Doubting the Efficacy/Effectiveness of Electroencephalographic Neurofeedback in Treating Children With Attention-Deficit/Hyperactivity Disorder Is As Yet Unjustified

To the Editor: Van Dongen-Boomsma et al.1 assessed the efficacy of electroencephalographic (EEG) neurofeedback in children with attention-deficit/hyperactivity disorder by comparing an experimental group (n = 22) receiving EEG neurofeedback to a placebo group (inactive neurofeedback) (n = 19). Results did not show significant differences between the 2 groups. The authors report that, post hoc, the “sample had 80% power to detect a treatment effect of 0.90” and that “it is unlikely that the negative results were due to limited statistical power,”1(p826) concluding that their study sheds serious doubt on the effectiveness of EEG neurofeedback as a way of treating children with attention-deficit/hyperactivity disorder.

First, we want to acknowledge the robust design and methodology of this study. However, we are not sure what the authors meant by stating in their discussion that “the sample had 80% power to detect a treatment effect of 0.90,” and we cannot find an explanation of the rationale behind such a statement. How would statistical power be an issue if the effect size was as high as 0.90 in a sample of 41 participants? We ran a post hoc power analysis, using G*Power Version 3.1.5,2,3 for a sample size of 41 participants, α of 5% and an effect size of 0.90 resulting from repeated-measures analysis of variance within-group and between-groups interaction tests for 2 groups and 2 measuring times, and found that this sample would have achieved 100% power to detect a treatment effect size of 0.90, not 80% as reported by the authors. Therefore, there is a 100% chance of finding statistically significant differences (P ≤ .05) between the EEG neurofeedback group and the placebo group, although none were found. Impossible, unless the effect size equaled zero. We suspect that the authors made a calculation error.

To verify this, we first calculated η² effect size for all the comparisons in the group x time effect (interaction) using the
following formula:\(^5\):
\[
F = \frac{(F_{\text{df}_1} 	imes df_{\text{J}})}{(F_{\text{df}_1} + df_{\text{J}})}
\]
We found that 0.13% ≤ η\(^2\) ≤ 5.48% (Table 1). Then, with these effect sizes, we ran a post hoc power analysis for each comparison, following the same procedure, using G*Power Version 3.1.5.\(^2\) Our calculations show that the achieved power varies between 5% and 11% (Table 1). On the basis of these findings, the risk of a type II error in this study varies between 89% and 95%. Hence, the authors’ conclusion that it is unlikely that the negative results were due to a lack of statistical power is conceptually wrong.

We read van Dongen-Boomsma and colleagues’\(^1\) study with great interest, and, given the η\(^2\) effect sizes we calculated, we agree that the differences between the effect of EEG neurofeedback and the placebo neurofeedback are small, and we believe that they probably would not be much higher in a larger sample if inclusion criteria were based on the same demographic and clinical characteristics. Nonetheless, given the lack of statistical power we showed above, it is not methodologically sound to describe the study as one with negative results. Clarification from the authors of the rationale and what they meant in saying “the sample had 80% power to detect a treatment effect of 0.90 would be useful. Further research with sound statistical power and validity is needed to corroborate van Dongen-Boomsma and colleagues’ conclusion doubting the efficacy/effectiveness of EEG neurofeedback, which has so far been demonstrated in other methodologically sound studies.\(^6,7\)

TABLE 1. Effect Sizes and Post Hoc Power Calculation

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group × Time Effect</th>
<th>F</th>
<th>df</th>
<th>P Value</th>
<th>η(^2) (%)</th>
<th>Achieved Power (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD-RS-IV investigator-rated score Total symptoms</td>
<td>0.36</td>
<td>1,39</td>
<td>0.55</td>
<td>0.91</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Inattention symptoms</td>
<td>0.17</td>
<td>1,39</td>
<td>0.68</td>
<td>0.43</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Hyperactivity/impulsivity symptoms</td>
<td>2.26</td>
<td>1,39</td>
<td>0.14</td>
<td>5.48</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>ADHD-RS-IV teacher-rated score   Total symptoms</td>
<td>0.45</td>
<td>1,37</td>
<td>0.50</td>
<td>1.20</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Inattention symptoms</td>
<td>0.25</td>
<td>1,37</td>
<td>0.62</td>
<td>0.67</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Hyperactivity/impulsivity symptoms</td>
<td>0.53</td>
<td>1,37</td>
<td>0.47</td>
<td>1.41</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>CGI-I score</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGAS score</td>
<td>1.96</td>
<td>1,38</td>
<td>0.16</td>
<td>4.90</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>SDQ score</td>
<td>0.05</td>
<td>1,37</td>
<td>0.81</td>
<td>0.13</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>PSERS score</td>
<td>0.10</td>
<td>1,39</td>
<td>0.75</td>
<td>0.26</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: ADHD-RS-IV = ADHD Rating Scale IV, CGAS = Children’s Global Assessment Scale, CGI-I = Clinical Global Impressions-Improvement scale, PSERS = Pittsburgh Side Effects Rating Scale, SDQ = Sleep Disorders Questionnaire.

REFERENCES


Letters to the Editor

Dr van Dongen-Boomsma Replies

To the Editors: My colleagues and I thank Dagenais and colleagues for their detailed letter, in which they ask for clarification of the following statement: “Post hoc, our sample had 80% power to detect a treatment effect of 0.90 . . . it is unlikely that our negative results were due to limited statistical power.”\(^1\)\(^8\)\(^2\)\(^6\) Hopefully, this letter will answer their questions.

First, we appreciate Dagenais and colleagues’ acknowledgment regarding the robust methodological design of our study. Further, we thank them for pointing out an error in the text. Although we actually did have an 80% power to detect a treatment effect of 0.90, the power to detect rather small differences between EEG neurofeedback and placebo neurofeedback indeed was much more limited, namely between 5% and 11%, as they show. Thus, our statement “it is unlikely that our negative results were due to limited statistical power” was wrong. In other words, only a substantially larger sample size would have had enough power to establish such small effect sizes. However, it is questionable whether such small treatment effects would be clinically relevant and meaningful.

To overcome the power problem, future research demands larger samples and an even more robust design than our study. Therefore, it is a pleasure that the near future will bring such studies, as discussed by Kerson and the Collaborative Neurofeedback Group.\(^2\)

Martine van Dongen-Boomsma, MD
m.vandongen-boomsma@karakter.com

Dr van Dongen-Boomsma Replies

To the Editor: My colleagues and I thank Dagenais and colleagues for their detailed letter, in which they ask for clarification of the following statement: “Post hoc, our sample had 80% power to detect a treatment effect of 0.90 . . . it is unlikely that our negative results were due to limited statistical power.”\(^1\)\(^8\)\(^2\)\(^6\) Hopefully, this letter will answer their questions.

First, we appreciate Dagenais and colleagues’ acknowledgment regarding the robust methodological design of our study. Further, we thank them for pointing out an error in the text. Although we actually did have an 80% power to detect a treatment effect of 0.90, the power to detect rather small differences between EEG neurofeedback and placebo neurofeedback indeed was much more limited, namely between 5% and 11%, as they show. Thus, our statement “it is unlikely that our negative results were due to limited statistical power” was wrong. In other words, only a substantially larger sample size would have had enough power to establish such small effect sizes. However, it is questionable whether such small treatment effects would be clinically relevant and meaningful.

To overcome the power problem, future research demands larger samples and an even more robust design than our study. Therefore, it is a pleasure that the near future will bring such studies, as discussed by Kerson and the Collaborative Neurofeedback Group.\(^2\)

Martine van Dongen-Boomsma, MD
m.vandongen-boomsma@karakter.com

REFERENCES


© 2014 COPYRIGHT PHYSICIANS POSTGRADUATE PRESS, INC. NOT FOR DISTRIBUTION, DISPLAY, OR COMMERCIAL PURPOSES.

© 2014 COPYRIGHT PHYSICIANS POSTGRADUATE PRESS, INC. NOT FOR DISTRIBUTION, DISPLAY, OR COMMERCIAL PURPOSES.