Asian Journal of Psychiatry 44 (2019) 29–34

Contents lists available at ScienceDirect

Asian Journal of Psychiatry

journal homepage: www.elsevier.com/locate/ajp

Statistical power estimation in non-invasive brain stimulation studies and its clinical implications: An exploratory study of the meta-analyses

Sayantanava Mitra[a,b,*], Urvakhsh Meherwan Mehta[c], Bhaskarapillai Binukumar[d], Ganesan Venkatasubramanian[c], Jagadisha Thirthalli[c]

[a] Central Queensland Mental Health, Alcohol and Other Drugs Services (CQMHAODS), Base Hospital, Rockhampton, Queensland, Australia
[b] Faculty of Medicine, The University of Queensland Rural Clinical School, Rockhampton, Queensland, Australia
[c] Department of Psychiatry, National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore, India
[d] Department of Biostatistics, National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore, India

ARTICLE INFO

Keywords:
Non-Invasive brain stimulation
rTMS
ECT
tDCS
Statistical power
Sample size

ABSTRACT

Background: Non-invasive brain stimulation (NIBS) techniques have emerged as a promising tool for understanding and treating psychiatric disorders, necessitating a caution in terms of interpreting research results.
Objective: This study aimed at systematically evaluating a representative sample of research conducted using NIBS interventions in neuro-psychiatric conditions, and assessing the power these studies achieved, given their sample sizes.
Methods: A database search was conducted with defined keyword combinations. Using reported summary effects of the meta-analyses as estimate of the true effects, we calculated achieved power of each individual study to detect the effect indicated by the corresponding meta-analysis.
Results: Findings suggest that mean and median powers in the field of NIBS were 0.50, with a mode at 0.83 (range 0.05–1.00). When analysed separately, the median powers were 0.27 for tDCS, 0.70 for TMS and 0.97 for ECT. These studies had a mean total sample size of 22.2 ± 24.9 subjects and the median reported effect size across all studies was 0.61.
Conclusion: According to our findings, studies conducted in NIBS miss around 50% of true positive results. Further, it appears that most of the researchers in this field chase statistical significance with small sample sizes, thus compromising the quality of their conclusions.

1. Introduction

Mental health, today, is at a riveting crossroad. Centuries of 'intuition-based approach' to treating mental illnesses is being replaced by 'evidence-based practice' (Geddes, 2000) at an unprecedented rate, and newer conceptualizations are finally providing us with a control over these ancient human maladies. Present-day psychiatry requires an integration of scientific principles and attitude with the art of healing (Wallace, 2011), which necessitates systematic assimilation of a massive amount of data in a format that is easily comprehendible to the clinicians. Exercises in critical appraisal and statistical analysis form an important cornerstone in this regard (Wallace, 2011). Inferential statistics provides us with the tools to formulate a general theory about a population based on sample characteristics, and to make evidence-based predictions (Lindley, 1990). Text

While Bayesian approach to statistical inference might inspire the user to think more probabilistically, tradition and ease enable classical methods to still hold sway (Cordani, 2010). Null-hypothesis statistical significance testing (NHST) is among the most widely used methods of data analysis in conventional statistics, and has been around for well over 7 decades. It states that "the experimental group and the control group are not different with respect to [a specified property of interest] and that any difference found between their means is due to sampling fluctuation" (Carver, 1978). A null-hypothesis is rejected if the test of significance returns a value which is less than a predefined level of significance (α), usually set at 0.05. In this, there are two ways of failing (Nickerson, 2000): rejecting it when it is true (Type I error) and not rejecting it when it is false (Type II error). There are two general NHST methods in vogue: the acceptance-support (AS) and the rejection-support (RS) NHST. The latter is more common of the two, where the "null hypothesis represents what the experimenter does not believe", and its rejection is taken as a vindication of the "experimenter's theoretical

* Corresponding author at: Central Queensland Mental Health, Alcohol and Other Drugs Services (CQMHAODS), Rockhampton, Queensland, Australia.
E-mail address: sayantanava@gmail.com (S. Mitra).

https://doi.org/10.1016/j.ajp.2019.07.006
Received 31 March 2019; Received in revised form 7 June 2019; Accepted 5 July 2019

position" (Nickerson, 2000). Therefore, it is imperative that the null-hypothesis be rejected correctly, so that the results reflect the real-world scenario.

Within this statistical framework, it is becoming increasingly clearer that a large number of initial findings are never replicated - or are outrightly refuted - during subsequent experiments. There is a growing concern that majority of the published results in modern scientific literature are false, and is the fallout of a convenient way of assuming that any relationship that crosses below the magical significance level of 0.05 is "significant" (Ioannidis, 2005, 2017). Contrary to this practice, a multitude of factors determine whether the claims being made are actually true – the prior probability of it being true, the significance level of the statistics being uses and the statistical power of the study (Ioannidis, 2005, 2017). The literature, in general, is strongly biased against committing a type-I error; and the power of a study (probability of not committing type-II error) is one of the less-considered factors in mental health research (Nickerson, 2000; LeMire, 2010). This, in combination with a highly competitive environment, encourages researchers to focus on the 'α' and forces them to engage in certain practices which increase the likelihood of a quick publication. According to Button et al. (2013), these practices include using flexible study designs and statistical analyses, conducting small-sampled and low-powered studies. A low statistical power translates into a lower probability of finding a true effect, a poorer positive predictive value of the results (Button et al., 2013) and an inflation of the magnitude of reported effects (Winner's curse or Type M error; Button et al., 2013; Gelman and Carlin, 2014). Further, such studies are more prone to methodological flaws as 'vibration of effects', publication biases, or selective data analyses and reporting (Button et al., 2013). Nevertheless, in fields where researchers are attempting to pursue low likelihood hypothesis with expectations of significant breakthroughs, the perils of a low sample size and power are bound to be apparent (Krzywinski and Altman, 2013). From another perspective, such studies might be wasteful of the available resources and in fact could border on being unethical (Button et al., 2013; Krzywinski and Altman, 2013). Interestingly, there has been a recent suggestion to reduce *P* value from 0.05 to 0.005 for NHST, to better match the Bayes factor – the ratio of the evidence from data to prior probability – and thus increasing the reproducibility (Benjamin et al., 2018). However, an increase in the risk of false-negative rates and an unfavourable cost-benefit outcome is clearly discernible in this proposal.

Non-invasive brain stimulation (NIBS) techniques have, of late, emerged as a promising tool for understanding and treating psychiatric disorders, through a more nuanced application of basic neurophysiological principles (Mehta, 2016). However, as with any emerging field of study, reports in NIBS might have suffered from the winner's curse and demonstrated the proteus phenomenon (Ioannidis, 2005; Button et al., 2013), the two statistical antitheses of clinical and methodological rigour.

Additionally, researches in neuroscience deal with considerably greater heterogeneity and uncertainties in the input variables as compared to more "hard" sciences, which might further add to the differences in outcome (Mitra et al., 2018). The implications of such deficiencies would be far reaching, since present-day evidence-based medical practice depends heavily on the quality of these data. This exercise is aimed at evaluating the powers of studies conducted in the field of TMS, ECT and tDCS - the three major NIBS techniques in clinical practice - and investigating the translatability of the available information into clinical practice.

2. Methods

A database search was conducted by SM, with defined key-words, to identify relevant published literature. Since this was a token exercise, the search was restricted to only one database – the PubMed (www.ncbi.nlm.nih.gov/pubmed/). All studies published in their final form, and listed on the database between 01.01.2016 and 31.12.2016, were screened. The following keyword combinations were used for the purpose: {(electroconvulsive therapy [tiab] OR ECT [tiab]) AND meta analysis [pt]} for studies on ECT; {(transcranial magnetic stimulation [tiab] OR TMS [tiab]) AND meta analysis [pt]} for studies on rTMS; and {(transcranial direct current stimulation [tiab] OR tdcs [tiab]) AND meta analysis [pt]} for studies on tDCS. Articles were excluded if no abstracts were available, the main text was not available in English, no meta-analysis had been conducted or if the paper did not provide/allow for the extraction of enough data for present analysis. After screening the returned search results for exclusion criteria, the following data were extracted from the remaining papers: the name of the first author and summary effect size estimate of the meta-analysis, and the first author, publication year and sample size (by groups) of the contributing studies.

Using the reported summary effects of the meta-analyses as the estimate of the true effects, the study calculated achieved power of each individual study to detect the effect indicated by the corresponding meta-analysis. When several meta-analyses were performed in a given paper, we selected the one with the most number of contributing data-sets, or (if they had equal number of data-sets) the one which looked at the primary outcome. If they reported completely different findings/aims of intervention, more than one meta-analysis was included from a paper.

The main outcome measure of the study was the achieved power of the individual studies to detect the estimated summary effect of the corresponding meta-analysis to which they contributed, assuming an alpha of 5%. The power was calculated using G* Power software, version 3.1 (Faul et al., 2007). Subsequently, the mean and median power across all studies was calculated, and further analyses were performed, using SPSS v.22 (IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.)

3. Results

A total of 31 articles were returned in the above described database search. Twenty-nine records were counted in, after the duplicates have been removed, and included in the assessment for eligibility. One article was in French, one was an RCT and four others were not related to NIBS methods. Two articles did not allow for the extraction of required data. The final quantitative synthesis was conducted with 21 studies, and 39 meta-analyses (Fig. 1). There was a total of 435 contributing data sets.

Among the 21 included studies, nine had a median power of less than 0.2 (42.8%), and 18 had a median power of less than 0.5 (85.7%) in their constituents. Three reported possible medium to high publication biases in their included articles, 5 did not comment on this, and 13 meta-analyses (61.09%) confidently reported a low publication bias (Table 1).

Our findings show that the mean and the median powers in the field of brain stimulation were 0.50, with a mode at 0.83 (range 0.05–1.00). When analysed separately, the median powers were 0.27 for tDCS, 0.70 for TMS and 0.97 for ECT (Table 2). Among the studies reporting on TMS, the two analyses performed by Chung et al. (2016) had included the largest number of data-sets (n = 178) and reported a median power of 0.8. However, these authors also found a moderate degree of bias in the contributing literature, and when this paper was removed, the median power dropped to 0.21 in TMS (mean 0.23, mode 0.06).

The brain stimulation techniques have both investigational and therapeutic applications, and they independently further the relevant science. In order to tease out the differences between their applications and achieved power of the corresponding studies, we conducted a method-wise analysis. Investigational applications of TMS had a median achieved power of 0.8 (Chung et al., 2016), while those of tDCS touched 0.15. Clinical studies in tDCS achieved a median power of 0.29 (n = 90), and those in TMS could garner 0.21.

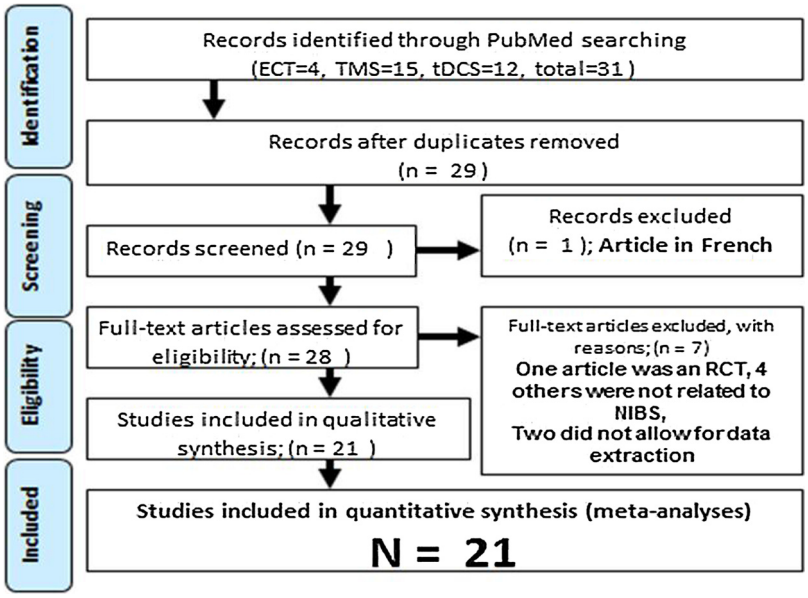

Fig. 1. The article selection process flowchart.

The studies on brain stimulation were conducted with a mean total sample size of 22.2 ± 24.9 subjects. Of the 435 contributing data sets, only 35 (8.04%) had a total sample size of equal to or greater than 50. We calculated the effect size required to achieve a power of 0.8 (considered appropriate by convention; Bacchetti, 2010) by these studies, assuming an allocation ratio of 1 between two independent groups. We found that we were chasing a figure of 1.25, which is very high (Cohen, 1988) and which assumes that 88% of the population in one group would be below the average subject in the other. Only two (of 39) meta-analyses reported an effect size which was greater than or equal to this value. In fact, the median reported effect size across all studies (weighted for the number of studies contributing to one particular meta-analysis) was found to be 0.61 (mean 0.57) (Figs. 2 and 3).

4. Discussion

Our findings suggest that studies conducted in NIBS miss around 50% of true positive results. They also indicate that most of the researchers in this field chase statistical significance with small sample sizes (22.2 ± 24.9 subjects), thus compromising the quality of their conclusions (Button et al., 2013). The effect sizes that we used for this study were based on the effect sizes reported in the corresponding meta-analyses. It is a well-documented fact that scientific literature, in today's significance-driven era of NHST, suffers from significant

Table 1
Characteristics of the included meta-analyses.

Sl.	Author	Number of contributing meta-analyses	Median power achieved across all contributing meta-analyses	Random/ Fixed effects	Author reported publication bias	Number of contributing datasets with more than 50 total subjects
Transcranial Direct Current Stimulation						
1	Barham et al. (2016)	2	0.23 (0.06-0.74)	Random	Not reported	0
2	Chhatbar et al. (2016)	3	0.06 (0.05-0.17)	Random	Low	2
3	Elsner et al. (2016a)	2	0.70 (0.36-0.88)	Random	Low	4
4	Elsner et al. (2016b)	2	0.19 (0.11-0.51)	Random	Low	6
5	Elsner et al. (2016c)	1	0.06 (0.05-0.06)	Random	Low	0
6	Hashemirad et al. (2016)	12	0.34 (0.05-0.99)	Fixed	Not reported	0
7	Hill et al. (2016)	1	0.13 (0.08-0.19)	Random	Low	0
8	Kang et al. (2016)	1	0.35 (0.21-0.87)	Random	Low	2
9	Pisegna et al. (2016)	1	0.30 (0.22-0.36)	Random	Low	0
10	Summers et al. (2016)	1	0.46 (0.26-0.82)	Random	Low	2
11	Tedesco Triccas et al. (2016)	1	0.07 (0.07-0.12)	Fixed	Not reported	1
Transcranial Magnetic Stimulation						
1	Chung et al. (2016)	2	0.80 (0.29-1.00)	Random	Moderate	5
2	Dollfus et al. (2016)	1	0.26 (0.13-0.44)	Random	Not reported	0
3	Dutra et al. (2016)	1	0.21 (0.16-0.49)	Fixed	Not reported	2
4	Enokibara et al. (2016)	1	0.06 (0.05-0.07)	Random	High	1
5	Knijnik et al. (2016)	1	0.18 (0.11-0.24)	Fixed	Low	0
6	Health Quality Ontario (2016)	1	0.10 (0.06-0.81)	Random	Low	4
7	Soleimani et al. (2016)	1	0.18 (0.09-0.19)	Fixed	Low	3
8	Trevizol et al. (2016)	1	0.44 (0.36-0.47)	Random	Low	0
Electro-Convulsive Therapy						
1	Fond et al. (2016)	1	0.22 (0.15-0.42)	Random	Possible	2
2	Takekita et al. (2016)	2	0.99 (0.67-1.00)	Random	Low	1

Table 2
Method-wise analysis of the contributing studies/data-sets.

	Combined	tDCS	TMS	ECT
Contributing data sets	435	169	245	21
Mean Power	0.50	0.29	0.62	0.79
Median Power	0.50	0.27	0.70	0.97
Mode Power	0.83	0.13	0.83	0.99
Minimum Power	0.05	0.05	0.05	0.15
Maximum Power	1.00	0.99	1.00	1.00
Mean sample size (N) of the contributing datasets	22.15 ± 24.9	25.78 ± 20.7	18.82 ± 26.7	31.61 ± 28.4

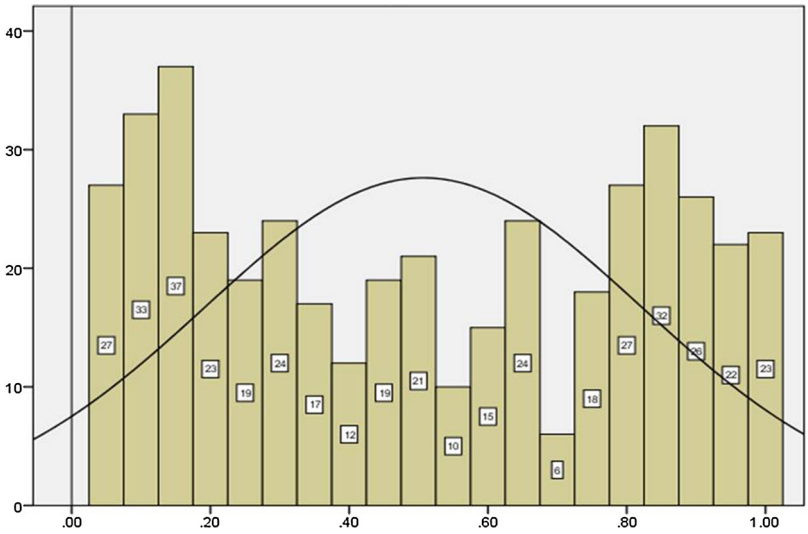

Fig. 2. Power distribution across the whole sample. X-Axis represents the power of the individual datasets, and Y-Axis corresponds to the frequency.

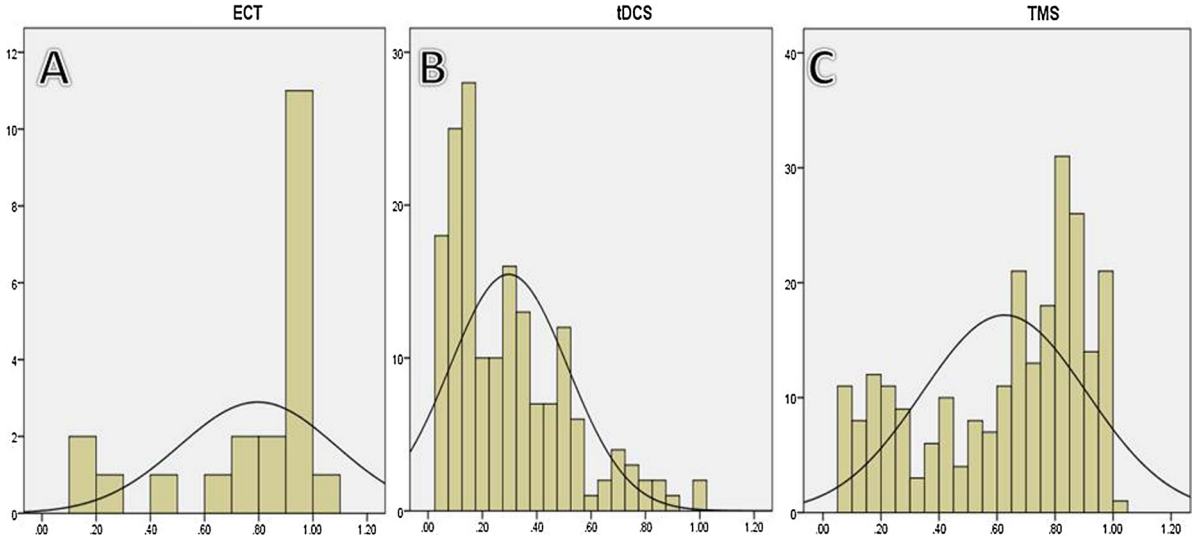

Fig. 3. Method-wise power distribution for ECT (3A), tDCS (3B) and TMS (3C). X-Axis represents the power of the individual datasets, and Y-Axis corresponds to the frequency.

publication biases (Ioannidis and Trikalinos, 2007; Button et al., 2013). The positive findings have a higher probability of getting published, and the negative results might never get circulated, or are published with significant time delays (time lag bias, Ioannidis, 1998). Further, studies conducted with low sample sizes often resort to selective data analysis and outcome reporting (Button et al., 2013) or even data fabrication (Weiss et al., 2001). Therefore, there are reasons for us to believe that a similar scenario could exist in the field of brain stimulation, and thus the summary effect sizes reported by the corresponding meta-analyses could actually be inflated. If that be the case, our calculated powers might actually be overestimations and the real powers could be lower.

There are counterviews to the assumption of 80% power as the gold-standard in research (Bacchetti, 2010), that has been criticized for artificially considering a “threshold” below which studies become meaningless. A careful consideration of these arguments show that they hold true in an ideal-world, where the papers divulge broad statistical data and go beyond the interpretation of p-values (Bacchetti, 2010). While sample-sizes and powers surely cannot be the sole determinants of the adequacy of a scientific study, we believe they go a long way in

mellowing down the inherent uncertainties in "soft-biology" research. Accordingly, several ill effects of low powered studies have been listed in literature. They miss genuine effects, and give us an impression of non-superiority of the intervention of interest (Type II error). In the field of brain stimulation, we conclude that this would have happened at least half the time (with a power of 0.5). Even when there is a report of a positive effect, the chances of it being a true one go down with the power. Termed the Positive Predictive Value (PPV), it depends on the power as: $PPV = ([1 - \beta] \times R) / ([1 - \beta] \times R + \alpha)$. In this equation, β and α are probability of type II and type I error rates respectively, and R is the pre-study odds that a probed effect is indeed non-null. '1 – β' is the power (Button et al., 2013) here. The equation clearly demonstrates that for any given values of α and R, power and PPV are positively correlated. This means that in the field of brain stimulation, assuming R of 0.2, every 3rd claim of a successful discovery is incorrect. Additionally, in pursuance of statistical significance, a study with a low power inflates the magnitude of the effect which it claims to be measuring (winner's curse; Ioannidis, 2008). This means, the findings become non-replicable at subsequent attempts with similar sample sizes (Button et al., 2013), and an erroneous conclusion can be arrived at. Interestingly, this non-reproducibility is overwhelmingly apparent in even the most methodologically and ethically-sound researches of today (Ioannidis, 2017), and is not limited to neuroscience in general and NIBS in particular.

The implication of all these is worth scientific consideration, particularly in a field like brain stimulation. Psychiatry has had its tryst with pharmacological and psychosocial interventions, and results have mostly been sub-optimal. Brain stimulation provides the field with a much-needed tool for neurobiologically-informed approach to understand and treat mental illnesses. However, in this era of evidence-driven medical practice, it is imperative that adequate and appropriate evidences be available. In order to garner a Level 1(b) evidence status, a novel intervention requires a significant result in one randomized controlled study (Masic and Miokovic, 2008). Therefore, such methods have considerable incentives to indulge in researches that would increase their chances at stumbling upon statistically significant NHST results and establish them as interventions of choice. Since most of these experiments in the field of brain stimulation are pioneering, the sample sizes are empirically calculated or arrived at through pilot studies. Both have their inherent flaws, which are perpetuated through the ensuing exploration. There are two possibilities: the study arrives at a positive result or it has a negative outcome. The correct interpretation of both these outcomes would require a good statistical power. If this is lacking, as evident in our analysis, the chances that the positive results are indeed true would be low and the likelihood of an inflated effect size would increase. On the other hand, if the results are negative, it might just be the consequence of a type-II statistical error, again due to a low power (Ioannidis, 2005). In either case, the findings would be spurious and doubtful (Krzywinski and Altman, 2013). It, therefore, seems illogical to conduct studies with such poor power. Underpowered studies waste time and resources, and their other methodological constraints (Ioannidis, 2005; Button et al., 2013) often make them out rightly unethical.

There are certain factors which drive research with low power in the field of brain stimulation. These have been elaborated by past authors, and include economic and ethical constraints put on the research groups by funding agencies and the ethics committees, a general incentive in science to publish positive results, pursuance of low-likelihood hypothesis (Button et al., 2013) for 'making ground-breaking discoveries', and opportunistic statistical manipulation of the data. But then, this field does not appear to be much different from the related fields of neuroscience (Button et al., 2013), psychiatry, neurology or somatic diseases (Dumas-Mallet et al., 2017). Multiple measures have been suggested to come out of these practices and the vicious cycle of low effect size, effect inflation and subsequent non-replicability (Ioannidis, 2005; Button et al., 2013; Krzywinski and Altman, 2013). What is required of a clinician is to be aware of such a possibility, and take findings with a pinch of salt. It would require them to read the fine-prints in a research article, and then arrive at a conclusion based on appropriate numbers.

It might not be surprising to note the possibility of non-robust meta-analyses contributing to deflated effect sizes, and a false-negative conclusion about some of the NIBS methods, in some of the cases. For example, a 2015 quantitative-review by Horvath et al. (2015) suggested a non-significant effect of single-session tDCS on cognitive functions in healthy individuals. The review was subsequently criticized (Price and Hamilton, 2015) for inconsistent data selection, mischaracterization of methodologies and faulty use of statistics; consequently, deflating the tDCS effects to non-significant levels. In fact, a rebuttal by Horvath (2015) to these comments explicitly accepts the concerns that a small number of available studies could jeopardize the conclusions of a meta-analysis, influencing the apparent effect-sizes (Horvath, 2015). On similar lines, Tremblay et al. (2014) have warned against analysing tDCS data without an a priori hypothesis and considerations for intermediate variables prompting the outcomes. There remains a possibility, therefore, that such flaws in meta-analyses data might have affected our conclusions.

Our analysis had a number of limitations. We presumed that the summary effect sizes reported in each meta-analysis were same as the mean effect sizes in the population. In face of publication biases, this might be a false assumption. Our analyses focused on a representational year of 2016 and thus, for example, missed a paper by Chen et al. (2017) comparing ECT to rTMS for major depression in pregnancy. Hence, we reiterate that encompassing multiple years for selection would have been more comprehensive. Though the papers were reviewed by a single investigator, our inclusion criteria for these were straight-forward ('any paper in which a meta-analysis has been performed'). Hence the possibility of inadvertent selection-errors creeping in is minimal. Lastly, constraints as "English language articles only" and use of PubMed as the sole search engine adds to its depictive nature. We suggest future studies using multiple search engines, with a wider time frame for inclusion of papers and screening of the included by at least two authors, to improve upon the quality and reliability of this paper.

In conclusion, the field of NIBS is fast emerging as a viable alternative (as well as a complementary method) to existent pharmacological and psychological interventions in psychiatry. However, given the novelty, the territory is largely uncharted and is littered with methodological black-holes – ready to disrupt and pull researchers into false complacency and improper conclusions. We believe that the readers need to be sceptical of fantastic claims, and the guidelines should await a clearer picture before formulating recommendations.

Financial disclosure

UMM was supported by the Wellcome Trust / DBT India Alliance Early Career Fellowship, Grant/Award Number: IA/E/12/1/500755. UMM also serves as an Associate Editor at Schizophrenia Research and receives an honorarium from Elsevier for the same. SM is currently an employee of Queensland Health, Australia.

Declaration of Competing Interest

None of the authors have any conflict of interest with respect to the current manuscript.

Acknowledgement

GV is supported by the Swarnajayanti Fellowship Grant (DST/SJF/LSA-02/2014-15) by the Department of Science and Technology, Government of India.

References

Bacchetti, P., 2010. Current sample size conventions: flaws, harms, and alternatives. BMC Med. 8, 17–24.

Barham, M.P., Enticott, P.G., Conduit, R., Lum, J.A., 2016. Transcranial electrical stimulation during sleep enhances declarative (but not procedural) memory consolidation: evidence from a meta-analysis. Neurosci. Biobehav. Rev. 63, 65–77.

Benjamin, D.J., Berger, J.O., Johannesson, M., et al., 2018. Redefine statistical significance. Nat. Hum. Behav. 2, 6–10.

Button, K.S., Ioannidis, J.P.A., Nosek, B.A., et al., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nature review Neuroscience 14 (365).

Carver, R.P., 1978. The case against statistical significance testing. Harv. Educ. Rev. 8, 378–399.

Chen, J.J., Zhao, L.B., Liu, Y.Y., et al., 2017. Comparative efficacy and acceptability of electroconvulsive therapy versus repetitive transcranial magnetic stimulation for major depression: a systematic review and multiple-treatments meta-analysis. Behav. Brain Res. 1 (March (320)), 30–36.

Chhatbar, P.Y., Ramakrishnan, V., Kautz, S., George, M.S., Adams, R.J., Feng, W., 2016. Transcranial direct current stimulation post-stroke upper extremity motor recovery studies exhibit a dose-response relationship. Brain Stimul. 9 (1), 16–26.

Chung, S.W., Hill, A.T., Rogasch, N.C., Hoy, K.E., Fitzgerald, P.B., 2016. Use of theta-burst stimulation in changing excitability of motor cortex: a systematic review and meta-analysis. Neurosci. Biobehav. Rev. 63, 43–64.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cordani, L.K., 2010. The very beginning of a class on inference: classical vs bayesian. Reading, C. (Ed.), Data and Context in Statistics Education: Towards an Evidence-Based Society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July,).

Dollfus, S., Lecardeur, L., Morello, R., Etard, O., 2016. Placebo response in repetitive transcranial magnetic stimulation trials of treatment of auditory hallucinations in schizophrenia: a meta-analysis. Schizophr. Bull. 42 (2), 301–308.

Dumas-Mallet, E., Button, K.S., Boraud, T., Gonon, F., Munafo, M.R., 2017. Low statistical power in biomedical science: a review of three human research domains. R. Soc. Open Sci. 4, 160254.

Dutra, T.G., Baltar, A., Monte-Silva, K.K., 2016. Motor cortex excitability in attention-deficit hyperactivity disorder (ADHD): a systematic review and meta-analysis. Res. Dev. Disabil. 56, 1–9.

Elsner, B., Kugler, J., Pohl, M., Mehrholz, J., 2016c. Transcranial direct current stimulation (tDCS) for idiopathic Parkinson's disease. Cochrane Database Syst. Rev. 7, CD010916.

Elsner, B., Kugler, J., Pohl, M., Mehrholz, J., 2016a. Transcranial direct current stimulation for improving spasticity after stroke: a systematic review with meta-analysis. J. Rehabil. Med. 48 (7), 565–570.

Elsner, B., Kugler, J., Pohl, M., Mehrholz, J., 2016b. Transcranial direct current stimulation (tDCS) for improving activities of daily living, and physical and cognitive functioning, in people after stroke. Cochrane Database Syst. Rev. 3, CD009645.

Enokibara, M., Trevizol, A., Shiozawa, P., Cordeiro, Q., 2016. Establishing an effective TMS protocol for craving in substance addiction: Is it possible? Am. J. Addict. 25 (1), 28–30.

Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39, 175–191.

Fond, G., Bennabi, D., Haffen, E., Brunel, L., Micoulaud-Franchi, J.A., Loundou, A., Lançon, C., Llorca, P.M., Auquier, P., Boyer, L., 2016. A Bayesian framework systematic review and meta-analysis of anesthetic agents effectiveness/tolerability profile in electroconvulsive therapy for major depression. Nature Sci Rep. 6 (19847).

Geddes, J., 2000. Evidence-based practice in mental health. In: Trinder, L., Reynolds, S. (Eds.), Evidence Based Practice. Blackwell Science, pp. 66–88.

Gelman, A., Carlin, J., 2014. Beyond power calculations: assessing type S (Sign) and type m (Magnitude) errors. Perspect. Psychol. Sci. 9 (6), 641–651.

Hashemirad, F., Zoghi, M., Fitzgerald, P.B., Jaberzadeh, S., 2016. The effect of anodal transcranial direct current stimulation on motor sequence learning in healthy individuals: a systematic review and meta-analysis. Brain Cogn. 102 (February), 1–12.

Health Quality Ontario, 2016. Repetitive transcranial magnetic stimulation for treatment-resistant depression: a systematic review and meta-analysis of randomized controlled trials. Ont. Health Technol. Assess. Ser. 16 (5), 1–66.

Hill, A.T., Fitzgerald, P.B., Hoy, K.E., 2016. Effects of anodal transcranial direct current stimulation on working memory: a systematic review and meta-analysis of findings from healthy and neuropsychiatric populations. Brain Stimul. 9 (2), 197–208.

Horvath, J.C., 2015. New quantitative analyses following price & hamilton's critique do not change original findings of Horvath et al. Brain Stimul. 8, 665–666.

Horvath, J.C., Forte, J.D., Carter, O., 2015. Quantitative review finds No evidence of cognitive effects in healthy populations from single-session transcranial direct current stimulation (tDCS). Brain Stimul. 8 (3), 535–550.

Ioannidis, J.P., 2017. Acknowledging and overcoming nonreproducibility in basic and preclinical research. JAMA 317 (10), 1019–1020.

Ioannidis, J.P., 1998. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 279, 281–286.

Ioannidis, J.P., 2005. Why most published research findings are false. PLoS Med. 2, e124.

Ioannidis, J.P., Trikalinos, T.A., 2007. An exploratory test for an excess of significant findings. Clin. Trials 4, 245–253.

Ioannidis, J.P., 2008. Why most discovered true associations are inflated. Epidemiology 19, 640–648.

Kang, N., Summers, J.J., Cauraugh, J.H., 2016. Transcranial direct current stimulation facilitates motor learning post-stroke: a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry. 87 (April (4)), 345–355.

Knijnik, L.M., Dussán-Sarria, J.A., Rozisky, J.R., Torres, I.L., Brunoni, A.R., Fregni, F., Caumo, W., 2016. Repetitive transcranial magnetic stimulation for fibromyalgia: systematic review and meta-analysis. Pain Pract. 16 (3), 294–304.

Krzywinski, M., Altman, N., 2013. Power and sample size. Nat. Methods 10 (12), 1139.

LeMire, S.D., 2010. An argument framework for the application of null hypothesis statistical testing in support of research. J. Stat. Educ. 18 (2), 1–23.

Lindley, D.V., 1990. The 1988 Wald Memorial Lectures: The present position in Bayesian statistics. Stat. Sci. 5 (1), 82–83.

Masic, I., Miokovic, M., 2008. Muhamedagic B. Evidence based medicine – new approaches and challenges. Acta Inform. Med. 16 (4), 219–225.

Mehta, U.M., 2016. The promises and challenges of brain stimulation in psychiatry. Asian J. Psychiatr. 21, A1.

Mitra, S., Kavoor, A.R., Mahintamani, T., 2018. The butterflies in the brain − What would it take to understand the genetic basis of psychiatric disorders? Asian J. Psychiatr. 31, 13–14.

Nickerson, R.S., 2000. Null hypothesis significance testing: a review of an old and continuing controversy. Psychol. Methods 5 (2), 241–301.

Pisegna, J.M., Kaneoka, A., Pearson Jr, W.G., Kumar, S., Langmore, S.E., 2016. Effects of non-invasive brain stimulation on post-stroke dysphagia: a systematic review and meta-analysis of randomized controlled trials. Clin. Neurophysiol. 127 (1), 956–968.

Price, A.R., Hamilton, R.H., 2015. A Re-evaluation of the cognitive effects from single-session transcranial direct current stimulation. Brain Stimul. 8 (3), 663–665.

Soleimani, R., Jalali, M.M., Hasandokht, T., 2016. Therapeutic impact of repetitive transcranial magnetic stimulation (rTMS) on tinnitus: a systematic review and meta-analysis. Eur. Arch. Otorhinolaryngol. 273 (7), 1663–1675.

Summers, J.J., Kang, N., Cauraugh, J.H., 2016. Does transcranial direct current stimulation enhance cognitive and motor functions in the ageing brain? A systematic review and meta- analysis. Ageing Res. Rev. 25, 42–54.

Takekita, Y., Suwa, T., Sunada, N., Kawashima, H., Fabbri, C., Kato, M., Tajika, A., Kinoshita, T., Furukawa, T.A., Serretti, A., 2016. Remifentanil in electroconvulsive therapy: a systematic review and meta-analysis of randomized controlled trials. Eur. Arch. Psychiatry Clin. Neurosci. 266 (8), 703–717.

Tedesco Triccas, L., Burridge, J.H., Hughes, A.M., Pickering, R.M., Desikan, M., Rothwell, J.C., Verheyden, G., 2016. Multiple sessions of transcranial direct current stimulation and upper extremity rehabilitation in stroke: a review and meta-analysis. Clin. Neurophysiol. 127 (January (1)), 946–955.

Tremblay, S., Lepage, J.F., Latulipe-Loiselle, A., et al., 2014. The uncertain outcome of prefrontal tDCS. Brain Stimul. 7, 773–783.

Trevizol, A.P., Barros, M.D., Silva, P.O., Osuch, E., Cordeiro, Q., Shiozawa, P., 2016. Transcranial magnetic stimulation for posttraumatic stress disorder: an updated systematic review and meta-analysis. Trends Psychiatry Psychother. 38 (1), 50–55.

Wallace, J., 2011. The practice of evidence-based psychiatry today. Adv. Psychiatr. Treat. 17 (5), 389–395.

Weiss, R.B., Gill, G.G., Hudis, C.A., 2001. An on-site audit of the South African trial of high-dose chemotherapy for metastatic breast cancer and associated publications. J. Clin. Oncol. 19, 2771–2777.