>0.35</td>
<td>2.13</td>
<td>5.56</td>
<td>5.91</td>
<td>0.38</td>
<td>0.36</td>
<td>0.56</td>
</tr>
<tr>
<td>WMFT-FAS</td>
<td>0.951</td>
<td>0.90</td>
<td>0.30</td>
<td>0.23</td>
<td>0.73</td>
<td>1.30</td>
<td>0.41</td>
<td>1.85</td>
</tr>
</tbody>
</table>

FMA indicates Fugl-Meyer Assessment; ARAT, Action Research Arm Test; WMFT, Wolf Motor Function Test performance time (WMFT-TIME) and functional ability scores (WMFT-FAS).
is equivalent to $d'$ and both indices will vary with the size of $r$. Hsieh et al have adjusted their ES but may have overlooked the fact that their method of adjustment of SRM to an ES is incomplete. As shown in the Table, correct transformation of their SRM estimates into ES suitable for Cohen’s classification resulted in a moderate effect for Fugl-Meyer Assessment and small effects for Action Research Arm Test, Wolf Motor Function Test–TIME, and Wolf Motor Function Test–FAS, respectively, which seems in sharp contrast with the conclusion of Hsieh et al. Extreme high correlations between baseline and follow-up (except Wolf Motor Function Test–TIME) are mainly responsible for overestimation of effect size by the method of Hsieh et al. Minor deviations from published indices are due to calculations without individual data.

**Disclosures**

None.

---

**Eric van Sonderen, PhD**

*University Medical Center Groningen*

*Department of Health Sciences*

*Subdivision of Health Psychology*

*University of Groningen*

*Groningen, The Netherlands*

Responsiveness and Validity of 3 Outcome Measures of Motor Function After Stroke Rehabilitation
Berrie Middel and Eric van Sonderen

Stroke. 2010;41:e463-e464; originally published online May 13, 2010; doi: 10.1161/STROKEAHA.110.580365

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/41/7/e463