

RESEARCH ARTICLE

Within-session test-retest reliability of pressure pain threshold and mechanical temporal summation in healthy subjects

Catherine Mailloux¹✉, Louis-David Beaulieu^{2‡}, Timothy H. Wideman^{3‡}, Hugo Massé-Alarie¹✉*

1 Centre Interdisciplinaire de Recherche en Réadaptation et Intégration Sociale (CIRRIS), Université Laval, Quebec, Canada, **2** BioNR Research Lab, Université du Québec à Chicoutimi, Chicoutimi, Canada, **3** School of Physical and Occupational Therapy, McGill University, Montreal, Canada

✉ These authors contributed equally to this work.

‡ These authors contributed also equally to this work.

* hugo.masse-alarie@fmed.ulaval.ca



OPEN ACCESS

Citation: Mailloux C, Beaulieu L-D, Wideman TH, Massé-Alarie H (2021) Within-session test-retest reliability of pressure pain threshold and mechanical temporal summation in healthy subjects. PLoS ONE 16(1): e0245278. <https://doi.org/10.1371/journal.pone.0245278>

Editor: Alison Rushton, Western University, CANADA

Received: September 1, 2020

Accepted: December 27, 2020

Published: January 12, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0245278>

Copyright: © 2021 Mailloux et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All files are available from the Scholar Portal Dataverse database (<https://doi.org/10.5683/SP2/URLJLN>), Scholars

Abstract

Objective

To determine the absolute and relative intra-rater within-session test-retest reliability of pressure pain threshold (PPT) and mechanical temporal summation of pain (TSP) at the low back and the forearm in healthy participants and to test the influence of the number and sequence of measurements on reliability metrics.

Methods

In 24 participants, three PPT and TSP measures were assessed at four sites (2 at the low back, 2 at the forearm) in two blocks of measurements separated by 20 minutes. The standard error of measurement, the minimal detectable change (MDC) and the intraclass correlation coefficient (ICC) were investigated for five different sequences of measurements (e.g. measurement 1, 1–2, 1-2-3).

Results

The MDC for the group (MDC_{gr}) for PPT ranged from 28.71 to 50.56 kPa across the sites tested, whereas MDC_{gr} for TSP varied from 0.33 to 0.57 out of 10 (numeric scale). Almost all ICC showed an excellent relative reliability (between 0.80 and 0.97), except when only the first measurement was considered (moderate). Although minimal differences in absolute PPT reliability were present between the different sequences, in general, using only the first measurement increase measurement error. Three TSP measures reduced the measurement error.

Discussion

We established that two measurements of PPT and three of TSP reduced the measurement error and demonstrated an excellent relative reliability. Our results could be used in future

Portal Dataverse, V1, UNF:6:g
+Lz6E4lvbaXly5wuxkzPQ==).

Funding: This study is financed by a grant from the Quebec Pain Research Network (QPRN; <https://qprm.ca/fr>), the Réseau Provincial de Recherche en Adaptation-Réadaptation (REPAR; <https://repar.ca>) and the Canadian Musculoskeletal Rehab Research Network (<http://mskrehabnet.com>). HMA is supported by salary awards from Fonds de recherche du Québec - Santé (281961 - FRQS; <http://www.frqs.gouv.qc.ca>). CM received grants from the Canadian Institute of Health Research (CIHR; <https://cihr-irsc.gc.ca/e/193.html>) and the Ordre professionnel de la physiothérapie du Québec (OPPQ; <https://oppq.qc.ca>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

pain research to confirm the presence of true hypo/hyperalgesia for paradigms such as conditioned pain modulation or exercise-induced hypoalgesia, indicated by a change exceeding the measurement variability.

1. Introduction

The experience of pain is highly variable and influenced by biological, psychological and social factors [1]. One essential feature of the experience of pain is the capacity of the nervous system to modulate pain through the interplay of multiple areas and mechanisms [2]. The complexity of pain modulation mechanisms makes it difficult to evaluate. In humans, psychophysical experimental paradigms have been developed in research as proxy of pain modulation. Conditioned pain modulation (CPM) and exercise-induced hypoalgesia (EIH), for instance, are used to approximate the efficacy of pain inhibition [3–5]. Recently, these paradigms have been increasingly used in research to determine if individuals with chronic pain have an altered pain inhibition response. For example, some studies reported an alteration of pain inhibition in individuals with chronic low back pain (CLBP) using CPM [6–9] and EIH [10,11].

CPM refers to the decrease in pain sensitivity after the administration of a painful stimulus on a remote body part (e.g. cold water immersion of the hand [12]). EIH represents the decrease in pain sensitivity that occurs following an isometric, resistance or aerobic exercise [13]. Opioidergic, serotonergic and noradrenergic systems contribute to both CPM [2,14–17] and EIH [18–24]. Experimental protocols developed to assess EIH and CPM take generally part in a within-session design, including pain sensitivity measures collected before and after a conditioning stimulus, and they are determined by the change in pain sensitivity between test and retest. The pressure pain threshold (PPT) is the most used pain sensitivity measure for these paradigms [13,25,26]. PPT is a static measure of pain and would reflect the basal state of pain perception [4,27]. Temporal summation of pain (TSP) has also been used as a pain sensitivity measure to quantify CPM/EIH [28–30] and rather constitutes a dynamic measure of pain sensitivity that refers to the perception of increasing pain in response to stable (continuous or repeated) noxious stimuli [31–33]. To ensure the validity of these paradigms, it is essential that pain sensitivity measures chosen as testing stimulus (PPT and TSP) present small measurement error. This permits to define a minimal level of true change, exceeding the measurement variability, to confirm the presence of hyper- or hypoalgesia following a conditioning stimulus. In addition, reliability needs to be considered in respect of the site tested. Therefore, with a view of studying CLBP population, measuring variability of PPT and TSP at the low back area and at a remote site is essential to provide a general view of pain sensitivity and pain modulation functioning. Some studies documented PPT reliability at the low back [34–37], but the sample size tested in these studies remains limited. For TSP, no study measured the reliability at the lower back area. Moreover, the number of measures needed to reach an acceptable level of test-retest reliability for TSP is not known. This is essential to ensure the applicability and the feasibility of measuring PPT and TSP in research and clinical practice, and to reduce the number of painful stimulus needed to obtain reliable and valid data.

The objectives are to 1) determine the absolute and relative intra-rater within-session test-retest reliability of PPT and mechanical TSP at the low back and at the forearm in healthy participants and 2) test the influence of the number and sequence of measurements for PPT and TSP on reliability metrics. Our study focused on absolute reliability that will provide the minimal level to reach a real change in pain sensitivity, which is essential to interpret the presence/absence of hyper/hypoalgesia using pain modulation paradigms such as CPM and EIH.

2. Materials and methods

2.1 Participants

Twenty-four healthy subjects (12 females | 12 males; 28.3 ± 11.0 years old) aged between 18 and 65 years old were recruited between December 2018 and June 2019 (see [Table 1](#) for descriptive statistics). Sampling method was by convenience with emails sent to Laval University community (60,000 individuals comprising students and employees) and by solicitation at the research center. Determination of selection criteria was based on consensus statement by the EUROPAIN and NEUROPAIN consortia for quantitative sensory testing (QST)-based studies to ensure validity of the data [38]. Exclusion criteria were: 1) pain lasting three months or longer, located anywhere in the body, 2) severe health problem (such as cancer, major rheumatoid, cardiac, neurologic or psychiatric disease), 3) low back pain lasting more than 7 days in the last 6 months, 4) consultation with a health professional because of low back pain in the last 6 months, 5) current bilateral wrist or forearm pain and 6) current pregnancy and/or gave birth in the last year. Subjects who were currently taking medication like antidepressants, opioids, neuroleptics, anticonvulsive drugs or steroids were also excluded. This study was approved by local ethic committee (CIUSSS-Capitale Nationale, project #2019–1547) and all participants provided informed written consent prior experimentation. The body mass index was calculated for each participant and the Global Physical Activity Questionnaire (GPAQ) was self-administered to rate the level of physical activity [39].

2.2 Study design

All measurements were collected in a single session at the research center by the same rater (CM). The rater was a physical therapist with five years of clinical experience who had undertaken a QST training with experienced researchers.

PPT and TSP were tested in two blocks, lasting approximately 15 minutes each, separated by a pause of 20 minutes ([Fig 1](#)). The 20 min-pause was established in consideration of the time interval duration between blocks of testing when assessing EIH and CPM (e.g. time interval between the measure of PPT pre to post conditioning stimulus) [26,40–43]. This was done to ensure that PPT and TSP reliability were tested in a design reproducing CPM/EIH protocols.

Table 1. Descriptive statistics (n = 24).

| | |
|------------------------------|--------------------------|
| Age, years, $\bar{X} \pm SD$ | 28.3 ± 11.0 |
| Sex (male: female) | 12: 12 |
| Race, % | |
| Caucasian | 92 |
| Asian | 4 |
| Black | 4 |
| BMI | 23.0 ± 3.12 [†] |
| GPAQ | 2763 ± 2018 [‡] |
| Dominance, Right, % | 83 |
| Dominant site assessed, % | 42 |

\bar{X} mean, *SD* standard deviation, *BMI* Body Mass Index, *GPAQ* Global Physical Activity Questionnaire.

[†] Missing data for 2 participants.

[‡] Missing data for one participant.

<https://doi.org/10.1371/journal.pone.0245278.t001>

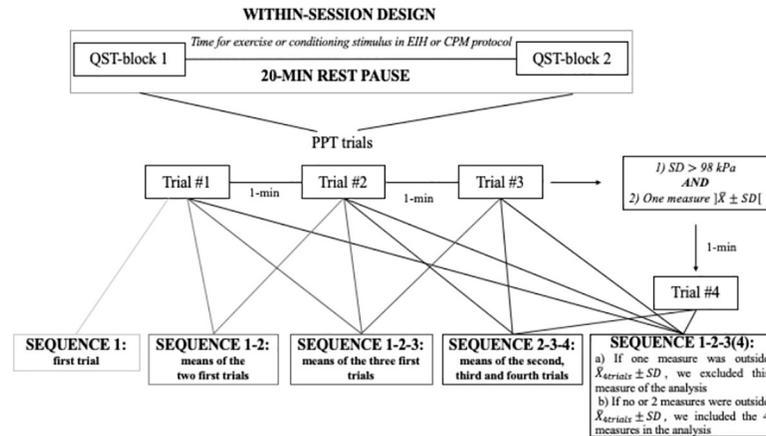


Fig 1. Study procedure and sequences of measurements analyzed for PPT.

<https://doi.org/10.1371/journal.pone.0245278.g001>

2.3 Quantitative sensory testing (QST)

PPT and TSP measures were collected in the same environment (same room with stable conditions regarding light, temperature and noise) and QST testing order within each testing block was randomized. Side was randomized independently of dominance and measures. If one side (wrist or low back) presented impairments (not related to pain) other than exclusion criteria, the opposite side was tested (1 participant had a limited wrist extension due to a scaphoid fracture 10 years ago [non-painful] and 1 participant had a low back lipoma). QST measures were first tested on the calf or the thigh for familiarization with the procedure.

2.3.1 Pressure pain threshold. PPT were assessed with a handheld digital algometer (1-cm² probe–FPIX, Wagner Instruments, Greenwich, CT, USA) for 22 participants and with a handheld dial algometer (1-cm² probe–FPK, Wagner Instruments, Greenwich, CT, USA) for 2 participants. Since FPK algometer does not allow to measure between 0 and 1 kg/cm², FPIX algometer was used for the remaining participants. To determine if the use of FPK algometer affected the reliability, we performed the analysis with and without the first two participants. Considering both analyses provided similar results, we included the first two participants in our analysis. A rate of ~0.5 kg/cm²/s was applied, at two back and two upper limb sites: i) lumbar erector spinae (LES), 2–3 cm laterally to L4/L5 ii) S1 spinous process, iii) dorsal aspect of the wrist on capitatum (WD) and iv) wrist flexors muscles (WF), 10 cm distally from medial humeral epicondyle on a line from medial epicondyle to styloid process of ulna, over the muscles bulk of the wrist flexors. All sites were located and marked before testing. Assessment of the back and the forearm were done in a prone lying position with a pillow under the abdomen and in sitting position with arm supported, respectively. Standardized verbal instructions were given based on German Research Network on Neuropathic Pain (DFNS) recommendations: “This is a test of your sensitivity to deep pain. Now I will press this pressure meter against your back/wrist/forearm and will gradually increase the pressure. Please say ‘Now’ as soon as the pressure starts to be painful. Remember that this is not a pain tolerance test, it is a pain threshold test” [32]. Instructions were freely translated into French. PPT were measured three times with one-minute break between measurements. To reduce the variability of the measurement and reduce the impact of potential outlier, a fourth measure was taken if the following two conditions were met: 1) the standard deviation (SD) of the three measures was larger than 1 kg/cm² and 2) one measure was outside the mean±SD interval (Fig 1). The same criteria were applied to the four values to determine if three or four data were kept for analysis. Because of the accuracy limit of the FPIX algometer, PPT data above 11.5 kg/cm² were considered as 11.5

kg/cm² (2 trials at S1). For the FPK, PPT between 0 and 1 kg/cm², were considered as 1 kg/cm² (4 trials at WD– 2 first participants). This may have caused a small overestimation of the reliability of the PPT measurement at S1 and WD. PPT were transformed from kg/cm² to kPa (1 kg/cm² = 98.07 kPa) to facilitate comparisons with the PPT literature.

2.3.2 Temporal summation of pain. TSP was tested using a pinprick stimulator (256 mN, MRC Systems GmbH, Heidelberg, Germany) and a series of ten punctuate stimuli at 1 Hz over (i) L4/L5 interspinous process line and (ii) hand dorsum. The 1-Hz frequency was displayed to the rater by a light metronome out of participant's sight. TSP was considered as the difference between the highest numeric rating scale (NRS) (anchored between 0 [no pain] and 10 [worst imaginable pain]) pain through the ten trials and the pain after a single stimulus [27,44]. Standardized verbal instructions were given based on DFNS recommendations: "This is a test of repeated pinpricks. I will now apply a single pinprick. Please give a number between '0' and '10' for the pain of that stimulus. I will now apply a series of 10 pinpricks in a row. Please give a number between '0' and '10' for the highest pain over that whole series of 10 pinpricks. This procedure will be repeated 3 times, with a 30-s break between" [32].

2.4 Statistical analysis

SPSS software was used for statistical analysis (IBM SPSS 25 for Mac, Armonk, NY, USA). First, Shapiro-Wilk's test and visual appreciation of histogram and Normal Q-Q Plots were used to assess normality of PPT and TSP distributions. Second, presence of outliers was assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box (representing a 99.3% confidence interval). Third, homoscedasticity i.e. absence of correlation between the size of error and the magnitude of the observed scores [45,46] was assessed. This was done by visual inspection of Bland-Altman plots of differences between the two values collected at test and retest for each participant by the means of these two values [47,48]. The correlation (R^2) between the absolute differences and the means values was also calculated for each variable by a linear regression analysis [49]. In presence of the two following conditions: 1) R^2 values greater than 0.1 [49] and 2) $p < 0.05$ for the linear regression model, heteroscedasticity was considered present. Natural logarithmic transformation was applied to data in the presence of non-normality or heteroscedasticity and normality/homoscedasticity were re-tested to ensure that previous assumptions were met.

2.4.1 Absolute reliability. Absolute reliability was assessed using the standard error of the measurement (SEM_{eas}) [50]. SEM_{eas} represents the within-subject variation and is defined as "the standard deviation of errors of measurement that is associated with the test scores for a specific group of test takers" [50,51]. SEM_{eas} was estimated as the square root of the mean square error (WMS) calculated by a one-way ANOVA ($ANOVA_{\text{RM}}$) applied on test and retest measurements ($SEM_{\text{eas}} = \sqrt{WMS}$) [45,52].

To evaluate measures responsiveness, the minimal detectable change for each individual (MDC_{ind}) was estimated for PPT and TSP with the following formula: $MDC_{\text{ind}} = 1,96 \times SEM_{\text{eas}} \times \sqrt{2}$ [53]. MDC for the group (MDC_{gr}) was also computed: $MDC_{\text{gr}} = MDC_{\text{ind}} / \sqrt{n}$, where "n" represents the sample size [54–56]. The values of SEM_{eas} , MDC_{ind} and MDC_{gr} were also presented as percentage of pooled means (average of test and retest measurements) [50,52] designated as % SEM_{eas} , % MDC_{ind} and % MDC_{gr} . % SEM_{eas} and % MDC allow comparisons between studies and facilitate interpretation, since they represent a dimension-less/unit-less measure [50,52]. If SEM_{eas} and MDC were obtained from a log-transformation of the data, an antilog (i.e. exponential (e^x)) was applied, resulting in a multiplication/division (\times/\div) factor to the test data. This factor indicates the random error component of the mean bias [45]. A constant was added to the distributions before the log-transformation to obtain positive and non-null value.

2.4.2 Relative reliability. The relative reliability was quantified using the intraclass correlation coefficient (ICC). ICC's model 2 was used since each subject was assessed by the same rater and the rater represents the population of possible raters [57]. The ICC was calculated using the Fisher's test of a two-way ANOVA_{RM}. "Random effects, absolute agreement, average measures" (2,k) model, for sequences including more than one measurement, and "single-measure" model (2,1) for the sequence of one measurement only were used (see next section for the description of the different sequences of measurements). ICC permits to determine systematic bias between test and retest (difference between means) [46] and ICC 95% confidence interval was calculated to represent ICC variability. An ICC > 0.80 was considered "excellent", between 0.61 and 0.80; "good", between 0.41 and 0.60; "moderate", between 0.21 and 0.40; "acceptable" and 0–0.20 represents "poor" reliability [58].

Considering that the ICC is influenced by the variability of the measurement, the coefficient of variation (CV) of the data was also calculated using the formula: $CV = (SD \div \bar{X}) \times 100$, where SD is the standard deviation of test-retest data and \bar{X} is the mean of test-retest data. CV measures the relative spread of data and helps to consider ICC results in function of its variability [50,52].

2.4.3 Comparison of different sequences of measurements. To determine how many measurements provided the best reliability metrics, five sequences of measurements were considered for PPT (Fig 1): first measurement only (1), the first two measurements (1–2), the first three measurements (1–2–3), the first three or the first fourth when applicable (1–2–3(4)) and the last two trials including the fourth when applicable (2–3–4). The sequence 2–3–4 was included because some studies suggest that the first measure of PPT tends to be higher [36,37,59], but this remains debated [34]. For TSP, three sequences were evaluated: 3 measurements (1–2–3), the first two (1–2) and the first only (1). For all sequences, relative and absolute within-subject short-term test-retest reliabilities were assessed.

3. Results

3.1 Pain sensitivity outcomes

3.1.1 Missing data. One participant TSP data were excluded of analysis because of a technical problem with the pinprick during the data collection (n = 23). For PPT, twenty-four participants were included in the analysis.

3.1.2 Assumptions validation and data transformation. Two distributions required natural logarithmic (ln) transformation because of non-normality: 1) TSP at the back at retest for sequence 1–2 ($p = 0.02$) and 2) TSP at the hand at test for sequence 1 ($p = 0.03$). Data were distributed normally after the transformation. One or two outliers were detected for 20 and 3 out of 52 data distributions, respectively. To determine if outliers affected the reliability analysis, we calculated the reliability with and without outliers. Considering both analyses provided similar reliability that was not affecting the interpretation of the results, outliers were included. All data distributions were homoscedastic.

3.1.3 Raw PPT and TSP data. Group means and SD for PPT and TSP at test and retest are reported in Table 2. The results of the one-way ANOVA_{RM} showed that there was no statistically significant difference between test and retest occasions for PPT and TSP at all sites for all sequences (all $p > 0.076$ – Table 2).

3.2 Reliability analysis

3.2.1 Absolute reliability for PPT and TSP. Considering that (i) MDC_{gr} is proportional to SEM_{eas} and MDC_{ind} , and that (ii) $\%MDC_{gr}$, $\%SEM_{eas}$ and $\%MDC_{ind}$ results present the same pattern between sequences, only MDC_{gr} and $\%MDC_{gr}$ will be described to facilitate the

Table 2. Means of PPT and TSP at test and retest occasions for each sequence of measurements.

| Outcomes | Sites | Sequences of measurements | Test | | Retest | | Test versus retest | |
|------------------------|-------|---------------------------|--------|--------|--------|--------|--------------------|--------|
| | | | Mean | SD | Mean | SD | (F) | (p) |
| PPT ^a (kPa) | S1 | 1-2-3(4) | 519.50 | 205.61 | 533.48 | 214.72 | 0.608 | 0.443 |
| | | 1-2-3 | 522.87 | 206.88 | 534.07 | 211.79 | 0.351 | 0.559 |
| | | 2-3-4 | 517.20 | 205.77 | 532.58 | 212.41 | 0.548 | 0.466 |
| | | 1-2 | 521.79 | 216.07 | 530.32 | 210.10 | 0.289 | 0.596 |
| | | 1 | 536.77 | 222.51 | 527.45 | 209.86 | 0.248 | 0.623 |
| | LES | 1-2-3(4) | 547.29 | 199.65 | 551.32 | 169.86 | 0.027 | 0.871 |
| | | 1-2-3 | 551.09 | 201.84 | 557.17 | 172.31 | 0.059 | 0.810 |
| | | 2-3-4 | 549.62 | 207.49 | 559.74 | 175.09 | 0.154 | 0.698 |
| | | 1-2 | 549.52 | 196.11 | 548.79 | 174.74 | 0.001 | 0.976 |
| | | 1 | 553.79 | 199.52 | 553.18 | 181.51 | 0.001 | 0.981 |
| | WD | 1-2-3(4) | 371.67 | 149.28 | 355.95 | 123.96 | 1.072 | 0.311 |
| | | 1-2-3 | 370.54 | 149.03 | 358.91 | 129.38 | 0.631 | 0.435 |
| | | 2-3-4 | 365.02 | 155.98 | 357.37 | 130.20 | 0.225 | 0.640 |
| | | 1-2 | 373.63 | 142.29 | 355.61 | 127.02 | 1.285 | 0.269 |
| | | 1 | 381.89 | 141.27 | 359.97 | 134.01 | 1.821 | 0.190 |
| | WF | 1-2-3(4) | 368.55 | 121.92 | 354.45 | 115.34 | 0.916 | 0.348 |
| | | 1-2-3 | 369.47 | 122.57 | 353.77 | 115.16 | 1.089 | 0.307 |
| | | 2-3-4 | 361.41 | 127.11 | 349.83 | 115.80 | 0.509 | 0.483 |
| | | 1-2 | 371.86 | 126.16 | 356.19 | 115.64 | 0.969 | 0.335 |
| | | 1 | 389.94 | 135.22 | 365.73 | 127.69 | 1.518 | 0.230 |
| TSP ^b (NRS) | Hand | 1-2-3 | 2.29 | 1.45 | 2.61 | 1.51 | 3.461 | 0.076 |
| | | 1-2 | 2.21 | 1.43 | 2.47 | 1.52 | 1.731 | 0.202 |
| | | 1 | 2.24** | 1.63** | 2.28** | 1.71** | 0.011* | 0.917* |
| | Back | 1-2-3 | 1.92 | 1.32 | 1.95 | 1.58 | 0.015 | 0.905 |
| | | 1-2 | 1.85** | 1.31** | 1.98** | 1.69** | 0.016* | 0.902* |
| | | 1 | 1.96 | 1.32 | 1.91 | 1.71 | 0.022 | 0.884 |

PPT pressure pain threshold, TSP temporal summation of pain, NRS numeric rating scale, SD standard deviation, S1 spinous process of S1, LES lumbar erector spinae, WD dorsal aspect of the wrist, WF wrist flexors muscles.

* ANOVA were done on natural log-transformation of the data distributions.

** Means and SD of the non-transformed raw data are presented to facilitate interpretation.

F Fisher’s test from a one-way ANOVA_{RM}.

p one-way ANOVA_{RM}.

^a n = 24.

^b n = 23.

<https://doi.org/10.1371/journal.pone.0245278.t002>

presentation of the results. Each metric is detailed in [Table 3](#). In general, differences in the measures of absolute reliability for PPT are small across the sequences and seem to be dependent on the site tested ([Fig 2](#)). For S1, MDC_{gr} varied from 31.10 kPa for sequence 1–2 to 40.71 kPa for sequence 2-3-4 (%MDC_{gr}: 5.91% (1–2) to 7.76% (2-3-4)). For LES, MDC_{gr} ranged from 46.13 kPa (1–2) to 50.56 kPa (1) (%MDC_{gr}: 8.40% (1–2) to 9.14% (1)). For WD, MDC_{gr} varied from 28.71 kPa (1-2-3) to 31.84 kPa (1) (%MDC_{gr}: 7.87% (1-2-3) to 8.58% (2-3-4)). For WF, MDC_{gr} ranged from 28.87 kPa (1-2-3(4)) to 38.52 kPa (1) (%MDC_{gr}: 7.99% (1-2-3(4)) to 10.19% (1)). Sequences 1-2-3(4) and 1-2-3 presented similar results at WD and WF and seem to present a smaller measurement error compared to other sequences regarding absolute reliability. At the low back (S1 and LES), sequence 1–2 presented the lowest MDC_{gr} compared to

Table 3. Within-session reliability parameters for PPT for each site using different sequences of measurements (n = 24).

| Sites PPT | Sequences | ICC (95%CI) | CV (%) | SEM _{cas} | | MDC _{ind} | | MDC _{gr} | |
|-----------|-----------|------------------|--------|--------------------|-------|--------------------|-------|-------------------|-------|
| | | | | kPa | % | kPa | % | kPa | % |
| S1 | 1-2-3(4) | 0.96 (0.90–0.98) | 39.52 | 62.08 | 11.79 | 172.08 | 32.68 | 35.12 | 6.67 |
| | 1-2-3 | 0.95 (0.89–0.98) | 39.21 | 65.49 | 12.39 | 181.54 | 34.35 | 37.06 | 7.01 |
| | 2-3-4 | 0.94 (0.86–0.97) | 39.44 | 71.95 | 13.71 | 199.45 | 38.00 | 40.71 | 7.76 |
| | 1–2 | 0.97 (0.92–0.99) | 40.09 | 54.97 | 10.45 | 152.37 | 28.96 | 31.10 | 5.91 |
| | 1 | 0.91 (0.81–0.96) | 40.22 | 64.83 | 12.18 | 179.90 | 33.77 | 36.68 | 6.89 |
| LES | 1-2-3(4) | 0.89 (0.74–0.95) | 33.39 | 85.21 | 15.51 | 236.20 | 43.00 | 48.21 | 8.78 |
| | 1-2-3 | 0.89 (0.73–0.95) | 33.51 | 86.56 | 15.62 | 239.93 | 43.30 | 48.98 | 8.84 |
| | 2-3-4 | 0.88 (0.73–0.95) | 34.25 | 89.36 | 16.11 | 247.70 | 44.66 | 50.56 | 9.12 |
| | 1–2 | 0.90 (0.76–0.96) | 33.46 | 81.53 | 14.85 | 225.98 | 41.15 | 46.13 | 8.40 |
| | 1 | 0.79 (0.57–0.90) | 34.09 | 89.37 | 16.15 | 247.71 | 44.76 | 50.56 | 9.14 |
| WD | 1-2-3(4) | 0.92 (0.82–0.97) | 37.38 | 52.62 | 14.46 | 145.85 | 40.09 | 29.77 | 8.18 |
| | 1-2-3 | 0.93 (0.84–0.97) | 37.89 | 50.74 | 13.91 | 140.66 | 38.57 | 28.71 | 7.87 |
| | 2-3-4 | 0.92 (0.82–0.97) | 39.37 | 55.92 | 15.48 | 154.99 | 42.91 | 31.64 | 8.76 |
| | 1–2 | 0.91 (0.79–0.96) | 36.68 | 55.07 | 15.10 | 152.66 | 41.87 | 31.16 | 8.55 |
| | 1 | 0.83 (0.65–0.92) | 36.84 | 56.28 | 15.17 | 155.99 | 42.05 | 31.84 | 8.58 |
| WF | 1-2-3(4) | 0.90 (0.77–0.96) | 32.54 | 51.02 | 14.11 | 141.42 | 39.12 | 28.87 | 7.99 |
| | 1-2-3 | 0.89 (0.76–0.95) | 32.61 | 52.12 | 14.41 | 144.47 | 39.95 | 29.49 | 8.15 |
| | 2-3-4 | 0.88 (0.73–0.95) | 33.86 | 56.22 | 15.81 | 155.84 | 43.82 | 31.81 | 8.95 |
| | 1–2 | 0.88 (0.74–0.95) | 32.96 | 55.14 | 15.15 | 152.84 | 41.99 | 31.20 | 8.57 |
| | 1 | 0.73 (0.47–0.87) | 34.59 | 68.08 | 18.02 | 188.71 | 49.94 | 38.52 | 10.19 |

ICC intraclass correlation coefficient, with a 95% confidence interval, CV coefficient of variation, SEM_{cas} standard error of the measurement, MDC_{ind} individual minimal detectable change, MDC_{gr} group minimal detectable change, S1 spinous process of S1, LES lumbar erector spinae, WD dorsal aspect of the wrist, WF wrist flexors muscles.

<https://doi.org/10.1371/journal.pone.0245278.t003>

other sequences although the difference remains minimal. In terms of the different sites, S1 presented the lowest measurement error compared to the three other sites that were very similar. Again, these differences were small.

For TSP, absolute reliability is depicted in Table 4. MDC_{gr} at the hand varied from 0.33 for sequence 1-2-3 to 0.40/10 for sequence 1-2 (%MDC_{gr}: 13.58% (1-2-3) to 17.29% (1-2)). At the back, MDC_{gr} ranged from 0.46 for sequence 1-2-3 to 0.57/10 for sequence 1 (%MDC_{gr}: 23.52% (1-2-3) to 29.22% (1)). An increase in the number of measurements for TSP leads to an increased absolute reliability. %MDC_{gr} was lower at the hand compared to the back. However, although we presented the non-transformed data in Table 4, comparisons with other sequences must be done with caution. For log-transformation data, calculation of %SEM_{cas} and %MDC was not applicable because transformed results do not correspond to the same ratio scale [54,55]. Antilog factor (\times/\div) derived from SEM_{cas} and MDC_{ind} were also calculated and presented in Table 4.

3.2.2 Relative reliability for PPT and TSP. For PPT, almost all ICC values are above 0.80, between 0.80 and 0.97, denoting that almost all sequences presented an excellent relative reliability, except sequence 1 for PPT at LES and WF, for which an acceptable reliability was observed with ICC of 0.79 and 0.73 respectively (Table 3). Sequence 1 presented lower ICCs for all sites, associated to large confidence intervals at LES, WD and WF. As illustrated in Fig 3, ICCs of sequence 1–2 are larger at S1 and LES than others. ICCs for WD and WF were similar among sequences, except for sequence 1 that presents with lower ICCs.

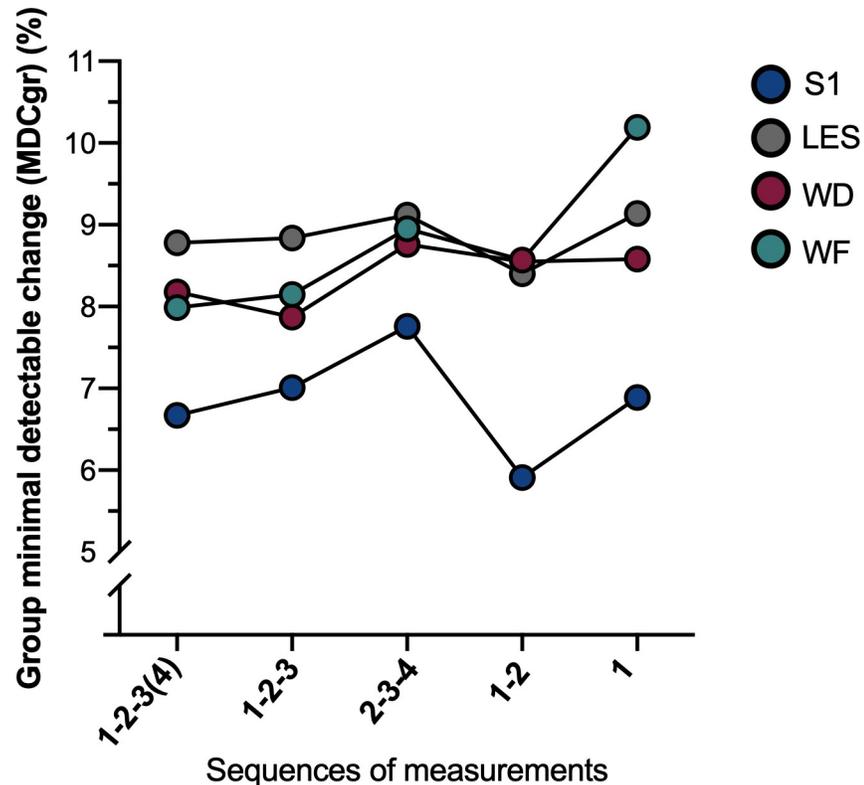


Fig 2. %MDC_{gr} for PPT at all sites for the five sequences of measurements tested.

<https://doi.org/10.1371/journal.pone.0245278.g002>

For TSP, results are depicted in Table 4. Sequences 1–2 and 1-2-3 showed an excellent reliability with ICC ranged between 0.81 and 0.91 and narrow confidence intervals. Visually, larger ICC confidence intervals were present at the back compared to the hand. As illustrated

Table 4. Within-session reliability parameters for TSP for each site using different sequences of measurements (n = 23).

| Sites TSP | Sequences of measurements | ICC (95%CI) | CV (%) | SEM _{cas} | | MDC _{ind} | | MDC _{gr} | |
|-----------|---------------------------|------------------|--------|--------------------|--------|--------------------|--------|-------------------|-------|
| | | | | NRS | % | NRS | % | NRS | % |
| Hand | 1-2-3 | 0.91 (0.79–0.96) | 60.29 | 0.58 | 23.50 | 1.59 | 65.16 | 0.33 | 13.58 |
| | 1–2 | 0.88 (0.71–0.95) | 62.74 | 0.70 | 29.91 | 1.94 | 82.91 | 0.40 | 17.29 |
| | 1* (non-transformed) | 0.73 (0.46–0.88) | 73.09 | 0.86 | 38.03 | 2.38 | 105.39 | 0.50 | 21.98 |
| | 1 (log _n) | 0.78 (0.54–0.90) | NA | 0.15 | NA | 0.41 | NA | 0.09 | NA |
| | 1 (antilog) | | | ×/÷1.16 | | ×/÷1.51 | | | |
| Back | 1-2-3 | 0.83 (0.59–0.93) | 74.47 | 0.79 | 40.70 | 2.18 | 112.81 | 0.46 | 23.52 |
| | 1-2* (non-transformed) | 0.81 (0.55–0.92) | 78.32 | 0.85 | 44.60 | 2.37 | 123.63 | 0.49 | 25.78 |
| | 1–2 (log _n) | 0.81 (0.55–0.92) | NA | 0.17 | NA | 0.47 | NA | 0.10 | NA |
| | 1–2 (antilog) | | | ×/÷1.18 | | ×/÷1.60 | | | |
| | 1 | 0.58 (0.23–0.80) | 78.12 | 0.98 | 50.56% | 2.71 | 140.15 | 0.57 | 29.22 |

ICC intraclass correlation coefficient, with a 95% confidence interval, CV coefficient of variation, SEM_{cas} standard error of the measurement, MDC_{ind} individual minimal detectable change, MDC_{gr} group minimal detectable change, NA: not applicable.

* Reliability outcomes are presented for non-transformed data distributions in order to facilitate interpretation but considerations of these data must be done cautiously because the assumption of the normality of the distribution was violated.

<https://doi.org/10.1371/journal.pone.0245278.t004>

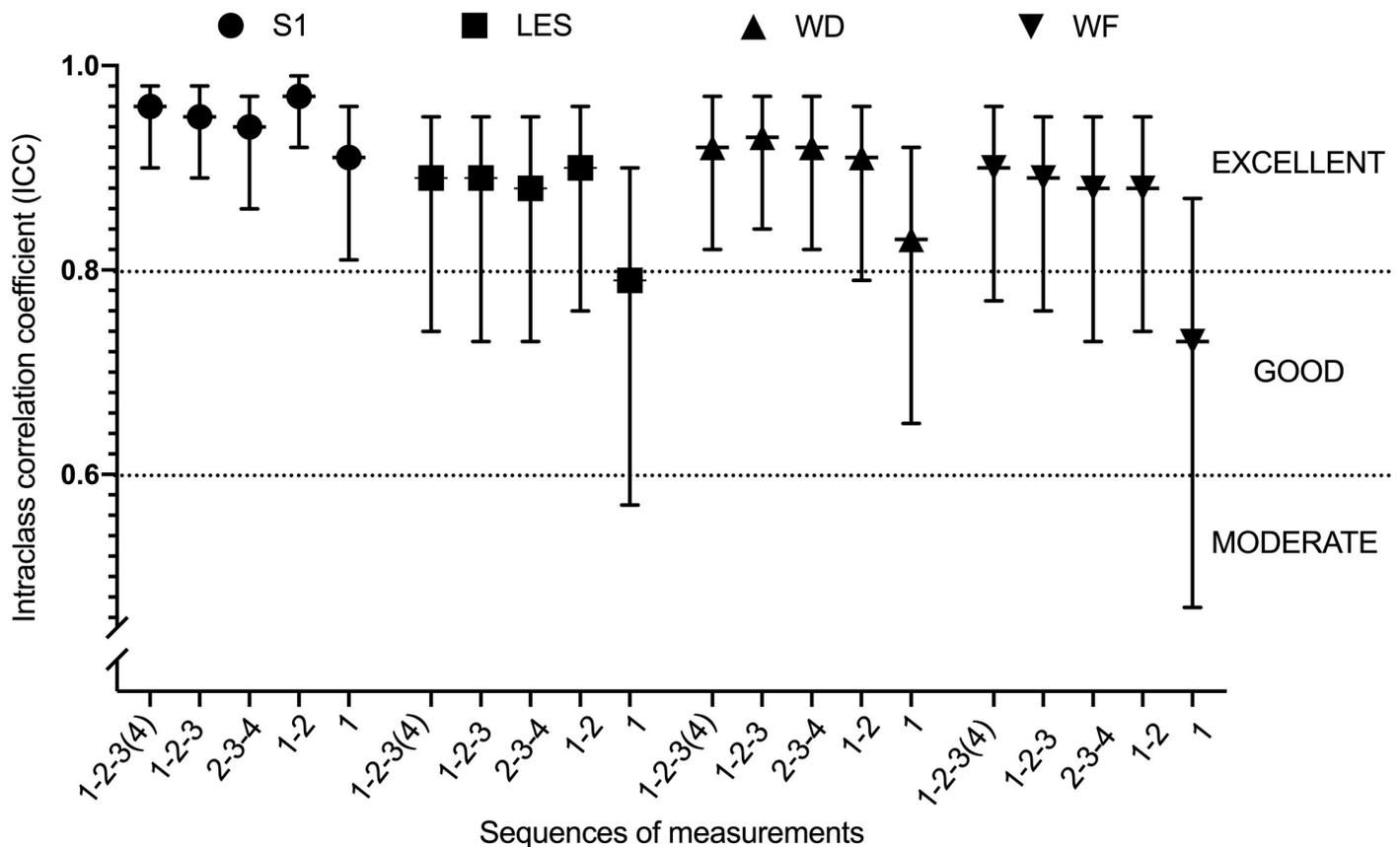


Fig 3. ICC for PPT for all sites by each sequence of measurements.

<https://doi.org/10.1371/journal.pone.0245278.g003>

in Fig 4, there was a large difference between sequence 1 and 1-2-3 for ICC at the two sites, with sequence 1-2-3 presenting larger ICCs compared to sequence 1.

4. Discussion

The objectives of this study were to determine the reliability of PPT and TSP and to determine the sequence of measurements providing the best reliability metrics. The design of the study was established to measure the minimal change exceeding the variability of PPT and TSP technique within a session (20 minutes apart). Our results could be used in future pain research to confirm the presence of ‘true’ hypo- or hyperalgesia (i.e. change over normal variability) for paradigms such as CPM or EIH. In addition, we established that two measures of PPT and three measures of TSP reduced the measurement error and demonstrated an excellent relative reliability.

4.1 PPT reliability

Four studies investigated the intra-rater reliability of PPT at the low back and at the wrist in healthy participants [34–37], but only two of them [34,37] evaluated the within-session test-retest reliability of PPT at the low back. The two other studies based their reliability analysis on the comparison of two or three consecutive measurements from a single block of measurements. Therefore, their results are not suitable in a context of pain modulation evaluation such as CPM/EIH and cannot be directly compared to our results. Balaguier et al. [34] evaluated

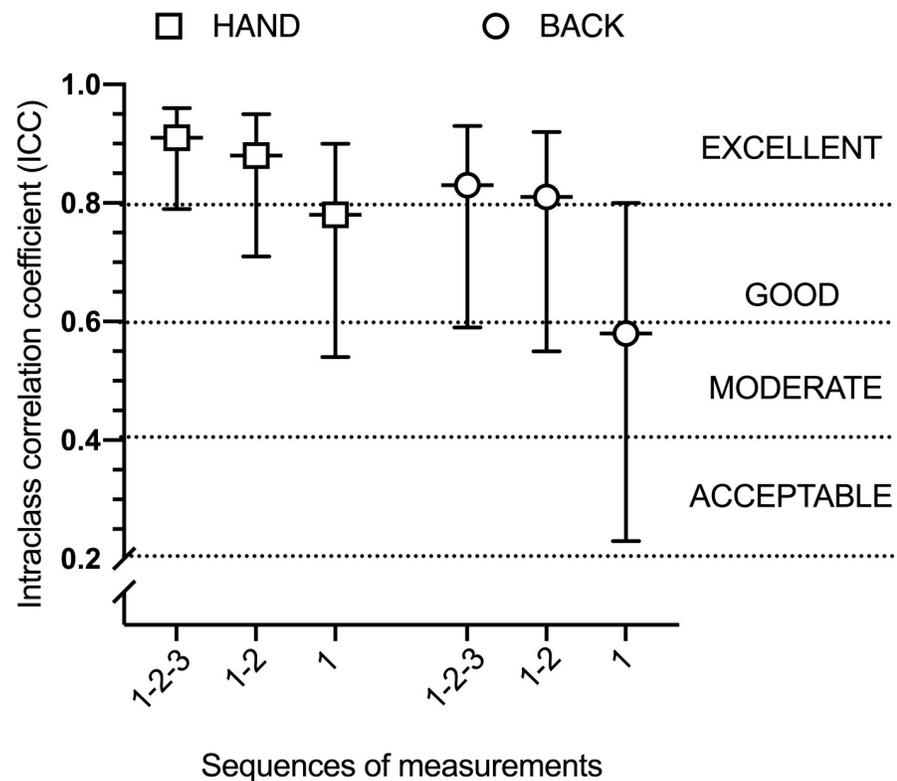


Fig 4. ICC for TSP at the hand and at the back by each sequence of measurements.

<https://doi.org/10.1371/journal.pone.0245278.g004>

absolute reliability at the low back over two sessions separated by one hour and observed a MDC_{ind} ranging from 94 to 253 kPa. These results are similar to our findings for the low back that varied from 152.37 to 247.71 kPa. In addition, our ICCs for PPT are consistent with the literature. One study reported excellent relative reliability (ICCs: 0.86 to 0.99) at 14 anatomical locations at the low back and another study reported good to excellent reliability (ICC: 0.40 to 0.99) for three sites over lumbar *erector spinae* (L1, L3, L5) [37]. Previous results must be interpreted with caution because of the limited sample size ($n = 5$ [37] and $n = 15$ [34]).

4.2 TSP reliability

Only three studies investigated mechanical TSP reliability [60–62] but none assessed the low back area. In two studies [60,62], 5 consecutive series of ten punctuate stimuli were used instead of two blocks of 3 series tested 20-min apart. One study [60] observed a poor reliability for mechanical TSP at the face, hand and foot and another study reported poor to good intrarater reliability at the tongue, face and gingiva [62]. The third study investigated the test-retest reliability of mechanical TSP at the hand within a two-weeks periods and observed an acceptable reliability in younger adults and a moderate to good reliability in older adults [61]. These findings are inconsistent with the excellent reliability observed for mechanical TSP at back and hand sites in our study. This discrepancy can be caused by differences in sites tested, study designs and calculation of TSP. In two of the above-mentioned studies [60,62], TSP reliability was analyzed using the wind-up ratio calculation ($WUR = \frac{\text{mean pain ratings of trains}}{\text{mean pain rating to single stimulus}}$). WUR cannot be calculated if the single pinprick stimulus is rated as non-painful (NRS = 0/10, meaning a null denominator), leading to an undefined division and limiting the number of patients included in the analysis (e.g. up

to ~27% participants excluded of the analysis [60]). We considered that the subtraction method used for TSP also reflect the facilitation process of nociceptive inputs but with the advantage to include more participants into the analysis.

Antilog factors were calculated for two TSP sequences considering that a log-transformation was applied in presence of normality violation. For example, an antilog factor for the MDC_{ind} of $\times/\div 1.60$ for method 1–2 at the back was obtained. This factor is applied on the test score to obtain lower/higher limits representing the level below/above which the retest has to change to be considered as a real change. For example, if a TSP of 2/10 is measured at test, it implies that a $TSP \geq 3.20/10$ (2×1.60) or $TSP \leq 1.25/10$ ($2 \div 1.60$) at retest for the same participant are considered as a true change [45,47,54]. Some methodological limitations derived from the application of the antilog factor, for example, the antilog factor could not be used if TSP at test was rated 0/10. Also, considering that TSP represents ordinal data, interpretation appears to be less precise. Therefore, antilog calculation constitutes an alternative way to analyze reliability in presence of normality violation, but its concrete and clinical applicability remains challenging with TSP data.

4.3 Comparison of sequences of measurements

For PPT, there was substantial heterogeneity in terms of the sequences of measurements producing the least measurement error, which differs in function of the site tested. Absolute reliability with sequence 1–2 was superior for S1 and LES, sequence 1-2-3 for WD and 1-2-3-(4) for WF. However, these differences were minimal across sequences. In addition, all the sequences tested for PPT demonstrated an excellent relative reliability at all sites, except sequence 1 which showed a good relative reliability at LES and WF. Our results are generally in accordance with current literature suggesting that the mean of two or three consecutive measurements are enough to provide reliable PPT [34,36]. However, one study proposed that using only the first measure constitutes an excellent method [34] in opposition to our findings. Some studies observed that the first PPT measurement tends to be significantly higher compared to the subsequent ones and recommended to exclude it from the analysis to reduce the variability [36,37,59]. Our results suggest that excluding the first measurement (2-3-4) did not reduce the measurement error. However, when the first measure was used alone, usually, reliability worsen, meaning that adding only one more measurement seemed sufficient to stabilize the measure. Disparity regarding the first measure can be caused by the fact that some studies did not include familiarization trial prior to PPT evaluation [36,37] or evaluated PPT at different sites than those evaluated in the present study (i.e. *biceps brachii*) [59]. Thus, we recommend to measure PPT at least twice in each block of testing to improve reliability.

For TSP, our results suggest that an increase in the number of measurements leads to an increased absolute reliability. Two and three measurements showed an excellent relative reliability at the hand and the back, in contrast a single measurement demonstrated a good relative reliability at the hand and moderate at the low back. These results may have been influenced by the sample heterogeneity, as represented by high coefficients of variation for TSP. Although some research groups recommend to measure 5 series of ten punctuate stimuli [32,60,62], our results suggest that a sequence of 3 series separated by 30 seconds provides an excellent reliability. Taking 3 measures instead of 5 may reduce the duration of the test and minimize skin irritation due to repeated pinprick stimulations. Therefore, we suggest to take at least three TSP measurement to improve the TSP reliability.

4.4 Applicability of absolute reliability in research

The MDC obtained in our study provides specific threshold to determine if change in PPT or TSP in a within-session design exceeds the variability of these measures for a group (MDC_{gr})

or an individual (MDC_{ind}). In research, MDC can serve as a cut-off value to determine if the pain sensitivity change (as measured by PPT or TSP) exceeds measurement error and can be considered as true hypo- or hyperalgesia following the conditioning stimulus in pain modulation paradigm (e.g. CPM and EIH). For instance, a previous study investigating EIH reported a significant within-session PPT change of 29.78 kPa at the low back in healthy controls following a repetitive lifting task [11]. We determined that the lowest MDC_{gr} at the low back was 46.13 kPa. Thus, a change inferior to this value cannot be considered as a real change despite statistical significance.

Also, the response to conditioning stimulus (e.g. CPM) has been used to stratify healthy and chronic pain participants in function of the change in pain sensitivity (decrease vs. increase) [9,63,64] suggesting a bias toward inhibitory or facilitatory descending control (e.g. anti- vs pro-nociceptive). However, this stratification is usually done without considering the measurement error. This may result in stratifying participants for whom changes remain within the measurement error (i.e. pain sensitivity did not change following conditioned stimulus). Our MDC_{ind} (or $\%MDC_{ind}$) could be used in future studies using a similar design (CPM/EIH) as cut-off values to subgroup participants. For example, an individual change of ~35% is necessary to exceed the $\%MDC_{ind}$ at S1. Considering that this change is large, it questions the validity of stratification methods using PPT.

4.5 Methodological considerations

Our study did not conduct the reliability analysis in different groups for each sequence and this could have underestimated PPT and TSP measurement variability. Also, considering that this study was conducted in a healthy pain-free population, current results are not generalizable to other populations. It is acknowledged that factors such as race/ethnicity, sex (e.g. phases of menstrual cycle [65]) and age could influence pain sensitivity [66,67], but considering that we measured reliability of pain sensitivity in a single session rather than pain sensitivity *per se*, the effect of these factors on our results remain limited. Future studies must be conducted in chronic pain participants, such as chronic low back pain.

4.6 Summary

This study observes that PPT and TSP at back and hand sites have small error of measurement and an excellent relative reliability using a within-session test-retest design. Our results also suggest that at least two PPT and three TSP consecutive measures are needed to optimize reliability and these recommendations may be used in future research and in clinical practice. Our results also provide cut-off values that may be used with pain modulation paradigms such as CPM and EIH to confirm that changes following conditioning stimulus exceed PPT and TSP measurement error (true hypo-/hyperalgesia). Further studies are warranted to investigate the within-session test-retest reliability of these parameters in chronic pain populations.

Author Contributions

Conceptualization: Catherine Mailloux, Louis-David Beaulieu, Timothy H. Wideman, Hugo Massé-Alarie.

Data curation: Catherine Mailloux.

Formal analysis: Catherine Mailloux.

Funding acquisition: Hugo Massé-Alarie.

Investigation: Catherine Mailloux.

Methodology: Catherine Mailloux, Louis-David Beaulieu, Timothy H. Wideman, Hugo Massé-Alarie.

Project administration: Catherine Mailloux, Hugo Massé-Alarie.

Resources: Catherine Mailloux, Hugo Massé-Alarie.

Software: Hugo Massé-Alarie.

Supervision: Louis-David Beaulieu, Timothy H. Wideman, Hugo Massé-Alarie.

Validation: Hugo Massé-Alarie.

Visualization: Hugo Massé-Alarie.

Writing – original draft: Catherine Mailloux, Hugo Massé-Alarie.

Writing – review & editing: Louis-David Beaulieu, Timothy H. Wideman, Hugo Massé-Alarie.

References

1. Moseley GL, Butler DS. Explain pain supercharged: The clinician's manual. First ed: NOIgroup Publications; 2017. 225 p.
2. Ossipov MH, Dussor GO, Porreca F. Central modulation of pain. *J Clin Invest*. 2010; 120(11):3779–87. <https://doi.org/10.1172/JCI43766> PMID: 21041960
3. Starkweather AR, Heineman A, Storey S, Rubia G, Lyon DE, Greenspan J, et al. Methods to measure peripheral and central sensitization using quantitative sensory testing: A focus on individuals with low back pain. *Appl Nurs Res*. 2016; 29:237–41. <https://doi.org/10.1016/j.apnr.2015.03.013> PMID: 26856520
4. Arendt-Nielsen L, Yarnitsky D. Experimental and clinical applications of quantitative sensory testing applied to skin, muscles and viscera. *J Pain*. 2009; 10(6):556–72. <https://doi.org/10.1016/j.jpain.2009.02.002> PMID: 19380256
5. Graven-Nielsen T, Arendt-Nielsen L. Assessment of mechanisms in localized and widespread musculoskeletal pain. *Nat Rev Rheumatol*. 2010; 6(10):599–606. <https://doi.org/10.1038/nrrheum.2010.107> PMID: 20664523
6. Neelapala YVR, Bhagat M, Frey-Law L. Conditioned Pain Modulation in Chronic Low Back Pain: A Systematic Review of Literature. *Clin J Pain*. 2020; 36(2):135–41. <https://doi.org/10.1097/AJP.0000000000000778> PMID: 31764164
7. Correa JB, Costa LO, de Oliveira NT, Sluka KA, Liebano RE. Central sensitization and changes in conditioned pain modulation in people with chronic nonspecific low back pain: a case-control study. *Exp Brain Res*. 2015; 233(8):2391–9. <https://doi.org/10.1007/s00221-015-4309-6> PMID: 25963754
8. Aoyagi K, He J, Nicol AL, Clauw DJ, Kluding PM, Jernigan S, et al. A Subgroup of Chronic Low Back Pain Patients with Central Sensitization. *Clin J Pain*. 2019. <https://doi.org/10.1097/AJP.0000000000000755> PMID: 31408011
9. Rabey M, Poon C, Wray J, Thamajaree C, East R, Slater H. Pro-nociceptive and anti-nociceptive effects of a conditioned pain modulation protocol in participants with chronic low back pain and healthy control subjects. *Man Ther*. 2015; 20(6):763–8. <https://doi.org/10.1016/j.math.2015.02.011> PMID: 25795107
10. Falla D, Gizzi L, Tschapek M, Erlenwein J, Petzke F. Reduced task-induced variations in the distribution of activity across back muscle regions in individuals with low back pain. *Pain*. 2014; 155(5):944–53. <https://doi.org/10.1016/j.pain.2014.01.027> PMID: 24502841
11. Kuithan P, Heneghan NR, Rushton A, Sanderson A, Falla D. Lack of Exercise-Induced Hypoalgesia to Repetitive Back Movement in People with Chronic Low Back Pain. *Pain Pract*. 2019. <https://doi.org/10.1111/papr.12804> PMID: 31187932
12. Yarnitsky D, Bouhassira D, Drewes AM, Fillingim RB, Granot M, Hansson P, et al. Recommendations on practice of conditioned pain modulation (CPM) testing. *European Journal of Pain*. 2015; 19(6):805–6. <https://doi.org/10.1002/ejp.605> PMID: 25330039
13. Naugle KM, Fillingim RB, Riley JL, 3rd. A meta-analytic review of the hypoalgesic effects of exercise. *J Pain*. 2012; 13(12):1139–50. <https://doi.org/10.1016/j.jpain.2012.09.006> PMID: 23141188

14. Sprenger C, Bingel U, Buchel C. Treating pain with pain: supraspinal mechanisms of endogenous analgesia elicited by heterotopic noxious conditioning stimulation. *Pain*. 2011; 152(2):428–39. <https://doi.org/10.1016/j.pain.2010.11.018> PMID: 21196078
15. Yarnitsky D, Granot M, Granovsky Y. Pain modulation profile and pain therapy: between pro- and anti-nociception. *Pain*. 2014; 155(4):663–5. <https://doi.org/10.1016/j.pain.2013.11.005> PMID: 24269491
16. Pud D, Granovsky Y, Yarnitsky D. The methodology of experimentally induced diffuse noxious inhibitory control (DNIC)-like effect in humans. *Pain*. 2009; 144(1–2):16–9. <https://doi.org/10.1016/j.pain.2009.02.015> PMID: 19359095
17. Gebhart GF. Descending modulation of pain. *Neurosci Biobehav Rev*. 2004; 27(8):729–37. <https://doi.org/10.1016/j.neubiorev.2003.11.008> PMID: 15019423
18. Da Silva Santos R, Galdino G. Endogenous systems involved in exercise-induced analgesia. *J Physiol Pharmacol*. 2018; 69(1):3–13. <https://doi.org/10.26402/jpp.2018.1.01> PMID: 29769416
19. Rice D, Nijs J, Kosek E, Wideman T, Hasenbring MI, Koltyn K, et al. Exercise-Induced Hypoalgesia in Pain-Free and Chronic Pain Populations: State of the Art and Future Directions. *J Pain*. 2019. <https://doi.org/10.1016/j.jpain.2019.03.005> PMID: 30904519
20. Sluka KA, Frey-Law L, Hoeger Bement M. Exercise-induced pain and analgesia? Underlying mechanisms and clinical translation. *Pain*. 2018; 159 Suppl 1:S91–S7. <https://doi.org/10.1097/j.pain.0000000000001235> PMID: 30113953
21. Paungmal A, Joseph LH, Punturee K, Sitalertpisan P, Pirunsan U, Uthaihpun S. Immediate Effects of Core Stabilization Exercise on beta-Endorphin and Cortisol Levels Among Patients With Chronic Non-specific Low Back Pain: A Randomized Crossover Design. *J Manipulative Physiol Ther*. 2018; 41(3):181–8. <https://doi.org/10.1016/j.jmpt.2018.01.002> PMID: 29459120
22. Crombie KM, Brellenthin AG, Hillard CJ, Koltyn KF. Endocannabinoid and Opioid System Interactions in Exercise-Induced Hypoalgesia. *Pain Med*. 2018; 19(1):118–23. <https://doi.org/10.1093/pm/pnx058> PMID: 28387833
23. Tour J, Lofgren M, Mannerkorpi K, Gerdle B, Larsson A, Palstam A, et al. Gene-to-gene interactions regulate endogenous pain modulation in fibromyalgia patients and healthy controls-antagonistic effects between opioid and serotonin-related genes. *Pain*. 2017; 158(7):1194–203. <https://doi.org/10.1097/j.pain.0000000000000896> PMID: 28282362
24. Koltyn KF, Brellenthin AG, Cook DB, Sehgal N, Hillard C. Mechanisms of exercise-induced hypoalgesia. *J Pain*. 2014; 15(12):1294–304. <https://doi.org/10.1016/j.jpain.2014.09.006> PMID: 25261342
25. Yarnitsky D, Arendt-Nielsen L, Bouhassira D, Edwards RR, Fillingim RB, Granot M, et al. Recommendations on terminology and practice of psychophysical DNIC testing. *Eur J Pain*. 2010; 14(4):339. <https://doi.org/10.1016/j.ejpain.2010.02.004> PMID: 20227310
26. Kennedy DL, Kemp HI, Ridout D, Yarnitsky D, Rice AS. Reliability of conditioned pain modulation: a systematic review. *Pain*. 2016; 157(11):2410–9. <https://doi.org/10.1097/j.pain.0000000000000689> PMID: 27559835
27. Coronado RA, Bialosky JE, Robinsen ME, George SZ. Pain sensitivity subgroups in individuals with spine pain: potential relevance to short-term clinical outcome. *Physical Therapy*. 2014; 94(8):1111–22. <https://doi.org/10.2522/ptj.20130372> PMID: 24764070
28. Nahman-Averbuch H, Yarnitsky D, Granovsky Y, Gerber E, Dagul P, Granot M. The role of stimulation parameters on the conditioned pain modulation response. *Scand J Pain*. 2013; 4(1):10–4. <https://doi.org/10.1016/j.sjpain.2012.08.001> PMID: 29913877
29. Marchand S, Arsenault P. Spatial summation for pain perception: interaction of inhibitory and excitatory mechanisms. *Pain*. 2002; 95:201–6. [https://doi.org/10.1016/s0304-3959\(01\)00399-2](https://doi.org/10.1016/s0304-3959(01)00399-2) PMID: 11839419
30. Tousignant-Laflamme Y, Page S, Goffaux P, Marchand S. An experimental model to measure excitatory and inhibitory pain mechanisms in humans. *Brain Res*. 2008; 1230:73–9. <https://doi.org/10.1016/j.brainres.2008.06.120> PMID: 18652808
31. Price DD, Hu JW, Dubner R, Gracely RH. Peripheral suppression of first pain and central summation of second pain evoked by noxious heat pulses. *Pain*. 1977; 3:57–68. [https://doi.org/10.1016/0304-3959\(77\)90035-5](https://doi.org/10.1016/0304-3959(77)90035-5) PMID: 876667
32. Rolke R, Baron R, Maier C, Tolle TR, Treede RD, Beyer A, et al. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. *Pain*. 2006; 123(3):231–43. <https://doi.org/10.1016/j.pain.2006.01.041> PMID: 16697110
33. Herrero JF, Laird JMA, Lopez Garcia JA. Wind-up of spinal cord neurones and pain sensation: much ado about something? *Progress in neurobiology*. 2000; 61:169–203. [https://doi.org/10.1016/s0301-0082\(99\)00051-9](https://doi.org/10.1016/s0301-0082(99)00051-9) PMID: 10704997

34. Balaguier R, Madeleine P, Vuillerme N. Is One Trial Sufficient to Obtain Excellent Pressure Pain Threshold Reliability in the Low Back of Asymptomatic Individuals? A Test-Retest Study. *PLoS One*. 2016; 11(8):e0160866. <https://doi.org/10.1371/journal.pone.0160866> PMID: 27513474
35. Waller R, Straker L, O'Sullivan P, Sterling M, Smith A. Reliability of pressure pain threshold testing in healthy pain free young adults. *Scand J Pain*. 2015; 9(1):38–41. <https://doi.org/10.1016/j.sjpain.2015.05.004> PMID: 29911647
36. Lacourt TE, Houtveen JH, van Doornen LJP. Experimental pressure-pain assessments: Test-retest reliability, convergence and dimensionality. *Scand J Pain*. 2012; 3(1):31–7. <https://doi.org/10.1016/j.sjpain.2011.10.003> PMID: 29913770
37. Farasyn A, Meeusen R. Pressure pain thresholds in healthy subjects: influence of physical activity, history of lower back pain factors and the use of endermology as a placebo-like treatment. *Journal of Bodywork and Movement Therapies*. 2003; 7(1):53–61.
38. Gierthmuhlen J, Enax-Krumova EK, Attal N, Bouhassira D, Cruccu G, Finnerup NB, et al. Who is healthy? Aspects to consider when including healthy volunteers in QST—based studies—a consensus statement by the EUROPAIN and NEURO-PAIN consortia. *Pain*. 2015; 156(11):2203–11. <https://doi.org/10.1097/j.pain.0000000000000227> PMID: 26075963
39. Riviere F, Widad FZ, Speyer E, Erpelding ML, Escalon H, Vuillemin A. Reliability and validity of the French version of the global physical activity questionnaire. *J Sport Health Sci*. 2018; 7(3):339–45. <https://doi.org/10.1016/j.jshs.2016.08.004> PMID: 30356654
40. Vaegter HB, Handberg G, Emmeluth C, Graven-Nielsen T. Preoperative Hypoalgesia After Cold Pressor Test and Aerobic Exercise is Associated With Pain Relief 6 Months After Total Knee Replacement. *Clin J Pain*. 2017; 33(6):475–84. <https://doi.org/10.1097/AJP.0000000000000428> PMID: 27526332
41. Vaegter HB, Dorge DB, Schmidt KS, Jensen AH, Graven-Nielsen T. Test-Retest Reliability of Exercise-Induced Hypoalgesia After Aerobic Exercise. *Pain Med*. 2018; 19(11):2212–22. <https://doi.org/10.1093/pm/pny009> PMID: 29425326
42. Vaegter HB, Lyng KD, Yttereng FW, Christensen MH, Sorensen MB, Graven-Nielsen T. Exercise-Induced Hypoalgesia After Isometric Wall Squat Exercise: A Test-Retest Reliability Study. *Pain Med*. 2019; 20(1):129–37. <https://doi.org/10.1093/pm/pny087> PMID: 29788440
43. Gehling J, Mainka T, Vollert J, Pogatzki-Zahn EM, Maier C, Enax-Krumova EK. Short-term test-retest reliability of conditioned pain modulation using the cold-heat-pain method in healthy subjects and its correlation to parameters of standardized quantitative sensory testing. *BMC Neurol*. 2016; 16:125. <https://doi.org/10.1186/s12883-016-0650-z> PMID: 27495743
44. Anderson RJ, Craggs JG, Bialosky JE, Bishop MD, George SZ, Staud R, et al. Temporal summation of second pain: variability in responses to a fixed protocol. *Eur J Pain*. 2013; 17(1):67–74. <https://doi.org/10.1002/j.1532-2149.2012.00190.x> PMID: 22899549
45. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (Reliability) in variables relevant to sports medicine. *Sports Med*. 1998; 26(4):217–38. <https://doi.org/10.2165/00007256-199826040-00002> PMID: 9820922
46. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*. 2005; 19(1):231–40. <https://doi.org/10.1519/15184.1> PMID: 15705040
47. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*. 1986:307–10. PMID: 2868172
48. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies*. 2010; 47(8):931–6.
49. Damron LA, Dearth DJ, Hoffman RL, Clark BC. Quantification of the corticospinal silent period evoked via transcranial magnetic stimulation. *J Neurosci Methods*. 2008; 173(1):121–8. <https://doi.org/10.1016/j.jneumeth.2008.06.001> PMID: 18588914
50. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med*. 2000; 1(30):1–15. <https://doi.org/10.2165/00007256-200030010-00001> PMID: 10907753
51. Harvill LM. Standard error of measurement. *Instructional Topics in Educational Measurement*. 1991:33–41.
52. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. *Am J Phys Med Rehabil*. 2005; 84(9):719–23. <https://doi.org/10.1097/01.phm.0000176452.17771.20> PMID: 16141752
53. Beckerman H, Roebroek ME, Lankhorst GJ, B J.G., Bezemer PD, Verbreek ALM. Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research*. 2001; 10:571–8. <https://doi.org/10.1023/a:1013138911638> PMID: 11822790

54. Beaulieu LD, Masse-Alarie H, Ribot-Ciscar E, Schneider C. Reliability of lower limb transcranial magnetic stimulation outcomes in the ipsi- and contralesional hemispheres of adults with chronic stroke. *Clin Neurophysiol.* 2017; 128(7):1290–8. <https://doi.org/10.1016/j.clinph.2017.04.021> PMID: 28549277
55. Schambra HM, Ogden RT, Martinez-Hernandez IE, Lin X, Chang YB, Rahman A, et al. The reliability of repeated TMS measures in older adults and in patients with subacute and chronic stroke. *Front Cell Neurosci.* 2015; 9:335. <https://doi.org/10.3389/fncel.2015.00335> PMID: 26388729
56. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007; 60(1):34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012> PMID: 17161752
57. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin.* 1979; 86(2):420–8. <https://doi.org/10.1037//0033-2909.86.2.420> PMID: 18839484
58. Portney LG, Watkins MP. *Foundations of clinical research applications to practice.* 3rd ed. Upper Saddle River (New Jersey): Pearson Education Inc.; 2009. 912 p.
59. Nussbaum EL, Downes L. Reliability of clinical pressure-pain algometric measurements obtained on consecutive days. *Physical Therapy.* 1998; 78(2).
60. Geber C, Klein T, Azad S, Birklein F, Gierthmuhlen J, Hugel V, et al. Test-retest and interobserver reliability of quantitative sensory testing according to the protocol of the German Research Network on Neuropathic Pain (DFNS): a multi-centre study. *Pain.* 2011; 152(3):548–56. <https://doi.org/10.1016/j.pain.2010.11.013> PMID: 21237569
61. Ohlman T, Miller L, Naugle K. (169) Comparison of Temporal Stability of Conditioned Pain Modulation and Temporal Summation of Pain in Healthy Older and Younger Adults. *The Journal of Pain.* 2019; 20(4).
62. Pigg M, Baad-Hansen L, Svensson P, Drangsholt M, List T. Reliability of intraoral quantitative sensory testing (QST). *Pain.* 2010; 148(2):220–6. <https://doi.org/10.1016/j.pain.2009.10.024> PMID: 20022428
63. Fingleton C, Smart KM, Doody CM. Exercise-induced Hypoalgesia in People With Knee Osteoarthritis With Normal and Abnormal Conditioned Pain Modulation. *Clin J Pain.* 2017; 33(5):395–404. <https://doi.org/10.1097/AJP.0000000000000418> PMID: 27518487
64. O'Neill S, Manniche C, Graven-Nielsen T, Arendt-Nielsen L. Association between a composite score of pain sensitivity and clinical parameters in low-back pain. *Clin J Pain.* 2014; 30(10):831–8. <https://doi.org/10.1097/AJP.000000000000042> PMID: 24121529
65. Pogatzki-Zahn EM, Drescher C, Englbrecht JS, Klein T, Magerl W, Zahn PK. Progesterone relates to enhanced incisional acute pain and pinprick hyperalgesia in the luteal phase of female volunteers. *Pain.* 2019; 160(8):1781–93. <https://doi.org/10.1097/j.pain.0000000000001561> PMID: 31335647
66. Ostrom C, Bair E, Maixner W, Dubner R, Fillingim RB, Ohrbach R, et al. Demographic Predictors of Pain Sensitivity: Results From the OPERA Study. *J Pain.* 2017; 18(3):295–307. <https://doi.org/10.1016/j.jpain.2016.10.018> PMID: 27884689
67. Riley JL 3rd, Cruz-Almeida Y, Glover TL, King CD, Goodin BR, Sibille KT, et al. Age and race effects on pain sensitivity and modulation among middle-aged and older adults. *J Pain.* 2014; 15(3):272–82. <https://doi.org/10.1016/j.jpain.2013.10.015> PMID: 24239561