Mana Biabani*, Michael Farrell, Maryam Zoghi, Gary Egan and Shapour Jaberzadeh

Crossover design in transcranial direct current stimulation studies on motor learning: potential pitfalls and difficulties in interpretation of findings

https://doi.org/10.1515/revneuro-2017-0056
Received July 20, 2017; accepted October 6, 2017

Abstract: Crossover designs are used by a high proportion of studies investigating the effects of transcranial direct current stimulation (tDCS) on motor learning. These designs necessitate attention to aspects of data collection and analysis to take account of design-related confounds including order, carryover, and period effects. In this systematic review, we appraised the method sections of crossover-designed tDCS studies of motor learning and discussed the strategies adopted to address these factors. A systematic search of 10 databases was performed and 19 research papers, including 21 experimental studies, were identified. Potential risks of bias were addressed in all of the studies, however, not in a rigorous and structured manner. In the data collection phase, unclear methods of randomization, various lengths of washout period, and inconsistency in the counteracting period effect can be observed. In the analytical procedures, the stratification by sequence group was often ignored, and data were treated as if it belongs to a simple repeated-measures design. An inappropriate use of crossover design can seriously affect the findings and therefore the conclusions drawn from tDCS studies on motor learning. The results indicate a pressing need for the development of detailed guidelines for this type of studies to benefit from the advantages of a crossover design.

***Corresponding author: Mana Biabani,** Department of Physiotherapy, School of Primary Health Care, Faculty of Medicine, Nursing and Health Sciences, Monash University, Frankston, Melbourne, Victoria 3199, Australia; and Monash Biomedical Imaging, Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, Melbourne, Victoria 3800, Australia, e-mail: mana.biabanimoghadam@monash.edu
Michael Farrell and Gary Egan: Monash Biomedical Imaging, Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, Melbourne, Victoria 3800, Australia
Maryam Zoghi: Discipline of Physiotherapy, Department of Rehabilitation, Nutrition and Sport, School of Allied Health, La Trobe University, Bundoora, Melbourne, Victoria 3086, Australia
Shapour Jaberzadeh: Department of Physiotherapy, School of Primary Health Care, Faculty of Medicine, Nursing and Health Sciences, Monash University, Frankston, Melbourne, Victoria 3199, Australia

Keywords: carryover effect; order effect; period effect.

Introduction

Motor learning refers to the ability to acquire new skills or to improve the efficiency of movements (Dayan and Cohen, 2011). The effects of transcranial direct current stimulation (tDCS) on this behavior have been widely investigated during the last two decades. Putative changes in motor behavior associated with tDCS have been ascribed to the capacity of the technique to induce neural plasticity (Dayan and Cohen, 2011; Jaberzadeh and Zoghi, 2013), although there remains a lack of consensus regarding the effect of tDCS on motor learning. The controversy in findings has been attributed to a number of issues such as characteristics of tDCS applications, training tasks, and experimental paradigms (Hashemirad et al., 2016). Nevertheless, the significance of study design has received considerably less attention in the literature.

Crossover designs are longitudinal studies in which two or more experimental interventions are administered to the same group of participants in a specified or random order (Portney and Watkins, 2000). These designs are employed by the majority of tDCS studies. Unlike parallel designs, under which each participant only receives one intervention during the entire experiment, in crossover designs, the same participants are exposed to different types of treatments, and outcomes in the same individual are contrasted. Crossover designs not only help to avoid unbalanced confounding variables, such as age and sex, across study groups, but they can also deliver comparable statistical power to parallel designs but with smaller sample sizes (Stephen, 2002). Despite these advantages, which may improve the cost efficiency of the research (Woods et al., 1989; Chen et al., 2002), crossover designs are prone to biases that demand special attention during study planning, data collection, and analysis.

Design-related nuisance parameters are of particular concern in crossover studies, which can easily lead to the misinterpretation of the data and distort conclusions drawn from the findings. Three major confounding factors that are widely considered to affect the crossover studies are order, carryover, and period effects (Woods

et al., 1989; Cleophas and Zwinderman, 2012). Order effect refers to the influence of the temporal order of presenting different interventions to the participants on study results (Woods et al., 1989; Portney and Watkins, 2000; Cleophas and Zwinderman, 2012). Considering 'period' as the time in which an intervention is applied, the carryover effect refers to the remaining effect of an intervention in one period into the subsequent one, whereas the period effect refers to changes in the ability of participants to respond to interventions from one period to another (Portney and Watkins, 2000; Stephen, 2002). These factors can decrease the statistical power to detect the actual effect of interventions and have been suggested as potential contributing factors to the inconsistent outcomes across brain stimulation experiments (Fregni et al., 2005; Brunoni and Fregni, 2011; Hulst et al., 2017). The negative impact of crossover design-related confounds on the power of tDCS studies (Palm et al., 2013) and the association of this design with the lack of improvements in tDCS-induced behavioral changes have been discussed in the literature (Brunoni and Vanderhasselt, 2014). However, there is no study discussing the reasons for high susceptibility of these specific types of experiments to crossover confounds or outlining the methodological considerations that should particularly be taken into account in these studies to minimize the loss of power. In general, to reduce the risk of data contamination by these factors, it is recommended to employ an appropriate method of randomization of the order in which interventions are delivered, use sufficient intervals between periods, and stabilize participants' response levels between sessions (Reed, 2004). In addition to adopting procedural strategies during the design stage, biostatisticians also argue for specific analytic strategies to reduce the risk of confusing confounding factors with intervention effects (Woods et al., 1989; Senn, 2002).

To date, a number of reviews have been conducted to address different aspects of published papers on tDCS and motor learning (Foerster et al., 2015; Hashemirad et al., 2016). However, the significance of study design has received considerably less attention in the literature.

Therefore, this systematic review examines the method section of crossover studies, which have applied tDCS to enhance the effects of motor learning, and discusses the strategies adopted by these studies to address order, carryover, and period effects.

Materials and methods

A systematic search of the following databases was performed for relevant English studies published from their inception to April 2016: PubMed, Ovid Medline, Scopus, CINAHL, Cochrane Library, PROQuest, Physiotherapy Evidence Database (PEDro), SPORT Discuss, EBM reviews, and EMBASE. In addition, the reference lists of all retrieved papers were hand searched to find any other relevant studies unidentified by the first search strategy. Key search terms were ('transcranial direct current stimulation' OR tDCS) AND motor AND (learning OR performance). Appropriate studies were initially identified by screening the title and abstract of papers. Potentially relevant studies were evaluated thoroughly by reference to the full text. For inclusion, studies needed to use a crossover design, involve samples of healthy participants, compare the effects of at least two different stimulation conditions (anode, cathode and sham), incorporate a motor learning task, and report behavioral outcomes. Noncrossover studies (e.g. parallel designs) and studies on patient population or nonhuman participants were excluded from the review. We also did not consider experiments that did not apply tDCS during a learning task or those without assessment of behavioral outcome measures. Reviews, case reports, and letters were excluded from this systematic review. The strategies being used to address order, carryover, and period effects, during planning/data collection and analysis phases of the studies, were extracted.

Results

The search strategy identified 2578 articles including 1556 duplicates. Screening papers by titles and abstracts identified duplicates, and these 933 publications were excluded from further analysis. Of 89 papers fully examined, 19 papers containing 21 experimental studies met the inclusion criteria and were included in the review (Figure 1).

Characteristics of included studies

Of the 21 included studies, 11 studies compared two types of stimulation (i.e. anode or cathode versus sham; Hunter et al., 2009; Kantak et al., 2012; Zimerman et al., 2012, 2013; Pavlova et al., 2014; Sriraman et al., 2014; Amadi et al., 2015; Avila et al., 2015; Conley et al., 2015; Minarik et al., 2015; Rroji et al., 2015), whereas 8 studies assessed three current types (i.e. anode, cathode, and sham; Nitsche et al., 2003, 2010; Stagg et al., 2011; Karok and Witney, 2013; Shah et al., 2013; Pavlova et al., 2014; Spieser et al., 2015; Ambrus et al., 2016). Four studies compared different electrode montages (Kantak et al., 2012; Karok and

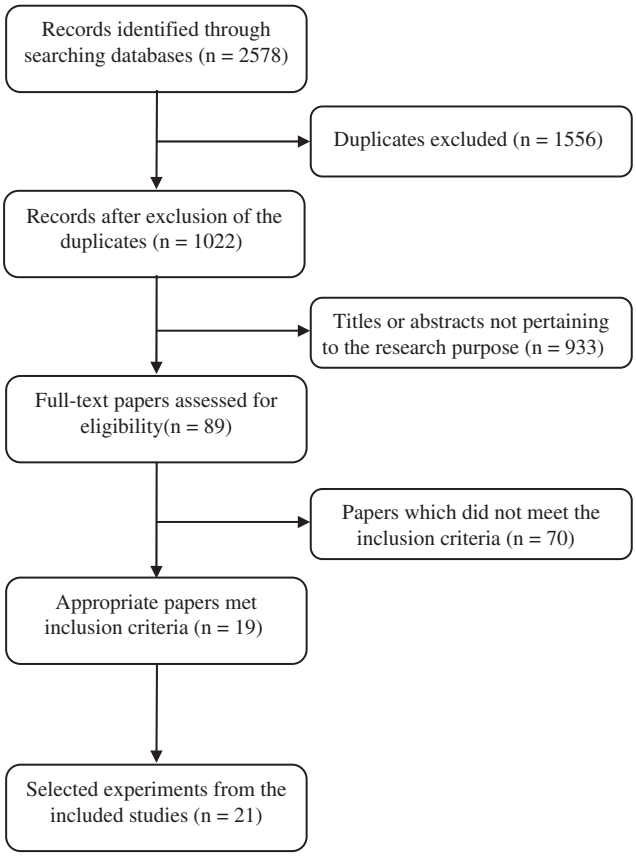

Figure 1: Systematic review of the flow diagram.

Witney, 2013; Shah et al., 2013; Pavlova et al., 2014), and one compared different current intensities (Cuypers et al., 2013; Table 1).

Strategies to minimize design-related confounding factors

Planning and data collection

Order effect

Except for two studies, which did not report the method of altering the sequence of intervention presented to participants (Zimerman et al., 2012, 2013), all other studies randomized the order of interventions (Nitsche et al., 2003, 2010; Hunter et al., 2009; Stagg et al., 2011; Kantak et al., 2012; Cuypers et al., 2013; Karok and Witney, 2013; Shah et al., 2013; Pavlova et al., 2014; Sriraman et al., 2014; Amadi et al., 2015; Avila et al., 2015; Conley et al., 2015; Minarik et al., 2015; Rroji et al., 2015; Spieser et al., 2015; Ambrus et al., 2016). Some studies briefly outlined their methods of randomization as block randomization (Sriraman et al., 2014), semibalancing (Pavlova et al., 2014), balancing (Nitsche et al., 2010; Pavlova et al., 2014), and counterbalancing (Stagg et al., 2011; Kantak et al., 2012; Karok and Witney, 2013; Amadi et al., 2015; Avila et al., 2015; Minarik et al., 2015; Rroji et al., 2015; Spieser et al., 2015); however, no study has clearly described these methods (Table 2).

Carryover effect

All reviewed studies reported a washout period; however, the length of this period ranged from 2 days (Spieser et al., 2015) to more than 3 months (Rroji et al., 2015). Although the majority of researchers used 7 days as washout period (Nitsche et al., 2003, 2010; Hunter et al., 2009; Stagg et al., 2011; Karok and Witney, 2013; Pavlova et al., 2014; Sriraman et al., 2014; Amadi et al., 2015; Minarik et al., 2015), washout periods 2 (Spieser et al., 2015), 3 (Cuypers et al., 2013), 4 (Shah et al., 2013; Ambrus et al., 2016), 5 (Galea and Celnik, 2009), 8 (Kantak et al., 2012), 10 (Zimerman et al., 2013), 21 (Conley et al., 2015), or between 3 and 7 days (Avila et al., 2015) were reported. No studies provided a reference or a clear rationale for choosing the reported length of washout period (Table 2).

Period effect

The strategies adopted by the retrieved studies to rule out period effect can be categorized into two groups. Three studies used practice before the commencement of each session (Karok and Witney, 2013; Pavlova et al., 2014; Conley et al., 2015), whereas seven papers exposed the participants to different versions of the evaluating tasks and used random trials (Nitsche et al., 2003; Kantak et al., 2012; Zimerman et al., 2012, 2013; Cuypers et al., 2013; Karok and Witney, 2013; Ambrus et al., 2016). One research study used a familiarization period at the beginning of each session and deadaptation trials between the stimulation sessions (Hunter et al., 2009). The methods used by the other nine studies were not clearly provided (Table 2).

Data analysis

Assessing the existence of nuisance factors

The majority of studies applied analytical procedures to model confounding factors (Nitsche et al., 2003, 2010; Hunter et al., 2009; Stagg et al., 2011; Kantak et al., 2012; Zimerman et al., 2012, 2013; Cuypers et al., 2013; Karok and Witney, 2013; Shah et al., 2013; Pavlova et al., 2014; Amadi et al., 2015; Avila et al., 2015; Minarik et al., 2015;

Table 1: Characteristics of the included studies.

Study	n	Stimulation	Task	Behavioral outcomes/evaluation items Reaction time
Avila et al. (2015)	10	A/S	Inward adaptation, outward adaptation	Saccadic gains, kinematics, adaptation gain
Minarik et al. (2015)	22	A/S	Grip force control	Time-on-target, deviation
Amadi et al. (2015)	13	A/S (4 sessions)	Sequential finger presses	Reaction time
Rroji et al. (2015)	14	A/S	Discrete ballistic thumb flexion	Peak velocity
Conley et al. (2015)	20 (Exp. 3)	A/S	Cued go/no-go task	Reaction time, error rate
Spieser et al. (2015)	24	A/C/S	Stimulus response compatibility task	Reaction time, error rate
Sriraman et al. (2014)	12	A/S (3 sessions)	Ankle tracking	Accuracy index
Pavlova et al. (2014)	12 (Exp. 1)	A/S	Strength dexterity	Absolute force, variability, compression number
	12 (Exp. 2)	A/C/S (5 sessions)		
Cuypers et al. (2013)	13	A/S (3 sessions)	Finger sequence task	Speed, accuracy
Karok and Witney (2013)	20	A/A-C/S	Finger sequence task	Response time, accuracy
Shah et al. (2013)	8	A/C/S (5 sessions)	Ankle tracking	Accuracy index
Zimerman et al. (2013)	10	A/S	Sequential finger tapping	Number of correct sequences, slope of improvement
Kantak et al. (2012)	13	A/S (3 sessions)	SRTT	Reaction time
Zimerman et al. (2012)	10	C/S	Sequential finger tapping	Number of correct sequences
	13			
Stagg et al. (2011)	7	A/C/S	Sequential finger presses	Reaction time
Nitsche et al. (2010)	12	A/C/S	Randomized sequential finger movement	Reaction time, error rate
Hunter et al. (2009)	14	A/S	Arm reaching task	Summed error, signed error, movement time, reaction time
Nitsche et al. (2003)	20	A/C/S	SRTT	Response time, error rate

n, number of participants; Exp., experiment; A, anode; C, cathode, S, sham; SRTT, serial reaction time task. All the included studies recruited healthy participants.

Rroji et al., 2015; Ambrus et al., 2016). Fourteen studies examined the presence of confounding factors by comparing baseline values before each intervention or the measured values of the first trial blocks, among which four articles applied paired *t*-tests (Zimerman et al., 2012, 2013; Cuypers et al., 2013; Avila et al., 2015) and 10 used repeated-measures analysis of variance (RM ANOVA; Nitsche et al., 2003, 2010; Hunter et al., 2009; Stagg et al., 2011; Kantak et al., 2012; Karok and Witney, 2013; Shah et al., 2013; Pavlova et al., 2014; Amadi et al., 2015; Ambrus et al., 2016). Five studies also incorporated session (Minarik et al., 2015; Rroji et al., 2015) or order (Hunter et al., 2009; Karok and Witney, 2013; Pavlova et al., 2014) factors in ANOVA models. In a different approach, Amadi et al. (2015) analyzed the sham-only period (without task) to identify changes associated with the passage of time, but only nonbehavioral outcomes were examined. The relevant statistical tests were not clearly stated in the remaining three studies (Sriraman et al., 2014; Conley et al., 2015; Spieser et al., 2015; Table 2).

Including nuisance factors in the analyses of stimulation effects

In 13 of the 21 included studies, the normalized data to baseline values were used for analysis (Nitsche et al., 2003; Hunter et al., 2009; Stagg et al., 2011; Zimerman et al., 2012; Cuypers et al., 2013; Shah et al., 2013; Pavlova et al., 2014; Sriraman et al., 2014; Amadi et al., 2015; Rroji et al., 2015; Ambrus et al., 2016). Almost all studies compared the effects of different types of interventions, using paired *t*-test or RM ANOVA, as if the data belong to a typical repeated-measures design without different sequence groups (Nitsche et al., 2003, 2010; Hunter et al., 2009; Stagg et al., 2011; Kantak et al., 2012; Zimerman et al., 2012, 2013; Cuypers et al., 2013; Karok and Witney, 2013; Shah et al., 2013; Pavlova et al., 2014; Sriraman et al., 2014; Amadi et al., 2015; Avila et al., 2015; Minarik et al., 2015; Spieser et al., 2015; Ambrus et al., 2016). Two studies considered the potential confounding parameters in analytical procedures estimating treatment effects and incorporated 'order' into their statistical model (Conley et al., 2015; Rroji et al., 2015; Table 2).

Table 2: Strategies adopted by studies to minimize the design-related nuisance parameters.

Study	Data collection phase			Data analysis phase	
	Order effect	Carryover effect[a]	Period effect	Nuisance factors assessment	Intervention effect estimation
Ambrus et al. (2016)	Randomized	≥4 days	Different task versions, random trials	RM ANOVA of block 1 values	RM ANOVA: Time*Stimulation[b]
Avila et al. (2015)	Pseudo-randomized and counterbalanced	3–7 days	–	Paired *t*-test of baseline values	RM ANOVA: Time*Stimulation
Minarik et al. (2015)	Counterbalanced	7 days	–	RM ANOVA: Session*Block	RM ANOVA: Time*Simulation*Difficulty
Amadi et al. (2015)	Counterbalanced	≥7 days	–	RM ANOVA of block 1 values, testing a nonbehavioral value during a sham-only time period	RM ANOVA: Time*Stimulation[b]
Rroji et al. (2015)	Counterbalanced	≥84 days	–	RM ANOVA of absolute values: Session*Block	RM ANOVA: Block*Stimulation*Order[b]
Conley et al. (2015)	Pseudo-randomized	≥21 days	Practice before each session	–	Four-way mixed design GLM: Stimulation*Cue*Response Hand*Order
Spieser et al. (2015)	Counterbalanced	≥2 days	–	–	RM ANOVA: Stimulation*Compatibility
Sriraman et al. (2014)	Block randomized	≥7 days	–	–	RM ANOVA: Time*Stimulation[b]
Pavlova et al. (2014)	Randomized and balanced	≥7 days	Practice before each session	RM ANOVA of baseline values	Paired *t*-test[b]
	Randomized and semibalanced			RM ANOVA of absolute values: Order*Stimulation	RM ANOVA: Time*Stimulation[b]
Cuypers et al. (2013)	Randomized and counterbalanced	≥3 days	Different task versions	Paired *t*-test of baseline values	Mixed model (SAS): Time*Stimulation[b] Paired *t*-test[b]
Karok and Witney (2013)	Randomized and counterbalanced	≥7 days	Practice before each session, different task versions	RM ANOVA of baseline values	RM ANOVA: Time*Stimulation ANCOVA: Time*Stimulation (covariate: order)
Shah et al. (2013)	Randomized	≥4 days	–	RM ANOVA of baseline values	RM ANOVA: Time*Stimulation[b]
Zimerman et al. (2013)	–	≥10 days	Different task versions	Paired *t*-test of baseline values	RM ANOVA: Time*Stimulation[b]
Kantak et al. (2012)	Randomized and counterbalanced	≥8 days	Different task versions, random trials	RM ANOVA of baseline values	RM ANOVA: Time*Stimulation
Zimerman et al. (2012)	–	≥10 days	Different task versions	Paired *t*-test of baseline values	RM ANOVA: Block*Stimulation
Stagg et al. (2011)	Counterbalanced	≥7 days (sham: 48 h)	–	RM ANOVA of block 1 values	RM ANOVA: Time*Stimulation[b]
Nitsche et al. (2010)	Randomized and balanced	≥7 days	–	RM ANOVA of block 1 values	RM ANOVA: Time*Stimulation[b]
Hunter et al. (2009)	Randomized	≥7 days	Movement familiarization, washout trials	RM ANOVA of baseline values RM ANOVA: Block*Order*Stimulation	RM ANOVA: Time*Stimulation[b]
Nitsche et al. (2003)	Randomized	≥7 days	Different task versions, random trials	RM ANOVA of block 1 values	RM ANOVA: Time*Stimulation[b]

[a]Length of washout period. [b]Data randomized to baseline values.

Discussion

In this study, we investigated the methodological aspects of crossover studies that examined the effects of tDCS on motor learning. We found that critical issues such as order, carryover, and period effects have not been consistently addressed by these studies. The different strategies reported by the included studies can be traced through their planning and data collection as well as analysis phase.

Planning and data collection

Order effect

The order in which stimulation types are presented to participants might be a potential source of bias (Portney and Watkins, 2000). The fixed order of intervention sessions, in which a specific type of stimulation is always preceded or followed by the exact same stimulation type, may be considered a confounding variable. This arrangement of stimuli makes it difficult to separate intervention effects from the effects of order and provides a context for extraneous parameters such as carryover and time effects to influence the outcomes (Portney and Watkins, 2000). To prevent these effects in crossover studies, researchers use different strategies to manipulate the order of interventions between the participants. One strategy is randomization, in which the order of intervention is randomized for each participant. There is a chance when randomly assigning participants that one sequence is repeated more frequently than others; consequently, this approach can create an unbalanced group of participants (Thomas and Hersen, 2011). Another method is complete counterbalancing, in which all of the possible intervention sequences are applied, and all sequences are repeated the same number of times (DePuy and Berger, 2005; Thomas and Hersen, 2011). There are also some incomplete counterbalancing methods that include elements of randomization and counterbalancing (DePuy and Berger, 2005; Thomas and Hersen, 2011). The literature indicates that randomization alone is not adequate to compensate for order effects, because it may produce sequence groups with an uneven number of participants (Thomas and Hersen, 2011). Alternatively, complete counterbalancing leads to the balanced distribution of order effect among interventions and also allows for the efficient statistical evaluation of the effects of order on findings (Thomas and Hersen, 2011).

The majority of the included studies in this review used randomization to categorize their participants into different groups, about half of which counterbalanced the order. Although nine papers have stated that the orders of stimulations were counterbalanced (Stagg et al., 2011; Kantak et al., 2012; Cuypers et al., 2013; Karok and Witney, 2013; Amadi et al., 2015; Avila et al., 2015; Minarik et al., 2015; Rroji et al., 2015; Spieser et al., 2015), the number of intervention types and sample sizes in some studies raise the question whether or not they have used the term 'counterbalancing' accurately (Stagg et al., 2011; Kantak et al., 2012; Cuypers et al., 2013; Amadi et al., 2015; Avila et al., 2015). To completely counterbalance the order of, for example, three stimulation types, six sequence groups would be required. To allocate the equal numbers of participants to each group, the number of participants is expected to be a product of 6. Thus, this is not clear how these studies (Stagg et al., 2011; Kantak et al., 2012; Cuypers et al., 2013; Amadi et al., 2015; Avila et al., 2015) have counterbalanced the stimulation conditions among their participants.

Carryover effect

A crucial step in designing a crossover trial is to consider the possibility of carryover effects, which means that the effect of one intervention may persist and affect the result of the subsequent intervention. To avoid this issue, intervention periods should be separated by a sufficiently long washout period, which is estimated based on the lasting effect of the applied interventions. If the length of lasting effect is unknown, a lengthy washout period is usually predefined based on the best available knowledge about the intervention; however, this method requires consideration. In addition to short washout periods, which are more likely to be confounded by carryover effects, long intervals would increase the likelihood of dropouts and the risk of introducing other confounding variables to the study such as change over time in participants' nutrition habits or physical activities, which may influence the outcome measures in the study.

The lasting effect of a single session tDCS on motor learning can vary across different stimulation parameters (e.g. site, duration, and intensity of stimulation; Nitsche and Paulus, 2001; Reinhart and Woodman, 2015; Reinhart et al., 2017), task variants (Saucedo Marquez et al., 2013; Buch et al., 2017), and the outcomes being measured (Reis et al., 2009; Buch et al., 2017). Changes in corticospinal excitability are the most commonly measured outcomes to determine the lasting effect of tDCS, which have been suggested to last up to a few hours depending on the simulation parameters (Nitsche

and Paulus, 2001; Jaberzadeh et al., 2012). However, the alteration of cortical excitability is not the only contributing factor to the neurophysiology of motor learning (Ungerleider et al., 2002; Sun et al., 2016). Learning a new task involves more persistent neurophysiological changes such as synaptic modifications and the formation of new connections between neurons, which have been shown to be perturbed by tDCS (Ungerleider et al., 2002; Polanía et al., 2011; Dayan et al., 2013). However, there is no solid evidence regarding the magnitude and duration of those interactions after one session of stimulation. Furthermore, the relationship between neurophysiological and behavioral after effects of tDCS is still unclear. From a behavioral perspective, the formation of motor skills occurs during (i.e. online motor learning) and after (i.e. offline motor learning or consolidation) motor practice (Vahdat et al., 2017). Studies suggest that the interaction of tDCS with the physiological process of learning persists beyond the termination of stimulation and is proposed to last during the consolidation phase (Reis et al., 2009, 2013). To the best of our knowledge, the furthest time point of skill reassessment after a single session motor training was 3 days, made by Walker et al. (2003), which revealed that motor skills still continued to increase, indicating that the consolidation phase is still ongoing. Therefore, the duration of washout requisite for the absence of further changes in motor skills remains unknown.

The lack of a profound understanding of motor learning processes and the high heterogeneity of the stimulation protocols make it difficult to determine the optimum washout time for tDCS studies on motor learning. This can explain the inconsistency in the length of washout periods and the application of additional statistical tests to rule out carryover effects in the included studies.

Period effect

The tendency of outcomes in the second period to differ systematically from those in the first, irrespective of interventions, is another potential confound of crossover studies (Woods et al., 1989). Many factors can give rise to period effects, such as familiarization with trial processes or improvement in a participant's skill (Díaz-Uriarte, 2002; Wellek and Blettner, 2012). Due to the nature of motor learning tasks, separating carryover effects from period effects is difficult. Therefore, any effort to defamiliarize the task or return the subject's performance to the baseline level before each period can address carryover effects of motor training as well as period effects.

In some studies of sequential motor skill learning, investigators applied different versions of the sequence tasks in each session (Nitsche et al., 2003; Kantak et al., 2012; Zimerman et al., 2012, 2013; Cuypers et al., 2013; Karok and Witney, 2013; Ambrus et al., 2016). Although this method can help to control the perceptual component of the task, it underestimates other skill parameters such as increase in visuospatial processing speed and hand-eye coordination that may improve in that experimental context during each period. However, in a number of studies, investigators introduced random sequences to the experimental procedures to elicit the pure motor aspects of responses and then contrasted responses to sequence trials against those of the random trials (Nitsche et al., 2003; Kantak et al., 2012; Ambrus et al., 2016). Although this approach has been used to minimize some confounding factors such as fatigue and motivation, it literally ignores the motor component of skill learning and could be more appropriate when the perceptual component is the only measure of interest (Robertson, 2007). There are some other measures that have been taken by some included studies to avoid the influence of period effect on study results. In addition to a common familiarization process, some investigators trained the participants before each stimulation sessions, which could be a strategy to maintain the same level of performance at baselines (Karok and Witney, 2013; Pavlova et al., 2014; Conley et al., 2015). Implementing practice trials would not guarantee comparable baseline levels, and the efficacy of this approach needs to be investigated with appropriate statistical methods, which will be discussed in the following section.

Data analysis

Assessing the existence of nuisance factors

Even complete counterbalancing does not ensure the elimination of the undesired effects of order (i.e. carryover and period effects), and this should be taken into account through proper data analysis (Campbell and Stanley, 1963).

As indicated in Table 2, the majority of the reviewed studies tested the presence of confounding factors using preliminary analytical procedures; however, different studies adopted different strategies. Some studies tested the specific effect of 'period' (session; Minarik et al., 2015; Rroji et al., 2015) or 'order' (Hunter et al., 2009; Pavlova et al., 2014) on outcomes using RM ANOVA. However, the most common method reported by the studies was

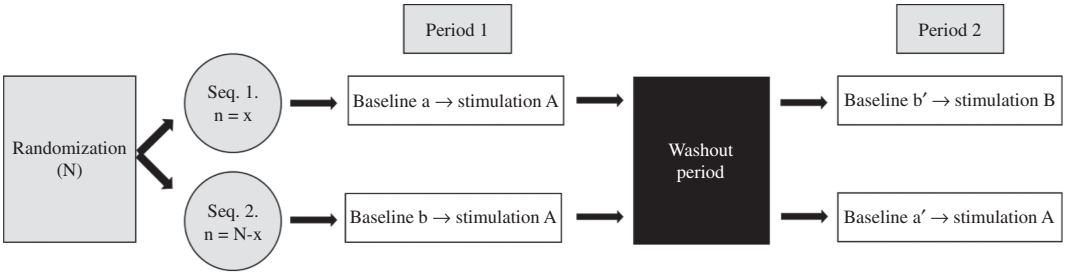

Figure 2: Design of a crossover trial with n participants.

a contrast of baseline values, taken at the beginning of or immediately before each stimulation, to investigate whether participants started each stimulation session from the same baseline level (Nitsche et al., 2003, 2010; Hunter et al., 2009; Stagg et al., 2011; Kantak et al., 2012; Zimerman et al., 2012, 2013; Cuypers et al., 2013; Karok and Witney, 2013; Shah et al., 2013; Pavlova et al., 2014; Amadi et al., 2015; Avila et al., 2015; Ambrus et al., 2016). Although the rationale behind this approach is clear, the way that data have been treated raises some concerns. To illustrate the point, consider a simple crossover design depicted in Figure 2.

The method that has been adopted in these studies is comparing baseline measures taken before stimulation A (a and a′) to those before stimulation B (b and b′) using paired *t*-test (or RM ANOVA). This method necessitates balanced randomization; otherwise, the results could be biased. In Figure 2, baseline b′ indicates the presence of the period effect plus carryover effect of stimulation A, whereas baseline a′ shows the period effect plus carryover effect of stimulation B. If there is an unequal number of participants in each group, the result of baseline comparisons can be influenced by the period effect and also by the carryover effect presented in a larger group (Díaz-Uriarte, 2002). In addition, as only the second baselines are confounded with nuisance parameters, lumping data from both periods (e.g. a and a′) can add an unnecessary between-subject variance, which decreases the power of the test. Alternative analytical procedures have been advocated by statisticians for use in such situations (Senn, 1994; Stephen, 2002).

Including nuisance factors in the analyses of stimulation effects

The primary aim of every crossover study is to compare effects of different types of interventions in the absence of any confounding factor. The first step taken by the majority of the included studies in this review was normalizing the outcomes to baseline values to reduce the effects of confounding parameters (Nitsche et al., 2003; 2010; Hunter et al., 2009; Stagg et al., 2011; Cuypers et al., 2013; Shah et al., 2013; Pavlova et al., 2014; Sriraman et al., 2014; Rroji et al., 2015; Zimerman et al., 2013; Ambrus et al., 2016). In addition to economical disadvantages of collecting extra data in every trial, this method has some statistical weaknesses that should be taken into account in interpreting the results. Willian and Pater have shown that, in spite of eliminating carryover effects, this method sacrifices the precision and power of the tests estimating intervention effects (Willian and Pater, 1986).

In the next step, to compare the effects of different types of stimulations, almost all reviewed studies applied statistical procedures for paired samples (paired *t*-test or RM ANOVA). In these tests, the overall outcome measures of stimulation A (Figure 2), regardless of their periods (i.e. 1 or 2), were compared to those of stimulation B. In this approach, data are treated as if it belongs to a simple repeated-measures study design (Wellek and Blettner, 2012). The first concern about this approach is that, due to the confounding effect of period, participants who receive treatment A in period 1 may show a different result from the ones who receive it in period 2. Therefore, by merging data from the participants in different periods, the between-subject variability resulting from period effects is ignored. This would introduce the period effect as a random error to the analysis that leads to a decrease in the power of the test (Díaz-Uriarte, 2002). This highlights the significance of minimizing period effect during the data collection phase of studies. More importantly, in this analytical procedure, we need to assume that the washout period was long enough to eliminate the carryover effect; otherwise, it simultaneously compares both the stimulation and carryover effects, and the results would not be pure treatment differences (Senn, 2002; Reed, 2004). This problem highlights the significance of enough washout periods between stimulation sessions (refer to Senn, 2002; Reed, 2004, for in-depth statistical discussions). Despite some arguments in favor of interpreting

the results under the condition of adopting those assumptions (Senn, 1994; Reed, 2004), some publications argue against the validity of such results stating that between-subject variance needs to be considered in the statistical analysis of crossover studies and they should not be analyzed as a simple repeated-measures design (Díaz-Uriarte, 2002; Wellek and Blettner, 2012). To emphasize the point, Wellek and Blettner believe that 'crossover trials in which the paired *t*-test (or any other procedure for paired samples) was used for analysis are methodologically flawed and do not contribute to evidence-based evaluation of the treatments concerned' (Wellek and Blettner, 2012). Among included studies in this review, only studies by Rroji et al. and Conley et al. considered this issue in their analysis, in which 'order' was put into the model for the estimation of treatment effect (Conley et al., 2015; Rroji et al., 2015). This method has been shown to accommodate the between-subject differences and directly tests the effects of treatments (Díaz-Uriarte, 2002). To date, several statistical methods have been introduced in the literature to address nuisance parameters in crossover studies. The presence of inconclusive evidence in favor or against all of these approaches in the literature, as well as the high vulnerability of our focused study types to get affected by nuisance parameters, suggests the need for introducing a well-developed analytical method for these types of studies.

Further considerations and limitations

An additional potential threat to the validity of tDCS findings involves the high within-subject variability of responses (Chew et al., 2015; López-Alonso et al., 2015), which can reduce the power of intervention comparisons. More research is required to understand the substrates of intraindividual variability mechanisms underpinning the effects of tDCS on order to optimize stimulation protocols. The other difficulty of reaching a valid conclusion about the effect of tDCS on motor learning is the low reproducibility of findings across different studies, which is attributed to the high methodological variabilities (Hashemirad et al., 2016; Buch et al., 2017). An important step toward reaching more conclusive results could be improving transparency and providing enough details about data collection and analysis, so similar procedures can be repeated and provide sufficient data to conduct robust meta-analyses. It should be noted that this review was a qualitative assessment of the published studies and indicated the possibility of biased outcomes. However, due to the wide diversity of the experimental protocols (e.g. various stimulation and task parameters) in the current literature, a precise quantification of the extent and direction of the possible distortions remains uncertain. Future well-designed experimental studies are encouraged to systematically investigate the impact of each confounding factor on the tDCS-induced changes, which can help establish methodological standards for this line of research. The other limitation of the present review is that only studies on healthy individuals were included, and it should be borne in mind that pathological conditions could demand specific methodological considerations accordingly.

Conclusion

Despite the concerns often expressed regarding the contamination of behavioral findings with crossover design-related confounds, these issues are often underestimated in the studies examining the effects of tDCS on motor learning. Most of the included studies have not provided enough information on how they dealt with order, time, and carryover effects in designing their experiments. In addition, although the optimal analytical method for crossover studies seems to be a matter of debate in the statistical literature, the approaches adopted by included studies in this review have not shown to meet the statistical concerns to the greatest possible degree. However, it should be noted that the aforementioned issues are identified based on the reported methods in the published articles and it is possible that some of these issues have already been addressed but not reported. Due to the negative effects of design-related nuisance parameters on the findings of crossover trials, researchers need to provide an explicit explanation about how they have dealt with these factors. An inappropriate use of crossover design may seriously affect the findings and therefore the conclusions drawn from tDCS studies on motor learning. This indicates a pressing need for the development of detailed conducting and reporting guidelines for this type of studies to benefit from the advantages of a crossover design.

References

Amadi, U., Allman, C., Johansen-Berg, H., and Stagg, C.J. (2015). The homeostatic interaction between anodal transcranial direct current stimulation and motor learning in humans is related to GABA(A) activity. Brain Stimul. *8*, 898–905.

Ambrus, G.G., Chaieb, L., Stilling, R., Rothkegel, H., Antal, A., and Paulus, W. (2016). Monitoring transcranial direct current stimulation induced changes in cortical excitability during the serial reaction time task. Neurosci. Lett. *616*, 98–104.

Avila, E., van der Geest, J.N., Kengne Kamga, S., Verhage, M.C., Donchin, O., and Frens, M.A. (2015). Cerebellar transcranial direct current stimulation effects on saccade adaptation. Neural Plast. *2015*, 968970.

Brunoni, A.R. and Fregni, F. (2011). Clinical trial design in non-invasive brain stimulation psychiatric research. Int. J. Methods Psychiatr. Res. *20*, e19–e30.

Brunoni, A.R. and Vanderhasselt, M.-A. (2014). Working memory improvement with non-invasive brain stimulation of the dorsolateral prefrontal cortex: a systematic review and meta-analysis. Brain Cognit. *86*, 1–9.

Buch, E.R., Santarnecchi, E., Antal, A., Born, J., Celnik, P.A., Classen, J., Gerloff, C., Hallett, M., Hummel, F.C., and Nitsche, M.A. (2017). Effects of tDCS on motor learning and memory formation: a consensus and critical position paper. Clin. Neurophysiol. *128*, 589–603.

Campbell, D.T. and Stanley, J.C. (1963). Experimental and Quasi-Experimental Designs for Research (Boston, MA: Houghton Mifflin).

Chen, X., Zhao, P.L., and Zhang, J. (2002). A note on ANOVA assumptions and robust analysis for a cross-over study. Stat. Med. *21*, 1377–1386.

Chew, T., Ho, K.-A., and Loo, C.K. (2015). Inter-and intra-individual variability in response to transcranial direct current stimulation (tDCS) at varying current intensities. Brain Stimul. *8*, 1130–1137.

Cleophas, T.J. and Zwinderman, A.H. (2012). Statistics Applied to Clinical Studies (Springer Science & Business Media).

Conley, A.C., Marquez, J., Parsons, M.W., Fulham, W.R., Heathcote, A., and Karayanidis, F. (2015). Anodal tDCS over the motor cortex on prepared and unprepared responses in young adults. PLoS One *10*, e0124509.

Cuypers, K., Leenus, D.J., van den Berg, F.E., Nitsche, M.A., Thijs, H., Wenderoth, N., and Meesen, R.L. (2013). Is motor learning mediated by tDCS intensity? PLoS One *8*, e67344.

Dayan, E. and Cohen, L.G. (2011). Neuroplasticity subserving motor skill learning. Neuron *72*, 443–454.

Dayan, E., Censor, N., Buch, E.R., Sandrini, M., and Cohen, L.G. (2013). Noninvasive brain stimulation: from physiology to network dynamics and back. Nat. Neurosci. *16*, 838–844.

DePuy, V. and Berger, V.W. (2005). Counterbalancing. Wiley StatsRef: Statistics Reference Online.

Díaz-Uriarte, R. (2002). Incorrect analysis of crossover trials in animal behaviour research. Anim. Behav. *63*, 815–822.

Foerster, Á., Rocha, S., Araújo, M.D.G.R., Lemos, A., and Monte-Silva, K. (2015). Effects of transcranial direct current stimulation on motor learning in healthy individuals: a systematic review. Fisioter. Mov. *28*, 159–167.

Fregni, F., Boggio, P.S., Nitsche, M., Bermpohl, F., Antal, A., Feredoes, E., Marcolin, M.A., Rigonatti, S.P., Silva, M.T., and Paulus, W. (2005). Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. Exp. Brain Res. *166*, 23–30.

Galea, J.M. and Celnik, P. (2009). Brain polarization enhances the formation and retention of motor memories. J. Neurophysiol. *102*, 294–301.

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

Hulst, T., John, L., Küper, M., van der Geest, J.N., Göricke, S.L., Donchin, O., and Timmann, D. (2017). Cerebellar patients do not benefit from cerebellar or M1 transcranial direct current stimulation during force-field reaching adaptation. J. Neurophysiol. *118*, 732–748.

Hunter, T., Sacco, P., Nitsche, M.A., and Turner, D.L. (2009). Modulation of internal model formation during force field-induced motor learning by anodal transcranial direct current stimulation of primary motor cortex. J. Physiol. *587*, 2949–2961.

Jaberzadeh, S. and Zoghi, M. (2013). Non-invasive brain stimulation for enhancement of corticospinal excitability and motor performance. Basic Clin. Neurosci. *4*, 257.

Jaberzadeh, S., Bastani, A., and Kidgell, D. (2012). Does the longer application of anodal-transcranial direct current stimulaton increase corticomotor excitability further? A pilot study. Basic Clin. Neurosci. *3*, 28–35.

Kantak, S.S., Mummidisetty, C.K., and Stinear, J.W. (2012). Primary motor and premotor cortex in implicit sequence learning – evidence for competition between implicit and explicit human motor memory systems. Eur. J. Neurosci. *36*, 2710–2715.

Karok, S. and Witney, A.G. (2013). Enhanced motor learning following task-concurrent dual transcranial direct current stimulation. PLoS One *8*, e85693.

López-Alonso, V., Fernández-del-Olmo, M., Costantini, A., Gonzalez-Henriquez, J.J., and Cheeran, B. (2015). Intra-individual variability in the response to anodal transcranial direct current stimulation. Clin. Neurophysiol. *126*, 2342–2347.

Minarik, T., Sauseng, P., Dunne, L., Berger, B., and Sterr, A. (2015). Effects of anodal transcranial direct current stimulation on visually guided learning of grip force control. Biology (Basel) *4*, 173–186.

Nitsche, M.A. and Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. Neurology *57*, 1899–1901.

Nitsche, M.A., Schauenburg, A., Lang, N., Liebetanz, D., Exner, C., Paulus, W., and Tergau, F. (2003). Facilitation of implicit motor learning by weak transcranial direct current stimulation of the primary motor cortex in the human. J. Cognit. Neurosci. *15*, 619–626.

Nitsche, M.A., Jakoubkova, M., Thirugnanasambandam, N., Schmalfuss, L., Hullemann, S., Sonka, K., Paulus, W., Trenkwalder, C., and Happe, S. (2010). Contribution of the premotor cortex to consolidation of motor sequence learning in humans during sleep. J. Neurophysiol. *104*, 2603–2614.

Palm, U., Reisinger, E., Keeser, D., Kuo, M.-F., Pogarell, O., Leicht, G., Mulert, C., Nitsche, M.A., and Padberg, F. (2013). Evaluation of sham transcranial direct current stimulation for randomized, placebo-controlled clinical trials. Brain Stimul. *6*, 690–695.

Pavlova, E., Kuo, M.F., Nitsche, M.A., and Borg, J. (2014). Transcranial direct current stimulation of the premotor cortex: effects on hand dexterity. Brain Res. *1576*, 52–62.

Polanía, R., Nitsche, M.A., and Paulus, W. (2011). Modulating functional connectivity patterns and topological functional organization of the human brain with transcranial direct current stimulation. Hum. Brain Mapp. *32*, 1236–1249.

Portney, L.G. and Watkins, M.P. (2000). Foundations of Clinical Research: Applications to Practice (Upper Saddle River, NJ, USA: Prentice Hall).

Reed, J.F. (2004). Analysis of two-treatment, two-period crossover trials in emergency medicine. Ann. Emerg. Med. *43*, 54–58.

Reinhart, R.M. and Woodman, G.F. (2015). The surprising temporal specificity of direct-current stimulation. Trends Neurosci. *38*, 459–461.

Reinhart, R.M., Cosman, J.D., Fukuda, K., and Woodman, G.F. (2017). Using transcranial direct-current stimulation (tDCS) to understand cognitive processing. Atten. Percept. Psychophys. *79*, 3–23.

Reis, J., Schambra, H.M., Cohen, L.G., Buch, E.R., Fritsch, B., Zarahn, E., Celnik, P.A., and Krakauer, J.W. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. Proc. Natl. Acad. Sci. U. S. A. *106*, 1590–1595.

Reis, J., Fischer, J.T., Prichard, G., Weiller, C., Cohen, L.G., and Fritsch, B. (2013). Time-but not sleep-dependent consolidation of tDCS-enhanced visuomotor skills. Cereb. Cortex *25*, 109–117.

Robertson, E.M. (2007). The serial reaction time task: implicit motor skill learning? J. Neurosci. *27*, 10073–10075.

Rroji, O., van Kuyck, K., Nuttin, B., and Wenderoth, N. (2015). Anodal tDCS over the primary motor cortex facilitates long-term memory formation reflecting use-dependent plasticity. PLoS One *10*, e0127270.

Saucedo Marquez, C.M., Zhang, X., Swinnen, S.P., Meesen, R., and Wenderoth, N. (2013). Task-specific effect of transcranial direct current stimulation on motor learning. Front Hum. Neurosci. *7*, 333.

Senn, S. (1994). The AB/BA crossover: past, present and future? Stat. Methods Med. Res. *3*, 303–324.

Senn, S. (2002). The AB/BA Design With Normal Data. Cross-over Trials in Clinical Research, 2nd ed. (Chichester, UK: John Wiley & Sons), pp. 35–88.

Shah, B., Nguyen, T.T., and Madhavan, S. (2013). Polarity independent effects of cerebellar tDCS on short term ankle visuomotor learning. Brain Stimul. *6*, 966–968.

Spieser, L., van den Wildenberg, W., Hasbroucq, T., Richard Ridderinkhof, K., and Burle, B. (2015). Controlling your impulses: electrical stimulation of the human supplementary motor complex prevents impulsive errors. J. Neurosci. *35*, 3010–3015.

Sriraman, A., Oishi, T., and Madhavan, S. (2014). Timing-dependent priming effects of tDCS on ankle motor skill learning. Brain Res. *1581*, 23–29.

Stagg, C.J., Jayaram, G., Pastor, D., Kincses, Z.T., Matthews, P.M., and Johansen-Berg, H. (2011). Polarity and timing-dependent effects of transcranial direct current stimulation in explicit motor learning. Neuropsychologia *49*, 800–804.

Stephen, S. (2002). Crossover Trials in Clinical Research (Chichester, UK: John Wiley).

Sun, Y., Lipton, J.O., Boyle, L.M., Madsen, J.R., Goldenberg, M.C., Pascual-Leone, A., Sahin, M., and Rotenberg, A. (2016). Direct current stimulation induces mGluR5-dependent neocortical plasticity. Ann. Neurol. *80*, 233–246.

Thomas, J.C. and Hersen, M. (2011). Understanding Research in Clinical and Counseling Psychology (Taylor & Francis).

Ungerleider, L.G., Doyon, J., and Karni, A. (2002). Imaging brain plasticity during motor skill learning. Neurobiol. Learn. Mem. *78*, 553–564.

Vahdat, S., Albouy, G., King, B., Lungu, O., and Doyon, J. (2017). Online and offline modulators of motor learning. Front. Hum. Neurosci. *11*, 69.

Walker, M.P., Brakefield, T., Seidman, J., Morgan, A., Hobson, J.A., and Stickgold, R. (2003). Sleep and the time course of motor skill learning. Learn. Mem. *10*, 275–284.

Wellek, S. and Blettner, M. (2012). On the proper use of the crossover design in clinical trials. Dtsch Arztebl Int. *109*, 276–281.

Willian, A.R. and Pater, J.L. (1986). Using baseline measurements in the two-period crossover clinical trial. Controlled Clin. Trials *7*, 282–289.

Woods, J.R., Williams, J.G., and Tavel, M. (1989). The two-period crossover design in medical research. Ann. Intern. Med. *110*, 560–566.

Zimerman, M., Heise, K.F., Hoppe, J., Cohen, L.G., Gerloff, C., and Hummel, F.C. (2012). Modulation of training by single-session transcranial direct current stimulation to the intact motor cortex enhances motor skill acquisition of the paretic hand. Stroke *43*, 2185–2289.

Zimerman, M., Nitsch, M., Giraux, P., Gerloff, C., Cohen, L.G., and Hummel, F.C. (2013). Neuroenhancement of the aging brain: restoring skill acquisition in old subjects. Ann. Neurol. *73*, 10–15.