A Methodological Evaluation of Meta-Analyses in tDCS - Motor Learning Research

Master thesis

Submitted by:

Taym Alsalti

Matriculation number: 5399182

14.12.2021

Primary supervisor: Prof. Dr. Steffi Pohl

Secondary supervisor: Univ.-Prof. Dr. med. Felix Blankenburg

Master Program Social Cognitive and Affective Neuroscience

Department of Education and Psychology

Freie Universität Berlin

Abstract

With transcranial direct-current stimulation's (tDCS) rising popularity both in motor learning research and as a commercial product, it is becoming increasingly important that the quality of evidence on its effectiveness be evaluated. Special attention should be paid to meta-analyses, as they usually have a larger impact on research and clinical practice than other types of studies. There is evidence for several methodological issues in the tDCS motor learning literature such as limited reproducibility, largely untested replicability, and marked hetereogeneity in the technical paramaters employed. These issues inevitably impact the quality of the meta-analyses synthesising these studies, which display methodological problems of their own. The aim of this project was to evaluate the methodological quality of meta-analyses estimating the effect of tDCS on motor learning with respect to three main aspects: reporting quality, reproducibility, and publication bias control. Akin to previous reviews with similar aims, we show that the methods and results sections of meta-analyses are severely underreported, which compromises the ability to judge the soundness of the methodogloical procedure adopted as well as its reproducibility. Furthermore, only one of the three meta-analyses reported having taken non-statistical approaches to control for publication bias and all three used older statistical methods which are known to produce results of limited validity. These results reemphasise the need to evaluate the quality of meta-analyses as they have the largest impact on research, clinical practice, and even consumer behaviour.

A Methodological Evaluation of Meta-Analyses in tDCS - Motor Learning Research

Introduction

Transcranial direct-current stimulation

Transcranial direct-current stimulation (tDCS) refers to a non-invasive brain stimulation technique which involves delivering constant, low current to the brain via electrodes fixed on the scalp (Gazzaniga, Ivry, & Mangun, 2018). tDCS is believed to be capable of both temporarily and permanently altering brain function by increasing or decreasing excitability in targeted cortical regions depending on the type of stimulation implemented (Stagg & Nitsche, 2011). Positive current delivered via the anodal electrode causes neurons just beneath the region stimulated to depolarise, that is, to increase their excitability, rendering them more likely to start off an action potential. Neurons targeted by negative stimulation, on the other hand, become hyperpolerised, i.e., less likely to "fire" (Gazzaniga et al., 2018). An important third type of stimulation is sham stimulation, which involves delivering an initial brief current that quickly fades off, leaving an inactive electrode for the remainder of the experiment. Sham stimulation is used as the control condition in experimental settings. tDCS is known to be safe and to cause negligible side-effects, if any (Gianni et al., 2021; Nitsche et al., 2008).

The typical set-up of a tDCS protocol involves no more than a handful of inexpensive components and relatively simple steps (Gebodh et al., 2019; Woods et al., 2016). A standard tDCS device consists of an anodal (positively charged) and a cathodal (negatively charged) electrode attached to a battery-powered device which controls the type and intensity of the current as well as the duration of the stimulation. To ensure that the administered current actually reaches the target area, a few measures are usually taken to minimise resistance between the skin and the active electrodes. This includes using a sponge and an electrolyte-based contact medium such as gel and any materials necessary to

make sure the electrodes remain fixed in their positions. Electroencephalography (EEG) or magnetic resonance imaging (MRI) is sometimes used to more accurately locate the region of interest or to monitor the physiological effects of tDCS in real time (Nitsche et al., 2008; Woods et al., 2016). More recent variations of tDCS used for at-home clinical interventions are even simpler (Riggs et al., 2018).

When considering these advantages, tDCS's rapidly increasing popularity over the last two decades is not surprising (Buch et al., 2017). In clinical research, its efficacy for treating or attenuating depression (Brunoni et al., 2016), memory deficits in Alzheimer's patients (Bennabi et al., 2014), pain (Luedtke et al., 2012), schizophrenia (Liu et al., 2021), and others, has been investigated. Interest in tDCS was not restricted to clinical settings: studies on healthy subjects have been conducted to test its effects on cognitive abilities such as language and memory (Horvath, Forte, & Carter, 2015), affective states (Austin et al., 2016), and motor skills, such as surgery (Hung et al., 2021a) and musical performance (Rosen et al., 2016).

Indeed, tDCS has become so established as a research tool that it is already approved for clinical use (with some restrictions) in several countries around the world (Fregni et al., 2015). Furthermore, an ever-richer variety of commercial tDCS products has become available on the market (Davis, 2016; Wexler, 2018; Zettler, 2017), prompting experts (Wurzman, Hamilton, Pascual-Leone, & Fox, 2016) to voice concerns over the increasing prevalence of this "do-it-yourself" use of tDCS. Out of the 449 such at-home tDCS consumers surveyed by Wexler (2018), 52% (237) did so for enhancement purposes. It appears likely that this "hype" surrounding tDCS products is partly due to considerable media coverage (Dubljević, Saigle, & Racine, 2014; Steenbergen et al., 2016).

tDCS and motor learning

One increasingly prevalent implementation of tDCS is to improve motor learning, that is, the acquisition of new motor skills or the improvement of existing ones. Here,

complexity is compounded by the multifacetedness of motor skills and the variety of ways to measure them (Buch et al., 2017). They can range from the most mundane daily activities like walking or doing the dishes to finer skills that can require years to master such as playing a musical instrument or performing surgery. tDCS is most often used as a supplement to specialised training (Buch et al., 2017). The two main types of improvement that can occur are referred to as "online learning" (improvement over short time periods such as within a single training session or a day) and consolidated learning (over several hours, days, or training sessions).

Experimental paradigms aiming to evaluate such patterns of improvement typically take the following form (Buch et al., 2017; Reis et al., 2009): following baseline measurements of different outcomes on the first day, study participants undergo a series of training sessions in a specific motor task (e.g., squeezing a hand-held force transducer to move the cursor on the computer screen quickly and accurately between a starting position and a target position). The participants receive tDCS (or sham, depending on group and time point) during this training. Performance is measured subsequently (online) and/or 1 to several weeks or months later (consolidation/retention). Measuring motor learning is a complex task, too, as motor skills require an optimisation of a speed-accuracy trade-off. Different tasks and measurement instruments take this into account in different ways depending on context. For example, some tasks involve instructing participants to focus on either speed or accuracy. Another approach is to stick to neutral instructions and instead explicitly model this trade-off, as done by Reis et al. (2009):

$$\text{skill} = \frac{1 - \text{error rate}}{\text{error rate}(\ln(\text{duration})^b)} \tag{1}$$

Where b is a fixed constant.

Issues in tDCS research

Although tDCS might hold promise for those and other applications, treating tDCS's positive effects as an established fact is deemed by experts in the field as premature (Buch et al., 2017). The technique's apparent simplicity might lead researchers to disregard aspects which are crucial to obtaining reliable results (Gebodh et al., 2019; Woods et al., 2016). When conducting a tDCS experiment, the researcher must select the type of stimulation, electrode size, shape and material; the duration of the stimulation; the location of the electrodes; and the current intensity, all of which are factors that have a substantial impact on stimulation efficacy.

The wide range of different possible tDCS settings appears to be well utilised by researchers as there is evidence of considerable heterogeneity in the literature with respect to tDCS-related parameters, tasks, and outcomes (Buch et al., 2017). tDCS research also appears to suffer from the same methodological problems as neighbouring fields in the behavioural and cognitive (neuro)sciences, such as suboptimal reproducibility, untested replicability, underpowered studies (Minarik et al., 2016), and publication bias (Button et al., 2013). These issues make it difficult to evaluate the evidence or reach conclusions based on it. They also hamper cumulative science as it is difficult to synthesise the results of studies for the purpose of a meta-analysis when they vary too much in quality and methodological approach (Borenstein, Hedges, Higgins, & Rothstein, 2009b).

Meta-analysis

Meta-analysis refers to a method of quantitatively synthesising the results of a set of studies while weighing individual studies differently depending on certain criteria (Borenstein & Hedges, 2019). The most common measure used for weighing effect sizes derived from primary studies is the sampling variance, which is largely dependent on the study's sample size. For example, should one wish to estimate the effect on an intervention

based on multiple studies with a control group, a common procedure would be to calculate a standardised mean difference (SMD) in the form of a Cohen's d:

$$d = \frac{\bar{x}_t - \bar{x}_c}{s} \tag{2}$$

$$s = \sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{n_t + n_c - 2}} \tag{3}$$

Where $\bar{x}_t$ is the mean of the treatment group, $\bar{x}_c$ is the mean of the control group, s is the pooled standard deviation, n_t and n_c are the sample sizes of the treatment and control groups, respectively, and s_t and s_c are the standard deviations of the treatment and control groups, respectively.

The next step would be to calculate the inverse-variance weight $w = \frac{1}{v_i}$ for each d, where the sampling variance v_i is conventionally calculated via:

$$v_i = \frac{n_t + n_c}{n_t n_c} v \tag{4}$$

Where v represents the pooled variance of the two groups. One can then calculate the pooled Cohen's d using the primary ds and their corresponding weights:

$$\bar{d} = \frac{\sum_{i=1}^{k} w_i d_i}{\sum_{i=1}^{k} w_i} \tag{5}$$

This would be a so-called fixed-effects meta-analytic model, under which one assumes the primary effect sizes to stem from a single common true effect size (Konstantopoulos & Hedges, 2019). On the other hand, the true effect effect size is itself random and has its own distribution under a random-effects model. Computationally, the only difference between a fixed-effects and a random-effects model is that for the purpose of the latter, the sampling variances v_i^* are formed by adding to the fixed-effects sampling variance a between-study variance component τ^2. There are about a dozen different methods to

estimate the between-study heterogeneity, see Veroniki et al. (2016) for an overview.

Meta-analysis studies merit special scrutiny and should be invariably executed to the highest standards as they are usually cited more frequently than primary studies about the same topics, are commonly assumed to provide the most accurate estimate of an effect, and have a large impact on theory development as well as policy and clinical practice (Gopalakrishnan & Ganeshkumar, 2013; Gøtzsche, Hróbjartsson, Marić, & Tendal, 2007; Ioannidis, 2016; Lakens et al., 2017; Morganti, 2007). However, meta-analyses can and often *do* suffer from methodological problems which go beyond those of the primary studies they synthesise (Ioannidis, 2016). Here, we focused on the issues of reporting quality, reproducibility, and publication bias control.

Reporting quality

One crucial issue is reporting completeness describing the used methods and the results. Researchers must make a great number of decisions at multiple stages of a study's timeline. These decisions impact the outcome of the study to various degrees and only by providing a full, justified account of which path was taken at each fork can other researchers evaluate the soundness of the methodological approach (Botvinik-Nezer et al., 2020; Gelman & Loken, 2013; Hyatt et al., 2020; Simmons, Nelson, & Simonsohn, 2011). Meta-analysts in particular are faced with a plethora of different decisions with regards to which databases to search and using which strings; primary study selection and exclusion; data extraction; statistical methods, among others (Ada, Sharman, & Balkundi, 2012; Geyskens, Krishnan, Steenkamp, & Cunha, 2009; Guzzo, Jackson, & Katzell, 1987; Nieminen, Nicklin, McClure, & Chakrabarti, 2011; Schalken & Rietbergen, 2017; Valentine, Pigott, & Rothstein, 2010; Voracek, Kossmeier, & Tran, 2019). Although the extent to which these "researcher degrees of freedom" (Simmons et al., 2011) impact the conclusions of the meta-analysis is not a completely uncontroversial issue[1], there is consensus regarding

[1] One review of meta-analyses in the organisational sciences found that 21 different analytical decisions have a negligible impact on the pooled ES estimate (Aguinis, Dalton, Bosco, Pierce, & Dalton, 2011)

the importance of transparently reporting these decisions (Aguinis, Pierce, Bosco, Dalton, & Dalton, 2011), as it is difficult to assess the quality and trustworthiness of that which one does not have access to (Page, McKenzie, et al., 2021).

Hundreds of different sets of reporting guidelines have been developed for different types of studies conducted across various fields ("Reporting guidelines | The EQUATOR Network," n.d.) including several for meta-analyses, for example, the Cochrane Handbook (Higgins et al., 2019), Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA, Moher et al., 2000; Moher, Liberati, Tetzlaff, Altman, & Group, 2009; Page, McKenzie, et al., 2021), Meta-Analysis Reporting Standards (MARS, American Psychological Association, n.d.), with PRISMA being the most widely known and adopted (Page & Moher, 2017).

PRISMA's predecessor, The Quality of Reports of Meta-analyses Statement (QUOROM, Moher et al., 2000), was developed in response to a rapidly rising number of meta-analyses with widely varying degrees of reporting completeness. The statement proceeded from a conference held to discuss these issues surrounding research synthesis, especially randomised controlled trials (RCTS). Its aim was to provide meta-analysts with help to better report their studies. This help came in the form of a flow diagram of the different stages of a meta-analysis and a checklist. The checklist includes 18 items about how the different sections of the meta-analytic study should look like.

The updated versions of the reporting guidelines (rebranded as PRISMA to encompass reviews without quantitative synthesis, Moher et al., 2009; Page, McKenzie, et al., 2021) elaborated on both the checklist and the flow chart (see Figure 1 for the 2009 edition). Here we focused on the 2009 version of PRISMA as we expected most meta-analyses we evaluated to be published after 2009 but before 2020.

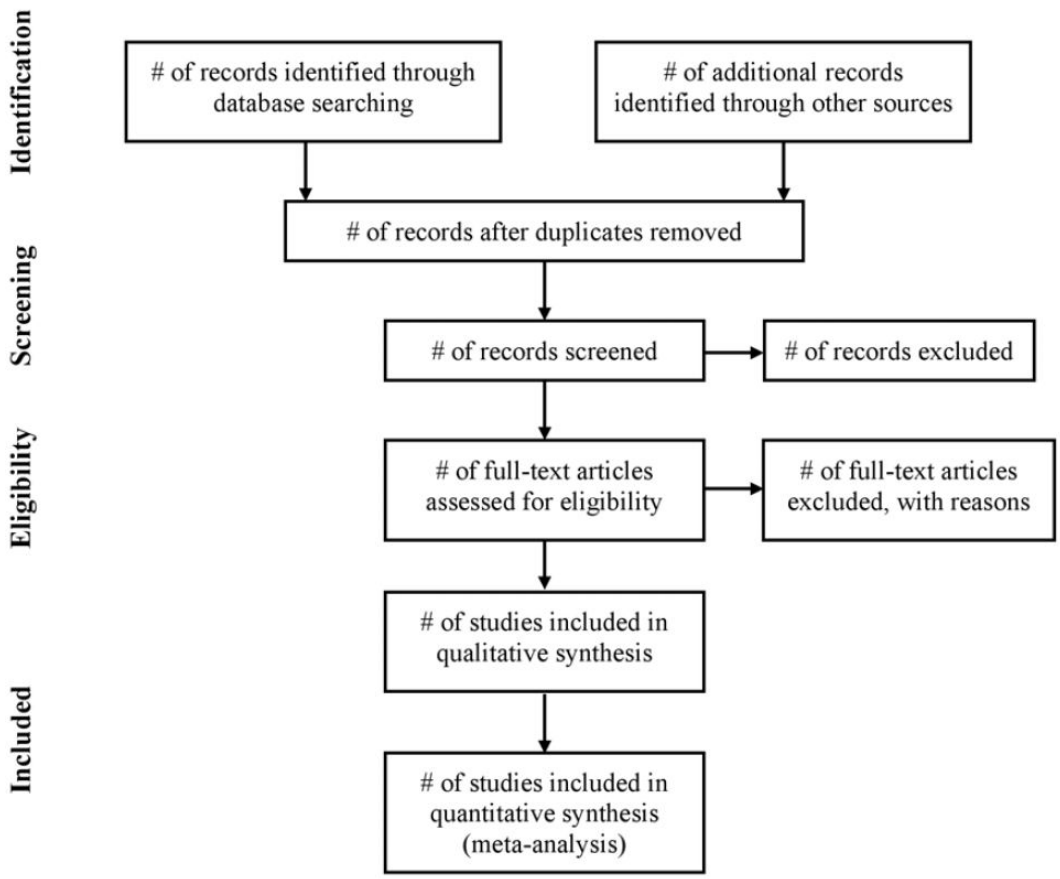


Figure 1. PRISMA flow chart of the meta-analytical process. Adapted as is from Moher et al. (2009).

The 2009 PRISMA (henceforth simply PRISMA) checklist included 27 items (see Table 1), with 12 of them pertaining to methods and 7 to results. We paid special attention to items 6 to 9 and 17 as they pertain to important methodological aspects (study selection) and are not covered by our evaluation of reproducibility.

Table 1
PRISMA (Moher et al., 2009) checklist.

Item number	Section	Original description of item
1	Title	Identify the report as a systematic review, meta-analysis, or both.
2	Abstract	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.
3	Introduction	Describe the rationale for the review in the context of what is already known.
4	Introduction	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).
5	Methods	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.
6	Methods	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.
7	Methods	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.
8	Methods	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.
9	Methods	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).
10	Methods	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.
11	Methods	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.
12	Methods	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.
13	Methods	State the principal summary measures (e.g., risk ratio, difference in means).
14	Methods	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., 12) for each meta-analysis.
15	Methods	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).
16	Methods	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.
17	Results	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.
18	Results	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.

Table 1
PRISMA (Moher et al., 2009) checklist. (continued)

Item number	Section	Original description of item
19	Results	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).
20	Results	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.
21	Results	Present results of each meta-analysis done, including confidence intervals and measures of consistency.
22	Results	Present results of any assessment of risk of bias across studies (see Item 15).
23	Results	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).
24	Discussion	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).
25	Discussion	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).
26	Discussion	Provide a general interpretation of the results in the context of other evidence, and implications for future research.
27	Funding	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.

Although QUOROM and other reporting guidelines have been available for over two decades, meta-scientific evaluations of adherence to PRISMA and other guidelines have shown that reporting standards of meta-analyses are generally suboptimal and that these guidelines are rarely fully adhered to (see Table 2 for an overview of such studies). For example, the most extensive review of adherence to PRISMA (Page & Moher, 2017), which aggregated the results of 57 meta-scientific reviews (which included a total of 6487 systematic reviews) with this research question, found that 11 items were adhered to by less than 67% of the systematic reviews. These include item 5 (methods: protocol and pre-registration, <10% of systematic reviews), item 8 (methods: search, <50%), item 11 (methods: data items, <65%), item 12 (methods: risk of bias in individual studies, <67%), item 15 (methods: risk of bias across studies, <50%), item 16 (methods: additional analyses, <50%), among others. Items 6 (methods: eligibility criteria), 7 (methods: information sources), 9 (methods: study selection), and 17 (results: study selection) were

adhered to in ~85%, ~85%, ~67%, and ~72% of the systematic reviews.

Table 2
Studies on reporting quality and adherence to reporting standards.

Review	Methods	Conclusions
Organisational sciences: Aytug et al. (2012); Kepes et al. (2013); Schalken & Rietbergen (2017)	Reviewed the reporting quality and adherence to MARS guidelines of meta-analyses.	Inadequate reporting and adherence to guidelines.
Educational science: Ahn et al. (2012)	Reviewed the methodological quality of meta-analyses.	"[the meta-analyses] followed general recommendations fairly well in problem formulation and data collection, but much improvement is needed in data evaluation and analysis."
Psychological science: Dieckmann et al. (2009); Hohn et al. (2020); Polanin et al. (2020)	Reviewed the quality of reporting and methods used in meta-analyses, respectively.	Considerable variability in transparency and methods used. Still insufficient despite improvements over the years.
Biomedical sciences: Page & Moher (2017); Pussegoda et al. (2017)	Summarised the findings of meta-research reviews on the adherence of systematic reviews and meta-analyses to reporting guidelines.	Reporting of several PRISMA items is suboptimal. Reporting quality varied substantially across items of 4 well known sets of reporting guidelines.

Reproducibility

A related issue is that the less information a meta-analyst provides about their analytical procedures, the harder it is to reproduce their meta-analysis (Aguinis et al., 2011; Gøtzsche et al., 2007; Lakens et al., 2017; Maassen, Assen, Nuijten, Olsson-Collentine, & Wicherts, 2020). Reproducibility (and relatedly, replicability) remain

a hot topic in the behavioural and biomedical sciences, with various fuzzy or inconsistent definitions still floating around (Goodman, Fanelli, & Ioannidis, 2016; Plesser, 2018). Both terms have been somewhat indiscriminately used to refer to the ability to obtain the same results of a given study by repeating its described procedure. One widely adopted way to distinguish them is by defining reproducibility as the ability to obtain the numerical findings of the original work using their methods (and if applicable, data pre-processing and analysis code) and the original data. (Direct) replication, on the other hand, refers to repeating the entire study using *new data*. (Broman et al., 2017).

This definition, however, does not do full justice to the nuances of the concept of reproducibility. Importantly, it fails to account for the different types of reproducibility and emphasises one type in particular: *results reproducibility* (Goodman et al., 2016). Two further types can be differentiated: *methods reproducibility*, which refers to "the provision of enough detail about study procedures and data so the same procedures could, in theory or in actuality, be exactly repeated" (Goodman et al., 2016, p. 2), and *inferential reproducibility*, which describes the ability to draw "qualitatively similar conclusions from either an independent replication of a study or a reanalysis of the original study" (Goodman et al., 2016, p. 4). For example, a study might report its methods extensively enough for it to be easily reproducible methodologically, but for which one obtains different results from the original upon a reproduction attempt. This could occur due to erroneous descriptions of methods or mistakes in the implementation of the described methods.

Similarly, inferential reproducibility can depend on the magnitude of the effect reported in a given study as two interpreters of the same results might be in stark disagreement regarding the inferences which can be drawn from them. A p-value of 0.04, for example, might have an entirely different "significance" for a hard frequentist than for a more Bayes-inclined researcher. Another limitation to the definition above is its implication of a dichotomous nature of reproducibility. Especially with regards to methods and inferential reproducibility, it is somewhat imprudent to think of reproducibility as a

binary feature since there are many factors (see methodological decisions discussed above) that might play a role in deeming a study reproducible or not (Broman et al., 2017). Concretely, although it is possible to construct a framework which allows one to make an unambiguous yes or no decision when the results of a study are reproducible (e.g., see Steiner, Wong, & Anglin, 2019), methods reproducibility is better served by considering it as a function of how much information is provided about the methodological procedure.

Applied to meta-analysis, evaluating all types of reproducibility involves attempting to repeat 1. study search, 2. study screening and selection, 3. data extraction, 4. computing ES estimates for each primary study included (primary ES), 5. computing the main pooled ES estimate (pooled ES), and often 6. computing "corrected" pooled ESs, 7. computing multiple pooled ESs based on study characteristics and/or conducting subgroup analyses (Cooper, Hedges, & Valentine, 2009). Here, the issue of reproducibility acquires yet another layer of complexity: reproducibility with regards to elements 3 to 7 depends not only on what is reported in the meta-analysis, but also what is reported in the included primary studies. How hard it is to reproduce a primary ES, for example, is a function of the amount and precision of the information the meta-analysts reported about how they computed the ES *and* the degree to which this information corresponds to data that is accessible in the primary study.

Several reviews testing the reproducibility of meta-analyses have been conducted (Gøtzsche et al., 2007; Lakens et al., 2017; Maassen et al., 2020, see Table 3)[2]. They emphasised somewhat different methodological aspects (e.g., complete reporting vs. computational correctness) but were uniform in their focus on reproducing data extraction and ES computation (both primary and pooled). Their conclusions about the

[2] Three further relevant reviews in the biomedical sciences were identified after the study was done and could not have influenced our approach: two evaluating the methodological reproducibility of meta-analyses (Page et al., 2018; i.e., without attempting to actually reproduce them, Wayant, Page, & Vassar, 2019) and one testing the reproducibility of entire meta-analyses (i.e., including study search and screening, Ford, Guyatt, Talley, & Moayyedi, 2010). These reviews, too, reached the conclusion that meta-analysis reproducibility is mostly limited.

reproducibility of meta-analyses were, although of varyingly grave consequences, also similar: the reproducibility of meta-analyses was severely limited due to under-reporting of methods and results as well as errors. Despite the growing popularity of tDCS and the ensuing increase in number of meta-analyses estimating its effect on motor learning, no evaluation of the reporting quality or reproducibility of meta-analyses in this field has been conducted as far as we are aware.

Table 3
Studies on reproducibility of meta-analyses.

Review	Methods	Findings	Reasons for Reproducibility
Gøtzsche et al. (2007)	Attempted to reproduce some of the SMDs reported in 27 biomedical meta-analyses with the objective of testing whether whether SMDs in meta-analyses are accurate. Contacted authors of meta-analyses to acquire unreported information.	"In total, 17 meta-analyses (63%) had errors for at least 1 of the 2 trials examined. For the 10 meta-analyses with errors of at least 0.1, we checked the data from all the trials and conducted our own meta-analysis, using the authors' methods. Seven of these 10 meta-analyses were erroneous (70%); 1 was subsequently retracted, and in 2 a significant difference disappeared or appeared."	"Common problems were erroneous number of patients, means, SDs [standard deviations], and sign for the effect estimate."
Lakens et al. (2017)	Attempted to reproduce all primary ESs as well as the pooled ES reported in 20 psychological meta-analyses. More emphasis on transparent reporting of methods and results than on correctness of analysis: classified meta-analyses that did not report primary ESs for each included study as not reproducible.	- "Five of the 20 randomly selected meta-analyses we attempted to reproduce could not be reproduced at all [. . .]" - "[. . .] differences between the reported and reproduced ES or sample size were common"	"[meta-analyses could not be reproduced] due to lack of access to raw data, no details about the ESs extracted from each study, or a lack of information about how ESs were coded

Table 3
Studies on reproducibility of meta-analyses. (continued)

Review	Methods	Findings	Reasons for Reproducibility
Maassen et al. (2020)	Attempted to reproduce 500 primary ESs reported in 33 psychological meta-analyses. Only included meta-analyses that reported primary ESs for each included primary study. Distinguished between irreproducibility of a primary ES due to lack of information and irreproducibility due to incorrect calculations.	234 out of the 500 primary ESs could not be reproduced.	54 out of the 234 could not be reproduced due to lack of necessary information (e.g., standard deviations in primary study not reported), 74 were incorrect (reproduced ES different to reported), 96 were "ambiguous", that is, "it was unclear what procedure was followed by the meta-analysts."

Publication bias control

Beyond transparency in reporting and any effect it may have on reproducibility, a further issue that merits consideration when evaluating the methodology of meta-analyses is how the meta-analysts dealt with publication bias. Publication bias, that is, the tendency of researchers to suppress findings that do not go their way ("negative" results) and journals to selectively publish studies that report significant results, has been known to distort the scientific literature for over six decades, having been first discussed in 1959 by Sterling. It remains a major concern, especially in the "softer" social sciences (Fanelli, 2010), where the proportion of studies reporting "positive" results is markedly higher than in the physical sciences. This remarkably elevated rate of positive results combined with the meagre average power of studies in the behavioural and neurosciences give a clear indication of the presence of publication bias (Button et al., 2013; Szucs & Ioannidis, 2017; Szucs & Ioannidis, 2020).

Since this apparent overrepresentation of positive results can (and often does) inflate the ES estimates of meta-analyses in a certain direction (Borenstein, Hedges, Higgins, &

Rothstein, 2009a; Friese & Frankenbach, 2020; Vevea, Coburn, & Sutton, 2019), meta-analysis reporting guidelines such as PRISMA have consistently recommended reporting any attempts to account for publication bias (Moher et al., 2000, 2015; Page, McKenzie, et al., 2021). Meta-analysts have at their disposal several measures they can take to mitigate the ubiquitous effects of publication bias. Locating as many unpublished studies as possible is likely to be the most effective method and searching repositories of potentially unpublished results are mandatory (e.g., ClinicalTrials.gov) when conducting a Cochrane systematic review (Higgins et al., 2019). Other such non-statistical approaches to minimising the effect of publication bias include searching pre-print and theses repositories, contacting authors of relevant studies to inquire about potentially file-drawered studies, not restricting the study search to articles written in English, etc..

This endeavour proving fruitless, the meta-analyst may next (or in addition) attempt to detect the presence of publication bias and/or adjust the meta-analytic ES estimate for it statistically. Many procedures for this purpose have been developed in the last three decades displaying varying performance in general and under certain conditions (for extensive reviews/comparisons of these methods see Carter, Schönbrodt, Gervais, & Hilgard, 2019; Harrer, Cuijpers, Furukawa, & Ebert, 2021; Marks-Anglin & Chen, 2020; McShane, Böckenholt, & Hansen, 2016; Renkewitz & Keiner, 2019; Vevea et al., 2019). In the following, we will briefly introduce the methods used either in the meta-analyses we reviewed or by us.

These statistical methods can be divided into two main categories: small-study effects-based and p-values-based. Small-study effect methods have been available for more than two decades and are very widely used [at least among the ~50% of meta-analyses which conduct such analyses pageEvaluationsUptakeImpact2017; Vevea et al. (2019); Borenstein et al. (2009a); Harrer et al. (2021)]. In essence, they are based on the often observed positive correlation between the ES derived from a primary study and the ES's standard error (SE, which can be seen as the inverse of the study's total sample size, N).

This association between N and ES should not exist, theoretically, but does due to many potential reasons, one of which being publication bias. The main assumption here is that large studies mostly get published (and thus become easily findable by meta-analysts) regardless of their outcome, whereas small studies only get published when they are significant, which can (due to the small N) only happen when the ES is large. Some popular examples of methods in this category include:

- Trim-and-fill (Duval & Tweedie, 2000b): a simple non-parametric method which is based on an earlier, purely graphical, procedure, the funnel plot (Light & Pillemer, 1984). The trim-and-fill method attempts to correct for bias-induced asymmetry in the funnel plot by 1. *trimming* an arbitrary number k of small studies with large ESs, usually the k rightmost studies on the funnel plot, 2. computing the pooled ES based on this new set of studies, 3. imputing for each trimmed study a *filled* study which mirrors it on the other side of the pooled ES computed in the previous step (i.e., identical SE, filled ES = trimmed ES − pooled ES from step 2), 4. computing the pooled ES based on both the original studies and the trimmed ones. Example results of this procedure are depicted in Figure 2.

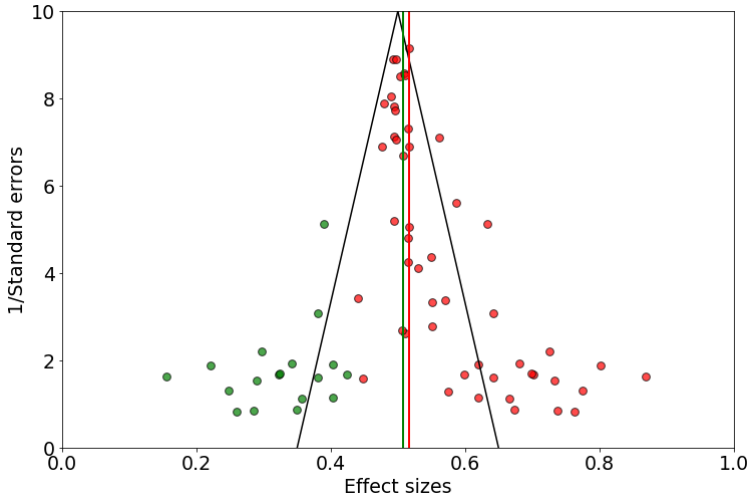


Figure 2. Funnel plot of simulated data illustrating asymmetry and the trim-and-fill method. The red dots represent the observed studies, the green dots the filled studies. The red and green lines are the fixed-effect pooled ES estimates of the observed studies only and the observed plus the filled studies, respectively.

- Begg and Mazumdar's (1994) rank correlation approach is another non-parametric test of funnel plot asymmetry. It involves calculating a normalised version of Kendall's Tau between the deviations of individual ESs from their fixed-effects pooled mean and their sampling variances. This statistic (the normalised Tau) is then tested for significance in reference to the standard normal distribution under the null hypothesis that no association between the effect and variance exists (Vevea et al., 2019).
- Egger's test (Egger, Smith, Schneider, & Minder, 1997) parametrically checks for funnel plot asymmetry by regressing the ratio of ES to SE on the inverse of SE:

$$\frac{\hat{\theta}_k}{\hat{\sigma}_{\hat{\theta}_k}} = \beta_0 + \beta_1 \frac{1}{\hat{\sigma}_{\hat{\theta}_k}} \tag{6}$$

  Where $\hat{\theta}_k$ is the ES estimate of the kth primary study, $\hat{\sigma}_{\hat{\theta}_k}$ is the corresponding SE, and β_0 and β_1 are the intercept and the regression coefficient, respectively. The resulting intercept indicates asymmetry if it is significantly larger than zero. This is assumed to occur when, due to bias, small studies with very large ESs are overrepresented, whereas the intercept (the expected scaled ES value as SE goes to infinity) should theoretically equal zero.

- PET-PEESE (T. Stanley, 2008; T. D. Stanley & Doucouliagos, 2014) is a more recent method that is gaining in popularity. Similarly to Egger's test, both the precision-effect test (PET) and precision-effect estimate with standard error (PEESE) regress the ES on a proxy of its precision, namely the standard error in the case of PET,

$$\hat{\theta}_k = \beta_0 + \beta_1 \hat{\sigma}_{\hat{\theta}_k} \tag{7}$$

  and the variance in the case of PEESE.

$$\hat{\theta}_k = \beta_0 + \beta_1 \hat{\sigma}^2_{\hat{\theta}_k} \tag{8}$$

However, the theoretical idea motivating this manoeuvre is quite different: here it is assumed that, because the intercept resulting from either one of these two regression analyses represents the ES when the sample variance (or SE) equals zero, this intercept should estimate the true ES.

The most prominent p-value based method is Rosenthal's (1979) Fail-Safe N method. Its purpose is to compute the number of additional studies reporting non-significant results needed to make the p-value corresponding to the pooled ES no longer significant. This "traditional" procedure remains one of the most widely employed for detecting publication bias despite several explicit recommendations against its use (e.g., Borenstein et al., 2009a; Higgins et al., 2019; Rothstein, Sutton, & Borenstein, 2005). Due to its many limitations of both mathematical and theoretical nature, the Fail-Safe N is "[...] now generally regarded as valueless" (Vevea et al., 2019, p. 390).

Two more recent p-value based methods are the so called p-curve (Simonsohn, Nelson, & Simmons, 2014b, 2014a; Simonsohn, Simmons, & Nelson, 2015) and p-uniform (van Assen, van Aert, & Wicherts, 2015). p-uniform is based on the notion that conditional p-values are uniformly distributed given a fixed true ES. p-curve analysis derives its logic from the fact that p-values are uniformly distributed when the true effect equals zero (the null hypothesis is true), and right skewed when there is a true effect (the null hypothesis is false). Importantly, this also applies to the range of p-values commonly defined as the region of significance in a given research field, e.g., $(0 < p < 0.05)$ in the behavioural, social, and biomedical sciences. Significant p-values stemming from a biased pool of studies should, on the contrary, exhibit a left-skewed distribution.

This left skewness is commonly attributed to selective reporting and publishing of effects whose p-values are over the significance threshold regardless of their magnitude. Such bodies of research are considered to possess little or no *evidential value* (Simonsohn et al., 2014a) as one assumes most of their significant effects to be the product of a

combination of p-hacking[3] and selective reporting/publication bias. p-curve analysis therefore involves testing for evidential value in a set of ESs by conducting 6 statistical tests on the subset containing exclusively significant ESs: one binomial test of right skewness, two parametric tests of right skewness, one binomial tests of flatness, and two parametric tests of flatness. There are two parametric tests each as the second one is conducted on half the curve only, i.e., the interval $(0, 0.025)$, in order to account for "ambitious p-hacking." The adjusted ES estimate produced by the analysis is, however, based on the full curve.

The last class of publication bias assessment tools, selections models, cannot be exclusively assigned to either category as they are a very versatile methods which allow their user to model *any* process assumed to generate bias in the meta-analytic estimates (Harrer et al., 2021; Vevea et al., 2019). The main purpose of these models is to adjust the observed pooled ES estimate by merging two functions: one describing the distribution of ESs in the absence of bias and the other describing the mechanism by which ESs are assumed to be selected for reporting or publication. Selection models have also been around for decades, although they have enjoyed much less use as they are considerably more complex conceptually and have been barely implemented in commercial point and click software (Vevea et al., 2019).

One relatively simple variety is the three-parameter selection model (McShane et al., 2016), which, as the name suggests, only estimates three parameters: the true average effect μ, the between-study variance τ^2, and the likelihood of obtaining non-significant ES in relation to obtaining a significant one, ω_2. What is fixed is the cut-off value α_1, which is usually the significance threshold for a one-sided p-value, 0.025. The relative likelihood of obtaining a p-value in this interval $(0 < \alpha_1 < 0.025)$, ω_1, is set to one. Hence, if the estimated ω_2 is close to one, the adjusted pooled ES estimate will deviate little from the one based on a standard random-effects model as this would indicate a similar likelihood of obtaining a non-significant ES to obtaining a significant one. Otherwise, the significant ESs

[3] "trying multiple analyses to obtain statistical significance" (Simonsohn et al., 2014a, p. 534)

will be downweighted if $\omega_2 < 1$ and upweighted if $\omega_2 > 1$, which diminishes or augments the pooled ES estimate, respectively.

Fail-Safe N's shortcomings were already mentioned. However, this should not be taken to mean that it is the only method with limitations. In fact, Fail-Safe N's meaninglessness is one of the very few things the publication bias assessment literature appears to be in agreement about (Carter et al., 2019; Harrer et al., 2021). The methods vary considerably in their assumptions and hence under what conditions they perform well. The different developers of such methods disagree *strongly* about which conditions are realistic in which disciplines and to what degree this should inform the methods' construction. For example, Simonsohn (2017), one of the developers of the p-curve, deems PET-PEESE to be worse than homoeopathy as it (PET-PEESE) can be actively harmful whereas homoeopathy is simply ineffective. Vevea et al. (2019)[4], on the other hand, see the p-curve and p-uniform methods as reinventions of the wheel and as "[. . .] modified versions of simplistic early weight-function models" (p.392). Duval and Tweedie (2000a)[5] quote DuMouchel and Harris (1997) as writing in reference to selection models: "attempts to assess publication bias beyond simple graphs like the funnel plot seem to involve a tour de force of modeling, and as such are bound to run up against resistance from those who are not statistical modeling wonks" (p. 95).

Given this intricate state of affairs when it comes to statistical methods of publication bias assessment and correction, it should have become clear that as of yet, there is neither a single method nor a single set of methods which can be recommended as an all-purpose tool (Harrer et al., 2021; Vevea et al., 2019). Simulations studies have shown that no single method outperforms all the others under all conditions (Carter et al., 2019). Another point on which a consensus seems to prevail is that these methods should be used

[4] Vevea being a developer of two selection model varieties (Vevea & Hedges, 1995; Vevea & Woods, 2005)

[5] The developers of the trim-and-fill method, which is widely regarded as being of little use for correcting bias, although potentially useful as a sensitivity analysis (Carter et al., 2019; Hilgard, Engelhardt, & Rouder, 2017; Moreno et al., 2009; Simonsohn et al., 2014b; van Assen et al., 2015).

as sensitivity analyses, both with respect to a single method (e.g., by varying the assumptions used) and with respect to which and how many tests one uses (Carter et al., 2019; McShane et al., 2016; Vevea et al., 2019). We are not aware of studies investigating how publication bias is accounted for in tDCS-motor learning research.

Research questions

In sum, 4 principal premises motivated our work: 1. tDCS appears to be remarkably popular as a research tool in basic and clinical research as well as in form of commercial gadgets, 2. meta-analyses of tDCS's effect on motor learning might be substantially impacting research and clinical practice and/or tDCS's uptake as a commercial product, 3. meta-analyses are often of suboptimal methodological quality, which compromises their credibility and informativeness, 4. no methodological evaluation of meta-analyses in tDCS-motor learning research has been conducted as of yet. Our aim was thus to provide such an evaluation while partitioning the concept of methodological quality of a meta-analysis into 3 main aspects:

1. Reporting quality: did the meta-analysis adhere to the PRISMA reporting guidelines? Which items were neglected, if any?
2. Reproducibility: how hard is it to reproduce the pooled ES estimate reported in the meta-analysis, if at all, based on the information provided therein? Which necessary pieces of information were missing, if any? If enough information was provided to reproduce the methods, does the reproduced pooled ES equal the reported pooled ES?
3. Consideration of publication bias: Did the meta-analysis report attempts to minimise the effects of publication bias? Which statistical or non-statistical methods were used? By conducting a publication bias analysis of our own, do we arrive at the same conclusion regarding the presence of publication bias as the meta-analysts?

Secondary methodological aspects we evaluated included:

1. Had the meta-analysis been pre-registered? If yes, did the pre-registration protocol adhere to meta-analysis pre-registration protocol reporting guidelines PRISMA-P (Moher et al., 2015)? Pre-registering the hypotheses and data analysis plan is essential for distinguishing between confirmatory and exploratory findings (Brian A. Nosek, Ebersole, DeHaven, & Mellor, 2018). In the context of reviews and meta-analyses, pre-registration protocols "[...] act as a guard against arbitrary decision making during review conduct, enable readers to assess for the presence of selective reporting against completed reviews, and, when made publicly available, reduce duplication of efforts and potentially prompt collaboration" (Shamseer et al., 2015, p. 1).
2. Did the meta-analysis discuss outliers? How was the presence of outliers addressed? Extreme ES values may affect the validity and robustness of meta-analytic results and there is a general consensus that meta-analyses should examine the presence of outliers and to what extent they influence conclusions (Viechtbauer & Cheung, 2010).

Methods

Although our methodological approach mostly followed the plan pre-defined in the thesis proposal (accessible on the thesis' Open Science Framework [OSF] project osf.io/uagvf/), there were important deviations from the plan, especially with respect to reproducibility testing. A document listing these deviations can also be found on the OSF project. For data wrangling, analysis and visualisation, we used `R` (Version 4.1.2, R Core Team, 2021) and the packages `dmetar` (Version 0.0.9000, Harrer, Cuijpers, Furukawa, & Ebert, 2019), `dplyr` (Version 1.0.7, Wickham, François, Henry, & Müller, 2021), `forcats` (Version 0.5.1, Wickham, 2021a), `ggplot2` (Version 3.3.5, Wickham, 2016), `MAd` (Version 0.8.2.1, Hoyt, 2014), `Matrix` (Version 1.3.4, Bates & Maechler, 2021), `meta` (Version 5.1.0,

Balduzzi, Rücker, & Schwarzer, 2019), `metafor` (Version 3.0.2, Viechtbauer, 2010), `purrr` (Version 0.3.4, Henry & Wickham, 2020), `readr` (Version 2.1.0, Wickham & Hester, 2021), `readxl` (Version 1.3.1, Wickham & Bryan, 2019), `stringr` (Version 1.4.0, Wickham, 2019), `tibble` (Version 3.1.6, Müller & Wickham, 2021), and `tidyr` (Version 1.1.4, Wickham, 2021b).

The package `renv` (Version 0.14.0, Ushey, RStudio, & PBC, 2021) was used to ensure long-term reproducibility of our analyses. This thesis was written using RMarkdown and the associated packages `papaja` (Version 0.1.0.9997, Aust & Barth, 2020), `kableExtra` (Version 1.3.4, Zhu, 2021), `knitr` (Version 1.36, Xie, 2014, 2015, 2021). Both writing and analysis were done using RStudio (Version 2021.9.1.372, RStudio Team, 2021). Data were extracted from figures using WebPlotDigitizer (apps.automeris.io/wpd/, Rohatgi, 2021). A video demonstration of how we extracted data from figures is available on the OSF project.

Sample of meta-analyses

Three meta-analyses (Hung et al., 2021b; Kang, Summers, & Cauraugh, 2016; Kang, Weingart, & Cauraugh, 2018) were selected based on the following eligibility criteria:

- Meta-analysis studies which quantitatively synthesise multiple (at least 3) primary studies on the effects of tDCS on motor learning.
- No restriction on primary outcomes (e.g., how speed or accuracy were measured), designs of primary studies (e.g., randomised vs. crossover designs), or participants (e.g., clinical or healthy subjects) in the primary studies were imposed.

Exclusion criteria:

- Reviews of any type without a quantitative synthesis
- Primary studies
- Reviews which did not report a “main” meta-analysis, but rather multiple meta-analyses of subgroups of studies.

All three meta-analyses were found via quick, non-systematic Google Scholar or Web of Science searches.

Data extraction

Elements extracted. There were two types of elements extracted:

1. Primary study level variables (e.g., sample sizes, means and SDs), were extracted from both the primary studies and the meta-analyses.
2. Meta-analysis level variables (e.g., pooled SMD, publication bias control related variables).

Both data sheets are also available on the OSF project page. They can be found by navigating to the GitHub thesis branch in the folder "data_thesis" under the names "Data_ps_raw_updated" and "Data_ma_raw," respectively. The codebook explaining column headers in the data sheets is also accessible on the OSF project (folder "Post_data_extraction").

Data extraction procedure. Data for all non-reproducibility-related variables were extracted directly to Google Sheets (no specialised data extraction software was used). The data extraction procedure for the purpose of testing reproducibility differed substantially from the rest and is described below.

Procedure, coding, and data analysis

Reporting quality. The adherence of each meta-analysis to the PRISMA reporting guidelines (Liberati et al., 2009; Moher et al., 2009) were checked. Concretely, for each of the 27 PRISMA items, we coded whether the meta-analysis reported the relevant information as recommended, regardless of whether the meta-analysis reported having adhered to any reporting guidelines. Besides coding whether the meta-analysis had been pre-registered, we checked whether the pre-registration protocol adhered to the PRISMA-P

reporting guidelines (Moher et al., 2015; Shamseer et al., 2015) for such protocols, provided that the meta-analysis had been pre-registered and published later than 2015.

Reproducibility. We based our treatment of meta-analysis reproducibility on the definition and principles of reproducibility put forward by the American Statistical Association (Broman et al., 2017): a meta-analysis is reproducible if its authors provided enough information to go through all the necessary procedures (search, screening, data extraction, calculation of primary ESs, calculation of the pooled ES. . .) to arrive at the same results. However, although we acknowledge the importance of all these steps, we, like Gøtzsche et al. (2007), Lakens et al. (2017), and Maassen et al. (2020), focused on data extraction and calculation of ES estimates.

Since reproducing the primary ESs is a necessary step towards reproducing the pooled SMD, we constructed a scheme for classifying their reproducibility status. The two variables in this scheme are A. whether the primary ES could be successfully reproduced or approximated numerically[6] (results reproducibility) and B. whether the procedure we followed in reproducing the ES strictly corresponded to the information given in the meta-analysis or to the procedure apparently adopted for at least two other ESs included in the meta-analysis (methods reproducibility).

Table 4 depicts the classification system in which we use SMD (for standardised mean difference) instead of ES since SMD was the only measure of ES employed in the meta-analyses we evaluated. "Procedure" here includes such analytic decisions as using p-values or test statistics in combination with sample sizes to estimate the SMD for a given primary study, using the raw means and SDs instead of means and SDs of changes in the outcome from baseline or vice versa, using follow-up means and SDs instead of post means and SDs, etc.. The second variable in the classification system was thus mainly adopted to

[6] An ES estimate was considered successfully reproduced (or reproducible) if the reproduced ES equalled the one reported in the meta-analysis at the second decimal (e.g., $0.3324543 = 0.33$) and approximated if the reproduced ES equalled the reported one at the first decimal (e.g., $0.1767492 \approx 0.22$).

capture cases where there was a discrepancy between how the meta-analysts reported having computed a primary ES and how they actually computed it.

Table 4
Reproducibility classification scheme.

.	Reproduced SMD equalled or approximated reported SMD	Reproduced SMD neither equalled nor approximated reported SMD
Strictly following information given in the meta-analysis or a standard procedure apparently adopted for at least two other successfully reproduced SMDs	1. Faithfully reproducible SMD	2. Faithfully irreproducible SMD
Following a procedure which either does not entirely correspond to the procedure the meta-analysts report having adopted OR does not (necessarily) produce an SMD that is comparable to what would result from following the procedure apparently adopted for at least two other successfully reproduced SMDs	3. Brute-force reproducible SMD	4. Brute-force irreproducible SMD

It is important to re-emphasise our distinction between "reproduced" and "reproducible." "Reproduced" simply means we calculated something and compared it with the original, whereas "reproducible" SMDs were those which were successfully reproduced computationally. Thus, "faithfully" and "brute-force" refer to whether we found data in the corresponding primary study that could be used at all and "reproducible" and "irreproducible" to whether the SMD resulting from our calculation matches (or approximates) the original. In other words, if a given SMD is methodologically irreproducible given the data reported in the primary study, then its faithful reproducibility cannot be tested and would have to be excluded when calculating the pooled SMD (see information about the different meta-analytic models we fit below). Furthermore, if we classify an SMD as "faithfully irreproducible," this does not imply that it is necessarily also

brute-force irreproducible because in most cases, we did not collect further values to test brute-force reproducibility if the values we chose to test faithful reproducibility very clearly corresponded to the meta-analysts' description of their procedure.

Reproducibility testing was an iterative process which involved several rounds of data extraction. The initial round of data extraction and testing reproducibility did not yield a lot of reproducible SMDs as most primary studies did not report the values necessary to directly compute an SMD and/or its sampling variance (e.g., for a between groups Cohen's d this would be the group means, SDs and sample sizes). Therefore, a less strict data extraction procedure was adopted, which involved the following steps:

1. For each primary SMD, we first looked for the raw means and SDs of the outcome reported as having been used by the meta-analysts. If the primary study reported multiple sets of means and SDs which can be seen as corresponding to the outcome described in the meta-analysts (e.g., the outcome in the meta-analysis for a given primary study is "Fugl-Myer Test" but the primary study reports values for "Fugl-Myer Test - upper limbs" and "Fugl-Myer Test - full"), all sets were extracted.
2. If no means and SDs for the relevant outcome were reported in the primary study, means and SDs were extracted from figures. If no figures were reported which contained means and SDs (or SEs or confidence intervals [CIs]), p and/or t-values for tests on the relevant outcome were extracted, which in combination with ns can be converted to SMDs.
3. Based on all extracted values, we computed each primary SMD using the estimator (Cohen's d or Hedges' g) reported as having been used by the meta-analysts. If this information was not given in the meta-analysis, we tried both formulas and for further analysis used the one which consistently approximated the reported SMDs better.
4. If none of the values extracted reproduced a given SMD, we double checked the correctness of the data extracted and, in some cases, extracted more values from the primary study (à la brute-force) and computed the SMD based on those.

5. We computed the sampling variances based on the SMDs and the corresponding ns.
6. For each meta-analysis, we fit three meta-analytic models (MAMs): 1. Based on the faithfully reproduced SMDs. 2. Based on the faithfully reproduced SMDs plus brute-force reproduced. 3. Based on the SMDs and sampling variances reported in the meta-analysis[7]. This was done to test for analytical reproducibility of the pooled SMD and to find out which between-study heterogeneity estimator was used.

The concrete procedure for data extraction thus differed for each single primary study. A detailed description of all values we extracted and how we analysed them is provided in the data analysis notebook, also available on our OSF project page.

Publication bias control. The non-statistical approaches to accounting for publication bias that we coded were:

1. Searched clinical trial registries (e.g., ClinicalTrials.org)
2. Searched thesis and dissertation repositories (e.g., ProQuest)
3. Did not restrict their search to studies written in English
4. Contacted known researchers in the field to inquire about unpublished results
5. Contacted authors of included studies to ask for raw data or unpublished results

Besides coding whether any statistical methods were used at all and which, we tested for publication bias in each meta-analysis using 3 different methods: PET-PEESE (T. Stanley, 2008; T. D. Stanley & Doucouliagos, 2014), p-curve (Simonsohn et al., 2014b, 2014a), and the three-parameter selection model (McShane et al., 2016). For each meta-analysis, we concluded that publication bias is a concern in the studies synthesised if at least two out of the three methods used yielded results indicating the presence of publication bias. Although such a "majority vote" approach is not recommended for actual publication bias assessment (Carter et al., 2019, p. 140), we adopted it so as to have an

[7] Sampling variances were not reported in any of the three meta-analyses. They were calculated based on SEs extracted from funnel plots in the case of the first two meta-analyses and based on CIs for the third meta-analysis

unambiguous decision rule whether our analysis agrees with that of the meta-analysts or not. For these analyses, we used the SMDs and the sample sizes reported in the meta-analyses along with the sampling variances extracted from the funnel plots in the case of the first two meta-analyses and calculated based on the CIs in the case of the third meta-analysis.

Outlier/influential study analysis. Besides coding whether each meta-analysis checked for the existence of outliers among the included studies and how this was done, we tested this ourselves for each meta-analysis using the leave-one-out method (Harrer et al., 2021; Tobias, 1999).

Results

General description of the meta-analyses

Table 5 gives an overview of the three meta-analyses we reviewed. Meta-analyses 1 and 2 aimed to estimate the effectiveness of tDCS for improving motor function in post-stroke patients, although meta-analysis 2 focused exclusively on the effects of cathodal tDCS. Meta-analysis 3 investigated effectiveness of tDCS for improving surgical performance of surgery trainees. Meta-analyses 2 and 3 reported more primary SMDs than primary studies because they drew two comparisons from some primary studies. In meta-analysis 1, the double comparisons represented cathodal vs. sham and anodal vs. sham pairs, whereas in meta-analysis 2 comparison pairs were based on two different outcomes. The first meta-analysis synthesised the results of 13 randomised controlled trials (RCTs) and 4 crossover trials, the second 6 RCTs and 9 crossover trials, the third 5 RCTs and one crossover trial. None of the three meta-analyses had been pre-registered. According to Google Scholar, the meta-analyses were cited 197, 12, and two times, respectively, as of 16.11.2021. The average total sample size N of the primary studies included in the three meta-analyses, that is, mean of $n_t + n_c$, were ~ 26, ~ 24, and ~ 32, respectively.

Table 5
Reviewed meta-analyses.

Meta-analysis ID	Authors	Year	Number of studies	Number of SMDs
1	Kang, Summers, & Cauraugh	2016	17	21
2	Kang, Weingart, & Cauraugh	2018	15	20
3	Hung et al.	2021	6	6

Reporting quality

Meta-analyses 1 did not report having adhered to any reporting guidelines. Despite this, it can be seen as having reported the content of 18 out of the 27 PRISMA (Moher et al., 2009) items. Meta-analysis 2 mentioned PRISMA and adhered to 19 out of the 27 items. Meta-analysis 3 reported having adhered to the most recent PRISMA guidelines (Page, McKenzie, et al., 2021) but we evaluated the adherence to the items of the 2009 version to ensure comparability to the other meta-analyses. Meta-analysis reported the content of 25 out of the 27 items. Table 6 summarises the three meta-analyses' adherence to the PRISMA items.

Table 6
PRISMA items adherence of the three meta-analyses

Item number	Original item label	1	2	3
1	Title	Reported	Reported	Reported
2	Structured Summary	Not reported	Reported	Not reported
3	Rationale	Reported	Reported	Reported
4	Objectives	Not reported	Reported	Reported

Table 6
PRISMA items adherence of the three meta-analyses (continued)

Item number	Original item label	1	2	3
5	Protocol and registration	Not reported	Not reported	Reported
6	Eligibility criteria	Reported	Reported	Reported
7	Information sources	Reported	Reported	Reported
8	Search	Not reported	Not reported	Reported
9	Study selection	Not reported	Not reported	Not reported
10	Data collection process	Not reported	Not reported	Reported
11	Data items	Reported	Reported	Reported
12	Risk of bias in individual studies	Not reported	Reported	Reported
13	Summary measures	Reported	Reported	Reported
14	Synthesis of results	Reported	Reported	Reported
15	Risk of bias across studies	Reported	Reported	Reported
16	Additional analyses	Reported	Reported	Reported
17	Study selection	Reported	Reported	Reported
18	Study characteristics	Reported	Reported	Reported
19	Risk of bias within studies	Not reported	Reported	Reported
20	Results of individual studies	Reported	Reported	Reported
21	Synthesis of results	Reported	Reported	Reported
22	Risk of bias across studies	Reported	Reported	Reported
23	Additional analysis	Reported	Reported	Reported
24	Summary of evidence	Reported	Reported	Reported
25	Limitations	Not reported	Not reported	Reported
26	Condusions	Reported	Reported	Reported

Table 6
PRISMA items adherence of the three meta-analyses (continued)

Item number	Original item label	1	2	3
27	Funding	Reported	Not reported	Reported

Note. Meta-analysis 3 indicated that it had not been pre-registered.

A more detailed description of the extent to which the three meta-analyses adhered to items 6 to 9 and 17 is provided in Table 7. All three meta-analyses adequately described their eligibility criteria but none of them described the actual process of how the criteria were enforced. All three meta-analyses did report the result of this process.

Table 7
Detailed description of adherence to PRISMA items 6-9 and 17

Item number	Original item label	1	2	3
6	Eligibility criteria	Target study characteristics as well as inclusion and exclusion criteria described but not following the PICO scheme (e.g., nothing about the outcomes). Rational unclear “Our literature search concentrated on tDCS studies that investigated long-term effects on motor functions post-stroke.”.	Target study characteristics as well as inclusion and exclusion criteria described but not following the PICO scheme (e.g., nothing about the outcomes). Rational unclear.	Target study characteristics as well as inclusion and exclusion criteria described following the PICO scheme but no actual outcomes were mentioned (only "change in surgical performance". Rational unclear.

Table 7
Detailed description of adherence to PRISMA items 6-9 and 17 (continued)

Item number	Original item label	1	2	3
7	Information sources	Databases used for "initial" search were listed. No mention of whether authors of primary authors were contacted. Dates of coverage described.	Databases used were listed. No mention of whether authors of primary authors were contacted. Dates of coverage described.	Databases used were listed. The meta-analysts mention in the appendix having contacted the authors of primary studies regarding data but not how the authors resonded. Dates of coverage described.
8	Search	Although the meta-analysts listed the search keywords used, they did not a describe a full search strategy for any of the databases.	Although the meta-analysts listed the search keywords used, they did not a describe a full search strategy for any of the databases.	The meta-analysts listed both the search keywords used and the full search strategy for PubMed and other databases.
9	Study selection	The meta-analysts did not describe how the different authors contributed to the different stages of eligibility checking or how potential disagreements were resolved.	The meta-analysts did not describe how the different authors contributed to the different stages of eligibility checking or how potential disagreements were resolved.	The meta-analysts did not describe how the different authors contributed to the different stages of eligibility checking or how potential disagreements were resolved.
17	Study selection	The meta-analysts reported how many studies were initially included and how many were subsequently excluded (including reasons). No PRISMA flow chart was provided.	The meta-analysts reported how many studies were initially included and how many were subsequently excluded in a PRISMA flow chart (including reasons).	The meta-analysts reported how many studies were initially included and how many were subsequently excluded in a PRISMA flow chart (including reasons).

Reproducibility

We faced difficulties in reproducing the primary SMDs and their sampling variances for all three meta-analyses, mainly due to limited reporting of the methods sections. None of the three meta-analyses shared their data or provided data analysis code. Although meta-analysis 3 stated that data "was available upon reasonable request" (p. 11), the corresponding author of the meta-analysis did not respond to our email requesting more information/data. The process necessitated several time-consuming rounds of data extraction and testing. Which relevant pieces of information were missing and how many primary SMDs could be successfully reproduced is reported below for each meta-analysis separately. The effect of irreproducible primary SMDs on the pooled SMDs is also reported. Since all three meta-analyses reported having fit a random-effects model using the Comprehensive Meta-Analysis software[8], which per default estimates between-study heterogeneity via the Der-Simonian-Laird method (DerSimonian & Laird, 1986), we used these settings for all our analyses, too.

Meta-analysis 1. The authors of this meta-analysis reported the following reproducibility-relevant information/data:

- Standardised ES measure is SMD
- Group sample sizes for each primary SMD calculated. In case of crossover studies, they reported the total sample size for both the control and treatment groups.
- All primary SMDs calculated
- CI lower and upper limits for each SMD
- The outcome measure for each SMD (e.g., "Total latency score in JHFT")
- What the two compared groups represented (e.g., control group: "sham before

[8] In the case of the first meta-analysis, we were informed by the first author of the meta-analysis that they used the Comprehensive Meta-Analysis software in response to an email asking for further information. Our email's text as well as all the reproducibility-relevant information contained in the meta-analysts' response (not the actual text of their responses) can be found in the document "Email_to_authors" on our OSF project.

inpatient daily rehabilitation at retention," treatment group: "ctDCS on cH before inpatient daily rehabilitation at retention")

- random-effects model
- The pooled SMD

The following information was missing:

- Type of SMD used (Cohen's d vs. Hedge's g)
- Whether a different method of standardisation was used for the crossover studies (since there are no two groups whose SDs can be pooled to standardise the mean difference)
- Sampling variances of the SMDs (they had to be extracted from the funnel plot)
- Whether sampling variances were calculated differently for the SMDs corresponding to crossover trials
- Which type of values were used to compute each SMD (e.g., means and SDs vs. p-value and ns)
- Which exact values were used and where they were found (e.g., p-value reported on page x line y or means and SDs reported in Figure z)
- Rationale for choosing outcomes
- Enough details about the outcome used so as to leave no room for ambivalence (e.g., "Upper Limb FMA" instead of just "FMA" when the primary study reported both "Upper Limb FMA" and "Total)
- Which between-study heterogeneity estimator was used
- Software used to run the meta-analysis

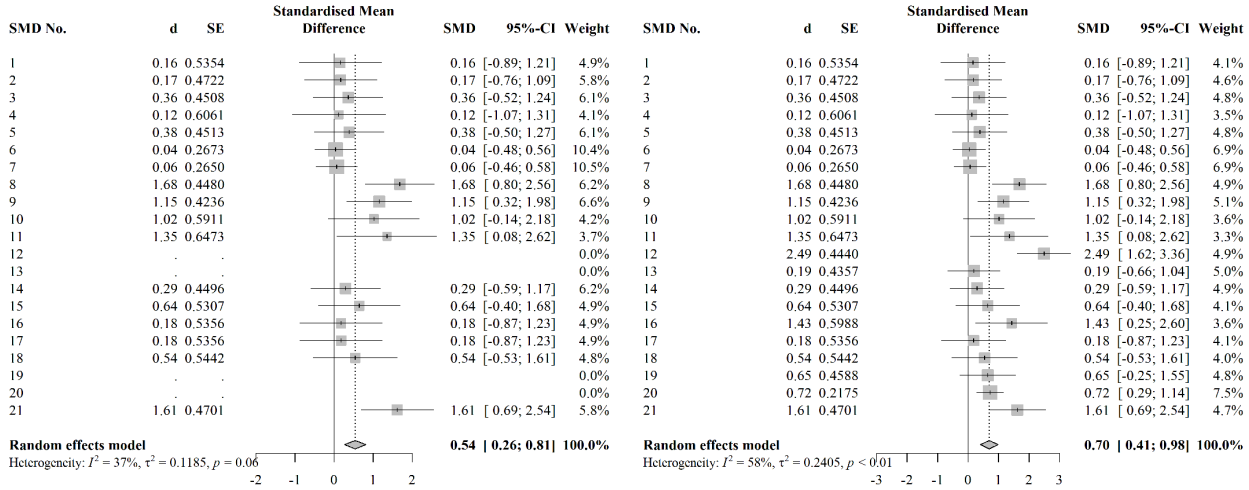
Table 7 summarises the reproducibility status and classification of each primary SMD. Six SMDs were successfully reproduced following the procedure as described in the meta-analysis or the seemingly standard procedure. Five further SMDs were approximated. Following a deviating procedure, three more SMDs could be reproduced.

Table 8
Reproducibility of primary SMDs, Meta-analysis 1

SMD no.	Reported SMD	Reproduced SMD	Reproducibility classification	Reason for irreproducibility
1	0.16	0.16	Faithfully reproducible	Not applicable.
2	0.18	0.17	Faithfully reproducible	Not applicable.
3	0.36	0.36	Faithfully reproducible	Not applicable.
4	0.08	0.12	Faithfully reproducible	Not applicable.
5	0.38	0.38	Faithfully reproducible	Not applicable.
6	0.04	0.04	Faithfully reproducible	Not applicable.
7	0.06	0.06	Faithfully reproducible	Not applicable.
8	1.59	1.68	Faithfully irreproducible	Not inferable.
9	1.08	1.15	Faithfully irreproducible	Not inferable.
10	1.05	1.02	Faithfully reproducible	Not applicable.
11	1.39	1.35	Faithfully reproducible	Not applicable.
12	0.93	NA	Brute-force irreproducible	Not inferable.
13	0.82	NA	Brute-force irreproducible	Not inferable.
14	0.29	0.29	Faithfully reproducible	Not applicable.
15	0.61	0.64	Faithfully reproducible	Not applicable.
16	1.43	0.18	Faithfully irreproducible, Brute-force reproducible	Outcome used does not correspond to description.
17	0.94	0.18	Faithfully irreproducible	Not inferable.
18	0.24	0.54	Faithfully irreproducible	Not inferable.
19	0.65	NA	Brute-force reproducible	Outcome used does not correspond to description.
20	0.72	NA	Brute-force reproducible	Sucessfully reproduced using a p-value derived from a medians test.
21	0.53	1.61	Faithfully irreproducible	Not inferable.

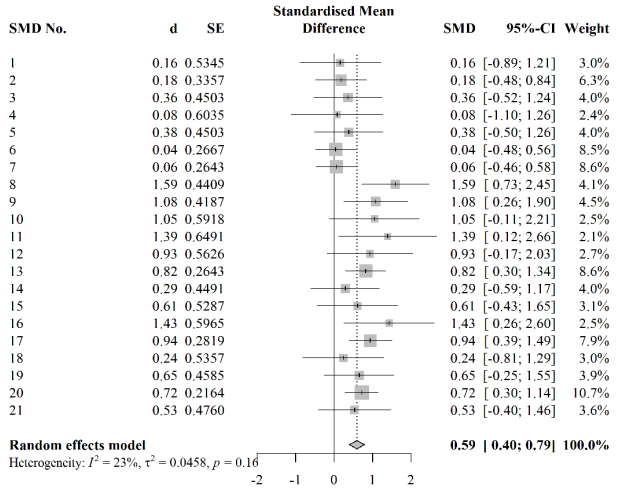
Note. The "Reproduced SMD" column lists the faithfully reproduced SMDs only (rounded to the second decimal).

Using the SMDs and the sample sizes reported in the meta-analysis along with the sampling variances extracted from the funnel plot (MAM 3), all meta-analytic estimates were successfully reproduced. This MAM was thus treated as the "reported" variety and compared to the two other MAMs. All three MAMs are depicted as forest plots in Figure 3. The pooled SMDs from MAMs 1 and 2 deviate from the reported pooled SMD by 0.05 and 0.11 SDs, respectively. More noteworthy differences are observable in the heterogeneity estimates I^2 and τ^2.



(a) MAM 1, based on faithfully reproduced SMDs.

(b) MAM 2, based on faithfully and brute-force reproduced SMDs.



(c) MAM 3, based on reported SMDs.

Figure 3. Forest plots of the three meta-analytic models, meta-analysis 1.

Meta-analysis 2. The following reproducibility-relevant information/data were reported:

- Standardised ES measure is SMD
- Group sample sizes (ns) for each primary SMD calculated. In case of crossover studies, they reported the total sample size (N) for both the control and treatment groups.
- All primary SMDs calculated
- CI lower and upper limits for each SMD
- The outcome measure for each SMD
- SMDs were based on comparisons between cathodal tDCS group and sham group at post
- random-effects model
- The pooled SMD
- Software used was Comprehensive Meta-Analysis

The following information was missing:

- Type of SMD used
- Whether a different method of standardisation was used for the crossover studies
- Sampling variances of the SMDs (depicted in the funnel plot)
- Whether sampling variances were calculated differently for the SMDs corresponding to crossover trials
- Which type of values were used to compute each SMD
- Which exact values were used and where they were found
- Rationale for choosing outcomes
- Enough details about the outcome used so as to leave no room for ambivalence
- Which between-study heterogeneity estimator was used

Table 8 summarises the reproducibility status and classification of each primary

SMD. One SMD was successfully reproduced following the procedure as described in the meta-analysis or the seemingly standard procedure. Three further SMDs were approximated. Following a deviating procedure, two more SMDs were reproducible and one approximated.

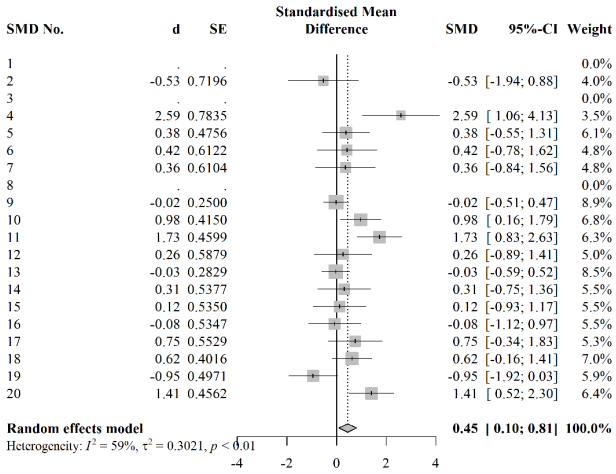
Table 9
Reproducibility of primary SMDs, Meta-analysis 2

SMD no.	Reported SMD	Reproduced SMD	Reproducibility classification	Reason for irreproducibility
1	0.96	NA	Brute-force irreproducible	Not inferable.
2	2.46	-0.53	Faithfully irreproducible	Not inferable.
3	0.68	NA	Brute-force reproducible	Brute-force reproducible but outcome used does not correspond to description.
4	1.56	2.59	Faithfully irreproducible	Not inferable.
5	1.25	0.38	Faithfully irreproducible	Not inferable.
6	0.28	0.26	Faithfully reproducible	Not applicable.
7	0.06	0.17	Faithfully irreproducible	Not inferable.
8	-0.14	NA	Brute-force reproducible	Successfully reproduced using a p-value derived from a Kruskal-Wallis test of differences between the three groups anodal, cathodal, and sham.
9	-0.11	-0.02	Faithfully irreproducible, Brute-force reproducible	Successfully reproduced using a p-value derived from a Kruskal-Wallis test of differences between the three groups anodal, cathodal, and sham. Means and SDs were reported in the primary study for the outcome used.
10	0.94	0.98	Faithfully irreproducible	Not inferable.
11	1.77	1.73	Faithfully irreproducible	Not inferable.

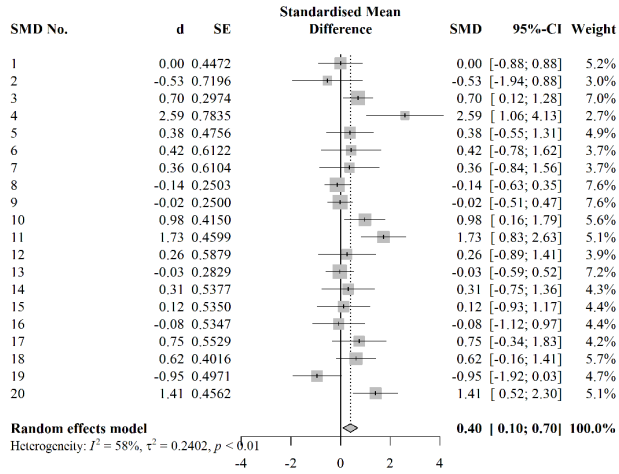
Table 9
Reproducibility of primary SMDs, Meta-analysis 2 (continued)

SMD no.	Reported SMD	Reproduced SMD	Reproducibility classification	Reason for irreproducibility
12	0.37	0.26	Faithfully irreproducible	Not inferable.
13	0.34	-0.03	Faithfully irreproducible	Not inferable.
14	2.10	0.31	Faithfully irreproducible	Not inferable.
15	1.18	0.12	Faithfully irreproducible	Not inferable.
16	0.08	-0.08	Faithfully irreproducible	Wrong sign.
17	0.75	0.75	Faithfully reproducible	Not applicable.
18	0.61	0.62	Faithfully reproducible	Not applicable.
19	-0.87	-0.95	Faithfully reproducible	Not applicable.
20	0.90	1.41	Faithfully irreproducible	Not inferable.

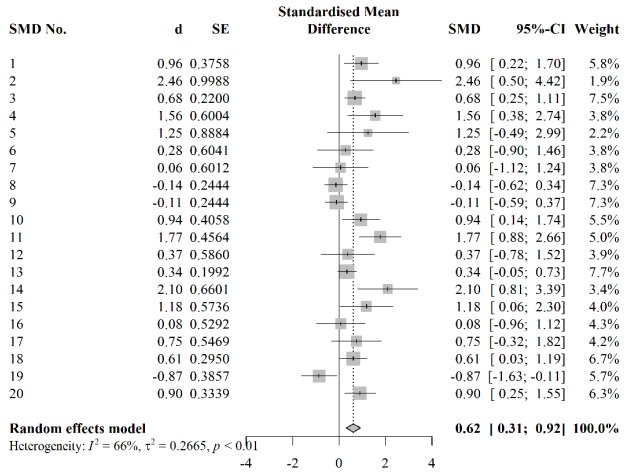
All meta-analytic estimates were successfully reproduced in MAM 3, whose output was again treated as "reported." The three MAMs are depicted in Figure 4. The pooled SMDs from MAMs 1 and 2 deviate from the reported pooled SMD by 0.17 and 0.22 SDs, respectively. All three models displayed a very high degree of heterogeneity.



(a) MAM 1, based on faithfully reproduced SMDs.



(b) MAM 2, based on faithfully and brute-force reproduced SMDs.



(c) MAM 3, based on reported SMDs.

Figure 4. Forest plots of the three meta-analytic models, meta-analysis 2.

Meta-analysis 3. The following reproducibility-relevant information/data were reported:

- Standardised ES measure is Hedges' g
- Group sample sizes (ns) for each primary SMD calculated. In case of crossover studies, they reported the total sample size (N) for both the control and treatment groups.

- All primary SMDs calculated
- CI lower and upper limits for each SMD
- The outcome measure used for each primary study was what the primary study defined as the primary outcome
- SMDs were based on comparisons between tDCS group and sham group at post
- The task learned by participants (which sometimes aided in finding the outcome used by the meta-analysts)
- The pooled SMD
- Software used was Comprehensive Meta-Analysis

The following information was missing:

- Whether a different method of standardisation was used for the crossover study
- Sampling variances of the SMDs
- Whether sampling variances were calculated differently for the SMD corresponding to the crossover trial
- Which type of values were used to compute each SMD
- Which exact values were used and where they were found
- Enough details about the outcome used so as to leave no room for ambivalence
- Which between-study heterogeneity estimator was used

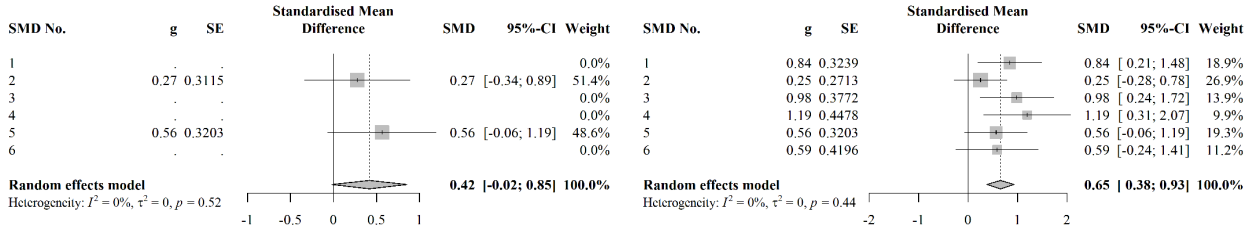
Table 9 summarises the reproducibility status and classification of each primary SMD. One SMD was approximated following the procedure as described in the meta-analysis or the seemingly standard procedure. Following a deviating procedure, all 5 remaining SMDs could be reproduced or approximated.

Table 10
Reproducibility of primary SMDs, Meta-analysis 3

SMD no.	Reported SMD	Reproduced SMD	Reproducibility classification	Reason for irreproducibility
1	0.84	NA	Brute-force reproducible	Successfully reproduced using a p-value (reported in the primary study as a range "<0.01") derived from a difference in medians test.
2	0.25	0.27	Faithfully irreproducible, Brute-force reproducible	Successfully reproduced using the values for one of the two outcomes indicated to have been used and doubling the tDCS group sample size.
3	0.98	NA	Brute-force reproducible	Sucessfully reproduced using a p-value (reported as a range "<0.01") derived from a medians test in combination with the total sample size in place of both treatment and control group sample sizes. Notably, this was not a journal article, but a conference abstract.
4	1.19	NA	Brute-force reproducible	Outcome used does not correspond to description.
5	0.59	0.56	Faithfully reproducible	Not applicable.
6	0.58	NA	Brute-force reproducible	Successfully reproduced using a p-value based on difference between tDCS and sham groups in change from baseline

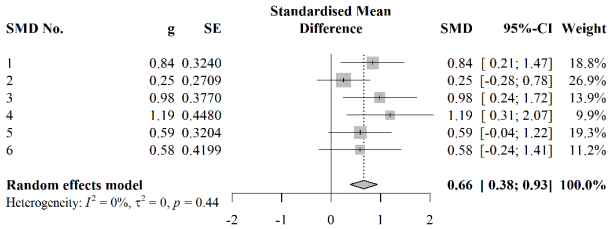
Note. The reported SMDs were rounded to the second decimal (3 decimals were reported).

Using the SMDs and the sample sizes reported in the meta-analysis along with the sampling variances calculated based on the reported CIs, all meta-analytic estimates were successfully reproduced. The output of this MAM (3) was thus again treated as the reported variety and used for comparison. The three MAMs are depicted in Figure 4. The pooled SMD from MAM 1 deviates from the reported pooled SMD by 0.24 SDs, whereas MAM 2 is almost identical to the MAM 3. Very little heterogeneity was observed in all three cases.



(a) MAM 1, based on faithfully reproduced SMDs.

(b) MAM 2, based on faithfully and brute-force reproduced SMDs.



(c) MAM 3, based on reported SMDs.

Figure 5. Forest plots of the three meta-analytic models, meta-analysis 3.

Publication bias control

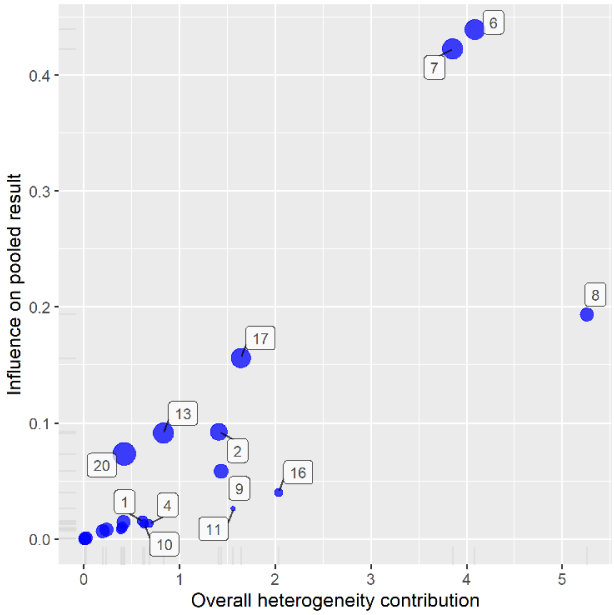
All three meta-analyses mentioned publication bias, although only meta-analysis 3 reported having taken measures to pre-emptively mitigate its effects: the authors report having searched ClinicalTrials.gov and ProQuest (which indexes theses and dissertations), contacted the authors of the primary studies to ask for more data, and not restricted their search to articles written in English.

All three meta-analyses report having inspected a funnel plot as a means to detect publication bias. Meta-analysis 1 additionally used Fail-Safe N. To correct for publication bias, meta-analysis 1 used trim-and-fill; meta-analysis 2 used trim-and-fill, Egger's test, and Begg and Maxumdar's rank correlation test (Begg & Mazumdar, 1994); meta-analysis 3 used Egger's test. The authors of meta-analysis 1 concluded that the findings of the tests they used "support a minor publication bias conclusion" (Kang et al., 2016, p. 348). Similarly, the conclusion in meta-analysis 2 (Kang et al., 2018, p. 5) was "minimal

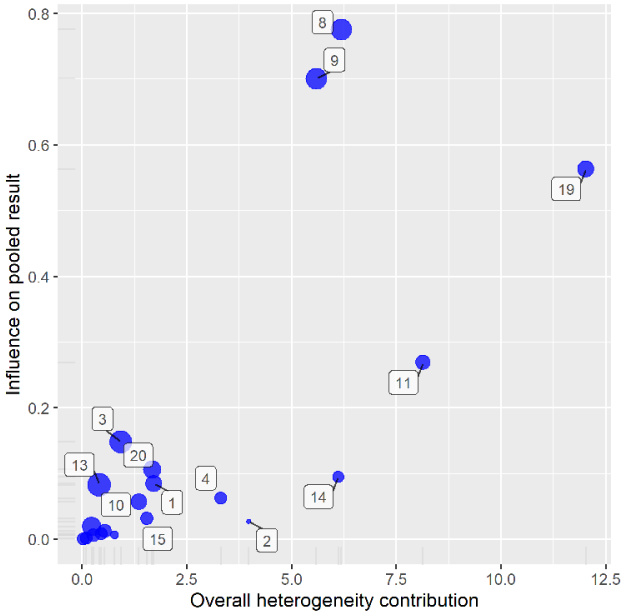
publication bias in the studies used." No clear conclusion was provided in meta-analysis 3.

Our publication bias testing routine (along with the associated decision rule pre-defined in the data analysis plan) indicated a conclusion concurring with that of the meta-analysts in the case of the first meta-analysis and the opposite conclusions for the two other meta-analyses. For meta-analysis 1, the estimate of the true effect produced by the three-parameter selection model (0.59) was virtually identical to the original. The estimate produced by the *p*-curve was larger (0.70). Only the PET-PEESE intercepts (0.34 and 0.45, respectively) indicated that the random-effects model based estimates might be overestimating the true effect. For meta-analysis 2, the selection model (0.20), *p*-curve (0.18), and PEESE (0.21) estimates were much smaller than the original (0.62). The PET intercept (-0.12) was negative. Similar results were observed for the last meta-analysis: the estimates produced by the the *p*-curve and PET-PEESE were 0.40, -0.71, and 0.01, respectively. Only the selection model yielded an estimate which is close to the one based on the random-effects model (0.62).

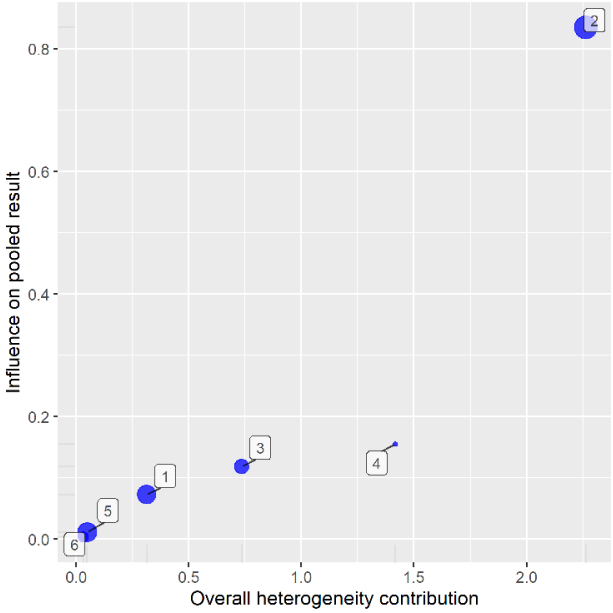
Outlier/influential study analysis. Only the third meta-analysis mentioned outliers or influential studies. They ran a leave-one-out analysis and reported that the main results did not change due to removing any one of the 6 studies they included. Our leave-one-out analysis indicated the presence of influential studies in all three meta-analyses (see Figure 6).



(a) Meta-analysis 1.



(b) Meta-analysis 2.



(c) Meta-analysis 3.

Figure 6. Baujat plots based on the leave-one-out method.

Discussion

The aim of this work was to evaluate the methodological quality of meta-analyses in tDCS-motor learning research with respect to reporting quality, reproducibility, and

publication bias control. We found the meta-analyses we studied to be lacking in all three points. Although the meta-analyses reported the content of most PRISMA (Moher et al., 2009) items, they described their methods so sparsely as to render reproducing their procedure impossible without lengthy detective work. None of the meta-analyses had pre-registration protocols, shared their data, or provided their data-analysis code. We failed to numerically reproduce the main pooled ES estimates reported in all meta-analyses when following the procedures they described. As to publication bias control, only meta-analysis 3 reported having searched the grey literature and all three meta-analyses mostly used "traditional" publication bias detection and correction tools without discussion of their assumptions or appropriateness.

To give a more comprehensive account of our results, with respect to reporting quality as measured by adherence to PRISMA items, our findings mostly follow the pattern of those reported in Table 2. Concretely, similarly to the large sample of systematic reviews investigated by Page and Moher (2017), only meta-analysis 3 mentioned pre-registration (item 5), included a reproducible search string for a database (item 8), and conducted an assessment of bias in the individual studies (item 12). Notable differences between our sample and that of Page and Moher (2017) is that all three meta-analyses we reviewed included a list of variables they sought data for (item 11), conducted statistical assessments of bias across studies (items 15 and 22), and reported the results of their study selection procedure (item 17), although none of them described the actual procedure (item 9).

With respect to reproducibility, the most notable finding is probably the high prevalence of discrepancies between how the meta-analysts reported having computed individual ESs and how they really did it. These discrepancies were most often in relation to the outcomes used. For example, there were multiple cases where the meta-analysis reported having used an outcome X whereas they actually used the outcome change in X from baseline. This was particularly perplexing when values for both outcomes were reported in the primary study. In general, all primary studies included in the

meta-analyses reported values/tests for several outcomes and meta-analysis 3 was the only one to provide a rationale, albeit a vague one[9], for why they chose the outcome they did for each primary study. In most cases, it was impossible to infer how these things came about as the authors of meta-analyses 1 and 2 did not respond to our request for data or data analysis code or protocol and the authors of meta-analysis 3 did not respond to our email at all. On the whole, we had similar difficulties in our reproducibility endeavour as Lakens et al. (2017) and Maassen et al. (2020).

However, although our reproduced pooled SMDs were calculated using less primary SMDs for all three meta-analyses (due to methodologically irreproducible SMDs), none of our reproductions led to a radical change in the pooled ES estimate like in Gøtzsche et al.'s (2007) or Ford et al.'s (2010) reviews. Indeed, our reproduced pooled SMDs for meta-analyses 1 and 2 were both significant in the same direction as the original ones, which is one of the many commonly used criteria for a successful replication (e.g., in large scale replication projects such as the one conducted by the Open Science Collaboration, 2015). This can be seen as an indication for inferential reproducibility of the meta-analyses' main conclusions, albeit a weak one, as this criterion for successful replication is often criticised and thus used in combination with other criteria (Heirene, 2021; Open Science Collaboration, 2015). A more thorough evaluation of inferential reproducibility would require a deeper dive into both methodological (e.g., marked differences in heterogeneity parameters between the reproduced and reported models) and tDCS-related aspects.

As to statistical publication bias analysis, like Banks, Kepes, and McDaniel (2012), we found the meta-analyses we reviewed to mostly use outdated publication bias control methods. However, this was not the critical issue, rather the fact the meta-analysts reported the results of their analyses very briefly and without discussion of their assumptions and their appropriateness for the data at hand. As mentioned before, publication bias testing remains an exceedingly active and contentious field of research and

[9] Namely that they used the outcome the respective primary study defined as their primary outcome.

no consensus exists as of yet regarding which tests to use under which conditions. Any single test or constellation of tests used is thus inevitably arbitrary. One major problem all current methods suffer from is their mediocre performance when applied to a heterogeneous set of ESs. Meta-analyses 1 and 2 synthesised such sets of ESs, meta-analysis 3 only included 6 primary studies.

Furthermore, the primary studies included in the meta-analyses are likely to have been selected for publication based on different values than the ones the meta-analysts (and we) used for the publication bias tests (i.e., results of omnibus tests on several outcomes and not the pair-wise comparisons in one outcome at a single time point). This is especially relevant for the *p*-curve and the selection model we used of which an important step is dividing the set of ESs into significant and non-significant ones. Therefore, the results of both our and the meta-analysts' statistical publication bias tests can only be taken with a grain of salt. Nevertheless, the fact that all three meta-analysts used more than one tool is commendable as "[a]ny assessment of publication bias is better than none, and using all methods available is better than using only one." (Vevea et al., 2019, p. 392). Their publication bias analysis would have been more informative had they used more recent methods as sensitivity analyses and provided clearer descriptions of the implications of tests' results.

Attempting to reproduce the meta-analyses revealed further methodological issues which do not directly pertain to the main methodological aspects we investigated: All three meta-analyses indiscriminately combined primary studies of different designs. For example, they computed SMDs using the same formula for both controlled and crossover designs, a procedure which neglects bias resulting from estimating sampling variances for crossover studies without accounting for carry-over effects or correlations between time points (Borenstein & Hedges, 2019; Madeyski & Kitchenham, 2018; Morris & DeShon, 2002). Furthermore, the fact the meta-analysts used the total sample size of the cross-over trial to replace the treatment and control group sample sizes is likely to have inflated the

power of these pair-wise comparisons (which is especially relevant for the publication bias tests). Another issue the meta-analysts neglected to account for is ES dependency (Gleser & Olkin, 2009), which is especially critical in the case of the first two meta-analyses as they derived multiple ESs from single studies.

Moreover, although there are several ways to calculate an SMD besides the classic formula based on means and SDs (for an overview see Borenstein & Hedges, 2019), it can be argued that the meta-analyses we reviewed used methods that went beyond what is plausible. For example, *p*-values based on median tests were used to calculate SMDs in two of the meta-analyses. Medians can replace means when they are derived from normally distributed data. However, primary studies usually report medians and interquartile ranges instead of means and SDs *because* their samples are not normally distributed (Higgins et al., 2019, p. 167). This is likely to have been the case as most studies in the meta-analyses we reviewed are small. Besides, one would still need to impute the corresponding SDs. Similarly, Higgins et al. (2019) advise against combining SMDs based on post values and SMDs based on change from baseline as the SDs used to standardise them estimate different entities (p. 253).

All in all, our results indicate that the methodological limitations prevalent in meta-analyses in neighbouring fields can also be observed in meta-analyses studying the effect of tDCS on motor learning. This is unfortunate since together, the three meta-analyses have been cited over 200 times and knowing the high status meta-analyses enjoy on the "hierarchy of evidence" (Evans, 2003), they are likely to be influential beyond the restricted realm of scientific publishing. Naive consumers of these meta-analyses might be mislead into believing that they provide a definitive "proof" of the tDCS' effectiveness and researchers will have trouble evaluating the validity of the meta-analyses' methodology due to the rather limited reporting in key aspects (e.g., how and why studies and outcomes were selected) and the lack of pre-registered protocols, raw data or data analysis code (Page, McKenzie, et al., 2021; Page, Moher, et al., 2021).

Although the meta-analyses we reviewed were adequate in some aspects (e.g., description of outcome used for each primary study), there is clearly some room for improvement. As a remedy for the compromised reporting quality and reproducibility, more detailed descriptions of the methodological procedure are called for, ideally accompanied by raw data and the data analysis code (Page, Moher, et al., 2021). For publication bias assessment, a more extensive testing procedure involving sensitivity analyses both within a single method and across the different methods is necessary (following a literature search which is as comprehensive as allocated resources permit, Vevea et al., 2019). When dealing with a highly heterogeneous set of ESs, van Aert, Wicherts, and van Assen (2016) recommend splitting the set into subgroups based on theoretical and methodological considerations before carrying out the bias testing procedure. Inzlicht, Gervais, and Berkman (2015) recommend presenting a range of plausible pooled ESs based on different publication bias tests (using different assumptions).

No self-evident solutions can be offered for the shortcomings related to the computational aspects of the meta-analyses. There is immense variation in reporting quality of primary studies and meta-analysts usually have no other choice but to work with the data reported in the primary study (that is, when contacting the authors of the primary study for more data is not feasible or proves to be fruitless, Higgins et al., 2019). It is thus entirely understandable that meta-analysts must sometimes resort to alternative means of calculating certain measures. However, there *are* limits to how divergent the alternative methods can be. For example, instead of converting a p-value derived from a medians test to an SMD, the meta-analysts could have estimated the means and SDs based on the medians and the corresponding ranges and or interquartile ranges calculated the SMD based on these means and SDs, which would have yielded a more comparable result to an SMD calculated on actual means and SDs (Hozo, Djulbegovic, & Hozo, 2005; Wan, Wang, Liu, & Tong, 2014).

Similarly, there are methods to compute the ESs and their corresponding sampling

variances when dealing with crossover studies (Madeyski & Kitchenham, 2018). Even estimating the ESs using the formula for within-subjects designs would have been better as these would have at least taken the correlations between time points, but not the carry-over effects, into account (Borenstein & Hedges, 2019). Instead of combining ESs derived from primary studies with different designs to compute a single pooled ESs, a practice strongly discouraged by meta-analysis pioneers like Borenstein and Hedges (2019) as such primary ESs cannot be seen as estimating the same parameter, meta-analysts can conduct multiple meta-analyses by grouping similar studies together. In some cases, meta-analysts must accept the loss of power by excluding primary studies, which are markedly different from the other studies in key aspects or which display a high or unclear risk of bias (Higgins et al., 2019, p. 192). Finally, it is paramount that the meta-analysts report how they computed each ES regardless of how they dealt with limited reporting in the primary study.

The fact that our publication bias testing routine is unlikely to have yielded more meaningful results than that of the meta-analysts is not the only limitation to our work. Others include:

- We have defined our exclusion criteria based on mostly practical considerations. Their high restrictiveness has probably led to a sample that is not representative of the field at large. Our non-systematic literature search and study selection strategy can only have exasperated this issue. It also goes without saying that three meta-analyses is too small a sample to draw far-reaching conclusions.
- Although we had a mechanism in place to minimise the probability of data extraction errors on our part when evaluating reproducibility, it cannot be excluded that potential errors when extracting data for other variables (e.g., for PRISMA adherence) influenced our results as data extraction and coding was not checked by others (Buscemi, Hartling, Vandermeer, Tjosvold, & Klassen, 2006; Jones, Remmington, Williamson, Ashby, & Smyth, 2005).
- We focused on reproducibility of data extraction and calculation of ESs because

previous reviews demonstrated a high prevalence of mistakes in these steps. However, it cannot be excluded that other aspects we neglected such as study search and selection have a larger impact.

- Throughout the process of reproduction, we had to make subjective decisions that cannot be guaranteed to have been faultless. Notably, it was not always trivial to judge whether our procedure for reproducing a certain primary SMD strictly followed the procedure purported to have been used by the meta-analysts. For example, we managed to approximate SMD no. 5 in meta-analysis 3 by averaging the means and SDs of two outcomes and computing an SMDs based on the average value. We subsequently classified this SMD as faithfully reproducible because the meta-analysts reported having used what each primary study defined as its primary outcome and this specific primary study defined both these outcomes as its primary outcomes. However, since the meta-analysts did not provide any information on how they computed the SMD or any indication that they took an average, it is almost certain that they calculated the SMD differently because otherwise we would have successfully reproduced the SMD to the third decimal like we did the other brute-force reproducible ones. Other such examples are documented in the data analysis notebook.
- Due to our rather limited expertise with regards to the technical and practical aspects of tDCS research, it cannot be excluded that we have overlooked or misunderstood relevant tDCS-related methodological facets.

Future similar works may thus aim for a more fine-grained and nuanced evaluation of reporting quality which goes beyond the minimal requirements set by reporting guidelines, a more comprehensive reproducibility testing which is not restricted to data extraction and ES calculation as well as a thorough investigation of robustness towards changes in analytical decisions (especially data selection and outcomes used), and a more principled approach towards evaluating publication bias assessment.

Despite these limitations, we hope to have provided consumers of tDCS-motor learning research with an incentive to evaluate meta-analyses in this field more critically when making decisions about the use of tDCS and meta-analysts with aspects to consider when conducting meta-analyses in the future. By no means do we want to imply that the solutions presented above are easy. Despite the abundance and comprehensiveness of guidelines such as PRISMA and tutorials on how to conduct a transparent and reproducible meta-analysis (e.g., Moreau & Gamble, 2020; Quintana, 2015), it is undeniable that meticulous adherence to guidelines and making one's meta-analysis reproducible require a substantial amount of time and effort. Likewise, it cannot be expected from substantive researchers to be adept at every methodological/statistical aspect relevant to conducting a meta-analysis. A discussion of potential measures to improve the situation on a higher, more structural level, such as ways to incentivise transparency-related practices (Bakker, van Dijk, & Wicherts, 2012; Higginson & Munafò, 2016; B. A. Nosek et al., 2015) or journals employing specialised statistical review (Hardwicke, Frank, Vazire, & Goodman, 2019), is beyond the scope of this work.

Nevertheless, given the epistemic weight meta-analyses usually carry, problems associated with how they are conducted and reported cannot be shrugged off. Efforts to improve them should be promoted. Besides the guidelines which are constantly being updated and advancements in statistical procedures for controlling publication bias, many methodological procedures have been developed in the recent years to systematically account for the impact of different analytical decisions (Taylor & Munafò, 2016; e.g., Voracek et al., 2019). More emphasis is also being placed on flagship open science practices such as pre-registration and data and code sharing (Maassen et al., 2020; Page, Moher, et al., 2021).

Finally, since meta-analyses can only be as good as the primary literatures they synthesise, meta-analysts cannot be expected to carry all the responsibility for improvement [Aguinis et al. (2011); Borenstein et al. (2009b). In the quest for a reliable

cumulative science, efforts to counteract the methodological issues found in the tDCS primary literature (e.g., heterogeneity in employed tDCS parameters, compromised reproducibility due to incomplete reporting Buch et al., 2017) must also be supported. Such efforts include mathematical models which can be used to systematically determine the optimal tDCS parameters to use (Lipka et al., 2021) and checklists outlining all aspects which should be disclosed when reporting the results of a tDCS study (Buch et al., 2017).

References

Ada, S., Sharman, R., & Balkundi, P. (2012). Impact of meta-analytic decisions on the conclusions drawn on the business value of information technology. *Decision Support Systems*, *54*(1), 521–533. https://doi.org/10.1016/j.dss.2012.07.001

Aguinis, H., Dalton, D. R., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (2011). Meta-Analytic Choices and Judgment Calls: Implications for Theory Building and Testing, Obtained Effect Sizes, and Scholarly Impact. *Journal of Management*, *37*(1), 5–38. https://doi.org/10.1177/0149206310377113

Aguinis, H., Pierce, C. A., Bosco, F. A., Dalton, D. R., & Dalton, C. M. (2011). Debunking Myths and Urban Legends About Meta-Analysis. *Organizational Research Methods*, *14*(2), 306–331. https://doi.org/10.1177/1094428110375720

Ahn, S., Ames, A. J., & Myers, N. D. (2012). A Review of Meta-Analyses in Education: Methodological Strengths and Weaknesses. *Review of Educational Research*, *82*(4), 436–476. https://doi.org/10.3102/0034654312458162

Aust, F., & Barth, M. (2020). *Papaja: Create APA manuscripts with R Markdown*. Manual.

Austin, A., Jiga-Boy, G. M., Rea, S., Newstead, S. A., Roderick, S., Davis, N. J., … Boy, F. (2016). Prefrontal Electrical Stimulation in Non-depressed Reduces Levels of Reported Negative Affects from Daily Stressors. *Frontiers in Psychology*, *7*, 315. https://doi.org/10.3389/fpsyg.2016.00315

Aytug, Z. G., Rothstein, H. R., Zhou, W., & Kern, M. C. (2012). Revealed or Concealed? Transparency of Procedures, Decisions, and Judgment Calls in Meta-Analyses. *Organizational Research Methods*, *15*(1), 103–133. https://doi.org/10.1177/1094428111403495

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The Rules of the Game Called Psychological Science. *Perspectives on Psychological Science*, *7*(6), 543–554. https://doi.org/10.1177/1745691612459060

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health*, (22), 153–160.

Banks, G. C., Kepes, S., & McDaniel, M. A. (2012). Publication Bias: A call for improved meta-analytic practice in the organizational sciences. *International Journal of Selection and Assessment*, *20*(2), 182–196. https://doi.org/10.1111/j.1468-2389.2012.00591.x

Bates, D., & Maechler, M. (2021). *Matrix: Sparse and dense matrix classes and methods*. Manual.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 1088–1101.

Bennabi, D., Pedron, S., Haffen, E., Monnin, J., Peterschmitt, Y., & Van Waes, V. (2014). Transcranial direct current stimulation for memory enhancement: From clinical research to animal models. *Frontiers in Systems Neuroscience*, *8*, 159. https://doi.org/10.3389/fnsys.2014.00159

Borenstein, M., & Hedges, L. V. (2019). Effect sizes for meta-analyses. In *The Handbook of Research Synthesis and Meta-Analysis,* (pp. 207–241).

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009a). Publication Bias. In *Introduction to Meta-Analysis* (pp. 277–292). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743386.ch30

Borenstein, M., Hedges, L., Higgins, J. P. T., & Rothstein, H. R. (2009b). Criticisms of meta-analysis. In *Introduction to Meta-Analysis* (pp. 377–387).

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., . . . Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84–88. https://doi.org/10.1038/s41586-020-2314-9

Broman, K., Cetinkaya-Rundel, M., Nussbaum, A., Paciorek, C., Peng, R., Turek, D., & Wickham, H. (2017). Recommendations to Funding Agencies for Supporting Reproducible Research.

Brunoni, A. R., Moffa, A. H., Fregni, F., Palm, U., Padberg, F., Blumberger, D. M., ... Loo, C. K. (2016). Transcranial direct current stimulation for acute major depressive episodes: Meta-analysis of individual patient data. *The British Journal of Psychiatry*, *208*(6), 522–531. https://doi.org/10.1192/bjp.bp.115.164715

Buch, E. R., Santarnecchi, E., Antal, A., Born, J., Celnik, P. A., Classen, J., ... Cohen, L. G. (2017). Effects of tDCS on motor learning and memory formation: A consensus and critical position paper. *Clinical Neurophysiology*, *128*(4), 589–603. https://doi.org/10.1016/j.clinph.2017.01.004

Buscemi, N., Hartling, L., Vandermeer, B., Tjosvold, L., & Klassen, T. P. (2006). Single data extraction generated more errors than double data extraction in systematic reviews. *Journal of Clinical Epidemiology*, *59*(7), 697–703. https://doi.org/10.1016/j.jclinepi.2005.11.010

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for Bias in Psychology: A Comparison of Meta-Analytic Methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115–144. https://doi.org/10.1177/2515245919847196

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The Handbook of Research Synthesis and Meta-Analysis* (2nd edition). New York: Russell Sage Foundation.

Davis, N. J. (2016). The regulation of consumer tDCS: Engaging a community of creative self-experimenters. *JL & Biosciences*, *3*, 304.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188.

Dieckmann, N. F., Malle, B. F., & Bodner, T. E. (2009). An Empirical Assessment of

Meta-Analytic Practice. *Review of General Psychology*, *13*(2), 101–115. https://doi.org/10.1037/a0015107

Dubljević, V., Saigle, V., & Racine, E. (2014). The Rising Tide of tDCS in the Media and Academic Literature. *Neuron*, *82*(4), 731–736. https://doi.org/10.1016/j.neuron.2014.05.003

Duval, S., & Tweedie, R. (2000a). A Nonparametric "Trim and Fill" Method of Accounting for Publication Bias in Meta-Analysis. *Journal of the American Statistical Association*, *95*(449), 89–98. https://doi.org/10.2307/2669529

Duval, S., & Tweedie, R. (2000b). Trim and Fill: A Simple Funnel-Plot–Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Evans, D. (2003). Hierarchy of evidence: A framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, *12*(1), 77–84. https://doi.org/10.1046/j.1365-2702.2003.00662.x

Fanelli, D. (2010). “Positive” Results Increase Down the Hierarchy of the Sciences. *PLOS ONE*, *5*(4), e10068. https://doi.org/10.1371/journal.pone.0010068

Ford, A. C., Guyatt, G. H., Talley, N. J., & Moayyedi, P. (2010). Errors in the conduct of systematic reviews of pharmacological interventions for irritable bowel syndrome. *The American Journal of Gastroenterology*, *105*(2), 280–288. https://doi.org/10.1038/ajg.2009.658

Fregni, F., Nitsche, M. A., Loo, C. K., Brunoni, A. R., Marangolo, P., Leite, J., . . . Bikson, M. (2015). Regulatory Considerations for the Clinical and Research Use of Transcranial Direct Current Stimulation (tDCS): Review and recommendations from an expert panel. *Clinical Research and Regulatory Affairs*, *32*(1), 22–35.

https://doi.org/10.3109/10601333.2015.980944

Friese, M., & Frankenbach, J. (2020). P-Hacking and publication bias interact to distort meta-analytic effect size estimates. *Psychological Methods*, *25*(4), 456–471. https://doi.org/10.1037/met0000246

Gazzaniga, M., Ivry, R. B., & Mangun, G. R. (2018). Methods of Cognitive Neuroscience. In *Cognitive Neuroscience: The Biology of the Mind* (Fifth edition). New York: W. W. Norton & Company.

Gebodh, N., Esmaeilpour, Z., Adair, D., Schestattsky, P., Fregni, F., & Bikson, M. (2019). Transcranial Direct Current Stimulation Among Technologies for Low-Intensity Transcranial Electrical Stimulation: Classification, History, and Terminology. In H. Knotkova, M. A. Nitsche, M. Bikson, & A. J. Woods (Eds.), *Practical Guide to Transcranial Direct Current Stimulation: Principles, Procedures and Applications* (pp. 3–43). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-95948-1_1

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*, *348*.

Geyskens, I., Krishnan, R., Steenkamp, J.-B. E. M., & Cunha, P. V. (2009). A Review and Evaluation of Meta-Analysis Practices in Management Research. *Journal of Management*, *35*(2), 393–419. https://doi.org/10.1177/0149206308328501

Gianni, E., Bertoli, M., Simonelli, I., Paulon, L., Tecchio, F., & Pasqualetti, P. (2021). tDCS randomized controlled trials in no-structural diseases: A quantitative review. *Scientific Reports*, *11*(1), 16311. https://doi.org/10.1038/s41598-021-95084-6

Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In *The handbook of research synthesis and meta-analysis, 2nd ed* (pp. 357–376). New York, NY, US: Russell Sage Foundation.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, *8*(341), 341ps12. https://doi.org/10.1126/scitranslmed.aaf5027

Gopalakrishnan, S., & Ganeshkumar, P. (2013). Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. *Journal of Family Medicine and Primary Care*, *2*(1), 9–14. https://doi.org/10.4103/2249-4863.109934

Gøtzsche, P. C., Hróbjartsson, A., Marić, K., & Tendal, B. (2007). Data Extraction Errors in Meta-analyses That Use Standardized Mean Differences. *JAMA*, *298*(4), 430–437. https://doi.org/10.1001/jama.298.4.430

Guzzo, R., Jackson, S., & Katzell, R. (1987). Meta-analysis analysis. *Undefined*.

Hardwicke, T. E., Frank, M. C., Vazire, S., & Goodman, S. N. (2019). Should Psychology Journals Adopt Specialized Statistical Review? *Advances in Methods and Practices in Psychological Science*, *2*(3), 240–249. https://doi.org/10.1177/2515245919858428

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing Meta-Analysis With R: A Hands-On Guide* (1st ed.). Boca Raton, FL and London: Chapman & Hall/CRC Press.

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. D. (2019). *Dmetar: Companion r package for the guide 'doing meta-analysis in r'*. Manual.

Heirene, R. M. (2021). A call for replications of addiction research: Which studies should we replicate and what constitutes a 'successful' replication? *Addiction Research & Theory*, *29*(2), 89–97. https://doi.org/10.1080/16066359.2020.1751130

Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools*. Manual.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Higginson, A. D., & Munafò, M. R. (2016). Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions. *PLOS Biology*, *14*(11), e2000995. https://doi.org/10.1371/journal.pbio.2000995

Hilgard, J., Engelhardt, C. R., & Rouder, J. N. (2017). Overstated evidence for short-term effects of violent games on affect and behavior: A reanalysis of Anderson et al.(2010).

Hohn, R. E., Slaney, K. L., & Tafreshi, D. (2020). An Empirical Review of Research and Reporting Practices in Psychological Meta-Analyses. *Review of General Psychology*, *24*(3), 195–209. https://doi.org/10.1177/1089268020918844

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 Stimulation*, *8*(3), 535–550. https://doi.org/10.1016/j.brs.2015.01.400

Hoyt, A. C. D. R. &. W. T. (2014). *MAd: Meta-analysis with mean differences*. Manual.

Hozo, S. P., Djulbegovic, B., & Hozo, I. (2005). Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology*, *5*(1), 13. https://doi.org/10.1186/1471-2288-5-13

Hung, C.-M., Zeng, B.-Y., Zeng, B.-S., Sun, C.-K., Cheng, Y.-S., Su, K.-P., . . . Tseng, P.-T. (2021b). The Efficacy of Transcranial Direct Current Stimulation in Enhancing Surgical Skill Acquisition: A Preliminary Meta-Analysis of Randomized Controlled Trials. *Brain Sciences*, *11*(6), 707. https://doi.org/10.3390/brainsci11060707

Hung, C.-M., Zeng, B.-Y., Zeng, B.-S., Sun, C.-K., Cheng, Y.-S., Su, K.-P., . . . Tseng, P.-T. (2021a). The Efficacy of Transcranial Direct Current Stimulation in Enhancing Surgical Skill Acquisition: A Preliminary Meta-Analysis of Randomized Controlled Trials. *Brain Sciences*, *11*(6), 707. https://doi.org/10.3390/brainsci11060707

Hyatt, C. S., Owens, M. M., Crowe, M. L., Carter, N. T., Lynam, D. R., & Miller, J. D. (2020). The quandary of covarying: A brief review and empirical examination of covariate use in structural neuroimaging studies on psychological variables. *NeuroImage*, *205*, 116225. https://doi.org/10.1016/j.neuroimage.2019.116225

Inzlicht, M., Gervais, W., & Berkman, E. (2015). Bias-Correction Techniques Alone Cannot Determine Whether Ego Depletion is Different from Zero: Commentary on

Carter, Kofler, Forster, & McCullough, 2015. https://doi.org/10.2139/ssrn.2659409

Ioannidis, J. P. A. (2016). The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly*, *94*(3), 485–514. https://doi.org/10.1111/1468-0009.12210

Jones, A. P., Remmington, T., Williamson, P. R., Ashby, D., & Smyth, R. L. (2005). High prevalence but low impact of data extraction and reporting errors were found in Cochrane systematic reviews. *Journal of Clinical Epidemiology*, *58*(7), 741–742. https://doi.org/10.1016/j.jclinepi.2004.11.024

Kang, N., Summers, J. J., & Cauraugh, J. H. (2016). Transcranial direct current stimulation facilitates motor learning post-stroke: A systematic review and meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, *87*(4), 345–355. https://doi.org/f8fwd8

Kang, N., Weingart, A., & Cauraugh, J. H. (2018). Transcranial direct current stimulation and suppression of contralesional primary motor cortex post-stroke: A systematic review and meta-analysis. *Brain Injury*, *32*(9), 1063–1070. https://doi.org/10.1080/02699052.2018.1481526

Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. (2013). Meta-analytic Reviews in the Organizational Sciences: Two Meta-analytic Schools on the Way to MARS (the Meta-analytic Reporting Standards). *Journal of Business and Psychology*, *28*(2), 123–143. https://doi.org/10.1007/s10869-013-9300-2

Konstantopoulos, S., & Hedges, L. V. (2019). Statistically analyzing effect sizes: Fixed-and random-effects models. In *The Handbook of Research Synthesis and Meta-Analysis,* (pp. 245–279).

Lakens, D., Page-Gould, E., Assen, M. A. L. M. van, Spellman, B., Schönbrodt, F., Hasselman, F., . . . Scheel, A. M. (2017). Examining the Reproducibility of Meta-Analyses in Psychology: A Preliminary Report. https://doi.org/10.31222/osf.io/xfbjf

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., ... Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, *339*, b2700. https://doi.org/10.1136/bmj.b2700

Light, R. J., & Pillemer, D. B. (1984). *Summing Up: The Science of Reviewing Research*. Harvard University Press.

Lipka, R., Ahlers, E., Reed, T. L., Karstens, M. I., Nguyen, V., Bajbouj, M., & Cohen Kadosh, R. (2021). Resolving heterogeneity in transcranial electrical stimulation efficacy for attention deficit hyperactivity disorder. *Experimental Neurology*, *337*, 113586. https://doi.org/10.1016/j.expneurol.2020.113586

Liu, Y., Gu, N., Cao, X., Zhu, Y., Wang, J., Smith, R. C., & Li, C. (2021). Effects of transcranial electrical stimulation on working memory in patients with schizophrenia: A systematic review and meta-analysis. *Psychiatry Research*, *296*, 113656. https://doi.org/10.1016/j.psychres.2020.113656

Luedtke, K., Rushton, A., Wright, C., Geiss, B., Juergens, T. P., & May, A. (2012). Transcranial Direct Current Stimulation for the Reduction of Clinical and Experimentally Induced Pain: A Systematic Review and Meta-analysis. *The Clinical Journal of Pain*, *28*(5), 452–461. https://doi.org/10.1097/AJP.0b013e31823853e3

Maassen, E., Assen, M. A. L. M. van, Nuijten, M. B., Olsson-Collentine, A., & Wicherts, J. M. (2020). Reproducibility of individual effect sizes in meta-analyses in psychology. *PLOS ONE*, *15*(5), e0233107. https://doi.org/10.1371/journal.pone.0233107

Madeyski, L., & Kitchenham, B. (2018). Effect sizes and their variance for AB/BA crossover design studies. *Empirical Software Engineering*, *23*(4), 1982–2017. https://doi.org/10.1007/s10664-017-9574-5

Marks-Anglin, A., & Chen, Y. (2020). Small-study effects: Current practice and challenges for future research. *Statistics and Its Interface*, *13*(4), 475–484. https://doi.org/10.4310/SII.2020.v13.n4.a5

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for Publication Bias in Meta-Analysis: An Evaluation of Selection Methods and Some Cautionary Notes. *Perspectives on Psychological Science*, *11*(5), 730–749. https://doi.org/10.1177/1745691616662243

Minarik, T., Berger, B., Althaus, L., Bader, V., Biebl, B., Brotzeller, F., . . . Sauseng, P. (2016). The Importance of Sample Size for Reproducibility of tDCS Effects. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00453

Moher, Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D. F. (2000). Improving the Quality of Reports of Meta-Analyses of Randomised Controlled Trials: The QUOROM Statement. *Oncology Research and Treatment*, *23*(6), 597–602. https://doi.org/10.1159/000055014

Moher, Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moher, Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*, 1. https://doi.org/10.1186/2046-4053-4-1

Moreau, D., & Gamble, B. (2020). Conducting a meta-analysis in the age of open science: Tools, tips, and practical recommendations. *Psychological Methods*, No Pagination Specified–No Pagination Specified. https://doi.org/10.1037/met0000351

Moreno, S. G., Sutton, A. J., Ades, A., Stanley, T. D., Abrams, K. R., Peters, J. L., & Cooper, N. J. (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Medical Research Methodology*, *9*(1), 2. https://doi.org/10.1186/1471-2288-9-2

Morganti, A. (2007). The impact of meta-analyses on clinical practice: The benefits. *Journal of Nephrology*, *20 Suppl 12*, S1–3.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105–125. https://doi.org/10.1037/1082-989X.7.1.105

Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames*. Manual.

Nieminen, L. R. G., Nicklin, J. M., McClure, T. K., & Chakrabarti, M. (2011). Meta-analytic Decisions and Reliability: A Serendipitous Case of Three Independent Telecommuting Meta-analyses. *Journal of Business and Psychology*, *26*(1), 105–121. https://doi.org/10.1007/s10869-010-9185-2

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., … Pascual-Leone, A. (2008). Transcranial direct current stimulation: State of the art 2008. *Brain Stimulation*, *1*(3), 206–223. https://doi.org/10.1016/j.brs.2008.06.004

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, Brian A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Page, M. J., Altman, D. G., Shamseer, L., McKenzie, J. E., Ahmadzai, N., Wolfe, D., … Moher, D. (2018). Reproducible research practices are underused in systematic reviews of biomedical interventions. *Journal of Clinical Epidemiology*, *94*, 8–18. https://doi.org/10.1016/j.jclinepi.2017.10.017

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Page, M. J., & Moher, D. (2017). Evaluations of the uptake and impact of the Preferred

Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: A scoping review. *Systematic Reviews*, *6*(1), 263. https://doi.org/10.1186/s13643-017-0663-8

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*, n160. https://doi.org/10.1136/bmj.n160

Plesser, H. (2018). Reproducibility vs. Replicability: A Brief History of a Confused Terminology. *Front. Neuroinform.* https://doi.org/10.3389/fninf.2017.00076

Polanin, J. R., Hennessy, E. A., & Tsuji, S. (2020). Transparency and Reproducibility of Meta-Analyses in Psychology: A Meta-Review. *Perspectives on Psychological Science*, *15*(4), 1026–1041. https://doi.org/10.1177/1745691620906416

Pussegoda, K., Turner, L., Garritty, C., Mayhew, A., Skidmore, B., Stevens, A., … Moher, D. (2017). Systematic review adherence to methodological or reporting quality. *Systematic Reviews*, *6*(1), 131. https://doi.org/10.1186/s13643-017-0527-2

Quintana, D. S. (2015). From pre-registration to publication: A non-technical primer for conducting a meta-analysis to synthesize correlational data. *Frontiers in Psychology*, *6*, 1549. https://doi.org/10.3389/fpsyg.2015.01549

R Core Team. (2021). *R: A language and environment for statistical computing*. Manual, Vienna, Austria: R Foundation for Statistical Computing.

Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., … Krakauer, J. W. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proceedings of the National Academy of Sciences*, *106*(5), 1590–1595. https://doi.org/10.1073/pnas.0805413106

Renkewitz, F., & Keiner, M. (2019). How to Detect Publication Bias in Psychological Research. *Zeitschrift Für Psychologie*, *227*(4), 261–279.

https://doi.org/10.1027/2151-2604/a000386

Reporting guidelines | The EQUATOR Network. (n.d.). https://www.equator-network.org/reporting-guidelines/.

Riggs, A., Patel, V., Paneri, B., Portenoy, R. K., Bikson, M., & Knotkova, H. (2018). At-Home Transcranial Direct Current Stimulation (tDCS) With Telehealth Support for Symptom Control in Chronically-Ill Patients With Multiple Symptoms. *Frontiers in Behavioral Neuroscience*, *12*, 93. https://doi.org/10.3389/fnbeh.2018.00093

Rohatgi, A. (2021). WebPlotDigitizer.

Rosen, D. S., Erickson, B., Kim, Y. E., Mirman, D., Hamilton, R. H., & Kounios, J. (2016). Anodal tDCS to Right Dorsolateral Prefrontal Cortex Facilitates Performance for Novice Jazz Improvisers but Hinders Experts. *Frontiers in Human Neuroscience*, *10*, 579. https://doi.org/10.3389/fnhum.2016.00579

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis*. Wiley Online Library.

RStudio Team. (2021). RStudio: Integrated development environment for r. Boston, MA: RStudio, PBC.

Schalken, N., & Rietbergen, C. (2017). The Reporting Quality of Systematic Reviews and Meta-Analyses in Industrial and Organizational Psychology: A Systematic Review. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01395

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., … Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ*, *349*, g7647. https://doi.org/10.1136/bmj.g7647

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as

Significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U. (2017). [59] PET-PEESE Is Not Like Homeopathy. *Data Colada*.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547. https://doi.org/10.1037/a0033242

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *9*(6), 666–681. https://doi.org/10.1177/1745691614553988

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller (2015).

Stagg, C. J., & Nitsche, M. A. (2011). Physiological basis of transcranial direct current stimulation. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *17*(1), 37–53. https://doi.org/10.1177/1073858410386614

Stanley, T. (2008). Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection*. *Oxford Bulletin of Economics and Statistics*, *70*(1), 103–127.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

Steenbergen, L., Sellaro, R., Hommel, B., Lindenberger, U., Kühn, S., & Colzato, L. S. (2016). “Unfocus” on foc. Us: Commercial tDCS headset impairs working memory. *Experimental Brain Research*, *234*(3), 637–643.

Steiner, P. M., Wong, V. C., & Anglin, K. (2019). A Causal Replication Framework for Designing and Assessing Replication Efforts. *Zeitschrift Für Psychologie*, *227*(4),

280–292. https://doi.org/10.1027/2151-2604/a000385

Sterling, T. D. (1959). Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance–Or Vice Versa. *Journal of the American Statistical Association*, *54*(285), 30–34. https://doi.org/10.2307/2282137

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, *15*(3), e2000797. https://doi.org/10.1371/journal.pbio.2000797

Szucs, D., & Ioannidis, J. PA. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage*, *221*, 117164. https://doi.org/10.1016/j.neuroimage.2020.117164

Taylor, A. E., & Munafò, M. R. (2016). Triangulating meta-analyses: The example of the serotonin transporter gene, stressful life events and major depression. *BMC Psychology*, *4*(1), 23. https://doi.org/10.1186/s40359-016-0129-0

Tobias, A. (1999). Assessing the influence of a single study in the meta-anyalysis estimate. *Stata Technical Bulletin*, *8*(47).

Ushey, K., RStudio, & PBC. (2021). *Renv: Project Environments*.

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How Many Studies Do You Need?: A Primer on Statistical Power for Meta-Analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215–247. https://doi.org/10.3102/1076998609346961

van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Conducting Meta-Analyses Based on p Values: Reservations and Recommendations for Applying p-Uniform and p-Curve. *Perspectives on Psychological Science*, *11*(5), 713–729. https://doi.org/10.1177/1745691616650874

van Assen, M. A. L. M., van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, *20*(3), 293–309. https://doi.org/10.1037/met0000025

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., . . . Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, *7*(1), 55–79. https://doi.org/10.1002/jrsm.1164

Vevea, J. L., Coburn, K., & Sutton, A. (2019). Publication Bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (pp. 383–430). Russell Sage Foundation. https://doi.org/10.7758/9781610448864.21

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419–435.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*(4), 428.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Voracek, M., Kossmeier, M., & Tran, U. S. (2019). Which data to meta-analyze, and how? A specification-curve and multiverse-analysis approach to meta-analysis. *Zeitschrift Für Psychologie*, *227*(1), 64–82. https://doi.org/10.1027/2151-2604/a000357

Wan, X., Wang, W., Liu, J., & Tong, T. (2014). Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Medical Research Methodology*, *14*(1), 135. https://doi.org/10.1186/1471-2288-14-135

Wayant, C., Page, M. J., & Vassar, M. (2019). Evaluation of Reproducible Research Practices in Oncology Systematic Reviews With Meta-analyses Referenced by National Comprehensive Cancer Network Guidelines. *JAMA Oncology*, *5*(11), 1550–1555. https://doi.org/10.1001/jamaoncol.2019.2564

Wexler, A. (2018). Who Uses Direct-to-Consumer Brain Stimulation Products, and Why? A Study of Home Users of tDCS Devices. *Journal of Cognitive Enhancement*, *2*(1), 114–134. https://doi.org/10.1007/s41465-017-0062-z

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations*. Manual.

Wickham, H. (2021a). *Forcats: Tools for working with categorical variables (factors)*. Manual.

Wickham, H. (2021b). *Tidyr: Tidy messy data*. Manual.

Wickham, H., & Bryan, J. (2019). *Readxl: Read excel files*. Manual.

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data manipulation*. Manual.

Wickham, H., & Hester, J. (2021). *Readr: Read rectangular text data*. Manual.

Woods, A. J., Antal, A., Bikson, M., Boggio, P. S., Brunoni, A. R., Celnik, P., ... Nitsche, M. A. (2016). A technical guide to tDCS, and related non-invasive brain stimulation tools. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *127*(2), 1031–1048. https://doi.org/10.1016/j.clinph.2015.11.012

Wurzman, R., Hamilton, R. H., Pascual-Leone, A., & Fox, M. D. (2016). An Open Letter Concerning Do-It-Yourself Users of Transcranial Direct Current Stimulation. *Annals of Neurology*, *80*(1), 1–4. https://doi.org/10.1002/ana.24689

Xie, Y. (2014). Knitr: A comprehensive tool for reproducible research in R. In V. Stodden, F. Leisch, & R. D. Peng (Eds.), *Implementing reproducible computational research*. Chapman and Hall/CRC.

Xie, Y. (2015). *Dynamic documents with R and knitr* (Second). Boca Raton, Florida: Chapman and Hall/CRC.

Xie, Y. (2021). *Knitr: A general-purpose package for dynamic report generation in r*. Manual.

Zettler, P. J. (2017). tDCS Research in a World With FDA Regulation. *AJOB Neuroscience*, *8*(1), 1–2. https://doi.org/10.1080/21507740.2017.1293192

Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax*. Manual.

Freie Universität Berlin

Declaration

Herewith I affirm that I wrote the work in hand autonomously and never have used another source or resource as declared.

.....……..

Date and signature