COMMENT

Reply to the comment on: "Reporting matters: Brain mapping with transcranial magnetic stimulation"

Hugo Massé-Alarie¹ | Michael Bergin² | Cyril Schneider^{1,3} | Siobhan Schabrun⁴ | Paul Hodges⁵

¹Rehabilitation, Université Laval Faculté de Médecine, Québec City, Québec, Canada

²Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital and Harvard Medical School, Charlestown, Massachusetts

³Rehabilitation, CHU de Québec Research Center, Neuroscience Unit (CHUL), Laboratory of Clinical Neuroscience and neuroStimulation, Université Laval, Québec, Canada

⁴Neuroscience Research Australia, Sydney, Australia

⁵Centre for Clinical Research Excellence in Spinal Pain, Injury and Health, The University of Queensland, St. Lucia, Queensland, Australia

Correspondence

Hugo Massé-Alarie, Rehabilitation, Université Laval Faculté de Médecine, Québec, Canada.
Email: hugo.masse-alarie@fmed.ulaval.ca

We thank Dr Héroux for the comments about our article published in *Human Brain Mapping* (Masse-Alarie, Bergin, Schneider, Schabrun, & Hodges, 2017) and the importance of reporting appropriately statistics. On reflection, we agree several statements were too strong regarding the interpretation of the results with respect to statistical trends and the small sample size of the study. We should have been more conservative in the discussion of some of the results. However, we also believe that the comments do not reflect the entirety of our work and overlooked large portion of our study to spin Dr Héroux's points. We have two issues of concern.

1 | INTERPRETATION OF STATISTICALLY AND NON-STATISTICALLY SIGNIFICANT P-VALUES

Our study is an original and exploratory studies testing novel hypotheses on the function of a whole system (the primary motor cortex organization and function). To test our hypothesis, we used *p*-values. The *p*-values inform about the probability of observing a value as extreme assuming that the null hypothesis is true and then help to interpret the validity of the results. This method is imperfect, but a widely accepted method to test a hypothesis. On the basis of this imperfection, its stringent use is not recommended (Curran-Everett & Benos, 2004; Wasserstein & Lazar, 2016). Dr Héroux seems to use a double standard in the interpretation of our results; he reproves our discussion of one finding of our study that did not reach the threshold of $p < .05$, but implied to be significant in the discussion (that we tend to agree with since we should have been more cautious in the

statements use) but thereby supports a very dichotomized analysis into statistically and non-statistically significance that does not account for the probabilistic nature of the *p*-value. If our *p*-values for the two analyses in question had been .003 and .01 smaller, then he would not have raised concern. He seems to reprove the report of *p*-values between $p = .05$ and $p = .10$ although he agrees that *P*-values are fickle. He even challenged us that we should report that *p*-values between $p = .01$ and $p = .05$ "tend to non-significance." These latter arguments do not help with the interpretation of results by introducing a polarized view, a certainty about results (null hypothesis is accepted or rejected; hypothesis is true or false) rather than using the probability than a set of outcomes that together supports or not the hypothesis, and inform about the function of the system tested.

Although we agree that we should have been more conservative with language in the discussion, the mistakes we made (as many authors do) were probably to use an $\alpha = 0.05$ instead of a more liberal $\alpha = 0.10$ because of the exploratory nature of our study (as recommended by the "Guideline for reporting statistics" [Curran-Everett & Benos, 2004]) and using terms as statistically and non-statistically significant as suggested by a statement released by American Statistical Association (AMA) on the use of *p*-values (Wasserstein & Lazar, 2016). It is important to note that although we used the latter terms we did not reduce our analysis to *significance*. The "Guideline for reporting statistics" rather suggests using statements as "may be a true effect" (for *p*-values between $p = .05$ and $p = .10$) and "good evidence of a true effect" (for *p*-values between $p = .01$ and $p = .05$). However, as authors it may be confronting to modify this strongly implanted way to report and interpret results especially because of the apprehension to be reviewed negatively by reviewers

who are not aware of recommendations. Perhaps implementation of statistical standards by journals' editors as those published by the *American Physiological Society* (APS) (Curran-Everett & Benos, 2004) might help to change how the scientific community report results and statistics. Clearly other methods that judge the size of effect would be valuable to report and aid interpretation.

2 | SEM INSTEAD OF SD TO REPRESENT DATA VARIABILITY

As reported by Dr. Héroux and many statistical-related comments and guidelines (Curran-Everett & Benos, 2004; Drummond & Vowler, 2011; Nagele, 2003), it would have been preferable to report and display standard deviations (SD) instead of standard error of the mean (SEM) as SD reflects variability in the study sample whereas SEM represents the precision of the sample compared with the underlying population. Dr Héroux is mainly reproaching that we want to *hide/conceal* true participant variability and mislead readers that are not aware than SEM is $SD/\sqrt{n}$ and representing different statistical concepts. However, we strongly defend our transparency in reporting of the variability in the data. First, in the Methods, we acknowledged the inherent variability of the data: "Inspection of the data revealed substantial variation in the location of the map relative to the grid (aligned to the vertex) between individuals. This produced high variation in parameters of the M1 map (CoG coordinate, volume, and distance) and problems with averaging of the data across the participant group." Second, we reported individual values in Table I (map volume, CoGx, CoGy, nb of peaks) and Table II (peak values), in Figure 3 (individual maps of 4 participants), and in Figure 5 (map contrast for all individuals). In the legend of Figure 3 we even mention that the presented maps were selected "to show diversity of map features." It is then not clear why Dr Héroux states: "Reporting results with standard deviations or other appropriate measures of variability provides the reader with a better sense of the underlying data which is important to appropriately interpret study results and figures" when addressing our results. In contrast to his claim, we believe that our presentation of the dataset provides an unusually clear account of the variability, reporting mostly all individual data in Tables and Figures as recommended by statisticians (Drummond & Vowler, 2011).

Finally, it is also important to note that the variability and lack of reproducibility in brain stimulation studies are not only related to statistical issues (Beaulieu, Flamand, Masse-Alarie, & Schneider, 2017; Chipchase et al., 2012). It is also likely to reflect a variable understanding of the measurement techniques, and the frequent suboptimal

control of factors that impact on the properties of the motor-evoked potential. Strict control over these factors are recommended to improve the overall quality of the studies (Chipchase et al., 2012).

Although we concede that greater care in some language was warranted, we strongly object to the notion that our intentions were to conceal the nature of our observations. In direct contrast to this view we have argued that our presentation of variability of our results aimed to be completely transparent. We believe that adequate reporting of statistics is of primary importance and as suggested by AMA and APS Guideline should not only be reduced to a mere dichotomous analysis of *significance* but also rather on accumulation of evidence that may inform about the a priori hypothesis.

ORCID

Hugo Massé-Alarie <https://orcid.org/0000-0003-4472-9656>

Siobhan Schabrun <https://orcid.org/0000-0002-9083-3107>

REFERENCES

- Beaulieu, L. D., Flamand, V. H., Masse-Alarie, H., & Schneider, C. (2017). Reliability and minimal detectable change of transcranial magnetic stimulation outcomes in healthy adults: A systematic review. *Brain Stimulation*, 10, 196–213.
- Chipchase, L., Schabrun, S., Cohen, L., Hodges, P., Ridding, M., Rothwell, J., ... Ziemann, U. (2012). A checklist for assessing the methodological quality of studies using transcranial magnetic stimulation to study the motor system: An international consensus study. *Clinical Neurophysiology*, 123, 1698–1704.
- Curran-Everett, D., & Benos, D. J. (2004). Guidelines for reporting statistics in journals published by the American Physiological Society. *Physiological Genomics*, 18, 249–251.
- Drummond, G. B., & Vowler, S. L. (2011). Show the data, don't conceal them. *The Journal of Physiology*, 589, 1861–1863.
- Masse-Alarie, H., Bergin, M. J. G., Schneider, C., Schabrun, S., & Hodges, P. W. (2017). "Discrete peaks" of excitability and map overlap reveal task-specific organization of primary motor cortex for control of human forearm muscles. *Human Brain Mapping*, 38, 6118–6132.
- Nagele, P. (2003). Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *British Journal of Anaesthesia*, 90, 514–516.
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 70, 129–133.

How to cite this article: Massé-Alarie H, Bergin M, Schneider C, Schabrun S, Hodges P. Reply to the comment on: "Reporting matters: Brain mapping with transcranial magnetic stimulation." *Hum Brain Mapp*. 2019;40:354–355. <https://doi.org/10.1002/hbm.24471>