Letter to the Editor

Data Synthesis in Meta-Analysis may Conclude Differently on Cognitive Effect From Transcranial Direct Current Stimulation

We read, with great interest, the newly published article in Brain Stimulation Journal by Dr. Horvath and his colleagues [1] suggesting that single-session transcranial direct current stimulation (tDCS) has no cognitive effect on healthy populations based on the meta-analysis of 59 published studies. The authors did a commendable job in searching literature and aggregating data. Their article summarized the cognitive effect from tDCS on humans. However, the approach used by the authors in data synthesis may have missed uncovering the potential cognitive effect by tDCS. Two issues are relevant for accurately calculating the summary effect size of stimulation.

First, the effect size (specifically, Hedge’s g) calculated from post-active (and post-sham) stimulation outcome scores do not always reflect the effects of active stimulation. The issue raised by using only post-active or post-sham stimulation outcome scores is two-fold:

a. Pre-stimulation outcome scores in either group is ignored, allowing between group variability as a confounder in the effect size calculations.

b. Variability in post-stimulation outcome scores is not representative of variability in the change scores (the difference between post and pre-stimulation outcome scores) across subjects. Use of the former, therefore, may inaccurately represent within-group variability and consequent Hedge’s g calculations.

To illustrate how this might affect the effect size calculation, we use subject-level data from three published tDCS studies in motor recovery [2–4] (Fig. 1A–C). Post-stimulation outcome scores inherit the variability of pre-stimulation outcome scores resulting in smaller Hedge’s g and subsequently a statistically non-significant summary effect (Fig. 1D). However, when such variability is compensated by using the change scores, the resulting Hedge’s g and summary effect size is bigger and more focused on reflecting the role of tDCS (Fig. 1E). Therefore, one must use change scores from both active and sham stimulation groups to calculate the Hedge’s g in order to properly represent the true effect of tDCS.

Second, aggregation of different behavioral outcome measures can introduce variability and decrease accuracy of summary effect size. While not an exhaustive list, below are some of the major factors that can contribute to diluting summary effect size in meta-analysis:

a. Different outcome measures use different evaluation protocols and different scoring systems.

b. Behavioral response to tDCS stimulation differs across various outcome measures.

These points can be illustrated by this included study [5] where the effect size of semantic verbal fluency is different from phonetic verbal fluency (1.00 versus 0.63). While hard to achieve optimally, (a) can be addressed by appropriate data normalization. Although there is no optimal solution to correct for the variability induced by (b), a researcher may mitigate such variability by weighing outcome measures (or only using sub-scores from a given outcome measure) when calculating the summary effect size. It is preferable not to mix different outcome measures when calculating summary effect size.

Extracting and synthesizing data for a meta-analysis can be difficult and time-consuming because required information is not always available in the published manuscripts. Resolving this omission may require contacting authors to provide required information. Studies may be excluded from meta-analysis when needed information cannot be obtained from the authors. Sub-optimal synthesis of data based on limited information available in the published literature (versus information actually required to accurately calculate effect sizes) may lead to different, and perhaps erroneous, effect size estimation.

We now discuss a specific example based on some results presented by Horvath and his colleagues (Fig. 2C and Table 4 in the article) to express our concerns.

1. The effect size for one study [6] is calculated using the change scores, but post-stimulation outcome scores (not the change scores) were used for the other two studies [7,8].

2. A phonetic verbal fluency task is chosen for the outcome measure in one study [6] while a semantic verbal fluency task is used in the other two studies [7,8].

While such data synthesis may not necessarily affect the results substantially, the use of the most relevant data can yield a greater confidence in summary effect size and subsequent conclusions.

Additionally, aggregating data over the same task across different tDCS montages may actually lead to a statistically significant (or near-significant) summary effect (Table 1). Use of change scores may further decrease the variance resulting in a statistically significant summary effect (e.g., in stroop task and verbal fluency task). Such results may dramatically change conclusion when the P-value is close to a significance level.
Figure 1. Data synthesis for the effect size calculation may change the results and conclusions. A–C, Jebsen-Taylor Hand Function Test (JTT) data from Boggio 2007 [2] (A) and Fregni 2005 [3] (B) as well as Fugl-Meyer Upper Extremity Score (FM-UE) data from Lindenberg 2010 [4] (C) are presented. Individual panels show data from a given stimulation group (active stimulation – Bihemispheric, Anodal, Cathodal – or sham stimulation). Pre-stimulation (open dots, denoted as “Pre”) and Post-stimulation (red dots, denoted as “Post”) outcome scores for a given subject are joined by a line and change scores (the difference between post- and pre-stimulation outcome scores) are shown separately (green dots, denoted as “Change”). Note that individual subjects have different Post, Pre and Change scores in a given study in a given group. More importantly, the standard deviation of Pre or Post scores is different from the one of Change scores. Error bars: mean ± SD. D, Calculation of summary effect size (Hedge’s g) using Post scores suggests that tDCS does not lead to improvement in outcomes, at individual study level as well as at the meta-analysis level. E, Similar calculation by using Change scores reveals the actual effect of tDCS in improving outcomes that is significant both at the individual studies level as well as at the meta-analysis level. The Change scores nullify the noise introduced by variability in Pre scores and incorporate actual variability in the improvement due to the intervention. Signs of mean scores are changed in D and E for FM-UE data from Lindenberg 2010 study to ensure correct directionality of the effect. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
To summarize, results from meta-analyses should be interpreted cautiously. Attention must be paid to the study inclusion criteria, selection of outcome measures, the method of data synthesis and calculation of the effect size.

Acknowledgments

We would like to thank Drs. James Sawers and Ilya Lipkovich for their generous input to the manuscript.

Drs Pratik Chhatbar and Wuwei Feng are both supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM109040. Dr. Feng is also supported by American Heart Association (14SDG1829003).

Pratik Y. Chhatbar
Wuwei Feng*
Department of Neurology and Neurosurgery
Medical University of South Carolina
USA

*Corresponding author. MUSC Stroke Center, 19 Hagood Avenue, HOT-501, Charleston, SC 29425, USA.
E-mail address: feng@musc.edu (W. Feng)

Received 7 May 2015
Available online xxx

http://dx.doi.org/10.1016/j.brs.2015.06.001

References


Table 1
Aggregating data over the same task may produce a significant summary effect.

<table>
<thead>
<tr>
<th>Task</th>
<th>Figure/Table in the article</th>
<th>SMD (95% CI)</th>
<th>Z value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop signal task</td>
<td>Fig. 1B–E/Table 1, 2</td>
<td>-0.39 (-0.68, -0.11)</td>
<td>-2.69</td>
<td>0.007</td>
</tr>
<tr>
<td>Stop signal task (SSRT only)</td>
<td>Fig. 1C–E/Table 1, 2</td>
<td>-0.37 (-0.69, -0.05)</td>
<td>-2.26</td>
<td>0.024</td>
</tr>
<tr>
<td>Stroop task</td>
<td>Fig. 1F–G/Table 2</td>
<td>0.51 (-0.01, 1.03)</td>
<td>1.92</td>
<td>0.055</td>
</tr>
<tr>
<td>Verbal fluency</td>
<td>Fig. 2C–D/Table 4</td>
<td>0.26 (0.00, 0.53)</td>
<td>1.94</td>
<td>0.052</td>
</tr>
</tbody>
</table>