

RESEARCH ARTICLE

# Is One Trial Sufficient to Obtain Excellent Pressure Pain Threshold Reliability in the Low Back of Asymptomatic Individuals? A Test-Retest Study

Romain Balaguier<sup>1,2</sup>, Pascal Madeleine<sup>2</sup>, Nicolas Vuillerme<sup>1,2,3</sup>\*

**1** Univ. Grenoble-Alpes, EA AGEIS, Grenoble, France, **2** Physical Activity and Human Performance group—SMI, Dept. of Health Science and Technology, Aalborg University, Aalborg, Denmark, **3** Institut Universitaire de France, Paris, France

✉ These authors contributed equally to this work.

\* [nicolas.vuillerme@agim.eu](mailto:nicolas.vuillerme@agim.eu)



OPEN ACCESS

**Citation:** Balaguier R, Madeleine P, Vuillerme N (2016) Is One Trial Sufficient to Obtain Excellent Pressure Pain Threshold Reliability in the Low Back of Asymptomatic Individuals? A Test-Retest Study. PLoS ONE 11(8): e0160866. doi:10.1371/journal.pone.0160866

**Editor:** Neil R. Smalheiser, University of Illinois-Chicago, UNITED STATES

**Received:** December 15, 2015

**Accepted:** July 26, 2016

**Published:** August 11, 2016

**Copyright:** © 2016 Balaguier 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:** With regard to data availability, ethical restrictions prevent deposition in a public repository. Data will be made available from Univ. Grenoble Alpes for researchers who meet the criteria for access to confidential data. Requests for data access should be made to the corresponding author (Dr. Nicolas Vuillerme).

**Funding:** This joint PhD project is financed by a grant from the French Ministry of Higher Education and Research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

The assessment of pressure pain threshold (PPT) provides a quantitative value related to the mechanical sensitivity to pain of deep structures. Although excellent reliability of PPT has been reported in numerous anatomical locations, its absolute and relative reliability in the lower back region remains to be determined. Because of the high prevalence of low back pain in the general population and because low back pain is one of the leading causes of disability in industrialized countries, assessing pressure pain thresholds over the low back is particularly of interest. The purpose of this study was (1) to evaluate the intra- and inter- absolute and relative reliability of PPT within 14 locations covering the low back region of asymptomatic individuals and (2) to determine the number of trial required to ensure reliable PPT measurements. Fifteen asymptomatic subjects were included in this study. PPTs were assessed among 14 anatomical locations in the low back region over two sessions separated by one hour interval. For the two sessions, three PPT assessments were performed on each location. Reliability was assessed computing intraclass correlation coefficients (ICC), standard error of measurement (SEM) and minimum detectable change (MDC) for all possible combinations between trials and sessions. Bland-Altman plots were also generated to assess potential bias in the dataset. Relative reliability for both intra- and inter- session was almost perfect with ICC ranged from 0.85 to 0.99. With respect to the intra-session, no statistical difference was reported for ICCs and SEM regardless of the conducted comparisons between trials. Conversely, for inter-session, ICCs and SEM values were significantly larger when two consecutive PPT measurements were used for data analysis. No significant difference was observed for the comparison between two consecutive measurements and three measurements. Excellent relative and absolute reliabilities were reported for both intra- and inter-session. Reliable measurements can be equally achieved when using the mean of two or three consecutive PPT measurements, as usually proposed in the literature, or with only the first one. Although reliability was almost perfect

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

regardless of the conducted comparison between PPT assessments, our results suggest using two consecutive measurements to obtain higher short term absolute reliability.

## Introduction

Pain is defined as an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or describe in terms of such damage [1]. According to the American Pain Society [2], pain is the fifth vital sign of medical examination. Pressure algometry (PA) performed with a handheld algometer is a method increasingly used since the 80s to assess mechanical pain sensitivity in different anatomical regions. When it is applied perpendicularly to the skin, the algometer creates a mechanical painful stimulation by activating group III and group IV muscle nociceptors [3]. Through pressure pain thresholds (PPT), PA provides a quantitative value related to deep structures sensitivity allowing clinicians or researchers to make comparison over time. In case of musculoskeletal pain, as recently proposed in a literature review by Arendt-Nielsen and Yarnitsky [4], PA seems particularly relevant to compare pain over time or between various normal, affected or treated anatomical regions.

It has been reported that pressure pain sensitivity is different between individual muscles [5] and also non uniformly distributed between muscle belly and tendons of a same muscle [6–9]. Thus, according to Anderssen and colleagues [6], the assessment of pain sensitivity in two adjacent sites can lead to two significantly different PPT's values. This difference could be explained by a change in muscle thickness and density of nociceptors. However, no difference are observed when PPT are assessed bilaterally over homologous body locations [5,10]. Among all the different anatomical locations, the low back region is particularly of interest for PPT's measurements since 70% of the population will experience Low Back Pain (LBP) at least once in his lifetime [11] and because LBP is often reported in relation to work related musculoskeletal disorders [12], disability and sickness absence from work [13,14]. The assessment of PPT can be used as a method to diagnose and monitor the effectiveness of various treatments or interventions over the lower back region [15–17].

According to a literature review by Arendt-Nielsen and Yarnitsky [4], PA seems relevant to compare pain sensitivity over time or between various normal, affected, or treated anatomical locations. In numerous studies, PA reported good to excellent intra- and inter- reliability to assess pain sensitivity in the low back [16,18–21]. Mokkink and colleagues [22] have defined relative reliability as the extent to which scores for subjects who have not changed are the same for repeated measurements, in our study assessed by one examiner on two different occasions. Relative reliability is commonly quantified using intraclass coefficient correlation (ICC) [23]. Absolute reliability also called “agreement” or “absolute measurement error” is defined as how close the score on repeated measures are [24] and it is quantified using standard error of measurement (SEM). Interestingly, PPT reliability studies have (1) generally assessed only two or four locations over the low back region and/or (2) assessed PPT's reliability only unilaterally. Further investigations are therefore needed to ensure that PA is a reliable method to assess PPT in numerous locations covering the low back region of asymptomatic individuals.

The purpose of this study was (1) to evaluate the intra- and inter- absolute and relative reliability of PPT within 14 locations covering the low back region of young asymptomatic individuals and (2) to determine the number of trials required to ensure reliable PPT assessments.

## Materials and Methods

### Subjects

Fifteen asymptomatic subjects (8 women and 7 men), described in [Table 1](#), volunteered to participate in this study. The subjects were recruited within the Grenoble community and consisted of students (11) and newly-hired workers (4). Inclusion criteria were being aged to 18 to 55 years, no musculoskeletal pain in the low back during the last week, no previous injury or/and surgery in the low back region and no pregnancy. This study was conducted in accordance with the Declaration of Helsinki and was approved by the national ethics committee (French society for independent-living technologies and gerontechnology). Subjects gave their informed written consent to the experimental procedure.

### Experimental protocol

A Somedic Algometer (Type 2, Sollentuna, Sweden) with a probe size of 1 cm<sup>2</sup> and calibrated before each session was used to assess PPT over two sessions separated by one hour and lasting approx. 30 minutes. The pressure was applied (1) by a single examiner, (2) over 14 anatomical locations in the lower back region with 7 locations on each side of the lumbar spinal processes L1-L5 and (3) at a rate of 30 kPa/s in line with previous studies. To avoid tissue injury [26–28], a 1 minute interval was observed between two consecutive PPT assessments over the same location to avoid temporal sensitization [29].

Subjects lying comfortably in a prone position were asked to press a button that locks the algometer when the pressure became painful. Then, the examiner noted the pressure indicated on the algometer display corresponding to the PPT. As in numerous studies [30–32] a training PPT measurement was realized prior recordings on the tibialis anterior [33], a remote site from the low back.

### Procedure to mark the 14 anatomical locations

After palpation, the examiner placed two marks at the level of the first (L1) and fifth (L5) vertebrae spinal processes and measures the distance between these two locations (d1). This distance allows the examiner to select one paper grid with 14 anatomical locations among 8 grids specially designed according to the average L1-L5 distance reported earlier [32–33]. Once selected, the examiner aligns the grid with the L1 and L5 marks over the skin and start the experiment.

To design these grids, we calculated d2, corresponding to the quarter of the distance L1-L5. A first column of 5 points was placed bilaterally at the distance (d2) from a fictive line joining L1 to L5. Then, a second column of 2 points was set bilaterally at 2 times the distance (d2) of L2 and L3 ([Fig 1](#)).

### Data analysis

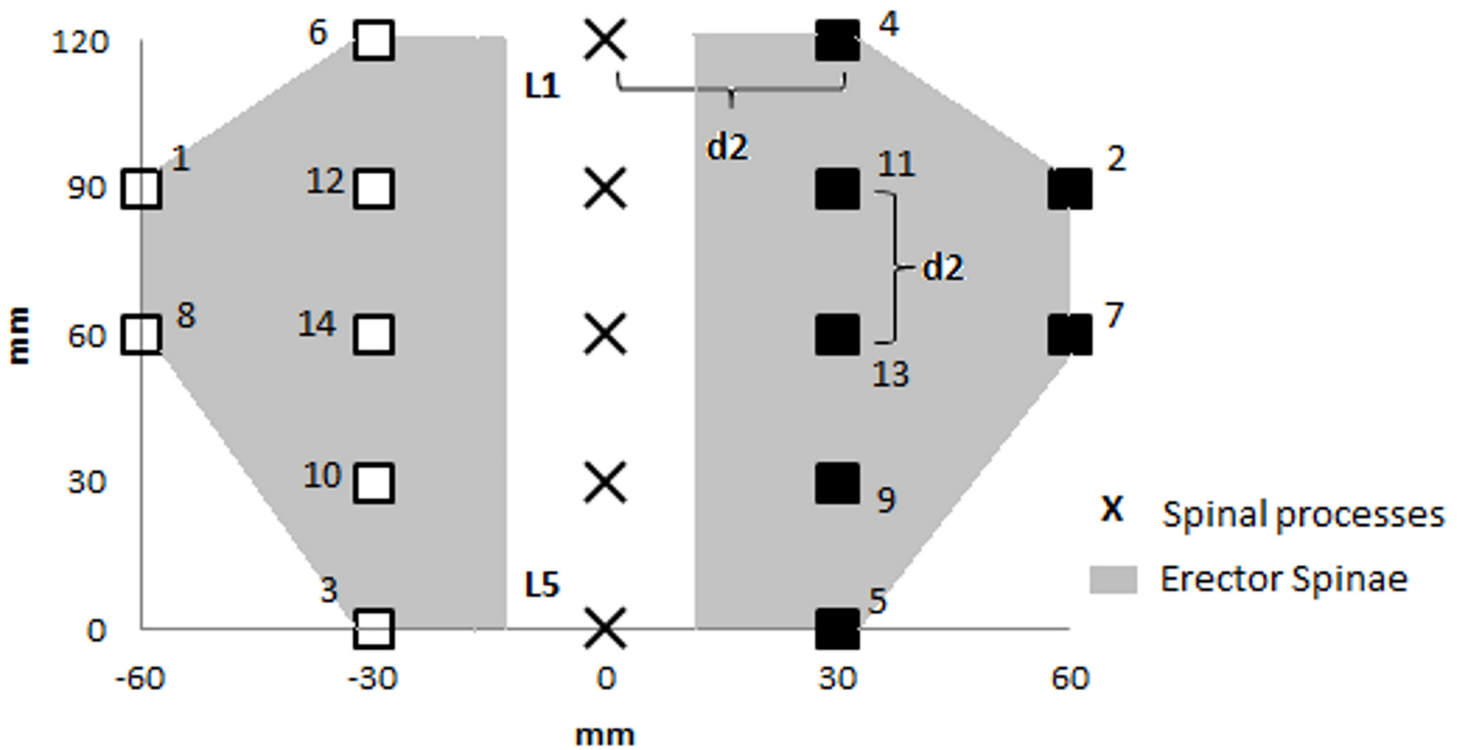
PPT measurements were found to be normally distributed (Shapiro-Wilk normality test).

**Table 1. Characteristics of the subjects (n = 15, 8 women and 7 men).**

Characteristics	Mean (SD)
Age (years)	27.2 (10.2)
Height (cm)	170.8 (11.4)
Weight (kg)	68.9 (15.6)
BMI (kg/m <sup>2</sup> )	23.3 (2.5)

doi:10.1371/journal.pone.0160866.t001

### Low Back recording grid



**Fig 1. Schematic representation of the low back PPT recording grid of the left (blank square) and right (black squares) erector spinae muscles.** The spinal processes of first (L1) to fifth (L5) vertebrae are represented by black crosses.

doi:10.1371/journal.pone.0160866.g001

On the one hand, the results of the first session were analyzed using a repeated measure of variance (ANOVA) to investigate the intra-session reliability, followed by Tukey post-hoc test to highlight differences between trials [34]. The relative and absolute reliability across the trials 1-2-3 were computed using ICC, SEM and minimum detectable change (MDC). The relative reliability was evaluated by calculating a 2-way fixed ICC<sub>2,1</sub> (for absolute agreement). Reliability coefficients (i.e. ICC values) were interpreted according to Landis and Koch [35] in which an ICC between 0.00–0.20 is considered poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial, and 0.81–1.00 is almost perfect. The SEM expressed in the same unit as pain sensitivity (kPa) quantifies the precision of PPT measurements of individual subjects [23,36]. The SEM was calculated as  $SD \sqrt{1 - ICC}$  where SD is the standard deviation of the scores from all subjects and ICC the relative reliability [36]. MDC calculated as  $SEM \times 1.96 \times \sqrt{2}$  provides information on the thresholds required to be confident that a difference can be considered as “real” [36]. Mean, standard deviation, ICC<sub>2,1</sub>, SEM, MDC and limits of agreement (LOA) values were calculated for the two sessions to investigate the inter-session reliability in relation to the following three comparisons [16, 37–39]:

1. trial 1 from session 1 *versus* trial 1 from session 2;
2. the mean of trials 1 and 2 (trials 1–2) from session 1 *versus* the mean of trials 1 and 2 (trials 1–2) from session 2;

- the mean of trials 1,2 and 3 (trials 1-2-3) from session 1 *versus* the mean of trials 1, 2 and 3 (trials 1-2-3) from session 2.

Furthermore, Bland and Altman plots of the differences between trials against their mean and LOA were used to assess the magnitude of disagreement between trials of the 2 sessions. Of note, a difference between trials outside the LOA can be considered as a real change [36].

Finally, one way ANOVA (with the number of trials as within-subject factor) followed by Tukey post-hoc test for pair-wise comparison was performed to compare reliability values (ICC and SEM) of the two sessions.

## Results

### Intra-session reliability

**Relative reliability of PPT in the low back.** As mentioned in Table 2, with values ranged from 0.85 to 0.99, the ICCs of the 14 anatomical locations (Table 2) and of the left, right and overall low-back ( $P_{\text{left}}$ ,  $P_{\text{right}}$ ,  $P_{\text{all}}$ ) were almost perfect regardless of the conducted comparison (trials 1-2-3).

**Absolute reliability of PPT in the low back.** Table 2 also reports that absolute reliability (*i.e.* SEM) remained non statistically different for all possible combinations. SEM values ranged from 26 to 91 kPa.

**Number of trials to ensure reliable measurements in the low back.** The mean PPT values at each anatomical locations were not significantly different between trials regardless of the three conducted comparisons (trial 1 *versus* trial 2, trial 1 *versus* trial 3, trial 2 *versus* trial 3), the p-values were ranged from 0.7220 to 1.000 (Table 3). Concerning the left, right and overall low-back ( $P_{\text{left}}$ ,  $P_{\text{right}}$ ,  $P_{\text{all}}$ ), p-values ranged from 0.9960 to 0.9995.

The comparison of means of ICC regardless of the conducted comparisons between trials (Table 2) showed a statistical difference for Trials 2–3 *versus* Trials 1–3 ( $p = 0.0338$ ). The same analysis for SEM further showed no significant difference between trials.

### Inter-session reliability

**Relative reliability of PPT in the low back.** The ICCs of the 14 locations were almost perfect regardless of the conducted comparison (Table 4 and Fig 2). ICC values ranged from 0.86 to 0.99.

The visual analysis of Bland and Altman's plots suggested no difference in PPT values between sessions because (1) zero was included in the 95% confidence interval and (2) all the subjects were inside the limits of agreement (Fig 3). Furthermore, this visual analysis also suggested narrowed LOA for the association trials 1–2 and trials 1-2-3 compared to the plot of the first trial.

**Number of trials to ensure reliable measurements in the low back.** The mean PPT values at each pressure pain location between sessions 1 and 2 were not significantly different regardless of the following three comparisons: (1) trial 1 from session 1 *versus* trial 1 from session 2, (2) the mean of trials 1 and 2 (trials 1–2) from session 1 *versus* the mean of trials 1 and 2 (trials 1–2) from session 2, (3) the mean of trials 1, 2 and 3 (trials 1-2-3) from session 1 *versus* the mean of trials 1, 2 and 3 (trials 1-2-3) from session 2, the p-values were ranged from 0.4137 to 0.9974.

When two consecutive measurements (Trials 1–2 or Trials 2–3) or all trials were used to calculate subjects' relative and absolute reliability, ICC and SEM values were significantly higher than when the first trial was used. Conversely, no statistical difference was observed between the two first consecutive trials and all trials, or between the two last consecutive trials and all

**Table 2. Intraclass correlation coefficients (ICC), standard error of measurement (SEM) and minimum detectable change (MDC) for pressure pain thresholds assessed during session 1 over 14 locations (P1 to P14) over the low back region, for left and right locations (Pleft and Pright) as well as overall low back (Pall) between the mean of the first and second trials (T1-T2), the first and the third trials (T1-T3), the second and the third trials (T2-T3) and the means of the three trials (T1-T2-T3).**

Trials	T1-T2				T1-T3				T2-T3				T1-T2-T3			
	ICC (95%CI)	SEM (kPa)	MDC (kPa)	MDC (%)	ICC (95%CI)	SEM (kPa)	MDC (kPa)	MDC (%)	ICC (95%CI)	SEM (kPa)	MDC (kPa)	MDC (%)	ICC (95%CI)	SEM (kPa)	MDC (kPa)	MDC (%)
P1	0.92 (0.72 - 0.97)	64	179	35	0.88 (0.67 - 0.96)	77	215	42	0.96 (0.88 - 0.99)	44	122	25	0.92 (0.81 - 0.97)	63	175	35
P2	0.85 (0.61 - 0.95)	85	234	46	0.90 (0.74 - 0.97)	60	167	33	0.93 (0.81 - 0.98)	59	164	32	0.90 (0.77 - 0.96)	68	190	37
P3	0.92 (0.78 - 0.97)	72	200	33	0.91 (0.76 - 0.97)	71	197	31	0.89 (0.71 - 0.96)	85	235	37	0.91 (0.80 - 0.96)	76	210	34
P4	0.94 (0.84 - 0.98)	50	138	21	0.86 (0.62 - 0.95)	76	212	33	0.91 (0.74 - 0.97)	67	185	29	0.90 (0.79 - 0.96)	65	180	28
P5	0.94 (0.78 - 0.98)	57	159	25	0.91 (0.75 - 0.97)	68	190	30	0.95 (0.86 - 0.98)	52	144	22	0.93 (0.85 - 0.97)	59	164	26
P6	0.88 (0.69 - 0.96)	83	229	37	0.88 (0.59 - 0.96)	84	232	37	0.94 (0.83 - 0.98)	60	166	26	0.90 (0.78 - 0.96)	76	210	34
P7	0.94 (0.84 - 0.98)	57	159	29	0.90 (0.71 - 0.97)	77	213	38	0.94 (0.83 - 0.98)	60	167	29	0.93 (0.84 - 0.97)	65	180	32
P8	0.91 (0.75 - 0.97)	81	224	41	0.91 (0.75 - 0.97)	74	204	37	0.93 (0.81 - 0.98)	65	180	32	0.91 (0.82 - 0.97)	73	202	36
P9	0.95 (0.87 - 0.98)	51	142	23	0.86 (0.65 - 0.95)	86	237	37	0.93 (0.81 - 0.98)	64	176	28	0.92 (0.82 - 0.97)	68	188	30
P10	0.95 (0.81 - 0.99)	51	141	24	0.96 (0.80 - 0.99)	46	127	22	0.97 (0.91 - 0.99)	40	111	18	0.96 (0.90 - 0.99)	46	126	21
P11	0.97 (0.91 - 0.99)	40	112	17	0.94 (0.83 - 0.98)	60	165	26	0.93 (0.81 - 0.98)	61	170	26	0.95 (0.88 - 0.98)	54	150	23
P12	0.95 (0.84 - 0.98)	61	168	26	0.98 (0.94 - 0.99)	34	94	15	0.94 (0.81 - 0.98)	62	172	27	0.95 (0.89 - 0.98)	53	148	24
P13	0.93 (0.78 - 0.97)	63	174	28	0.88 (0.68 - 0.96)	79	218	35	0.95 (0.87 - 0.98)	45	126	20	0.92 (0.82 - 0.97)	63	176	28
P14	0.87 (0.66 - 0.95)	91	253	42	0.89 (0.71 - 0.96)	78	217	37	0.94 (0.84 - 0.98)	57	158	27	0.90 (0.78 - 0.96)	76	212	36
Pleft	0.97 (0.91 - 0.99)	41	115	20	0.96 (0.90 - 0.99)	43	118	20	0.99 (0.96 - 1.00)	26	72	12	0.97 (0.94 - 0.99)	37	103	18
Pright	0.98 (0.95 - 1.00)	27	75	12	0.95 (0.86 - 0.98)	46	128	21	0.97 (0.93 - 0.99)	35	98	16	0.97 (0.93 - 0.99)	37	102	17
Pall	0.98 (0.94 - 0.99)	32	88	15	0.96 (0.88 - 0.99)	44	122	20	0.99 (0.96 - 1.00)	27	74	12	0.97 (0.94 - 0.99)	35	96	16

doi:10.1371/journal.pone.0160866.t002

trials. Finally, no statistical difference was observed between the two first and the two last consecutive measurements (Tables 5 and 6). Visual analysis of Bland and Altman plots showed that LOA values decreased when the two first and the three trials were analyzed (Fig 4).

### Discussion

Considering the importance of collecting reliable PPT over the lower back region, the purpose of the present experiment was (1) to evaluate the intra- and inter- absolute and relative reliability of PPT within 14 locations covering the low back region of asymptomatic individuals, and (2) to determine the number of trial required to ensure reliable PPT assessments. Intra-session results will be discussed before those of the inter-session.

First, the analysis of PPT measurements of the low back showed excellent relative reliability for the intra-session. ICCs values were almost perfect regardless of the conducted comparisons

**Table 3. Mean (SD) pressure pain thresholds (kPa) for session 1 assessed over 14 locations covering the low back region and level of significance (p-values) among trial.** See “[Procedure to mark the 14 anatomical locations](#)” for explanation concerning the locations of PPT assessments.

	Trial 1		Trial 2		Trial 3		T1-T2-T3		T1 - T2	T1 - T3	T2 - T3
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	p-value	p-value	p-value
<b>P1</b>	539.0	(237.3)	487.1	(219.5)	494.4	(209.6)	506.8	(218.2)	0.8114	0.8571	0.9958
<b>P2</b>	506.1	(190.2)	512.1	(252.5)	519.2	(205.0)	512.5	(212.2)	0.9971	0.9862	0.9958
<b>P3</b>	607.5	(237.7)	621.5	(275.1)	645.7	(244.6)	624.9	(247.2)	0.9883	0.9159	0.9653
<b>P4</b>	654.6	(193.9)	651.0	(228.3)	641.3	(214.6)	649.0	(207.6)	0.9989	0.9851	0.9920
<b>P5</b>	607.6	(219.8)	653.9	(241.9)	638.9	(233.9)	633.4	(227.2)	0.8583	0.9324	0.9840
<b>P6</b>	593.4	(231.9)	626.6	(249.0)	664.6	(249.7)	628.2	(239.5)	0.9312	0.7220	0.9107
<b>P7</b>	534.9	(243.0)	565.8	(250.7)	586.0	(246.9)	562.2	(241.7)	0.9413	0.8481	0.9745
<b>P8</b>	535.5	(259.3)	564.0	(278.9)	571.6	(227.3)	557.0	(250.2)	0.9534	0.9264	0.9966
<b>P9</b>	629.8	(227.8)	620.0	(259.1)	657.6	(241.3)	635.8	(237.6)	0.9938	0.9506	0.9118
<b>P10</b>	560.9	(246.8)	605.6	(238.4)	604.6	(230.5)	590.4	(233.7)	0.8734	0.8790	0.9999
<b>P11</b>	628.9	(242.4)	651.4	(216.1)	664.1	(251.9)	648.1	(231.9)	0.9659	0.9187	0.9890
<b>P12</b>	616.0	(263.2)	654.4	(262.7)	616.5	(229.4)	629.0	(246.7)	0.9145	1.0000	0.9166
<b>P13</b>	596.9	(243.6)	638.6	(222.5)	634.9	(212.1)	623.5	(221.7)	0.8773	0.8969	0.9990
<b>P14</b>	604.2	(252.5)	593.3	(264.2)	573.4	(221.4)	590.3	(240.9)	0.9925	0.9418	0.9753
<b>Pleft</b>	579.5	(235.0)	593.2	(240.4)	595.8	(218.2)	589.5	(225.8)	0.9865	0.9810	0.9995
<b>Pright</b>	594.1	(211.0)	613.2	(229.5)	620.3	(215.8)	609.2	(213.8)	0.9709	0.9463	0.9960
<b>Pall</b>	586.8	(220.8)	603.2	(233.6)	608.1	(214.5)	599.4	(217.8)	0.9793	0.9656	0.9982

doi:10.1371/journal.pone.0160866.t003

(trial 1 *versus* trial 2, trial 1 *versus* trial 3, trial 2 *versus* trial 3) suggesting no difference in PPTs’ measurements between trials and no systematic error in the data. Moderate to excellent relative reliability was also obtained in previous studies, assessing PPT in other anatomical locations such as tibia [40], calf, hand [41] and trapezius [42]. In a recent study assessing PPT in the lower back region of young healthy subjects, Waller and colleagues [43] reported ICC ranged from 0.94 to 0.99 and further conclude that intra-rater reliability was excellent in the low back. As Waller and colleagues [43] have assessed PPT only over one location in the low back (2 cm laterally from L4/L5). However, the generalization of such a finding is questionable considering that PPT can be different over the same muscle [6]. Moreover, the study population was small but sufficient to obtain substantial relative reliability values [35].

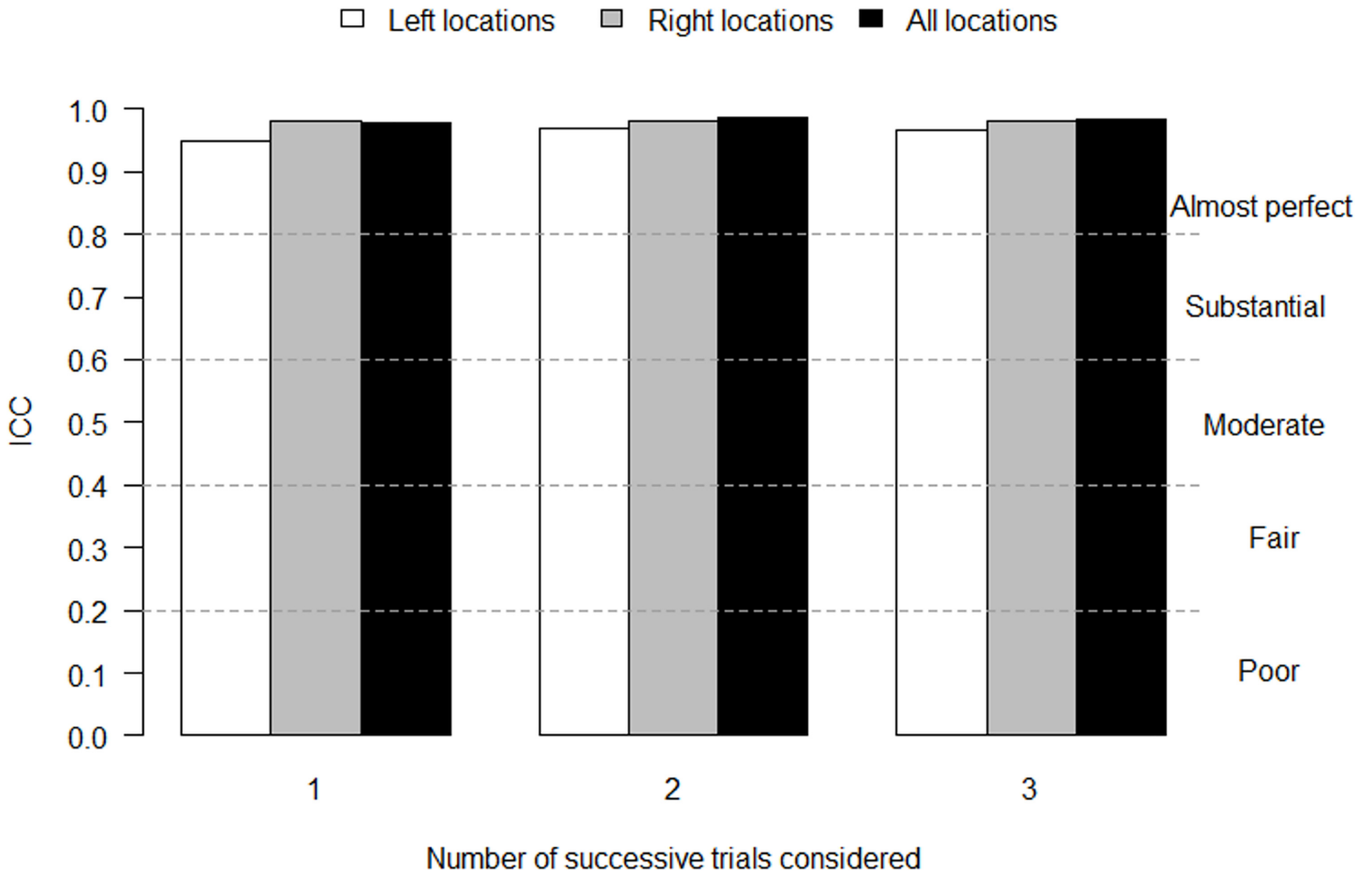
The inter-session relative reliability has also been shown to be excellent in our study. It is first important to note that no significant difference was observed for PPT measurements between session 1 and session 2. Then, the analysis trial-to-trial showed that ICCs values were also almost perfect regardless of the number of trials considered, confirming excellent reliability previously reported by Koo and colleagues [19] in the low back of healthy individuals. In the latter study, the six anatomical locations assessed (1) bilaterally in the low back perpendicularly to the spinal processes of L1, L3 and L5 and (2) over two sessions separated by 5 minutes led to ICC ranged from 0.86 to 0.91. Conversely to the intra-session’s results, we report higher relative and absolute inter-session reliability (*i.e.* ICC and SEM values) when two consecutive PPTs measurements were used for data analysis. Similar results have also been reported in the low back of healthy individuals by Chesterton and colleagues [29], *i.e.* higher intra-session reliability for the mean of three consecutive PPTs assessments than when only the first assessment was used for analysis. Even more interesting to note was that in our study, contrary to numerous studies [18,31,44], the first PPT assessment did not need to be discarded to obtain excellent reliability for both intra- and inter- session. Lacourt and colleagues [44], in a test-retest study have reported significant differences in PPTs values respectively, between the first and second

**Table 4. Mean ( $\pm$ SD); Standard Error of Measurement (SEM); Minimal Detectable Change (MDC); Intraclass Correlation Coefficient (ICC), 95% Confidence Interval (CI) of ICC, and 95% Limits of Agreement (LoA) for Pressure Pain Threshold (PPT) among 14 lumbar locations, left locations ( $P_{left}$ ), right locations ( $P_{right}$ ) point and all the locations ( $P_{all}$ ) when 1,2 or 3 trials during session 1 and session 2 are compared.**

Number of trials	Points	Mean	Mean	SEM	MDC	MDC	ICC	95% LOA	p-value
		Session 1 (SD)	Session 2 (SD)	(kPa)	(kPa)	(%)	(95% CI)		
1	Point 1	553.5 (235.4)	530.2 (167.7)	75	209	39	0.86 (0.64 - 0.95)	-190.89 - 237.42	0.7575
	Point 2	520.7 (191.8)	518.4 (194.7)	67	186	36	0.87 (0.67 - 0.96)	-192.77 - 197.43	0.9739
	Point 3	633.7 (250.5)	660.7 (251.3)	77	212	33	0.90 (0.74 - 0.97)	-243.45 - 189.32	0.7698
	Point 4	677.6 (207.1)	678.4 (233.3)	43	119	18	0.96 (0.89 - 0.99)	-125.62 - 124.02	0.9921
	Point 5	633.7 (234.8)	648.0 (254.9)	74	206	32	0.90 (0.74 - 0.97)	-228.74 - 200.21	0.2770
	Point 6	620.5 (246.9)	634.5 (249.9)	89	247	39	0.87 (0.65 - 0.95)	-271.13 - 243.13	0.8784
	Point 7	534.9 (234.1)	536.7 (221.2)	60	165	31	0.93 (0.80 - 0.98)	-175.36 - 171.63	0.9822
	Point 8	547.4 (254.1)	562.2 (222.2)	74	206	37	0.90 (0.73 - 0.97)	-228.71 - 199.11	0.8664
	Point 9	654.5 (239.4)	645.5 (234.1)	80	222	34	0.88 (0.68 - 0.96)	-223.79 - 241.66	0.9184
	Point 10	590.1 (263.5)	637.6 (240.0)	54	150	24	0.95 (0.79 - 0.99)	-173.19 - 78.25	0.6100
	Point 11	653.6 (252.5)	669.1 (232.5)	65	179	27	0.93 (0.80 - 0.97)	-201.48 - 170.42	0.8621
	Point 12	641.6 (272.3)	673.0 (258.6)	52	145	22	0.96 (0.88 - 0.99)	-169.82 - 107.02	0.7485
	Point 13	623.7 (256.8)	661.5 (237.2)	95	262	41	0.85 (0.62 - 0.95)	-302.81 - 227.21	0.6785
	Point 14	630.6 (263.9)	643.9 (238.9)	89	248	39	0.87 (0.66 - 0.95)	-271.63 - 244.97	0.8857
	Pleft	602.5 (243.3)	620.3 (219.3)	51	143	23	0.95 (0.86 - 0.98)	-163.59 - 127.94	0.8346
	Pright	614.1 (217.6)	622.5 (216.4)	37	102	17	0.91 (0.76 - 0.97)	-201.86 - 142.34	0.7008
	Pall	608.3 (228.5)	621.4 (216.8)	41	114	19	0.95 (0.85 - 0.98)	-159.01 - 111.42	0.7660
2	Point 1	533.5 (230.7)	532.2 (178.0)	53	147	28	0.93 (0.81 - 0.98)	-152.94 - 155.48	0.9867
	Point 2	532.7 (226.0)	536.1 (203.2)	55	153	29	0.93 (0.81 - 0.98)	-163.63 - 156.76	0.9654
	Point 3	637.8 (258.6)	646.9 (259.2)	52	146	23	0.96 (0.88 - 0.99)	-161.09 - 142.83	0.9237
	Point 4	675.9 (220.0)	700.1 (234.5)	47	130	19	0.96 (0.88 - 0.98)	-151.96 - 103.63	0.7731
	Point 5	655.3 (239.8)	659.4 (233.0)	63	175	27	0.93 (0.79 - 0.97)	-188.14 - 180.01	0.4137
	Point 6	636.0 (246.3)	657.3 (261.4)	64	177	27	0.93 (0.82 - 0.98)	-202.58 - 159.91	0.8197
	Point 7	552.7 (234.9)	549.1 (224.6)	52	143	26	0.95 (0.85 - 0.98)	-147.01 - 154.34	0.9655
	Point 8	562.1 (257.8)	566.0 (227.9)	45	124	22	0.97 (0.90 - 0.99)	-133.75 - 125.89	0.9650
	Point 9	649.9 (251.6)	661.2 (237.5)	50	137	21	0.96 (0.88 - 0.99)	-153.98 - 131.38	0.9002
	Point 10	611.0 (255.7)	621.7 (234.4)	32	89	15	0.98 (0.95 - 0.99)	-102.29 - 80.89	0.9057
	Point 11	664.1 (238.5)	670.7 (233.9)	65	179	27	0.92 (0.79 - 0.97)	-194.01 - 180.88	0.9399
	Point 12	659.5 (267.4)	665.6 (249.8)	41	112	17	0.97 (0.93 - 0.99)	-123.84 - 111.64	0.9490
	Point 13	643.2 (242.2)	673.0 (234.4)	62	171	26	0.93 (0.81 - 0.98)	-198.91 - 139.31	0.7345
	Point 14	625.5 (261.7)	641.9 (236.6)	57	158	25	0.95 (0.85 - 0.98)	-179.27 - 146.53	0.8587
	Pleft	609.3 (244.0)	618.8 (225.3)	33	92	15	0.98 (0.94 - 0.99)	-103.89 - 84.94	0.9128
	Pright	624.8 (227.1)	635.7 (218.2)	31	87	14	0.98 (0.94 - 0.99)	-99.62 - 78	0.7887
	Pall	617.1 (234.2)	627.2 (220.6)	30	84	14	0.98 (0.95 - 0.99)	-96.20 - 75.92	0.8514
3	Point 1	526.7 (222.6)	537.8 (177.0)	49	136	26	0.94 (0.83 - 0.98)	-152.32 - 130.14	0.8810
	Point 2	537.3 (223.6)	531.4 (193.2)	48	134	25	0.94 (0.84 - 0.98)	-134.61 - 146.47	0.9386
	Point 3	648.3 (252.7)	655.7 (255.8)	48	134	21	0.96 (0.89 - 0.99)	-147.88 - 133.03	0.9368
	Point 4	672.4 (217.6)	701.2 (235.8)	43	118	17	0.96 (0.88 - 0.99)	-138.51 - 80.73	0.7300
	Point 5	657.9 (238.5)	662.4 (234.2)	54	150	23	0.95 (0.85 - 0.98)	-161.65 - 152.58	0.5269
	Point 6	653.0 (247.2)	661.0 (257.8)	57	157	24	0.95 (0.85 - 0.98)	-172.42 - 156.28	0.9309
	Point 7	563.6 (232.6)	544.8 (213.2)	50	139	25	0.95 (0.86 - 0.98)	-122.04 - 159.60	0.8194
	Point 8	569.8 (244.6)	574.7 (227.3)	42	116	20	0.97 (0.91 - 0.99)	-126.58 - 116.63	0.9544
	Point 9	660.1 (246.4)	667.0 (230.8)	39	107	16	0.97 (0.92 - 0.99)	-118.44 - 104.66	0.9376
	Point 10	617.7 (251.1)	632.5 (232.9)	37	104	17	0.98 (0.93 - 0.99)	-119.91 - 90.17	0.8677
	Point 11	671.6 (242.2)	671.3 (239.3)	54	149	22	0.95 (0.85 - 0.98)	-156.27 - 156.85	0.9974
	Point 12	653.7 (258.0)	672.2 (247.5)	45	124	19	0.97 (0.91 - 0.99)	-144.05 - 107.16	0.8431
	Point 13	648.6 (233.7)	674.1 (229.4)	42	116	18	0.97 (0.90 - 0.99)	-136.42 - 85.35	0.7649
	Point 14	617.6 (252.6)	647.8 (237.9)	52	145	23	0.95 (0.86 - 0.98)	-170.55 - 110.15	0.7386
	Pleft	612.4 (238.0)	626.0 (224.2)	34	94	15	0.98 (0.94 - 0.99)	-108.19 - 81.03	0.8734
	Pright	630.2 (224.2)	636.0 (215.8)	26	71	11	0.99 (0.96 - 1.00)	-79.86 - 68.19	0.8708
	Pall	621.3 (229.5)	631.0 (219.0)	28	77	12	0.98 (0.95 - 0.99)	-88.38 - 68.96	0.8711

doi:10.1371/journal.pone.0160866.t004



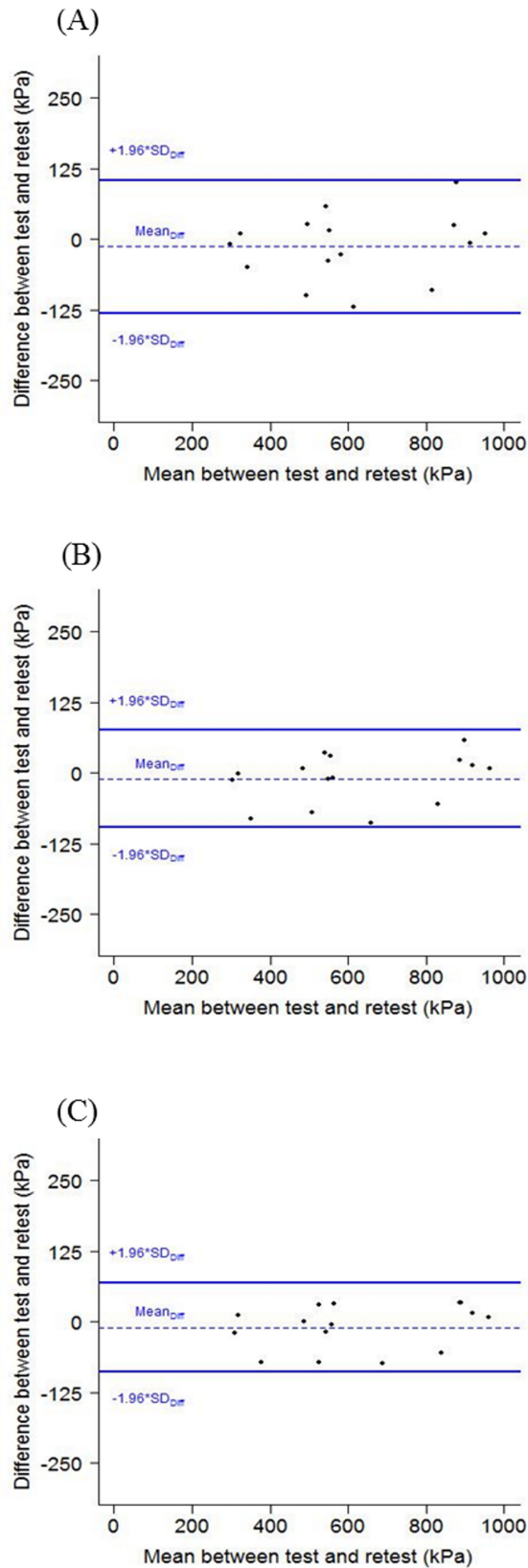


**Fig 2. Inter-session ICC values of the left, right and all locations depending on the number of successive trials considered.**

doi:10.1371/journal.pone.0160866.g002

PPT measurement and also between the first and third one. This result led them to use only the second and third PPT measurements for data analysis. Higher inter-session reliability was also found by Nussbaum and Downes [31], when the first PPT measurement was omitted. This could be explained by the effort made in the current study to familiarize the subject with PPT measurements (tests at a remote location, one practice session).

As ICC is largely influenced by between-subjects variability and does not provide information on typical error [36], it was necessary to complete our analysis by computing SEM and MDC. When the first PPT measurement was associated with the second or third one, SEM were generally below 65 kPa and MDC ranged from 11% to 27% (71 kPa to 179 kPa). In other words, this result suggests that (1) the true score of PPT was 65 kPa below or above the observed score and (2) that a clinical change will not be masked by measurement error if the observed score changed by more than 11 to 27%. The limited number of published studies assessing absolute reliability has made comparison difficult with the existing literature. However, after two sessions of PPTs measurements over one location on the trapezius muscle and the tibialis anterior separated by three to five days, Walton and colleagues [45] have reported a SEM value close to ours with a value of 49 kPa. Similar results were found by Fingleton and colleagues [25] in the lower limbs with SEM ranged from 16 to 39 kPa and by Chesterton and colleagues [29] in the back with SEM equal to 60 kPa.



**Fig 3.** Bland and Altman analyses plotted for worker's pressure pain threshold of the overall low back for the mean of the first trials (A), the first and second trials (B) and the three trials (C).

doi:10.1371/journal.pone.0160866.g003

**Table 5. Comparison of means of ICCs values regardless of the conducted comparisons between trials.**

Comparison	Difference (95% CI)	p-value
<b>Trials 1–2 vs. Trial 1</b>	0.040 (0.015 - 0.064)	0.0004***
<b>Trials 2–3 vs. Trial 1</b>	0.043 (0.018 - 0.068)	0.0002***
<b>Trials 1-2-3 vs. Trials 1</b>	0.043 (0.019 - 0.068)	0.0001***
<b>Trials 2–3 vs. Trials 1–2</b>	0.025 (0.023 - 0.028)	0.9937
<b>Trials 1-2-3 vs. Trials 1–2</b>	0.003 (0.021 - 0.028)	0.9431
<b>Trials 1-2-3 vs. Trials 2–3</b>	0.001 (0.024 - 0.026)	0.9998

\*p-value<0.05

\*\* p-value<0.01;

\*\*\* p-value<0.001.

Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval.

doi:10.1371/journal.pone.0160866.t005

In general, when looking at the number of trials required ensuring reliable measurement in the low back, our results are rather original. Indeed, we have reported almost perfect intra- and inter-session reliability on the first PPT measurement (ICC ranged from 0.85 to 0.99) suggesting that one training trial over the tibialis anterior would be sufficient to familiarize the participant with the PPT procedure. Hopkins in 2000 [46] assumed that the reliability of a test could be influenced by several factors such as motivation or boredom. For instance, during series of trials, the second one is often better than the first because participants want to improve their performance or because they benefit from the experience of the first one. Conversely, a decreased performance between the first trial and the following ones could be explained by fatigue or loss of motivation. In our study it seems that there is no learning effect between trials because PPT values did not change. Then, it seems that the cognitive and attentional resources needed to perform three consecutive PPT assessments over the low back do not generate boredom or loss of motivation and do not influence reliability. Furthermore, even though both relative and absolute relative reliabilities were significantly higher when two or three consecutive measurements were used for data analysis compared with the first measurement, no statistical difference was observed between the two first and two last PPT measurements. Therefore, this result suggests that using the two first PPT measurements for data analysis will not lead to lower relative and absolute reliabilities than using the two last or three PPT measurements.

**Table 6. Comparison of means of SEMs values regardless of the conducted comparisons between trials.**

Comparison	Difference (95% CI)	p-value
<b>Trials 1–2 vs. Trials 1</b>	-17.222 (-29.523 - -4.920)	0.0025**
<b>Trials 2–3 vs. Trial 1</b>	-18.835 (-31.136 - -6.634)	0.0008***
<b>Trials 1-2-3 vs. Trials 1</b>	-19.081 (-31.383 - -6.780)	0.0007***
<b>Trials 2–3 vs. Trials 1–2</b>	-1.613 (-13.915 - 10.688)	0.9856
<b>Trials 1-2-3 vs. Trials 1–2</b>	-1.859 (-13.460 - 9.741)	0.9207
<b>Trials 1-2-3 vs. Trials 2–3</b>	0.246 (-12.548 - 12.056)	0.9999

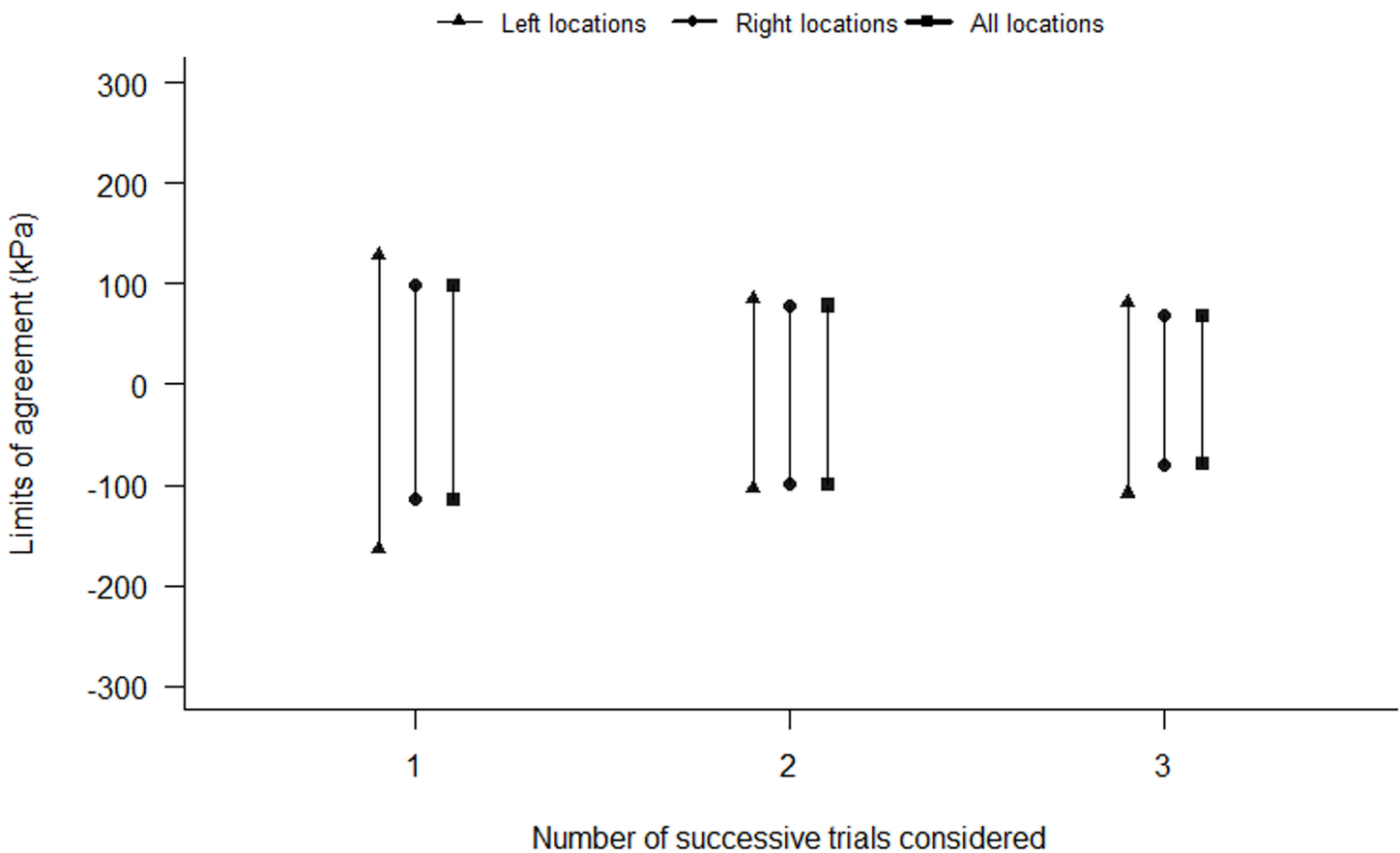
\*p-value<0.05

\*\* p-value<0.01;

\*\*\* p-value<0.001.

Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval.

doi:10.1371/journal.pone.0160866.t006



**Fig 4. Limits of Agreement (LoA) values of left, right and the overall back locations are averaged to calculate PPTs measurements.**

doi:10.1371/journal.pone.0160866.g004

Finally, in accordance with existing literature in the low back [16,45], MDC values were regularly between 100 and 200 kPa corresponding to approx. 10–20% of the PPT scale range considered as acceptable measurement error by Chiarotto and colleagues [47].

For instance, to be confident that a true change was observed in the low back of young football player after an intervention, Madeleine and colleagues [16] reported MDC value of 140 kPa. Walton and colleagues [45] assessing PPTs in the trapezius muscle among young healthy subjects reported MDC of 113 kPa. These results imply a small sensitivity to change and that a change in PPT measurement can be masked by the measurement error regardless of the absolute changes in PPT [16].

Recent studies have reported good to excellent PPTs' reliability between sessions, respectively separated by one day and assessed over 14 locations covering the abdominal region [48], two days and assessed over the 2 locations from the low back [20] and twenty-one days and assessed 1 location from the trapezius muscle [49]. Still, the current results need to be confirmed by assessing PPT's reliability over longer period of time. Then, the absence of significant difference for some important parameters such as ICC and SEM values between two or three consecutive PPT measurements could be explained by the relatively small sample size. Indeed, true significant effects might have been missed in our study because the sample size used might have not adequate power for detecting a true difference of a meaningful magnitude.

Finally, we recruited a mixed population of asymptomatic individuals classified as such since they did not report pain in the low-back within the last 7 days prior to the experiment and had no history of low back injury or/and surgery. However, pain is usually fluctuating as

reported by recent studies [50,51,52]. Further, the present results should not be generalized to specific population or gender as gender differences are reported in pressure pain sensitivity [33,53,54]. Still, Paungmali and colleagues [20] have reported almost perfect relative reliability for chronic non-specific low back pain individuals in line with our results. Future studies could address the relative and absolute variability of PPT assessed over 14 locations covering the low back in population suffering from LBP.

## Conclusions

Excellent relative and absolute reliability of PPT measured over 14 locations covering the low back of asymptomatic individuals were reported for both intra- and inter-session. Reliable measurements can be equally achieved when using the mean of consecutive PPT measurement or with only the first one. Although reliability was almost perfect regardless of the conducted comparison between PPT assessments, our results suggest using at least two consecutive measurements to obtain higher inter-session absolute reliability among asymptomatic participants in the low back region. Further studies are needed to enable a more global generalization of these findings.

## Acknowledgments

The presented work is part of the joint PhD thesis of Romain Balaguier at Univ. Grenoble Alpes (France) and Aalborg University (Denmark), who was supported by a grant of the French Ministry of Higher Education and Research. This work is also part of a larger pluri-disciplinary project called 'EWS' (Ergonomics at Work and in Sports). EWS project has benefited from support from the Blåtand French-Danish scientific cooperation program (Institut Français du Danemark), the Direction des Relations Territoriales et Internationales from Univ. Grenoble Alpes (France) and Aalborg University (Denmark) we acknowledge gratefully. The authors would like to thank anonymous reviewers for helpful comments and suggestions.

## Author Contributions

**Conceptualization:** RB PM NV.

**Data curation:** RB PM NV.

**Formal analysis:** RB PM NV.

**Funding acquisition:** RB PM NV.

**Investigation:** RB PM NV.

**Methodology:** RB PM NV.

**Project administration:** RB PM NV.

**Resources:** RB PM NV.

**Software:** RB PM NV.

**Supervision:** PM NV.

**Validation:** RB PM NV.

**Visualization:** RB PM NV.

**Writing - original draft:** RB PM NV.

**Writing - review & editing:** RB PM NV.

## References

1. Bonica JJ. International for the study of pain: pain definition. The need of a taxonomy. *Pain*. 1979; 6(3):247–248. PMID: [460931](#)
2. Merboth MK, Barnason S. Managing pain: the fifth vital sign. *Nurs Clin North Am*. 2000; 35(2):375–383. PMID: [10873249](#)
3. Mense S, Simons DG. Muscle pain. Understanding its nature, diagnosis, and treatment. Philadelphia: Lippincott Williams & Wilkins; 1992.
4. Arendt-Nielsen L and Yarnitsky D. Experimental and clinical applications of quantitative sensory testing applied to skin, muscles and viscera. *J Pain*. 2009; 10(6):556–572. doi: [10.1016/j.jpain.2009.02.002](#) PMID: [19380256](#)
5. Fischer AA. Pressure algometry over normal muscles. Standard values, validity and reproducibility of pressure threshold. *Pain*. 1987; 30(1):115–126. PMID: [3614975](#)
6. Andersen H, Arendt-Nielsen L, Danneskiold-Samsøe B and Graven-Nielsen T. Pressure pain sensitivity and hardness along human normal and sensitized muscle. *Somatosens Mot Res*. 2006; 23(3–4):97–109. PMID: [17178545](#)
7. Baker SJ, Kelly NM, Eston RG. Pressure pain tolerance at different sites on the quadriceps femoris prior to and following eccentric exercise. *Eur J Pain*. 1997; 1:229–233. PMID: [15102404](#)
8. Fridén J, Lieber RL. Eccentric exercise-induced injuries to contractile and cytoskeletal muscle fibre components. *Acta Physiol Scand*. 2001; 171(3):321–326. PMID: [11412144](#)
9. Nie H, Arendt-Nielsen L, Madeleine P, Graven-Nielsen T. Enhanced temporal summation of pressure pain in the trapezius muscle after delayed onset muscle soreness. *Exp Brain Res*. 2006; 170(2):182–190. PMID: [16328284](#)
10. Ylinen J. Pressure algometry. *Aust J Physiother*. 2007; 53(3):207. PMID: [17899675](#)
11. Andersson GB. Epidemiological features of chronic low-back pain. *The Lancet*. 1999; 354(9178):581–585.
12. Punnett L and Wegman DH. Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. *J Electromyogr Kinesiol*. 2004; 14(1):13–23. PMID: [14759746](#)
13. Freburger JK, Holmes GM, Agans RP, Jackman AM, Darter JD et al. The rising prevalence of chronic low back pain. *Arch Int Med*. 2009; 169(3):251–258.
14. Hoy D, Brooks P, Blyth F and Buchbinder R. The epidemiology of low back pain. *Best Pract Res Clin Rheumatol*. 2010; 24(6):769–781. doi: [10.1016/j.berh.2010.10.002](#) PMID: [21665125](#)
15. Andersen CH, Andersen LL, Zebis MK and Sjøgaard G. Effect of scapular function training on chronic pain in the neck/shoulder region: a randomized controlled trial. *J Occup Rehabil*. 2014; 24(2):316–324. doi: [10.1007/s10926-013-9441-1](#) PMID: [23832167](#)
16. Madeleine P, Hoj BP, Fernandez-de-Las-Peñas C, Rathleff MS and Kaalund S. Pressure pain sensitivity changes after use of shock-absorbing insoles among young soccer players training on artificial turf: a randomized controlled trial. *J Orthop Sports Phys Ther*. 2014; 44(8):587–594. doi: [10.2519/jospt.2014.5117](#) PMID: [25029914](#)
17. Ylinen J, Takala EP, Kautiainen H, Nykänen M, Häkkinen A, Pohjolainen T, Karppi SL and Airaksinen O. Effect of long-term neck muscle training on pressure pain threshold: A randomized controlled trial. *Eur J Pain*. 2005; 9(6):673–673. PMID: [16246820](#)
18. Farasyn A and 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. *J Bodyw Mov Ther*. 2003; 7(1):53–61.
19. Koo TK, Guo JY and Brown CM. Test-Retest Reliability, Repeatability, and Sensitivity of an Automated Deformation-Controlled Indentation on Pressure Pain Threshold Measurement. *J Manipulative Physiol Ther*. 2013; 36(2):84–90. doi: [10.1016/j.jmpt.2013.01.001](#) PMID: [23499143](#)
20. Paungmal A, Sitalertpisan P, Taneyhill K, Pirunsan U and Uthaihpup S. Intrarater reliability of pain intensity, tissue blood flow, thermal pain threshold, pressure pain threshold and lumbo-pelvic stability tests in subjects with low back pain. *Asian J Sports Med*. 2012; 3(1):8. PMID: [22461960](#)
21. Vanderweeen L, Oostendorp RAB, Vaes P and Duquet W. Pressure algometry in manual therapy. *Man Ther*. 1996; 1(5):258–265. PMID: [11440515](#)
22. Morkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010; 63(7):737–745. doi: [10.1016/j.jclinepi.2010.02.006](#) PMID: [20494804](#)
23. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005; 19(1):231–240. PMID: [15705040](#)

24. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007; 60(1):34–42. PMID: [17161752](#)
25. Fingleton CP, Dempsey L, Smart K and Doody CM. Intraexaminer and interexaminer reliability of manual palpation and pressure algometry of the lower limb nerves in asymptomatic subjects. *J Manipulative and Physiol Ther.* 2014; 37(2):97–104.
26. Graven-Nielsen T, Wodehouse T, Langford RM, Arendt-Nielsen L and Kidd BL. Normalization of widespread hyperesthesia and facilitated spatial summation of deep-tissue pain in knee osteoarthritis patients after knee replacement. *Arthritis & Rheum.* 2012; 64(9):2907–2916.
27. Ohrbach R and Gale EN. Pressure pain thresholds in normal muscles: reliability, measurement effects, and topographic differences. *Pain.* 1989; 37(3):257–263. PMID: [2755707](#)
28. Nie H, Graven-Nielsen T, Arendt-Nielsen L. Spatial and temporal summation of pain evoked by mechanical pressure stimulation. *Eur J Pain.* 2009; 13(6):592–599. doi: [10.1016/j.ejpain.2008.07.013](#) PMID: [18926745](#)
29. Chesterton LS, Sim J, Wright CC and Foster NE. Interrater reliability of algometry in measuring pressure pain thresholds in healthy humans, using multiple raters. *Clin J Pain.* 2007; 23(9):760–766. PMID: [18075402](#)
30. Mutlu EK and Ozdincler AR. Reliability and responsiveness of algometry for measuring pressure pain threshold in patients with knee osteoarthritis. *J Phys Ther Sci.* 2015; 27(6):1961. doi: [10.1589/jpts.27.1961](#) PMID: [26180358](#)
31. Nussbaum EL and Downes L. Reliability of clinical pressure-pain algometric measurements obtained on consecutive days. *Phys Ther.* 1998; 78(2):160–169. PMID: [9474108](#)
32. Binderup AT, Holtermann A, Sogaard K and Madeleine P. Pressure pain sensitivity maps, self-reported musculoskeletal disorders and sickness absence among cleaners. *Int Arch Occup Environ Health.* 2011; 84(6):647–654. doi: [10.1007/s00420-011-0627-6](#) PMID: [21400102](#)
33. Binderup AT, Arendt-Nielsen L and Madeleine P. Pressure pain sensitivity maps of the neck-shoulder and the low back regions in men and women. *BMC Musculoskelet Disord.* 2010; 11(1):234.
34. Schless SH, Desloovere K, Aertbeliën E, Molenaers G, Huenaerts C et al. The intra- and inter-rater reliability of an instrumented spasticity assessment in children with cerebral palsy. *PloS One.* 2015; 10(7): e0131011. doi: [10.1371/journal.pone.0131011](#) PMID: [26134673](#)
35. Landis JR and Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics.* 1977; 33(2):363–374. PMID: [884196](#)
36. Harvill LM. Standard error of measurement. *Educational Measurement: Issues and Practice.* 1991; 10(2):33–41.
37. Bruyneel AV and Bridon F. Inclinométrie du genou: comparaison de la reproductibilité d'un outil mécanique et d'une application sur smartphone. *Kinésithérapie, la revue.* 2015; 15(158):74–79.
38. Roren A, Fayad F, Roby-Brami A, Revel M, Fermanian J et al. Precision of 3D scapular kinematic measurements for analytic arm movements and activities of daily living. *Man Ther.* 2013; 18(6):473–480. doi: [10.1016/j.math.2013.04.005](#) PMID: [23726286](#)
39. Pinsault N, Fleury A, Virone G, Bouvier B, Vaillant J et al. Test-retest reliability of cervicocephalic relocation test to neutral head position. *Physiother Theory Pract.* 2008; 24(5):380–391. doi: [10.1080/09593980701884824](#) PMID: [18821444](#)
40. Aweid O, Gallie R, Morrissey D, Crisp T, Maffulli N et al. Medial tibial pain pressure threshold algometry in runners. *Knee Surg Sports Traumatol Arthrosc.* 2014; 22(7):1549–1555. doi: [10.1007/s00167-013-2558-0](#) PMID: [23740326](#)
41. Nikolajsen L, Kristensen AD, Pedersen LK, Rahbek O, Jensen TS et al. Intra- and interrater agreement of pressure pain thresholds in children with orthopedic disorders. *J Child Orthop.* 2011; 5(3):173–178. doi: [10.1007/s11832-011-0336-4](#) PMID: [22654978](#)
42. Park G, Kim CW, Park SB, Kim MJ and Jang SH. Reliability and usefulness of the pressure pain threshold measurement in patients with myofascial pain. *Ann Rehabil Med.* 2011; 35(3):412–417. doi: [10.5535/arm.2011.35.3.412](#) PMID: [22506152](#)
43. Waller R, Straker L, O'Sullivan P, Sterling M and Smith A. Reliability of pressure pain threshold testing in healthy pain free young adults. *Scand J Pain.* 2015; 9:38–41.
44. Lacourt TE, Houtveen JH and van Doornen LJ. Experimental pressure-pain assessments: test-retest reliability, convergence and dimensionality. *Scand J Pain.* 2012; 3(1):31–37.
45. Walton D, MacDermid J, Nielson W, Teasell R, Chiasson M et al. Reliability, standard error, and minimum detectable change of clinical pressure pain threshold testing in people with and without acute neck pain. *J Orthop Sports Phys Ther.* 2011; 41(9):644–650. doi: [10.2519/jospt.2011.3666](#) PMID: [21885906](#)

46. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med.* 2000; 30(1):1–15. PMID: [10907753](#)
47. Chiarotto A, Maxwell LJ, Terwee CB, Wells GA, Tugwell P et al. Roland-Morris Disability Questionnaire and Oswestry Disability Index: which has better measurement properties for measuring physical functioning in nonspecific low back pain? A Systematic Review and Meta-Analysis. *Phys Ther.* 2016. doi: [10.2522/ptj.20150420](#)
48. Montenegro MLLS, Braz CA, Mateus-Vasconcelos EL, Rosa-e-Silva JC, Candido-dos-Reis FJ et al. Pain pressure threshold algometry of the abdominal wall in healthy women. *Braz J Med Biol Res.* 2012; 45(7):578–582. PMID: [22527127](#)
49. Soee ABL, Thomsen LL, Tornoe B and Skov L. Reliability of four experimental mechanical pain tests in children. *J Pain Res.* 2013; 6:103–110. doi: [10.2147/JPR.S38514](#) PMID: [23403523](#)
50. Dunn KM, Jordan K and Croft PR. Characterizing the course of low back pain: a latent class analysis. *Am J Epidemiol.* 2006; 163(8):754–761. PMID: [16495468](#)
51. Downie AS, Hancock MJ, Rzewuska M, Williams CM, Lin CWC et al. Trajectories of acute low back pain: a latent class growth analysis. *Pain.* 2016; 157(1):225–234. doi: [10.1097/j.pain.000000000000351](#) PMID: [26397929](#)
52. Kongsted A, Kent P, Hestbaek L and Vach W. Patients with low back pain had distinct clinical course patterns that were typically neither complete recovery nor constant pain. A latent class analysis of longitudinal data. *Spine J.* 2015; 15(5):885–894. doi: [10.1016/j.spinee.2015.02.012](#) PMID: [25681230](#)
53. Fillingim RB, King CD, Ribeiro-Dasilva MC, Rahim-Williams B, and Riley JL. Sex, gender, and pain: A review of recent clinical and experimental findings. *J Pain.* 2009; 10:447–485. doi: [10.1016/j.jpain.2008.12.001](#) PMID: [19411059](#)
54. Ge HY, Madeleine P and Arendt-Nielsen L. Sex differences in temporal characteristics of descending inhibitory control: an evaluation using repeated bilateral experimental induction of muscle pain. *Pain.* 2004; 110(1):72–78.