

RESEARCH ARTICLE

Variability and effect sizes of intracranial current source density estimations during pain: Systematic review, experimental findings, and future perspectives

Juan Manuel Völker¹ | Federico Gabriel Arguissain¹ | Ole Kæseler Andersen¹ | José Biurrun Manresa^{1,2} 

¹Integrative Neuroscience Group, Center for Neuroplasticity and Pain (CNAP), Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

²Institute for Research and Development in Bioengineering and Bioinformatics (IBB), National Scientific and Technical Research Council (CONICET) and National University of Entre Ríos (UNER), Oro Verde, Argentina

Correspondence

José Biurrun Manresa, Integrative Neuroscience Group, Center for Neuroplasticity and Pain (CNAP), Department of Health Science and Technology, Aalborg University, Aalborg, Denmark.
Email: jbiurrun@hst.aau.dk

Funding information

Danmarks Grundforskningsfond, Grant/Award Number: DNR121

Abstract

Pain arises from the integration of sensory and cognitive processes in the brain, resulting in specific patterns of neural oscillations that can be characterized by measuring electrical brain activity. Current source density (CSD) estimation from low-resolution brain electromagnetic tomography (LORETA) and its standardized (sLORETA) and exact (eLORETA) variants, is a common approach to identify the spatiotemporal dynamics of the brain sources in physiological and pathological pain-related conditions. However, there is no consensus on the magnitude and variability of clinically or experimentally relevant effects for CSD estimations. Here, we systematically examined reports of sample size calculations and effect size estimations in all studies that included the keywords *pain*, and *LORETA*, *sLORETA*, or *eLORETA* in Scopus and PubMed. We also assessed the reliability of LORETA CSD estimations during non-painful and painful conditions to estimate hypothetical sample sizes for future experiments using CSD estimations. We found that none of the studies included in the systematic review reported sample size calculations, and less than 20% reported measures of central tendency and dispersion, which are necessary to estimate effect sizes. Based on these data and our experimental results, we determined that sample sizes commonly used in pain studies using CSD estimations are suitable to detect medium and large effect sizes in crossover designs and only large effects in parallel designs. These results provide a comprehensive summary of the effect sizes observed using LORETA in pain research, and this information can be used by clinicians and researchers to improve settings and designs of future pain studies.

KEYWORDS

EEG, LORETA, source localization, test-retest reliability

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Pain is a fundamental experience for survival, guiding behavior towards minimizing harm. It arises from the integration of sensory, cognitive, emotional, and motivational processes in the brain (Apkarian, Bushnell, Treede, & Zubieta, 2005). The complex interaction of these processes are the result of specific patterns of neural oscillations that can be directly characterized by measuring the electrical brain activity with magneto- and electroencephalography (M/EEG) (Ploner, Sorg, & Gross, 2017). Ultimately, researchers and clinicians want to identify the spatiotemporal dynamics of the brain sources in physiological and pathological pain-related conditions (Davis & Seminowicz, 2017; Ploner & May, 2018). This could help obtaining robust biomarkers from these measures with the aim of stratifying patients and improving medical diagnosis and treatments (Mouraux & Iannetti, 2018; Tracey, Woolf, & Andrews, 2019).

In this regard, source localization methods are a particular type of spatial filter that provide information about the activity and localization of the neural sources (Michel & Murray, 2012). A popular approach to estimate the brain sources from scalp EEG data is by low-resolution brain electromagnetic tomography (LORETA) (Babiloni et al., 2017; Pascual-Marqui et al., 1999). This method and its standardized (sLORETA) and exact (eLORETA) variants have been used to characterize the brain activity through the estimation of intracranial current source density (CSD), and localize sources in pain studies (González-Roldán, Cifre, Sitges, & Montoya, 2016; Hansen et al., 2017; Lelic, Hansen, Mark, Olesen, & Drewes, 2017; Moont, Crispel, Lev, Pud, & Yarnitsky, 2011; Nir, Sinai, Moont, Harari, & Yarnitsky, 2012; Shao, Shen, Yu, Wilder-Smith, & Li, 2012). This family of methods have been validated with standard methods for source localization, such as functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) (Cannon, Kerson, & Hampshire, 2011; Esslen, Kochi, Lehmann, & Pascual-marqui, 2002; Michel & Murray, 2012; Oakes et al., 2004; Pascual-Marqui, Esslen, Kochi, & Lehmann, 2002). However, most studies employing LORETA emphasize statistically significant differences between areas or experimental conditions, disregarding the variability and effect sizes behind these differences.

In order to constitute a robust biomarker, a measurement must be valid but also reliable, that is, its variability must be sufficiently low across experimental conditions (subjects, sessions, assessors) when experimental or clinical conditions remain unchanged (Downing, 2004). While test-retest reliability of LORETA studies was examined in different settings, for example during resting EEG (Cannon et al., 2012) and superficial painful stimulation (Hansen et al., 2017), the interpretation of the outcomes is usually carried out in oversimplified, binary terms, that is, the measurement is deemed reliable or not based on a comparison with a fixed threshold set using arbitrary criteria, which can be misleading if data and results are not critically assessed (Biurrun Manresa et al., 2014).

In this context, it is clear that there is no consensus in the literature on the magnitude of clinically or experimentally relevant effects for CSD estimations, and what is the magnitude of the variability

behind these measures. This information is crucial for properly planning future research designs but is currently unavailable. Thus, the aims of this study were (1) to systematically examine reports of sample size calculations and estimations of effect sizes in studies that used LORETA for pain research; (2) to assess the reliability of LORETA CSD estimations in experimental settings; (3) to estimate hypothetical sample sizes for future experiments using source localization of neural oscillations for the identification of potential biomarkers related to pain.

2 | METHODS

This study is divided in three parts: the first part is a systematic review of existing literature involving CSD estimation using LORETA, focusing on reports of effect sizes and variability in pain research. The second part describes methodological considerations and results from an experiment specifically carried out to determine the reliability of CSD estimations in controlled experimental conditions. Finally, the third part integrates the findings from the previous two to estimate future sample sizes for potential experiments involving these measurements.

2.1 | Part I: Systematic review of effect sizes and variability of CSD estimates in pain-related studies

2.1.1 | Search strategy

Electronic literature search was performed by the first author (J.M.V.) in SCOPUS and PubMed databases for articles published before July 20th, 2020. The search descriptors in the title, abstract or keywords were “pain” and at least one of the following “LORETA” or “eLORETA” or “sLORETA.”

2.1.2 | Inclusion and exclusion criteria

Articles were included in the analysis if they met the following criteria: (1) full articles, written in English and published in peer-reviewed journals; (2) human pain research studies; (3) included more than five participants (no case studies); (4) included EEG frequency analysis; (5) provided information about central tendency and dispersion of the CSD estimations for the sample.

2.1.3 | Data extraction

All studies involving pain research that used LORETA were evaluated to report the type of study (exploratory or confirmatory), the percentage of studies that reported sample size calculations, measures of central tendency and dispersion and, in case of experimental studies, estimation of effect sizes (Noordzij et al., 2010). If the effect size was not reported, it was estimated using the mean and standard deviation

of the CSD estimation and the number of participants (assuming a normal distribution after a logarithmic transformation) (Lakens, 2013). If the mean and standard deviation were not reported numerically, they were extracted from the graphics using the software Engauge Digitizer 12.1 (<http://digitizer.sourceforge.net>).

2.2 | Part II: Reliability of CSD estimates in controlled experimental conditions

2.2.1 | Endpoints

CSD is usually estimated in the θ (4–7.9 Hz), α (8–13.9 Hz), β (14–29.9 Hz), and γ (30–90 Hz) bands, which are commonly reported to be involved during tonic pain-related neural oscillations (Backonja et al., 1991; Chang, Arendt-Nielsen, Graven-Nielsen, Svensson, & Chen, 2001; Dowman, Rissacher, & Schuckers, 2008; Li et al., 2016; Nir et al., 2012; Nir, Sinai, Raz, Sprecher, & Yarnitsky, 2010; Ploner et al., 2017; Schulz et al., 2015; Shao et al., 2012). Furthermore, the CSD for these frequencies is extracted from specific regions of interest (ROIs) in the brain. A survey of the studies from the systematic review revealed that data from quantitative analysis was available for the anterior cingulate cortex (ACC), the anterior insula (AI), and the primary somatosensory cortex (S1) in the θ and α frequency bands. We also included the γ band in the analysis of our own experimental data, since cortical oscillations in these frequency bands have been associated with pain processing (Hauck, Lorenz, & Engel, 2007; Nickel et al., 2017). Moreover, common settings in experimental pain research involve controlled conditions for nociceptive stimulation, as well as baseline or reference recordings of non-nociceptive stimulation as well. For this experiment, we used a model of sustained deep-tissue pain (SDTP) elicited using cuff-pressure algometry, and we also performed sustained deep-tissue non-painful stimulation (SDTnP) and recorded resting-state EEG (REEG) for comparison purposes. Thus, we defined the primary endpoint as the test–retest reliability of CSD estimations in the aforementioned three brain regions and three frequency bands, during SDTP. Secondary endpoints were the test–retest reliability of CSD estimations at the same brain regions in the same frequency bands, for SDTnP and REEG.

2.2.2 | Setting

The experiment was carried out at the Department of Health Science and Technology, Aalborg University, Aalborg (Denmark). The study was conducted according to the Declaration of Helsinki and was approved by the ethical committee of Northern Jutland, Denmark (N-20170047).

2.2.3 | Participants

Twenty-one healthy volunteers participated in the experiment (mean age = 25.0 years, SD = 2.6; 14 females). Inclusion criteria were

(1) healthy men and women between 18 and 50 years old, able to speak and understand English. Exclusion criteria were (1) pregnancy or breast-feeding, (2) previous neurologic, musculoskeletal, or mental illnesses, (3) history of chronic pain or current acute pain, (4) skin allergies, (5) presence of wounds in the forearm, and (6) incapacity to provide informed consent. Furthermore, participants were asked to refrain from any pain medication at least 24 h before the experimental sessions.

2.2.4 | EEG recordings

EEG data were recorded using a medical-grade amplifier (g.Hlamp, g.tec-medical engineering GmbH, Austria) using 64 active electrodes. Before the EEG recording, the head of the participants was measured using the distance between the nasion and inion for appropriate placement of the electrodes according to the international 10/20 system. A flexible EEG cap (g.GAMMA cap2, g.tec-medical engineering GmbH, Austria) with chinstrap was used to maintain fixed electrode positions. The impedance of the electrodes was kept below 10 k Ω and the sampling rate was set at 1,200 Hz. The AFz electrode served as ground and the left earlobe (A1) served as reference. Recordings for each condition lasted at least 3 min, and conditions were randomized across participants, with a 5-min pause between conditions.

2.2.5 | Computer-controlled cuff pressure algometry

A computer-controlled pneumatic pressure algometry (NociTech, Denmark, and Aalborg University, Denmark) was used to deliver sustained deep tissue stimulation. A 10-cm wide silicone tourniquet cuff (VBM, Germany) was tied around the right forearm at a 3-cm distance of the cubital fossa and over the extensor carpi radialis brevis muscle belly. Participants were trained to score the perceived stimulation sensation using a computerized visual analogue scale (VAS) where 0 represented no sensation, 10 represented the most intense pain imaginable and 5 indicated the pain threshold. The average cuff pressure that elicited ratings of 3 and 7 in the VAS of three consecutive ramps (rate = 1 kPa/s, interstimulus interval = 5 min, pressure limit = 110 kPa) was used to estimate the pressure of SDTnP and SDTP, respectively.

2.2.6 | EEG pre-processing

Raw EEG data were off-line pre-processed using Matlab R2018b (The Mathworks Inc, Natick, MA) and EEGLAB toolbox v15.0b (Delorme & Makeig, 2004). Data were digitally band-pass-filtered by applying a Hamming-windowed *sinc* finite impulse response filter with an order of 7,920 and cutoff frequencies of 0.1 and 100 Hz. Additionally, a 50-Hz notch filter was used, and data were resampled to 500 Hz. Channels with low signal-to-noise ratio were removed from both

recordings of a subject. Filtered data were cleaned by visual inspection to exclude artifacts (non-cerebral source activities). Furthermore, blinks and muscle activity were removed using independent component analysis (Jung et al., 2000). Removed channels were interpolated afterwards using their neighboring channels. Finally, artifact-free EEG data were reduced to 120 s for each condition to ensure the same amount of data for all subjects.

2.2.7 | Estimation of cortical sources

The exact low-resolution brain electromagnetic tomography (eLORETA, LORETA-KEY software v20150415, <http://www.uzh.ch/keyinst/loreta.htm>) method was used to localize cortical brain activity (Pascual-Marqui, 2007). This inverse modeling method estimates the current density in the brain volume, thus providing an appraisal of where in the brain the scalp-recorded EEG is being generated (Pascual-Marqui, Michel, & Lehmann, 1994). The ROIs consisted of the collection of voxels that are within a 15-mm radius sphere around the seed point (Canuet et al., 2012). The coordinates of the seed point of the ROIs are in Table 1 (Julious, 2004). In particular, the left ACC and S1 regions were selected because they are contralateral to the stimulus, whereas activity in the right AI was quantified as there is evidence of a preponderance of this side in pain processing (Brooks, Nurmikko, Bimson, Singh, & Roberts, 2002; Craig, 2003; Ostrowsky et al., 2002). It is known that the CSD estimations obtained from LORETA do not show a uniformly normal distribution (Thatcher, North, & Biver, 2005), which is a requirement for the following reliability estimations (Julious, 2004). Hence, a logarithmic transformation was used to obtain a normal distribution.

2.2.8 | General methodological aspects

The experiment was designed to minimize the influence of external confounding factors in a test-retest reliability assessment (e.g., population age, time of the day where the experiment was carried out, and interval between sessions), which are known to affect EEG activity and pain sensitivity (Carrier, Land, Buysse, Kupfer, & Monk, 2001; Hjermstad et al., 2011). Therefore, only young adults were included, and the experiment was scheduled at the same time of the day for every volunteer. The experiment was carried out in two

TABLE 1 Spatial locations of regions of interest considered for the analysis

Region	BA	x	y	z
Left-anterior cingulate cortex (L-ACC)	32	-2	32	22
Left-primary somatosensory cortex (L-S1)	2	-32	-36	60
Right-anterior insula (R-AI)	13	36	12	8

Note: Coordinates are in Talairach space.

Abbreviations: BA, Brodmann area; x, medial-lateral; y, anterior-posterior; z, superior-inferior.

different sessions, separated by 7 ± 2 days. Participants were thoroughly familiarized with constant cuff-pressure stimulation to reduce any effects of arousal or anxiety before data acquisition in the first session. During the experimental sessions, the participants were seated on a comfortable chair in front of a computer screen (HP P17A ProDisplay, Hewlett-Packard Company, USA) that showed a vertical bar, whose length represented the perceived intensity rating of the VAS, anchored as described in Section 2.2.4. Participants also wore foam earplugs (Earplugs, TaperFit 3M, Minnesota) to mask ambient noise and to minimize any auditory bias introduced by the air compressor. All participants attended the second experimental session at the same time of the day (± 1 h of difference regarding the first experimental session), to rule out possible inference of the circadian rhythm in pain perception (Glynn, Lloyd, & Folkard, 1976). Furthermore, to rule out inter-assessor variation, the same researcher (J.M.V.) collected data from all volunteers.

2.2.9 | Data analysis and statistics

Between-session reliability of cuff pressure and CSD estimations was determined using Bland-Altman (BA) analysis. BA analysis consists of plotting the differences vs. the average of repeated measurements for each subject. Furthermore, the limits of agreement (LoA) can be derived, which express the average difference (*bias*) ± 1.96 times the standard deviation of the differences between repeated measurements (SD_{diff}). In this way, the difference between the upper and lower LoA delimit the range where 95% of the differences between two repeated measurements are expected to lie (Bland & Altman, 1986). The 95% confidence intervals (CI) are reported for both LoA and *bias* (Bland & Altman, 1999). The between-session reliability of the stimulus used for the experimental intervention (cuff-algometry) was calculated for the SDTP and SDTnP as well. The reliability of CSD during SDTP, SDTnP, and REEG was calculated for theta, alpha, and gamma bands and the three ROIs (L-ACC, L-S1, and R-AI). Values are expressed as mean \pm SD unless stated otherwise.

2.3 | Part III: Considerations for future research designs

2.3.1 | Sample size estimation

Hypothetical sample size estimation is a valid and complementary approach for assessing reliability (Shieh, 2014). It was calculated for both parallel (N_p) and crossover (N_c) study designs. N_p indicates the amount of subject needed for parallel studies (in which there are two groups, for example, patients and healthy volunteers), and N_c indicates the amount of subject needed for crossover studies (where the same group of participants is assessed in two different conditions). The estimation of N_p and N_c was calculated considering a type I error level of 5% and 80% power for an independent-sample or paired-sample t test, respectively.

2.4 | Relationship between reliability estimates, sample size, and effect sizes

All the reliability assessment methods used in this study are related to the within-subject standard deviation (SD_w). For a test–retest experiment, the SD_w can be calculated as $SD_w = SD_{diff}/\sqrt{2}$, and the LoA can be reformulated as $bias \pm 1.96SD_w\sqrt{2}$ (Burrin Manresa et al., 2014). Finally, sample size estimator can also be estimated from these relations, since $N_c = (15.6SD_w^2)/(\bar{X}_1 - \bar{X}_2)^2$ and $N_p = (15.6SD_w^2)/((1-r)(\bar{X}_1 - \bar{X}_2)^2)$, where \bar{X}_1 and \bar{X}_2 are the average measurements of the two groups in a parallel design, or the average values of the two conditions for the same group in a crossover design, and $r = 1 - (SD_w/SD)^2$ is the correlation among repeated measures (SD is the standard deviation measurements of the sample) (Julious, 2004). In this way, $(\bar{X}_1 - \bar{X}_2)$ is the simple effect size. Additionally, a standardized effect size can be derived using Cohen's $d = (\bar{X}_1 - \bar{X}_2)/\sqrt{(SD_1^2 - SD_2^2)/2}$ for parallel designs or $d_z = (\bar{X}_1 - \bar{X}_2)/\sqrt{(SD_1^2 + SD_2^2 - (2rSD_1SD_2))}$ for crossover designs, where SD_1 and SD_2 are the standard deviation measurements of the two groups/conditions (Lakens, 2013).

3 | RESULTS

3.1 | Part I: Systematic review of effect sizes and variability of CSD estimates in pain-related studies

A descriptive flowchart of the selection process for eligible studies can be seen in Figure 1. The systematic review resulted in 100 articles, from which 58 unique studies remained after removing duplicates found in both databases, and 46 studies were included in the analysis after checking for the first three items in the inclusion criteria. Afterwards, the main reason for exclusion was that 29/46 studies performed time-domain analysis (as opposed to frequency-domain analysis), for which we nonetheless still present results regarding reporting of estimates of central tendency and dispersion. Table S1 contains the bibliographic information about the studies included in the systematic review and Tables S2 and S3 contain details on the articles with frequency- and time-domain analysis, respectively, including research design and setting, and whether they include effect sizes and variability of CSD estimates.

Table 2 shows the percentage of studies that reported indexes of central tendency and dispersion for CSD estimates, as well as estimates of effect sizes and sample size calculations. Remarkably, no studies reported sample size calculations, although it should be clarified that 12 studies (26%) were typified as exploratory, in which a sample size calculation is not strictly necessary. After screening the papers that performed a frequency-domain analysis on the CSD estimates (17/46), we found that 14 of them did not provide information about central tendency or dispersion of the CSD at any ROI, which

prevents further estimations of sample size. Thus, the systematic review yielded three publications selected for further analysis in Part III.

Briefly, Ye, Yan, Yao, Lou, and Peng (2019) attempted to identify abnormalities of spontaneous cortical oscillations among patients with somatoform pain disorder, and observed differences in resting-state alpha oscillations between patients and controls. While statistically significant differences were found in CSD estimations in the parietal region, quantitative values for patients and controls (\bar{X}_1 and \bar{X}_2 , respectively, in Table 3) were also reported for other areas, such as the central region including S1. Meanwhile, Prinsloo et al. (2019) proposed to use LORETA-based neurofeedback to attempt to modify brain activity in patients with acute pain from head and neck cancer. Among other quantitative data, they reported CSD estimations in the ACC, S1, and AI at baseline and pain onset (\bar{X}_1 and \bar{X}_2 , respectively, in Table 3). Finally, Prichep, Shah, Merkin, and Hiesiger (2018) compared CSD estimations of chronic pain patients and age- and gender-matched controls (\bar{X}_1 and \bar{X}_2 , respectively, in Table 3) in several brain regions in the theta frequency band. A summary of statistical indexes of central tendency and dispersion for these articles can be found in Table 3.

3.2 | Part II: Reliability of CSD estimates in controlled experimental conditions

3.2.1 | Cuff pressure algometry

Individual average pressures of SDTP and SDTnP are shown in Figure 2. The average pressures that elicited a perceived intensity of 3 and 7 in the VAS scale were 27.7 ± 9.1 and 63.3 ± 14.0 kPa on session 1 and 24.7 ± 6.7 and 60.3 ± 12.2 kPa on session 2, respectively. A visual inspection of Bland–Altman plots (Figure 3) did not reveal clear signs of heteroscedasticity. Statistically significant differences in cuff pressure between sessions were observed during SDTnP ($t = 2.356$, $p = 0.029$, Cohen's $d_z = 0.51$), but not during SDTP ($t = 1.628$, $p = 0.12$, Cohen's $d_z = 0.36$).

3.2.2 | LORETA CSD during SDTP, SDTnP, and REEG

For reproducibility purposes, Table S4 includes, for each condition, ROI and frequency band, a summary of the mean, standard deviation, maximum and minimum values for CSD estimations, as well as quantitative data of bias and LoA with their corresponding CIs from the Bland–Altman analysis. Furthermore, Table S5 includes every individual CSD estimation for each subject, condition, ROI and frequency band. Individual CSD differences between sessions and Bland Altman plots are shown in Figures 4 and 5, respectively. Both figures represent conditions (SDTP, SDTnP, and REEG) in each ROI (L-ACC, L-S1, and R-AI) and frequency bands (theta, alpha, and gamma). A visual

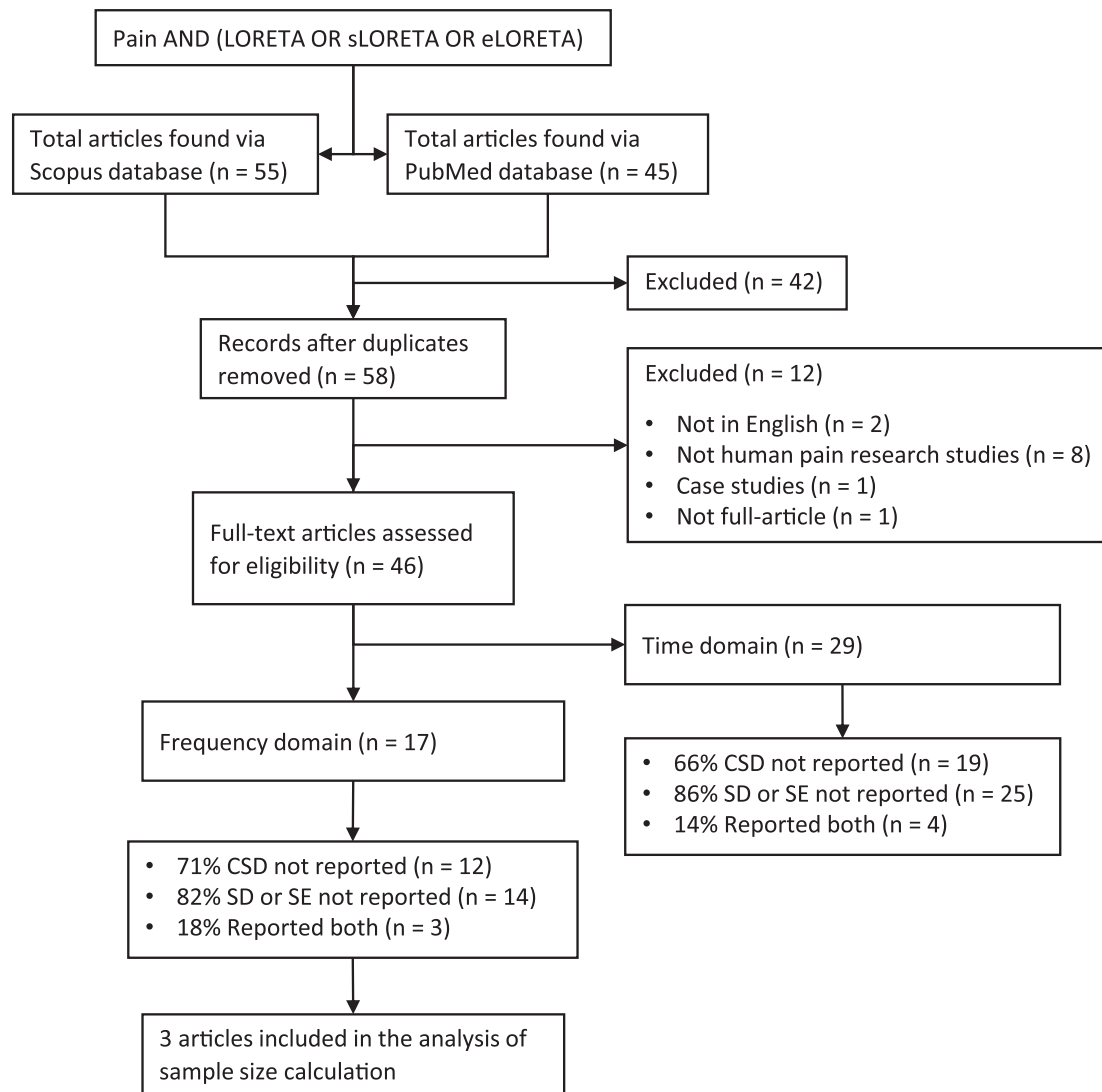


FIGURE 1 Flow chart for the systematic literature review. CSD, current source density; SD, standard deviations; SE, standard error

TABLE 2 Percentage of pain studies using LORETA that reported sample size calculation and parameters needed for future sample size calculations ($n = 46$)

Parameter	Reported
Sample size calculation	0 (0%)
Mean CSD at any ROI for any condition/group	15 (33%)
SD or SE of the mean CSD at any ROI for any condition/group	7 (15%)
Cohen's <i>d</i> effect size	13 (28%)

Abbreviations: CSD, current source density; SD, standard deviation; SE, standard error; ROI, region of interest.

inspection of Bland–Altman plots (Figure 5) did not reveal clear signs of heteroscedasticity. Systematic bias was only observed between session in the gamma band in the R-AI during SDTP and in the L-ACC during SDTnP.

3.3 | Part III: Considerations for future research designs

3.3.1 | Sample size calculation with data from the systematic review in Part I

Table 4 shows the hypothetical sample sizes that are needed to detect the effect sizes reported for the articles selected in the literature review, with a type I error of 5% and a power of 80%.

3.3.2 | Sample size calculation with data from the experimental findings in Part II

Hypothetical sample sizes are listed in Tables 5–7 for SDTP, SDTnP, and REEG, respectively. The estimation indicates the minimal number of participants that will be required to correctly reject the null hypothesis of no effect or no difference given a type I error of 5% and a power of

TABLE 3 A summary of empirical findings of information about central tendency and dispersion of the log CSD estimates in pain research using LORETA

Article	Parameters	ACC			S1			AI		
		θ	α	γ	θ	α	γ	θ	α	γ
Ye et al.	\bar{X}_1 (SD)	-	-	-	-	0.30 (1.14)	-	-	-	-
	\bar{X}_2 (SD)	-	-	-	-	-0.10 (0.76)	-	-	-	-
Prinsloo et al. ^a	\bar{X}_1 (SD)	-	1.48 (0.24)	-	-	1.26 (0.24)	-	-	1.09 (0.24)	-
	\bar{X}_2 (SD)	-	1.57 (0.32)	-	-	1.05 (0.33)	-	-	1.29 (0.33)	-
Prichet et al. ^b	\bar{X}_1 (SD)	0.67 (1.29)	-	-	1.38 (1.44)	-	-	-	0.43 (1.51)	-
	\bar{X}_2 (SD)	0.60 (1.23)	-	-	1.01 (1.59)	-	-	-	0.57 (1.34)	-

Abbreviations: ACC, anterior cingulate cortex cluster that included this region; AI, anterior insula cluster that included this region; S1, primary somatosensory cortex or cluster that included this region.

^aDid not report the units of CSD (suspected \log_{10} CSD $\mu\text{A}/\text{mm}^2$).

^bReported mean CSD and SD log values under z-transformation.

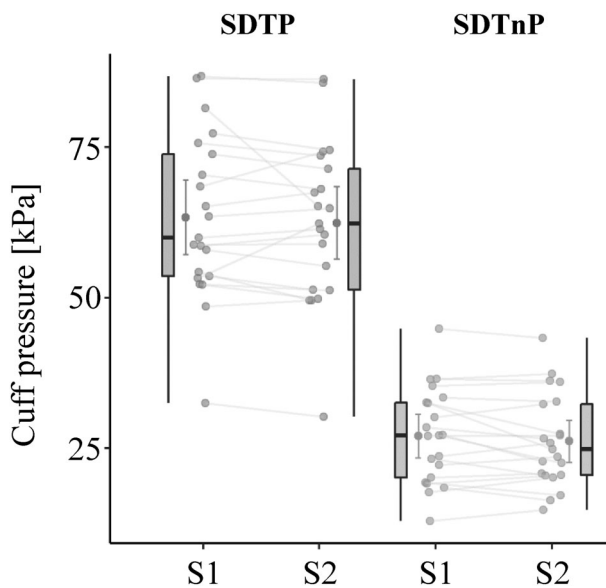


FIGURE 2 Magnitude of the pressure used to elicit a painful sensation (VAS = 7, SDTP, left) and non-painful sensation (VAS = 3, SDTnP, right) in both sessions. VAS, visual analog scale; SD, standard deviations; SDTP, sustained deep tissue pain; SDTnP, sustained deep tissue no-pain; S1, session 1; S2, session 2. The boxplots represent the median (central black mark), the edges of the box the 25th and 75th percentiles. Gray lines next to boxplots represents the mean (gray dot) and standard deviation (gray whiskers)

80%. After observing the variability in the experimental data, three different simple effect sizes were proposed, representing absolute differences of ± 0.05 , ± 0.10 , and ± 0.25 in the \log_{10} CSD estimates between groups (for parallel designs) or conditions (for crossover designs).

4 | DISCUSSION

The aim of the current study was to assess the variability and effect sizes of intracranial current source density (CSD) estimations in pain research. The first part of the study consisted of a systematic review

that was performed to examine the frequency that human pain studies reported sample size calculation, estimates of effect size and measures of central tendency and dispersion for CSD quantities. The review revealed that none of the 46 included studies reported sample size calculations and only 7 out of 46 reported the parameters needed for future sample size calculations. The second part of the study analyzed the test-retest reliability of CSD estimations during controlled experimental conditions. Here, EEG data collected during resting, non-painful, and painful cuff-pressure stimulation from 21 healthy participants were analyzed. Results show that LORETA CSD estimations are equally reliable for within-session comparisons among all conditions, ROIs and frequencies, although estimations during REEG at S1 in the alpha band showed less between-session variability than the rest of estimations. Finally, hypothetical sample size calculations were derived using effect sizes observed in the literature as well as effect sizes proposed following the observed variation in the experimental data.

4.1 | Systematic review of effect sizes and variability of CSD estimates in pain-related studies

None of the 46 included studies from the systematic review reported sample size calculations. This is in agreement with a previous systematic review in EEG studies, reporting that published articles rarely reported sample size calculations (Larson & Carbine, 2017). Still, this is not an exclusive trend in EEG, since studies using other neuroimaging techniques, such as fMRI, lack in reporting sample size calculation (Guo et al., 2014). Even though there has been a small improvement in this respect in other fields of medicine in recent years (Castellini, Gianola, Bonovas, & Moja, 2016), this is still far from optimal and journals should suggest and implement guidelines for appropriate sample size calculations (Lee & Tse, 2017).

Since the first study using LORETA was published in 1994, there should be plenty of data available to properly describe CSD estimations from different brain regions, as well as the expected differences due to experimental paradigms or clinical conditions. However, less

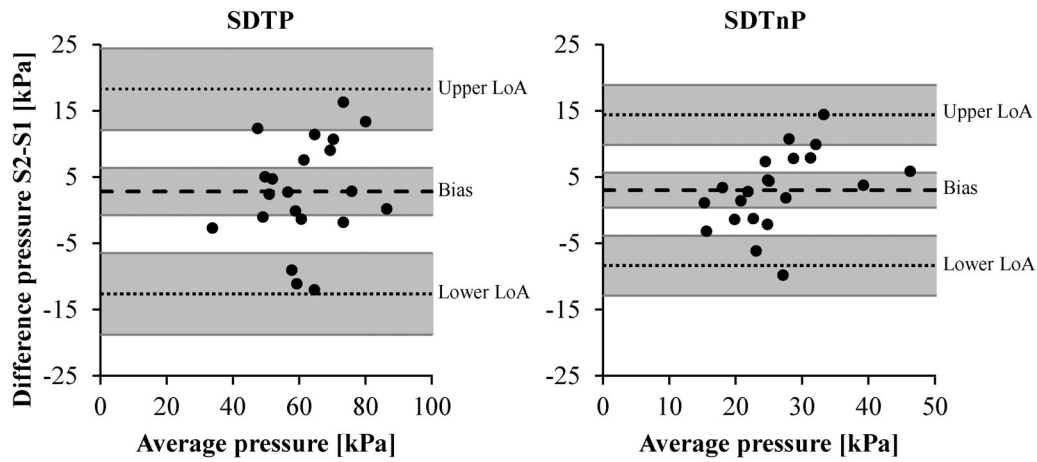


FIGURE 3 Bland-Altman plots of the magnitude of the pressure used in the cuff in the left: sustained deep tissue pain (SDTP) and right: sustained deep tissue no-pain (SDTnP) stimulations across days. The dashed line indicates the bias between sessions and the dotted lines the limits of agreement (LoA), calculated as ± 1.96 times the standard deviation (SD) of the differences in measurements between sessions. Shaded areas indicate the 95% confidence intervals of the bias and the limits of agreement

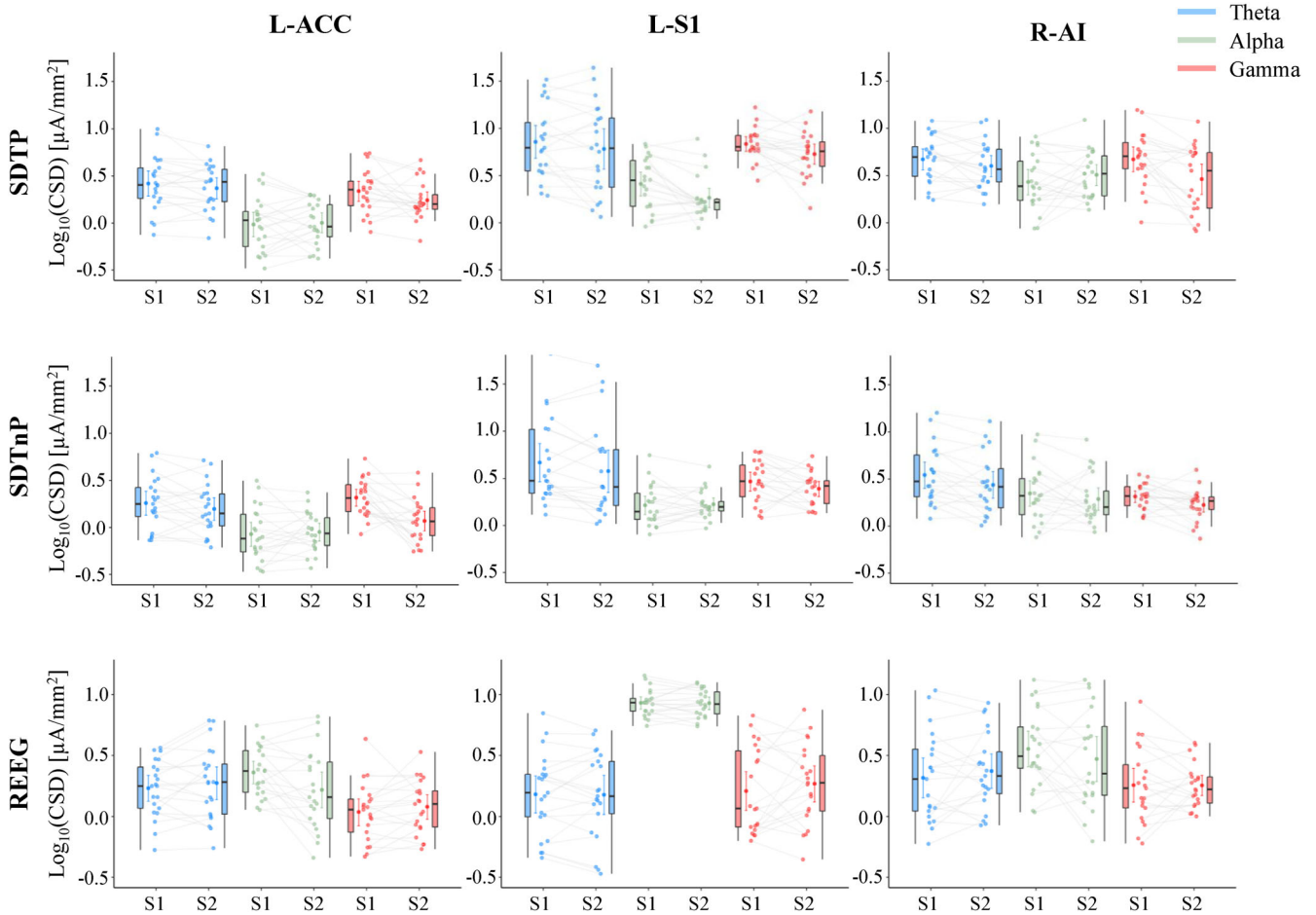


FIGURE 4 Log current source density (CSD) at each region of interest (ROI), for the theta, alpha and gamma band during sustained deep tissue pain (SDTP) and sustained deep tissue no-pain (SDTnP) and resting (REEG) in session 1 (S1) and session 2 (S2). L-ACC, left anterior cingulate cortex; L-S1, left primary somatosensory; R-AI, right anterior insula. The boxplots represent the median (central black mark), the edges of the box the 25th and 75th percentiles. Lines next to boxplots represent the mean (central dot) and standard deviation (SD) (whiskers)

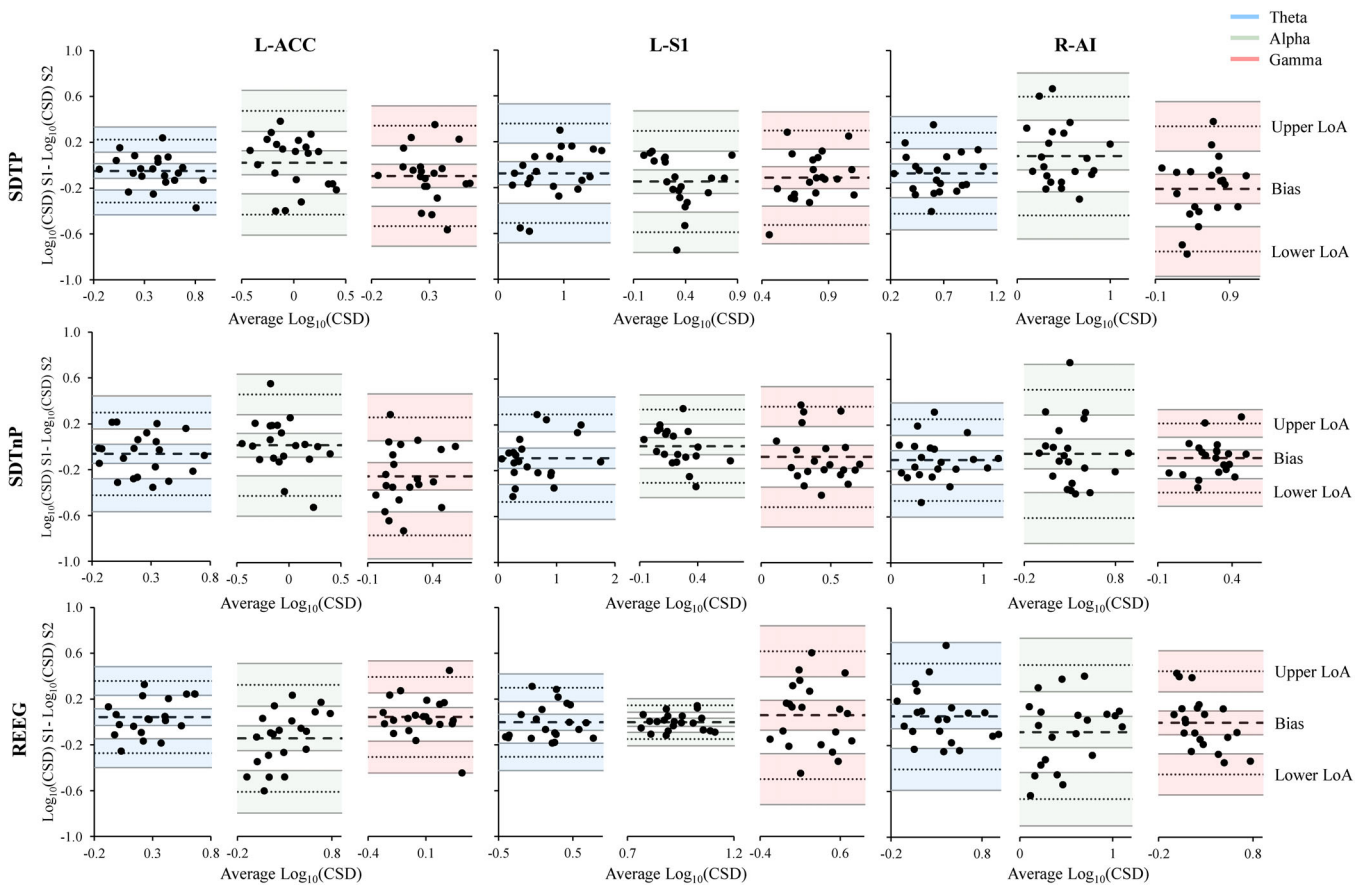


FIGURE 5 Bland–Altman plots of the difference between session 1 (S1) and session 2 (S2) of the \log_{10} current source density (CSD) for each region of interest, at theta, alpha and gamma band during sustained deep tissue pain (SDTP), sustained deep tissue no-pain (SDTnP) and resting (REEG). The dashed line indicates the bias between sessions and the dotted lines the limits of agreement (LoA), calculated as ± 1.96 times the standard deviation (SD) of the differences in measurements between sessions. Shaded areas indicate the 95% confidence intervals of the bias and the limits of agreement. L-ACC, left anterior cingulate cortex; L-S1, left primary somatosensory; R-AI, right anterior insula

TABLE 4 Hypothetical sample size calculation for the effect sizes obtained from the systematic review

Article		ACC			S1			AI		
		θ	α	γ	θ	α	γ	θ	α	γ
Ye et al. $N_1 = 17$ $N_2 = 17$	d	-	-	-	-	0.43	-	-	-	-
	N_p	-	-	-	-	84	-	-	-	-
Prinsloo et al. $N = 12$	d_z	-	0.36	-	-	0.98	-	-	0.88	-
	N_c	-	112	-	-	8	-	-	8	-
Pritchep et al. $N_1 = 77$ $N_2 = 77$	d	0.07	-	-	0.54	-	-	0.18	-	-
	N_p	4,245	-	-	56	-	-	450	-	-

Note: Effects size where significant differences were effectively reported are highlighted in bold ($p < .05$). Abbreviations: ACC, anterior cingulate cortex cluster that included this region; AI, anterior insula cluster that included this region; d , Cohen's d standardized effect size; N , total number of participants reported in the study; N_c , number of participants estimated for cross-over experiment design. N_p , number of participants in each group estimated for parallel-over experiment design; S1, primary somatosensory cortex or cluster that included this region.

than 20% of the articles included in the systematic review reported the minimum parameters required for future sample size calculations involving CSD estimations. In contrast, another systematic review on scalp EEG found that half of the articles provided information about

central tendency and dispersion of EEG scalp potentials (Larson & Carbine, 2017). These parameters are not only useful for planning future studies, but also help in the peer-review process to determine if the study was well-planned and adequately powered.

TABLE 5 Hypothetical sample size calculation for crossover (N_c) and parallel (N_p) sustained deep tissue pain (SDTP) experiment design as a function of the effect size (d and d_z)

Effect size ($\log_{10}(\text{CSD})$)		L-ACC			L-S1			R-AI		
		θ	α	γ	θ	α	γ	θ	α	γ
± 0.05	N_c	61	166	156	153	159	138	102	218	242
	d_z	0.33	0.20	0.21	0.25	0.21	0.29	0.28	0.17	0.21
	N_p	446	409	305	1,129	426	273	384	457	705
	d	0.17	0.18	0.21	0.13	0.18	0.29	0.20	0.17	0.17
± 0.10	N_c	15	42	39	38	40	34	25	54	61
	d_z	0.66	0.39	0.42	0.50	0.42	0.58	0.56	0.35	0.42
	N_p	111	102	76	282	107	68	96	114	176
	d	0.34	0.35	0.43	0.26	0.36	0.58	0.40	0.34	0.35
± 0.25	N_c	2	7	6	6	6	6	4	9	10
	d_z	1.65	0.98	1.06	1.26	1.04	1.45	1.39	0.87	1.05
	N_p	18	16	12	45	17	11	15	18	28
	d	0.86	0.88	1.07	0.65	0.90	1.46	1.01	0.85	0.87

Note: Even if some of the sample sizes are smaller than 10 participants, the minimum suggested sample size for a group should not be less than that (Hopkins, 2000).

Abbreviations: L-ACC, left anterior cingulate cortex; L-S1, left primary somatosensory; R-AI, right anterior insula.

Furthermore, the majority of the articles that did not report these parameters only reported images of voxel-by-voxel statistical maps, showing the locations of statistically significant variation of CSD estimations and not the actual effect sizes. Using these reported p -values or t , F , or z -statistics is not enough to calculate sample size for future studies (Sullivan & Feinn, 2012). More importantly, reporting the test statistics does not help in the evaluation of experimentally or clinically relevant effect sizes, which are crucial for the design of future experiments.

4.2 | Reliability of CSD estimates in controlled experimental conditions

Most reliability studies use a single measurement, often the intraclass correlation (ICC) and arbitrary cutoff thresholds to assess whether a measurement or method is reliable or not. However, a proper analysis of reliability is not as simple or straightforward (Bruton, Conway, & Holgate, 2000). In some cases, high ICC values may not necessarily reflect high reliability and vice versa (Biurrun Manresa et al., 2014). More importantly, ICC is a measure of relative reliability, but most researchers are actually interested in the absolute reliability of a measurement, with the key advantage that it can be interpreted in terms of the original measurement units. Finally, many authors suggest to study reliability in broader terms and not base the conclusion solely in one number as if reliability was a binary construct, that is, considering that “the measure is reliable/good or the measure is not reliably/poor” (Biurrun Manresa et al., 2014; Kottner et al., 2011; Rousson, Gasser, & Seifert, 2002). For this reason, reliability in this study was presented in terms of these experimental conditions and the

amount of measurement error that is acceptable given the application.

4.2.1 | The reliability of cuff pressure algometry

Controlled experimental conditions require that not only the outcomes are reliable, but also the experimental stimulus. Graphical analysis of Bland–Altman plots showed no clear evidence of heteroscedasticity in the between-session cuff pressure algometry. In other words, the differences in pressure values do not seem to increase with increasing mean pressure values between sessions. There is a small systematic mean difference between the six sessions for the SDTnP and SDTP that might be related to a learning process, considering that volunteers are naïve to the stimulation in the first session but not in the second. Still, this observed bias in the SDTP is small (3 kPa approx.) compared to the mean (4.7%) and may be irrelevant for pain research (Polianskis, Graven-Nielsen, & Arendt-Nielsen, 2002). On the other hand, the bias in the SDTnP compared to the mean (10.8%) is more relevant and should be taken into consideration in between-session experiments that involve similar cuff pressure stimulation. The specific reason for this effect is not clear. Previous work found systematic bias for the pain threshold between sessions of cuff pressure algometry applied in the forearm (Graven-Nielsen, Vaegter, Finocchietti, Handberg, & Arendt-Nielsen, 2015). One possibility is that volunteers become familiar with the stimulation and therefore less anxious about the experiment. Some authors suggest that at least one practice session is necessary to overcome this carry-over effect (McEvoy, Smith, & Gevins, 2000). Nevertheless, the main measure in this study was the EEG data during constant stimulation, which elicited a sensation of 3 and 7 in the VAS scale in both sessions, and both were achieved in the present study.

TABLE 6 Hypothetical sample size calculation for crossover (N_c) and parallel (N_p) sustained deep tissue no-pain (SDTnP) experiment design as a function of the effect size (d and d_z)

Effect size ($\log_{10}(\text{CSD})$)		L-ACC			L-S1			R-AI		
		θ	α	γ	θ	α	γ	θ	α	γ
±0.05	N_c	106	160	215	119	83	156	103	255	74
	d_z	0.26	0.19	0.24	0.27	0.26	0.20	0.28	0.16	0.38
	N_p	454	372	378	1,351	210	258	592	479	152
	d	0.18	0.18	0.25	0.11	0.23	0.22	0.16	0.17	0.37
±0.10	N_c	26	40	54	30	21	39	26	64	18
	d_z	0.52	0.39	0.47	0.54	0.52	0.41	0.56	0.33	0.76
	N_p	114	93	94	338	52	64	148	120	38
	d	0.36	0.36	0.50	0.23	0.46	0.45	0.33	0.34	0.75
±0.25	N_c	4	6	9	5	3	6	4	10	3
	d_z	1.31	0.96	1.18	1.34	1.30	1.02	1.39	0.82	1.90
	N_p	18	15	15	54	8	10	24	19	6
	d	0.90	0.89	1.26	0.56	1.16	1.12	0.82	0.85	1.87

Note: Even if some of the sample sizes are smaller than 10 participants, the minimum suggested sample size for a group should not be less than that (Hopkins, 2000).

Abbreviations: L-ACC, left anterior cingulate cortex; L-S1, left primary somatosensory; R-AI, right anterior insula.

TABLE 7 Hypothetical sample size calculation for crossover (N_c) and parallel (N_p) sustained resting EEG (REEG) experiment design as a function of the effect size (d and d_z)

Effect size ($\log_{10}(\text{CSD})$)		L-ACC			L-S1			R-AI		
		θ	α	γ	θ	α	γ	θ	α	γ
±0.05	N_c	81	177	99	74	18	253	173	278	165
	d_z	0.35	0.29	0.27	0.33	0.66	0.17	0.20	0.19	0.18
	N_p	435	475	323	721	74	703	692	812	385
	d	0.21	0.25	0.21	0.15	0.46	0.14	0.14	0.16	0.16
±0.10	N_c	20	44	25	18	4	63	43	70	41
	d_z	0.70	0.57	0.54	0.66	1.33	0.34	0.39	0.38	0.35
	N_p	109	119	81	180	19	176	173	203	96
	d	0.43	0.49	0.42	0.30	0.91	0.28	0.28	0.31	0.33
±0.25	N_c	3	7	4	3	1	10	7	11	7
	d_z	1.75	1.43	1.34	1.65	3.32	0.84	0.98	0.95	0.88
	N_p	17	19	13	29	3	28	28	32	15
	d	1.07	1.24	1.05	0.75	2.28	0.71	0.69	0.78	0.82

Note: Even if some of the sample sizes are smaller than 10 participants, the minimum suggested sample size for a group should not be less than that (Hopkins, 2000).

Abbreviations: L-ACC, left anterior cingulate cortex; L-S1, left primary somatosensory; R-AI, right anterior insula.

4.2.2 | The reliability of CSD estimations

Bland Altman plots showed no clear signs of heteroscedasticity in the CSD estimation between sessions. This was expected because CSD data were log-transformed, and in most cases, log-transformation will address heteroscedasticity (Schmidt, Germano, & Milani, 2019). Further analysis of the plots revealed that there is no clear sign of systematic bias during SDTP, SDTnP, and REEG (except in the AI during SDTP and the L-ACC during SDTnP both in the gamma band). This is

evident because the CI of the bias did not overlap the zero-log (CSD) difference between sessions (except in the two aforementioned cases).

Overall findings revealed that the reliability of CSD estimations is largely independent of the selected ROIs. This is evident because the LoA are not different among conditions, ROIs and frequencies. A subsequent analysis exposed higher mean CSD estimations and narrower LoA during REEG in S1 in the alpha band, which leads to a better reliability (Olofsen, Dahan, Borsboom, & Drummond, 2014). There is no

clear explanation of why alpha reliability is higher during resting at S1. As many researchers have reported, participants exhibited higher alpha activity in parietal and occipital regions during eyes-open resting state (Barry & De Blasio, 2017; Chen, Feng, Zhao, Yin, & Wang, 2008; Pitchford & Arnell, 2019; Smith et al., 2020). Therefore, the higher signal-to-noise ratio in the alpha band at S1 may be to a large extent responsible for the higher reliability (Elkum & Shoukri, 2008). (Martín-Buro, Garcés, & Maestú, 2016).

A few studies attempted to assess the reliability of CSD using the LORETA method during different conditions, nonetheless, their results may appear inconsistent (Cannon et al., 2012; Hansen et al., 2017; Segalowitz et al., 2010; Tenke et al., 2017). Whereas some studies report good to excellent reliability (Cannon et al., 2012; Hansen et al., 2017), others reported mixed results (Segalowitz et al., 2010; Tenke et al., 2017) depending on different factors, such as the ROI or the frequency band from which the CSD was estimated. Under this framework, it is not clear from the literature whether the LORETA method is reliable or not, or which are the causes of these inconsistencies. These inconsistencies may not be fully related to methodological aspects of the experiments, but also in which manner studies report and interpret reliability. In most cases, researchers judge reliability based on arbitrary thresholds and make a binary decision (reliable/not reliable) disregarding the width of the 95% CI of the estimations which again depends on the sample size. Furthermore, none of these studies examined heteroscedasticity, which is known to impact relative and absolute reliability (Brehm, Scholtes, Dallmeijer, Twisk, & Harlaar, 2012; Schmidt et al., 2019).

4.2.3 | Methodological considerations

Pain is a broad concept, encompassing a large number of subcategories, for example, experimental or clinical, acute or chronic, nociceptive or neuropathic, and tonic or phasic, just to name a few. While our original intention was to focus on experimental tonic pain, the preliminary analysis of the literature revealed a scarcity of quantitative data in this area, revealing one of the large issues addressed in this study. Only by broadening the scope to include all articles in pain research were we able to find reports of effect sizes and variability of CSD estimations, mainly from exploratory studies of clinical pain. A related issue arose with regards to the ROIs from which quantitative data were reported: although it is well-known that multiple areas in the brain (aside from those specifically mentioned in this study) contribute to pain processing, and this often affects neural oscillations in all frequency bands, measurable outcomes were often not reported in sufficient detail for these variables. In these regard, the purpose of the present article is not to present comprehensive results of CSD estimations from all possible brain areas and frequency bands during different types of pain, but instead to draw attention to current flaws in outcome reporting in the field, and to provide reliable (if limited) data on effect sizes and variability of CSD estimations from the literature and the present experimental findings, in order to contribute to the proper planning of future research designs using these tools.

Although there are other CSD estimation methods, such as those based on minimum norm estimates (Grech et al., 2008), we chose to focus on LORETA for the systematic review and experimental analysis because it was by far the most commonly used. Furthermore, eLORETA was used to estimate the CSD in the L-ACC, L-S1, and R-AI. Even though this technique has no localization bias (Pascual-Marqui, 2007), the spatial accuracy is highly dependent on the number of electrodes and whether individual MRI data or a template head model is used (Michel et al., 2004; Michel & Brunet, 2019). A whole-head, dense-array sampling (e.g., 256 channels), the individual MRI of each volunteer and the co-registration of the spatial locations of the electrodes will obtain a more accurate source localization. In this study, however, a 64-channel EEG amplifier and a standard head model were used. This approach reduces the complexity of the experiment and avoids measuring the electrode positions on every volunteer. Nonetheless, variability due to anatomical differences and the real electrode positions were not considered, and this certainly increases spatial variability. Additionally, participants were not presumably consuming any drugs interfering with EEG activity (in accordance with the inclusion criteria), but this cannot be verified with absolute certainty. Finally, the sample in this experiment was not gender balanced. Gender differences have previously been observed in resting state EEG (Shankman et al., 2011) and CSD estimations (Stewart, Bismark, Towers, Coan, & Allen, 2010); whereas reliability is unlikely to be affected by differences, authors are encouraged to take this into consideration for future experimental designs involving effect sizes of CSD estimations.

4.3 | Considerations for future research designs

Over the last decade, there was an increased concern about the “crisis of unreplicable research.” This crisis is not entirely about replication failures, but also due to the misinterpretation of statistical inferences (Amrhein, Trafimow, & Greenland, 2019). Several difficulties arise when researchers base their conclusions solely on arbitrary thresholds such as $p = 0.05$. For example, p -values smaller than 0.05 may arise from random variation, and $p = 0.06$ could be considered similar to $p = 0.04$ from a practical perspective, in which the only difference might be the sample size (Halsey, Curran-Everett, Vowler, & Drummond, 2015). Thus, a small sample size might negatively impact the quality of the conclusions that can be drawn from an analysis, particularly when nonsignificant results are often obtained. Nonsignificant results are the consequence of (1) absence of real effect of the treatment/condition (true negative) or (2) a real effect that is too small to be detected with the current sample size (false negative, type II error).

However, it is not always adequate to increase the number of participants. Since increasing the number of participants will allow obtaining smaller p -values even with small effects, it can be problematic in medicine where those small differences can be attributed to the natural variability among participant or to measurement error. To overcome this situation, it is important to have a proper estimation of

the effect size of the treatment/experimental condition, compared to the variability of the measurement (i.e., the measurement error). Therefore, researchers suggest that inferences should not be based only on p -values (Ioannidis, 2019), and the results and their implications should take into consideration the physiological and clinical relevance (Schober, Bossers, & Schwarte, 2018). In this context, the capacity of an experiment to truly identify a relevant effect is highly dependent on adequate sample size.

In this study, hypothetical sample size calculations revealed that the number of participants to sufficiently power the study seems to be independent of the conditions, ROI and frequency band. Following the rule of thumb for the interpretation of Cohen's d effect size (Sawilowsky, 2009), results show that a small effect size ($d < 0.20$) requires a large number of participants ($N_c \sim 100$ and $N_p > 200$), a medium effect size ($d < 0.50$) requires sample size of $N_c \sim 40$ and $N_p > 100$ and a large effect size ($d > 0.80$) requires a sample of $N_c \sim 10$ and $N_p \sim 20$. It is evident that for small effect sizes, especially in parallel designs, CSD estimations may not be suitable to truly detect a clinically relevant effect. A further revision of the 46 unique papers revealed that, in average, between 20 and 30 subjects were included per study (see Tables S2 and S3). Based on the results, these sample sizes are suitable to detect medium and large effect sizes in crossover studies and only large effects in parallel studies. From here on, it is up to researcher and clinicians to establish if these effect sizes are experimentally or clinically relevant in each particular setting, in order to properly plan their experiments or interventions before they are carried out.

4.4 | Conclusions

None of the studies included in the systematic review of CSD estimations during pain reported sample size calculations, and less than 20% of them reported absolute measures of central tendency and dispersion, which are necessary to estimate effect sizes. Based on these data and our own experimental results, we determined that sample sizes commonly used in pain studies using CSD estimations are suitable to detect medium and large effect sizes in crossover designs and only large effects in parallel designs. The results presented here provide a comprehensive summary of the effect sizes that can be observed using LORETA in pain research, that can be used by clinicians and researchers to improve settings and designs of future pain studies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Juan Manuel Völker: conceptualization, methodology, software, analysis, investigation, writing—original draft, reviewing, and editing. Federico Gabriel Arguissain: conceptualization, methodology, writing—reviewing and editing. Ole Kæseler Andersen: conceptualization, methodology, writing—original draft, reviewing and editing, supervision, funding acquisition. José Biurrun Manresa:

conceptualization, methodology, analysis, writing—original draft, reviewing and editing, supervision.

DATA AVAILABILITY STATEMENT

All data required to reproduce the results in this article is attached as supplementary material. Furthermore, raw EEG data collected for this study is available on request to the corresponding author, given the size of the database (> 10 GB). Per university policies, a formal data sharing agreement is required beforehand.

ETHICS STATEMENT

The present study was conducted according to the Declaration of Helsinki and was approved by the ethical committee of Northern Jutland, Denmark (N-20170047).

ORCID

José Biurrun Manresa  <https://orcid.org/0000-0003-4060-9665>

REFERENCES

- Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician*, *73*, 262–270. <https://doi.org/10.1080/00031305.2018.1543137>
- Apkarian, A. V., Bushnell, M. C., Treede, R.-D., & Zubieta, J.-K. (2005). Human brain mechanisms of pain perception and regulation in health and disease. *European Journal of Pain*, *9*, 463–484. <https://doi.org/10.1016/j.ejpain.2004.11.001>
- Babiloni, C., Del Percio, C., Lizio, R., Noce, G., Cordone, S., Lopez, S., ... Bonanni, L. (2017). Abnormalities of cortical neural synchronization mechanisms in patients with dementia due to Alzheimer's and Lewy body diseases: An EEG study. *Neurobiology of Aging*, *55*, 143–158. <https://doi.org/10.1016/j.neurobiolaging.2017.03.030>
- Backonja, M., Howland, E. W., Wang, J., Smith, J., Salinsky, M., & Cleland, C. S. (1991). Tonic changes in alpha power during immersion of the hand in cold water. *Electroencephalography and Clinical Neurophysiology*, *79*, 192–203. [https://doi.org/10.1016/0013-4694\(91\)90137-5](https://doi.org/10.1016/0013-4694(91)90137-5)
- Barry, R. J., & De Blasio, F. M. (2017). EEG differences between eyes-closed and eyes-open resting remain in healthy ageing. *Biological Psychology*, *129*, 293–304. <https://doi.org/10.1016/j.biopsycho.2017.09.010>
- Biurrun Manresa, J. A., Fritsche, R., Vuilleumier, P. H., Oehler, C., Mørch, C. D., Arendt-Nielsen, L., ... Curatolo, M. (2014). Is the conditioned pain modulation paradigm reliable? A test-retest assessment using the nociceptive withdrawal reflex. *PLoS One*, *9*, e100241. <https://doi.org/10.1371/journal.pone.0100241>
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*, 307–310.
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*, 135–160. <https://doi.org/10.1191/096228099673819272>
- Brehm, M. A., Scholtes, V. A., Dallmeijer, A. J., Twisk, J. W., & Harlaar, J. (2012). The importance of addressing heteroscedasticity in the reliability analysis of ratio-scaled variables: An example based on walking energy-cost measurements. *Developmental Medicine and Child Neurology*, *54*, 267–273. <https://doi.org/10.1111/j.1469-8749.2011.04164.x>
- Brooks, J. C. W., Nurmikko, T. J., Bimson, W. E., Singh, K. D., & Roberts, N. (2002). fMRI of thermal pain: Effects of stimulus laterality and attention. *NeuroImage*, *15*, 293–301. <https://doi.org/10.1006/nimg.2001.0974>

- Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: What is it, and how is it measured? *Physiotherapy*, *86*, 94–99. [https://doi.org/10.1016/S0031-9406\(05\)61211-4](https://doi.org/10.1016/S0031-9406(05)61211-4)
- Cannon, R., Kerson, C., & Hampshire, A. (2011). SLORETA and fMRI detection of medial prefrontal default network anomalies in adult ADHD. *Journal of Neurotherapy*, *15*, 358–373. <https://doi.org/10.1080/10874208.2011.623093>
- Cannon, R. L., Baldwin, D. R., Shaw, T. L., Diloreto, D. J., Phillips, S. M., Scruggs, A. M., & Riehl, T. C. (2012). Reliability of quantitative EEG (qEEG) measures and LORETA current source density at 30 days. *Neuroscience Letters*, *518*, 27–31. <https://doi.org/10.1016/j.neulet.2012.04.035>
- Canuet, L., Tellado, I., Couceiro, V., Fraile, C., Fernandez-Novoa, L., Ishii, R., ... Cacabelos, R. (2012). Resting-state network disruption and APOE genotype in Alzheimer's disease: A lagged functional connectivity study. *PLoS One*, *7*, 1–12. <https://doi.org/10.1371/journal.pone.0046289>
- Carrier, J., Land, S., Buysse, D. J., Kupfer, D. J., & Monk, T. H. (2001). The effects of age and gender on sleep EEG power spectral density in the middle years of life (ages 20–60 years old). *Psychophysiology*, *38*, 232–242. <https://doi.org/10.1017/S0048577201991838>
- Castellini, G., Gianola, S., Bonovas, S., & Moja, L. (2016). Improving power and sample size calculation in rehabilitation trial reports: A methodological assessment. *Archives of Physical Medicine and Rehabilitation*, *97*, 1195–1201. <https://doi.org/10.1016/j.apmr.2016.02.013>
- Chang, P. F., Arendt-Nielsen, L., Graven-Nielsen, T., Svensson, P., & Chen, A. C. N. (2001). Different EEG topographic effects of painful and non-painful intramuscular stimulation in man. *Experimental Brain Research*, *141*, 195–203. <https://doi.org/10.1007/s002210100864>
- Chen, A. C. N. N., Feng, W., Zhao, H., Yin, Y., & Wang, P. (2008). EEG default mode network in the human brain: Spectral regional field powers. *NeuroImage*, *41*, 561–574. <https://doi.org/10.1016/j.neuroimage.2007.12.064>
- Craig, A. D. (2003). Pain mechanisms: Labeled lines versus convergence in central processing. *Annual Review of Neuroscience*, *26*, 1–30. <https://doi.org/10.1146/annurev.neuro.26.041002.131022>
- Davis, K. D., & Seminowicz, D. A. (2017). Insights for clinicians from brain imaging studies of pain. *The Clinical Journal of Pain*, *33*, 291–294. <https://doi.org/10.1097/AJP.0000000000000439>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Downar, R., Rissacher, D., & Schuckers, S. (2008). EEG indices of tonic pain-related activity in the somatosensory cortices. *Clinical Neurophysiology*, *119*, 1201–1212. <https://doi.org/10.1016/j.clinph.2008.01.019>
- Downing, S. M. (2004). Reliability: On the reproducibility of assessment data. *Medical Education*, *38*, 1006–1012. <https://doi.org/10.1111/j.1365-2929.2004.01932.x>
- Elkum, N., & Shoukri, M. M. (2008). Signal-to-noise ratio (SNR) as a measure of reproducibility: Design, estimation, and application. *Health Services and Outcomes Research Methodology*, *8*, 119–133. <https://doi.org/10.1007/s10742-008-0030-2>
- Glynn, C. J., Lloyd, J. W., & Folkard, S. (1976). The diurnal variation in perception of pain. *Journal of the Royal Society of Medicine*, *69*, 369–372. <https://doi.org/10.1177/003591577606900532>
- González-Roldán, A. M., Cifre, I., Sitges, C., & Montoya, P. (2016). Altered dynamic of EEG oscillations in fibromyalgia patients at rest. *Pain Medicine (United States)*, *17*(6), 1058–1068. <https://doi.org/10.1093/pm/pnw023>
- Graven-Nielsen, T., Vaegter, H. B., Finocchietti, S., Handberg, G., & Arendt-Nielsen, L. (2015). Assessment of musculoskeletal pain sensitivity and temporal summation by cuff pressure algometry. *Pain*, *156*, 2193–2202. <https://doi.org/10.1097/j.pain.0000000000000294>
- Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., ... Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis. *Journal of Neuroengineering and Rehabilitation*, *5*, 1–33. <https://doi.org/10.1186/1743-0003-5-25>
- Guo, Q., Thabane, L., Hall, G., McKinnon, M., Goeree, R., & Pullenayegum, E. (2014). A systematic review of the reporting of sample size calculations and corresponding data components in observational functional magnetic resonance imaging studies. *NeuroImage*, *86*, 172–181. <https://doi.org/10.1016/j.neuroimage.2013.08.012>
- Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, *12*, 179–185. <https://doi.org/10.1038/nmeth.3288>
- Hansen, T. M., Mark, E. B., Olesen, S. S., Gram, M., Frøkjær, J. B., & Drewes, A. M. (2017). Characterization of cortical source generators based on electroencephalography during tonic pain. *Journal of Pain Research*, *10*, 1401–1409. <https://doi.org/10.2147/JPR.S132909>
- Hauck, M., Lorenz, J., & Engel, A. K. (2007). Attention to painful stimulation enhances γ -band activity and synchronization in human sensorimotor cortex. *The Journal of Neuroscience*, *27*, 9270–9277. <https://doi.org/10.1523/JNEUROSCI.2283-07.2007>
- Hjermstad, M. J., Fayers, P. M., Haugen, D. F., Caraceni, A., Hanks, G. W., Loge, J. H., ... European Palliative Care Research Collaborative (EPCRC). (2011). Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: A systematic literature review. *Journal of Pain and Symptom Management*, *41*, 1073–1093. <https://doi.org/10.1016/j.jpainsymman.2010.08.016>
- Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Journal of Sports Medicine*, *30*, 1–15.
- Ioannidis, J. P. A. (2019). What have we (not) learnt from millions of scientific papers with P values? *The American Statistician*, *73*, 20–25. <https://doi.org/10.1080/00031305.2018.1447512>
- Julious, S. A. (2004). Sample sizes for clinical trials with normal data. *Statistics in Medicine*, *23*, 1921–1986. <https://doi.org/10.1002/sim.1783>
- Jung, T., Makeig, S., Humphries, C., Lee, T., McKeown, M. J., Iragui, I., & Sejnowski, T. J. (2000). Removing electroencephalographic artefacts by blind source separation. *Psychophysiology*, *37*, 163–178. <https://doi.org/10.1111/1469-8986.3720163>
- Kottner, J., Audige, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., ... Streiner, D. L. (2011). Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *International Journal of Nursing Studies*, *48*, 661–671. <https://doi.org/10.1016/j.ijnurstu.2011.01.016>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 1–12. <https://doi.org/10.3389/fpsyg.2013.00863>
- Larson, M. J., & Carbine, K. A. (2017). Sample size calculations in human electrophysiology (EEG and ERP) studies: A systematic review and recommendations for increased rigor. *International Journal of Psychophysiology*, *111*, 33–41. <https://doi.org/10.1016/j.ijpsycho.2016.06.015>
- Lee, P. H., & Tse, A. C. Y. (2017). The quality of the reported sample size calculations in randomized controlled trials indexed in PubMed. *European Journal of Internal Medicine*, *40*, 16–21. <https://doi.org/10.1016/j.ejim.2016.10.008>
- Lelic, D., Hansen, T. M., Mark, E. B., Olesen, A. E., & Drewes, A. M. (2017). The effects of analgesics on central processing of tonic pain: A cross-over placebo controlled study. *Neuropharmacology*, *123*, 455–464. <https://doi.org/10.1016/j.neuropharm.2017.06.022>
- Li, L., Liu, X., Cai, C., Yang, Y., Li, D., Xiao, L., ... Qiu, Y. (2016). Changes of gamma-band oscillatory activity to tonic muscle pain. *Neuroscience Letters*, *627*, 126–131. <https://doi.org/10.1016/j.neulet.2016.05.067>
- Martín-Buro, M. C., Garcés, P., & Maestú, F. (2016). Test-retest reliability of resting-state magnetoencephalography power in sensor and source space. *Human Brain Mapping*, *37*, 179–190. <https://doi.org/10.1002/hbm.23027>

- McEvoy, L. K., Smith, M. E., & Gevins, A. (2000). Test-retest reliability of cognitive EEG. *Clinical Neurophysiology*, *111*, 457–463. <https://doi.org/10.1167/iov.03-0265>
- Michel, C. M., & Brunet, D. (2019). EEG source imaging: A practical review of the analysis steps. *Frontiers in Neurology*, *10*, 325. <https://doi.org/10.3389/fneur.2019.00325>
- Michel, C. M., & Murray, M. M. (2012). Towards the utilization of EEG as a brain imaging tool. *NeuroImage*, *61*, 371–385. <https://doi.org/10.1016/j.neuroimage.2011.12.039>
- Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave De Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology*, *115*(10), 2195–2222. <https://doi.org/10.1016/j.clinph.2004.06.001>
- Moont, R., Crispel, Y., Lev, R., Pud, D., & Yarnitsky, D. (2011). Temporal changes in cortical activation during conditioned pain modulation (CPM), a LORETA study. *Pain*, *152*, 1469–1477. <https://doi.org/10.1016/j.pain.2011.01.036>
- Mouraux, A., & Iannetti, G. D. (2018). The search for pain biomarkers in the human brain. *Brain*, *141*, 3290–3307. <https://doi.org/10.1093/brain/awy281>
- Nickel, M. M., May, E. S., Tiemann, L., Schmidt, P., Postorino, M., Ta Dinh, S., ... Ploner, M. (2017). Brain oscillations differentially encode noxious stimulus intensity and pain intensity. *NeuroImage*, *148*, 141–147. <https://doi.org/10.1016/j.neuroimage.2017.01.011>
- Nir, R. R., Sinai, A., Moont, R., Harari, E., & Yarnitsky, D. (2012). Tonic pain and continuous EEG: Prediction of subjective pain perception by alpha-1 power during stimulation and at rest. *Clinical Neurophysiology*, *123*, 605–612. <https://doi.org/10.1016/j.clinph.2011.08.006>
- Nir, R. R., Sinai, A., Raz, E., Sprecher, E., & Yarnitsky, D. (2010). Pain assessment by continuous EEG: Association between subjective perception of tonic pain and peak frequency of alpha oscillations during stimulation and at rest. *Brain Research*, *1344*, 77–86. <https://doi.org/10.1016/j.brainres.2010.05.004>
- Noordzij, M., Tripepi, G., Dekker, F. W., Zoccali, C., Tanck, M. W., & Jager, K. J. (2010). Sample size calculations: Basic principles and common pitfalls. *Nephrology, Dialysis, Transplantation*, *25*, 1388–1393. <https://doi.org/10.1093/ndt/gfp732>
- Oakes, T. R., Pizzagalli, D. A., Hendrick, A. M., Horras, K. A., Larson, C. L., Abercrombie, H. C., ... Davidson, R. J. (2004). Functional coupling of simultaneous electrical and metabolic activity in the human brain. *Human Brain Mapping*, *21*, 257–270. <https://doi.org/10.1002/hbm.20004>
- Olofsen, E., Dahan, A., Borsboom, G., & Drummond, G. (2014). Improvements in the application and reporting of advanced Bland–Altman methods of comparison. *Journal of Clinical Monitoring and Computing*, *29*, 127–139. <https://doi.org/10.1007/s10877-014-9577-3>
- Ostrowsky, K., Magnin, M., Rylvlin, P., Isnard, J., Guenot, M., & Mauguière, F. (2002). Representation of pain and somatic sensation in the human insula: A study of responses to direct electrical cortical stimulation. *Cerebral Cortex*, *12*, 376–385. <https://doi.org/10.1093/cercor/12.4.376>
- Pascual-Marqui, R. D. (2007). Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: Exact, zero error localization. *Neuroscience Letters*, *485*, 198–203. <https://doi.org/10.1016/j.neulet.2010.09.011>
- Pascual-Marqui, R. D., Esslen, M., Kochi, K., & Lehmann, D. (2002). Functional imaging with low-resolution brain electromagnetic tomography (LORETA): A review. *Methods and Findings in Experimental and Clinical Pharmacology*, *24*(Suppl C), 91–95.
- Pascual-Marqui, R. D., Lehmann, D., Koenig, T., Kochi, K., Merlo, M. C. G., Hell, D., & Koukkou, M. (1999). Low resolution brain electromagnetic tomography (LORETA) functional imaging in acute, neuroleptic-naive, first-episode, productive schizophrenia. *Psychiatry Research: Neuroimaging*, *90*, 169–179. [https://doi.org/10.1016/S0925-4927\(99\)00013-X](https://doi.org/10.1016/S0925-4927(99)00013-X)
- Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, *18*, 49–65. [https://doi.org/10.1016/0167-8760\(84\)90014-X](https://doi.org/10.1016/0167-8760(84)90014-X)
- Pascual-Marqui, R. D., Esslen, M., Kochi, K., & Lehmann, D. (2002). Functional imaging with low resolution brain electromagnetic tomography (LORETA): Review, new comparisons, and new validation. *Japanese Journal of Clinical Neurophysiology*, *30*, 81–94.
- Pitchford, B., & Arnell, K. M. (2019). Resting EEG in alpha and beta bands predicts individual differences in attentional breadth. *Consciousness and Cognition*, *75*, 102803. <https://doi.org/10.1016/j.concog.2019.102803>
- Ploner, M., & May, E. S. (2018). Electroencephalography and magnetoencephalography in pain research—Current state and future perspectives. *Pain*, *159*, 206–211. <https://doi.org/10.1097/j.pain.0000000000001087>
- Ploner, M., Sorg, C., & Gross, J. (2017). Brain rhythms of pain. *Trends in Cognitive Sciences*, *21*, 100–110. <https://doi.org/10.1016/j.tics.2016.12.001>
- Polianskis, R., Graven-Nielsen, T., & Arendt-Nielsen, L. (2002). Spatial and temporal aspects of deep tissue pain assessed by cuff algometry. *Pain*, *100*, 19–26. [https://doi.org/10.1016/S0304-3959\(02\)00162-8](https://doi.org/10.1016/S0304-3959(02)00162-8)
- Prichep, L. S., Shah, J., Merkin, H., & Hiesiger, E. M. (2018). Exploration of the pathophysiology of chronic pain using quantitative EEG source localization. *Clinical EEG and Neuroscience*, *49*, 103–113. <https://doi.org/10.1177/1550059417736444>
- Prinsloo, S., Rosenthal, D. I., Lyle, R., Garcia, S. M., Gabel-Zepeda, S., Cannon, R., ... Cohen, L. (2019). Exploratory study of low resolution electromagnetic tomography (LORETA) real-time Z-score feedback in the treatment of pain in patients with head and neck cancer. *Brain Topography*, *32*, 283–285. <https://doi.org/10.1007/s10548-018-0686-z>
- Rousson, V., Gasser, T., & Seifert, B. (2002). Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Statistics in Medicine*, *21*, 3431–3446. <https://doi.org/10.1002/sim.1253>
- Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*, 597–599. <https://doi.org/10.22237/jmasm/1257035100>
- Schmidt, D., Germano, A. M. C., & Milani, T. L. (2019). Subjective sensitivity data: Considerations to treat heteroscedasticity. *Cogent Medicine*, *6*, 1–10. <https://doi.org/10.1080/2331205X.2019.1673086>
- Schober, P., Bossers, S. M., & Schwarte, L. A. (2018). Statistical significance versus clinical importance of observed effect sizes: What do p values and confidence intervals really represent? *Anesthesia and Analgesia*, *126*, 1068–1072. <https://doi.org/10.1213/ANE.0000000000002798>
- Schulz, E., May, E. S., Postorino, M., Tiemann, L., Nickel, M. M., Witkovsky, V., ... Ploner, M. (2015). Prefrontal gamma oscillations encode tonic pain in humans. *Cerebral Cortex*, *25*, 4407–4414. <https://doi.org/10.1093/cercor/bhv043>
- Segalowitz, S. J., Santesso, D. L., Murphy, T. I., Homan, D., Chantziantonou, D. K., & Khan, S. (2010). Retest reliability of medial frontal negativities during performance monitoring. *Psychophysiology*, *47*, 260–270. <https://doi.org/10.1111/j.1469-8986.2009.00942.x>
- Shankman, S. A., Klein, D. N., Torpey, D. C., Olino, T. M., Dyson, M. W., Kim, J., ... Tenke, C. E. (2011). Do positive and negative temperament traits interact in predicting risk for depression? A resting EEG study of 329 preschoolers. *Development and Psychopathology*, *23*, 551–562. <https://doi.org/10.1017/S0954579411000022>
- Shao, S., Shen, K., Yu, K., Wilder-Smith, E. P. V., & Li, X. (2012). Frequency-domain EEG source analysis for acute tonic cold pain perception. *Clinical Neurophysiology*, *123*, 2042–2049. <https://doi.org/10.1016/j.clinph.2012.02.084>
- Shieh, G. (2014). Optimal sample sizes for the design of reliability studies: Power consideration. *Behavior Research Methods*, *46*, 772–785. <https://doi.org/10.3758/s13428-013-0396-0>
- Smith, E. E., Tenke, C. E., Deldin, P. J., Trivedi, M. H., Weissman, M. M., Auerbach, R. P., ... Kayser, J. (2020). Frontal theta and posterior alpha in resting EEG: A critical examination of convergent and discriminant validity. *Psychophysiology*, *57*, 1–20. <https://doi.org/10.1111/psyp.13483>

- Stewart, J. L., Bismark, A. W., Towers, D. N., Coan, J. A., & Allen, J. J. B. (2010). Resting frontal EEG asymmetry as an endophenotype for depression risk: Sex-specific patterns of frontal brain asymmetry. *Journal of Abnormal Psychology, 119*, 502–512. <https://doi.org/10.1037/a0019196>
- Sullivan, G. M., & Feinn, R. (2012). Using effect size—Or why the P value is not enough. *The Journal of Graduate Medical Education, 4*, 279–282. <https://doi.org/10.4300/jgme-d-12-00156.1>
- Tenke, C. E., Kayser, J., Pachtel, P., Webb, C. A., Dillon, D. G., Goer, F., ... Bruder, G. E. (2017). Demonstrating test-retest reliability of electrophysiological measures for healthy adults in a multisite study of biomarkers of antidepressant treatment response. *Psychophysiology, 54*, 34–50. <https://doi.org/10.1111/psyp.12758>
- Thatcher, R. W., North, D., & Biver, C. (2005). Parametric vs. non-parametric statistics of low resolution electromagnetic tomography (LORETA). *Clinical EEG and Neuroscience, 36*, 1–8. <https://doi.org/10.1177/155005940503600103>
- Tracey, I., Woolf, C. J., & Andrews, N. A. (2019). Composite pain biomarker signatures for objective assessment and effective treatment. *Neuron, 101*, 783–800. <https://doi.org/10.1016/j.neuron.2019.02.019>
- Ye, Q., Yan, D., Yao, M., Lou, W., & Peng, W. (2019). Hyperexcitability of cortical oscillations in patients with somatoform pain disorder: A resting-state EEG study. *Neural Plasticity, 2019*, 2687150–10. <https://doi.org/10.1155/2019/2687150>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Völker JM, Arguissain FG, Andersen OK, Biurrun Manresa J. Variability and effect sizes of intracranial current source density estimations during pain: Systematic review, experimental findings, and future perspectives. *Hum Brain Mapp.* 2021;42:2461–2476. <https://doi.org/10.1002/hbm.25380>