**Aalborg Universitet**

# Long-term consistency of clinical sensory testing measures for pain assessment

Bellosta-López, Pablo; Doménech-García, Víctor; Palsson, Thorvaldur Skuli; Herrero, Pablo; Christensen, Steffan Wittrup Mcphee

# KJP

**Experimental Research Article**

# Long-term consistency of clinical sensory testing measures for pain assessment

Pablo Bellosta-López[1], Víctor Doménech-García[1], Thorvaldur Skuli Palsson[3], Pablo Herrero[4], and Steffan Wittrup Mcphee Christensen[2,5]

[1]San Jorge University, Campus Universitario, Zaragoza, Spain
[2]Department of Health Science and Technology, Faculty of Medicine, Aalborg University, Aalborg, Denmark
[3]Department of Physiotherapy and Occupational Therapy, Aalborg University Hospital, Aalborg, Denmark
[4]Department of Physiatry and Nursing, Faculty of Health Sciences, University of Zaragoza, Zaragoza, Spain
[5]Department of Physiotherapy, University College of Northern Denmark, Aalborg, Denmark

**ABSTRACT**

**Background:** Understanding the stability of quantitative sensory tests (QSTs) over time is important to aid clinicians in selecting a battery of tests for assessing and monitoring patients. This study evaluated the short- and long-term reliability of selected QSTs.

**Methods:** Twenty healthy women participated in three experimental sessions: Baseline, 2 weeks, and 6 months. Measurements included pressure pain thresholds (PPT) in the neck, upper back, and leg; Pressure-cuff pain tolerance around the upper-arm; conditioned pain modulation during a pressure-cuff stimulus; and referred pain following a suprathreshold pressure stimulation. Intraclass correlation coefficients (ICC) and minimum detectable change (MDC) were calculated.

**Results:** Reliability for PPT was excellent for all sites at 2 weeks (ICC, 0.96–0.99; MDC, 22–55 kPa) and from good to excellent at 6 months (ICC, 0.88–0.95; MDC, 47–91 kPa). ICC for pressure-cuff pain tolerance indicated excellent reliability at both times (0.91–0.97). For conditioned pain modulation, reliability was moderate for all sites at 2 weeks (ICC, 0.57–0.74; MDC, 24%–35%), while it was moderate at the neck (ICC, 0.54; MDC, 27%) and poor at the upper back and leg at 6 months. ICC for referred pain areas was excellent at 2 weeks (0.90) and good at 6 months (0.86).

**Conclusions:** PPT, pressure pain tolerance, and pressure-induced referred pain should be considered reliable procedures to assess the pain-sensory profile over time. In contrast, conditioned pain modulation was shown to be unstable. Future studies prospectively analyzing the pain-sensory profile will be able to better calculate appropriate sample sizes.

**Keywords:** Humans; Hyperalgesia; Pain Measurement; Pain Perception; Pain, Referred; Pain Threshold; Reproducibility of Results; Sample Size.

Correspondence: Víctor Doménech-García
San Jorge University, Campus Universitario, Autov. A23 km 299, 50830, Villanueva de Gállego, Zaragoza, Spain
Tel: +34-976-060-100, Fax: +34-976-077-581, E-mail: vdomenech@usj.es

## INTRODUCTION

Pain is the most common reason for people with musculoskeletal (MSK) conditions to consult a clinician [1]. The number of people with MSK conditions requiring rehabilitation has increased by 63% in less than 30 years [2], and it is estimated that approximately 20%–33% of people globally live with painful MSK conditions [3]. Although the onset of MSK pain may have been initiated by tissue pathology [4], symptom persistence has been suggested to be influenced by altered central processing of noxious and non-noxious stimuli (also known as central sensitization) [5] and high levels of psychosocial factors such as pain catastrophizing or depression [6].

Evaluating the pain-sensory profile may be beneficial for some patients with persistent MSK pain (*e.g.*, in patients with low back pain [7], knee osteoarthritis [8], or whiplash-associated disorders [9]), as it may provide clinicians with valuable information to decide and plan rehabilitation, and help to establish prognosis and classify patients into subgroups [10,11].

Assessment of the pain-sensory profile covers a range of psychophysical procedures, usually referred to as quantitative sensory testing (QST), intended to quantify pain responses by applying standardized stimuli to different somatic structures [12]. Pressure pain thresholds (PPT) and conditioned pain modulation (CPM) are two of the most commonly used procedures to assess sensitivity to noxious stimuli in people with or without painful conditions [13]. PPT refers to the lowest intensity at which a given pressure stimulus is perceived as painful. PPT is contrasted by pressure pain tolerance (PPTol), which refers to the level of pain a subject is willing to endure for a given pressure stimulus [14]. CPM refers to the dampening effect that a competing heterotopic painful stimulus has on the sensitivity to noxious peripheral input and is used to assess the efficiency of endogenous analgesia [15]. Furthermore, assessing the distribution of perceived experimental pain following a supra-threshold pressure-stimulation can provide valuable information about nociceptive processing [16]. According to a previous definition [17], testing PPT, PPTol, CPM, and assessing the perceived distribution of pain might be useful to adopt in clinical practice. However, for such purposes, it is important to know the stability of such tests over time.

To date, only a few studies have investigated the long-term reliability of psychophysical test procedures, which is limited to a low number of body regions such as hand, forearm, lumbar spine, upper leg, or foot [18,19]. Additionally, it is not clear how and if such QST procedures are internally correlated. In case of a strong correlation and stability in measures over time, a full QST protocol might not be needed to sufficiently assess the pain-sensory profile, making it more appealing to employ in clinical research and practice. Overall, the identification of altered sensitivity to noxious stimuli in humans is challenging, and data supporting appropriate protocols for assessing MSK pain disorders is needed [11]. Therefore, the results of this study are expected to provide important knowledge of the reliability of the above-mentioned tests and thereby be of value for clinicians and clinical researchers who want to monitor the pain-sensory profile and any changes hereof over time.

## MATERIALS AND METHODS

### 1. Study design

This reliability study is a secondary analysis of data from a previously published prospective case-control study [20]. This manuscript followed the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [21].

### 2. Participants and procedure

Healthy individuals aged 18 to 50 years old were recruited between March and June 2019 from the local community through advertisements at the university campus and on social media. Participants were considered healthy if they were free of any pain condition in the previous six months, and had no pain-related pathology, previous history of major surgery or trauma, or regular medication intake. All participants provided written informed consent before participation, and all study procedures were approved by the ethics committee of Aragon (n-PI16/0132).

Participants attended test sessions on three different occasions: at baseline, at 2 weeks and 6 months. The same rater ran all test sessions following a standardized protocol. After verifying inclusion criteria and filling out questionnaires, PPT, cuff-PPTol, CPM, and pressure-induced referred pain were tested (**Fig. 1**).

All sessions were conducted in the same clinical setting, in a quiet room with ambient light, temperature, and humidity-controlled between 23ºC–24ºC and 30%–35%, respectively. Participants were assessed at approximately the same time of the day to increase repeatable testing conditions between days [22]. Prior to running the psychophysical tests, participants completed a battery of questionnaires on sleep disturbances with the Medical
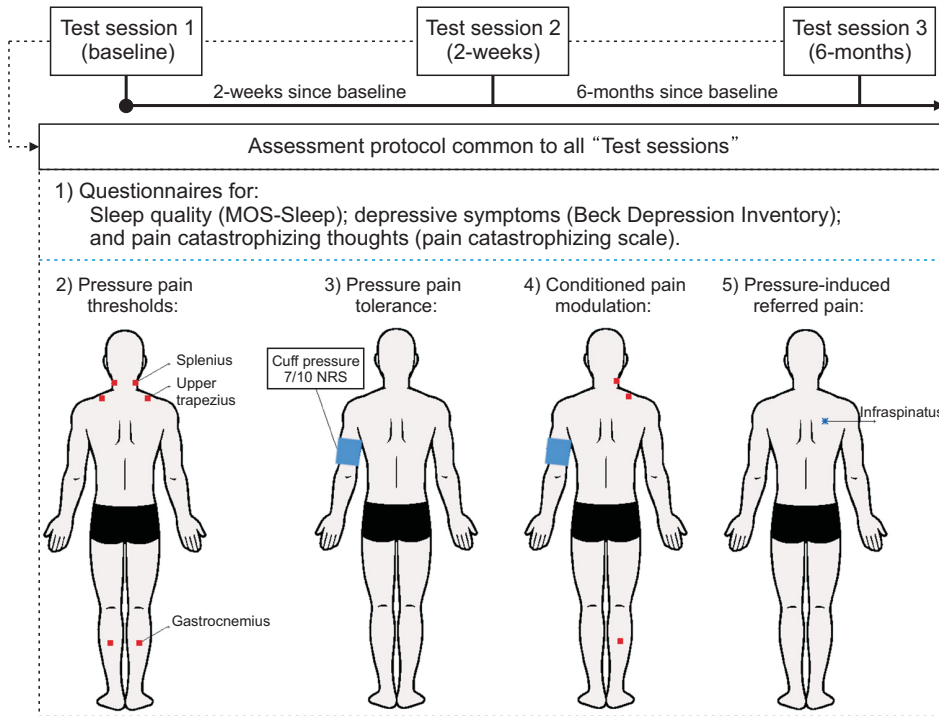
**Fig. 1.** Schematic overview of the timeline and the assessment protocol. MOS-Sleep: medical outcomes study sleep scale, NRS: numeric rating scale.

Outcomes Study Sleep Scale [23], depressive symptoms with the Beck Depression Inventory [24], and pain-related catastrophizing thoughts with the Pain Catastrophizing Scale [25]. The purpose of completing these questionnaires was to ensure that these three domains remained similar across the six months [20]. Additionally, before the second and third testing, participants were asked to confirm that they still fulfilled the inclusion criteria for the study. Each test session lasted approximately 20 minutes, and the rater was blinded to findings obtained in previous sessions.

## 3. PPT

A digital handheld pressure algometer (Somedic) connected by a cable to a pushbutton was used to assess PPT. The algometer was mounted with a 1 cm$^2$ probe and pressure was steadily increased at a constant rate of 30 kPa/s.

At the beginning of each session, the rater followed a similar familiarization protocol to ensure that instructions to the participants were similar between days. First, the rater explained to the participants that "Pressure pain threshold is the pressure reached when you first feel that the pressure becomes painful" [20]. Here, particular emphasis was placed on evaluating the PPT and not the tolerance. Therefore, the rater asked the participants to push the button at the first moment the pressure became

painful, *i.e.*, when it reached a score of 1 on a numeric rating scale (NRS), ranging from 0 "no pain" to 10 "the worst imaginable pain". Secondly, the assessor demonstrated the procedure to the participant on his/her own forearm, followed by a demonstration on the forearm of the participant. After the first demonstration, the participants were asked to close their eyes during the procedure to have a greater focus on their sensations. Finally, participants were asked to press the button when the stimulus became 3 in the NRS to better distinguish between what the first sensation of pain was and show that they need not tolerate any level of pain during the testing.

PPT was assessed bilaterally over three muscles (splenius capitis, upper trapezius, and soleus), with participants lying prone. The location of the PPT sites was identified by palpation of anatomical landmarks in addition to using a tape measure. PPT sites were marked with semipermanent ink [20].

The order of PPT sites was randomized for each participant before the first session. This order was maintained in the subsequent sessions [20]. Two measurements of PPT were recorded for each site with a 30-second interval before assessing the same site again [26]. The mean value for each site averaged across the two sides was extracted for reliability analysis.

### 4. PPTol & CPM

An inflatable pressure-cuff was mounted on the participants' non-dominant arm (Model DS54; Welch Allyn), approximately at the midpoint between the acromion and the olecranon. The pressure-cuff was manually inflated at a rate of 20 mmHg/s until participants indicated a score of 7 on the 11-point NRS. Cuff-PPTol was considered the exact mmHg value when NRS 7 was reached (cuff-PPTol-7).

After 30 seconds with the cuff inflated, PPT was recorded, as described above, on the contralateral side to the cuff [9]. Once all PPT recordings were conducted, the cuff was deflated. A CPM value was calculated by subtracting the initial PPT value, without the conditioning painful stimuli, from the PPT value recorded while the cuff was inflated. For the reliability analysis, CPM values were transformed into percentages according to the initial PPT values [27].

### 5. Pressure-induced referred pain

Referred pain was evoked by applying a suprathreshold pressure stimulation (STPS) over the infraspinatus muscle on the dominant side [28,29]. The infraspinatus site was located by identifying the intermediate point between the inferior angle of the scapula, the spine of the scapula, and the mid-point of the medial border of the scapula. First, the PPT at the infraspinatus site was determined, after which an STPS value was calculated as 1.2 times the PPT value [28,29]. The target pressure was achieved in 2–3 seconds, and it was maintained for one minute in all participants [28,29].

Before the stimulation, the rater illustrated on a hand-held PC tablet (Airis OnePad 750; Infinity System) how to draw the perceived area of pain on a digital body chart (anterior and posterior views) in an android application (Navigate Pain v1.0; Aalborg University) [30]. Immediately following the STPS, participants drew the pain areas. The area of perceived pain was then automatically extracted and transformed into pixels for further analysis.

### 6. Statistical analysis

Statistical analysis was performed using SPSS v.25 (IBM Co.), and a significance level of $P < 0.05$ was accepted.

To analyze for a confounding effect of psychosocial factors in QST measures over time, a repeated-measures analysis of covariance (RM-ANCOVA) with time (baseline, 2 weeks, 6 months) as within factors, and baseline values of sleep disturbances, depressive symptoms, and pain catastrophizing thoughts as covariables, was conducted.

The short and long-term reliability of PPT, cuff-PPTol-7, CPM, and pressure-induced referred pain area were evaluated for single measurement absolute agreement based on a two-way mixed model by computing intra-class correlation coefficients ($ICC_{3,1}$). An ICC above 0.90 was interpreted as excellent reliability, 0.75–0.90 good reliability, 0.50–0.75 moderate reliability, and less than 0.50 poor reliability [31]. The standard error of measurement (SEM) was calculated using the following formula: SEM = $SD_{pooled} \times \sqrt{(1\text{-}ICC)}$. The SEM represented the expected random variation in scores when no real change had happened [32]. The minimum detectable change (MDC) at 95% was calculated using the formula: MDC = $1.96 \times SEM \times \sqrt{2}$. MDC is considered the minimal change needed for being a real change rather than a random measurement error [32]. SEM and MDC were calculated when ICC analyses were significant.

Following a Shapiro–Wilk test, Pearson's or Spearman's correlation was used to investigate the relationship between pain-sensory variables that showed moderate to excellent reliability. Correlations were considered as very strong ($r \geq 0.90$), strong ($0.70 > r < 0.89$), moderate ($0.40 > r < 0.69$), and poor ($r < 0.39$). Bonferroni correction for multiple comparisons was applied (*i.e.*, $P$ value of 0.05 divided by 15 comparisons per variable). The average of correlation values obtained on the three occasions was presented for an easier interpretation.

## RESULTS

Twenty healthy participants (12 women) with an average age of 31 ± 7 years completed the study (**Table 1**). The test session took place at baseline, at 2 weeks (range 14 to 20 days) and 6 months (range 161 to 182 days). No changes were observed in sleep disturbances, depressive symptoms, or pain catastrophizing thoughts, as previously reported [20]. The RM-ANOVA, adjusted by the above-mentioned psychosocial factors, showed no interaction for any confounder.

**Table 2** presents $ICC_{3,1}$, SEM, and MDC for each psychophysical test at 2 weeks and 6 months compared to baseline. For PPT, reliability was excellent at the splenius capitis and upper trapezius sites (0.92–1.00) at 2 weeks and from good to excellent at 6 months (0.83–0.98). While, for the soleus site, PPT reliability was from good to excellent at 2 weeks (0.89–0.98) and from moderate to excellent at 6 months (0.72–0.95). For cuff-PPTol-7, reli-

**Table 1.** Characteristics of participants (n = 20)

| Variable | Mean ± Standard deviation | Range |
| --- | --- | --- |
| Age (yr) | 31 ± 7 | 18 to 44 |
| Female, n (%) | 12 (60) | - |
| Splenius PPT (kPa) | 170 ± 77 | 75 to 394 |
| Upper trapezius PPT (kPa) | 221 ± 76 | 105 to 385 |
| Soleus PPT (kPa) | 304 ± 99 | 131 to 460 |
| Cuff-PPTol-7 (mmHg) | 234 ± 60 | 110 to 340 |
| Splenius CPM (% PPT) | 19 ± 16 | –15 to 53 |
| Upper trapezius CPM (% PPT) | 18 ± 20 | –9 to 68 |
| Soleus CPM (% PPT) | 18 ± 23 | –9 to 84 |
| PIRP area (pixels) | 1,471 ± 1,748 | 164 to 6,429 |
| Infraspinatus STPS (kPa) | 394 ± 164 | 133 to 738 |
| MOS-Sleep (0–100) | 24 ± 13 | 7 to 56 |
| PCS (0–44) | 11 ± 6 | 0 to 23 |
| BDI (0–66) | 3 ± 2 | 0 to 10 |

Higher scores in MOS-Sleep, PCS, and BDI indicate worse levels.
PPT: pressure pain threshold, PPTol-7: pressure pain tolerance of 7 on a 11-point numeric rating scale, CPM: conditioned pain modulation, PIRP: pressure-induced referred pain, STPS: suprathreshold pain stimulation (120% of PPT), MOS-Sleep: medical outcomes study sleep scale, PCS: pain catastrophizing scale, BDI: Beck Depression Inventory.

ability was from good to excellent at both times points (0.78–0.99). For CPM, reliability was from poor to good at the three sites at 2 weeks (0.18–0.89). At 6 months, however, $ICC_{3,1}$ at the splenius capitis muscle indicated poor to good reliability (0.16–0.79), while the reliability for the upper trapezius and soleus muscles was negligible. Finally, for the areas of referred pain evoked by STPS at the infraspinatus muscle, reliability was from good to excellent at 2 weeks (0.77–0.96) and from moderate to excellent at 6 months (0.69–0.42).

**Table 3** presents average correlation coefficients for pain-sensory variables across the three test sessions. PPT at all sites showed a strong inter-correlation at all test sessions (r > 0.71). No significant correlation was found between the other variables following the Bonferroni correction.

## DISCUSSION

This study investigated the short- and long-term reliability of a battery of four simple and clinically applicable psychophysical tests, in a healthy cohort with test sessions over a 6-month period. The results showed that PPT and cuff-PPTol-7, and referred pain assessed after a STPS (120% of PPT), had good to excellent reliability in the short term (2 weeks) and moderate to excellent reliability in the long term (6 months). In contrast, the CPM showed poor to good reliability in the short-term and negligible to good reliability in the long term.

### 1. Short-term reliability

Several studies have previously investigated the short-term reliability of PPT over one to five days in healthy participants [33–35] and found excellent reliability (0.92–0.97), which is in line with the current results. Both the current and previous studies indicate that estimation of PPT is reliable regardless of the evaluated body region. In contrast, only a few studies have evaluated the reliability of cuff-PPTol in the short term, showing moderate to excellent reliability (0.74 to 0.96) [36,37]. This study reports similar ICC confidence intervals for cuff-PPTol-7, although the previous studies established a pain tolerance threshold with a computerized cuff at NRS 10 out of 10. Even though pressure-induced referred pain has been assessed in several previous studies [16,28,29,38], this study is the first to report reliability over time, showing good reliability. For CPM, numerous studies have analyzed short-term reliability, demonstrating different results depending on the conditioning stimulus [39]. While some conditioning stimuli, such as the cold pressor test, seem

**Table 2.** Reliability indicators of pressure pain thresholds, pressure pain tolerance, conditioned pain modulation, and pressure-induced referred pain compared to baseline (n = 20)

| Variable | 2-weeks | | | 6-months | | |
|---|---|---|---|---|---|---|
| | ICC (95% CI) | SEM (95% CI) | MDC (95% CI) | ICC (95% CI) | SEM (95% CI) | MDC (95% CI) |
| PPT (kPa) | | | | | | |
| Splenius | 0.99 (0.95 to 1.00) | 8 (6 to 18) | 22 (15 to 49) | 0.95 (0.88 to 0.98) | 17 (11 to 26) | 47 (30 to 72) |
| Upper trapezius | 0.97 (0.92 to 0.99) | 14 (8 to 23) | 38 (22 to 63) | 0.93 (0.83 to 0.97) | 21 (14 to 33) | 58 (38 to 91) |
| Soleus | 0.96 (0.89 to 0.98) | 20 (14 to 33) | 55 (39 to 91) | 0.88 (0.72 to 0.95) | 33 (21 to 50) | 91 (59 to 139) |
| Cuff-PPTol-7 (mmHg) | 0.97 (0.94 to 0.99) | 10 (6 to 14) | 27 (16 to 38) | 0.91 (0.78 to 0.96) | 17 (11 to 27) | 47 (31 to 74) |
| CPM (% PPT) | | | | | | |
| Splenius | 0.74 (0.46 to 0.89) | 9 (6 to 12) | 24 (16 to 35) | 0.54 (0.16 to 0.79) | 10 (7 to 13) | 27 (18 to 36) |
| Upper trapezius | 0.70 (0.37 to 0.87) | 10 (7 to 14) | 28 (18 to 40) | 0.28 (−0.19 to 0.64)[a] | - | - |
| Soleus | 0.57 (0.18 to 0.81) | 13 (8 to 15) | 35 (23 to 48) | 0.16 (−0.30 to 0.56)[a] | - | - |
| Pressure-induced referred pain area (pixels) | 0.90 (0.77 to 0.96) | 502 (317 to 761) | 1,390 (879 to 2,109) | 0.86 (0.69 to 0.94) | 552 (361 to 821) | 1,529 (1,001 to 2,275) |
| Infraspinatus STPS (kPa) | 0.92 (0.83 to 0.97) | 49 (30 to 71) | 13 (82 to 196) | 0.91 (0.79 to 0.96) | 46 (31 to 70) | 127 (85 to 195) |

CI: confidence interval, ICC: intraclass correlation coefficient, SEM: standard error of mean, MDC: minimum detectable change, PPT: pressure pain threshold, PPTol-7: pressure pain tolerance of 7 on a 11-point numeric rating scale, CPM: conditioned pain modulation, STPS: suprathreshold pain stimulation (120% of PPT).

[a]$P > 0.05$ after $ICC_{3,1}$ (two-way mixed model, single assessment, with absolute agreement), indicating lack of reliability.

**Table 3.** Correlation values between pain-sensory variables averaged for the three test sessions (n = 20)

| Pain sensory variables | PPT UT[a] | PPT SOL[a] | STPS[a] | PIRP[b] | Cuff-PPTol-7[a] |
|---|---|---|---|---|---|
| PPT SP | 0.81* | 0.78* | 0.84* | −0.22 | 0.33 |
| PPT UT | | 0.75* | 0.85* | −0.31 | 0.09 |
| PPT SOL | | | 0.80* | −0.48 | 0.16 |
| STPS | | | | −0.23 | 0.11 |
| PIRP | | | | | −0.12 |

PPT: pressure pain threshold, PPTol-7: pressure pain tolerance of 7 on a 11-point numeric rating scale, STPS: suprathreshold pressure stimulation (120% of PPT), PIRP: pressure-induced referred pain, SOL: soleus muscle, SP: splenius capitis muscle, UT: upper trapezius muscle.
[a]Pearsons's r. [b]Spearman's ρ.
*Significant correlation after Bonferroni correction (*P* < 0.003).

to offer good to moderate intra- and inter-session reliability [40,41], studies assessing the reliability of cuff pressure as a conditioning stimulus show good intra-session reliability but poor inter-session reliability [40]. However, despite superior inter-session reliability for CPM based on cold pressor test, this may not be easy to implement into clinical practice due to the considerable setup required.

## 2. Long-term reliability

The long-term reliability of psychophysical tests in healthy participants has previously been investigated. One study found good reliability of PPT at the lower back and hand sites 4 months after baseline [18], while another study found moderate reliability at the same sites 10 weeks after baseline [19]. The reasons behind the good to excellent reliability seen here after 6 months could relate to the simplicity of the assessment protocol. In contrast to previous studies assessing long-term reliability of QST parameters [18,19], the present study included a standardized procedure to introduce the algometer and PPT measures to the participants, as has been suggested previously [34]. Rigorous methodology is key to generating reliable outcomes in psychophysical testing, as many testing procedures are operator-dependent [42]. For this reason, the simplicity of procedures, such as that used for PPT here, may be highly relevant if sensory testing is to be implemented into clinical practice.

Psychophysical tests are influenced by affective and cognitive factors, which may change over time [42]. Marcuzzi and colleagues assessed psychosocial factors at baseline, but none of the previous long-term reliability studies investigated whether changes in psychosocial

variables occurred during their assessment period [18,19]. As psychosocial factors have been demonstrated to influence both the experience of pain and QST [43], the lack of changes in the psychosocial assessments used here could explain the temporal stability of the current results compared to those reported previously [20].

Finally, the long-term reliability values of CPM are similar to those reported by Marcuzzi et al. [18], indicating that this test paradigm may be less useful for monitoring changes over time. This may be due to the fact that CPM requires a more complex procedure than PPT, PPTol, or pressure-induced referred pain, and may introduce potential confounders like sensory modality or the intensity of the conditioning and test stimuli [18,39]. Interestingly, despite the methodological recommendations suggested to standardize the CPM assessment [27], a simple method adapted to a clinical setting is still lacking. In this study, the conditioning stimulus (*i.e.*, cuff-PPTol-7) was controlled and revealed excellent intra-individual reliability over time and could easily be used in clinical practice, but the CPM effect was not reliable enough to recommend adoption in the clinic.

## 3. Correlation between psychophysical tests

The current study showed a strong positive inter-correlation between PPT at all assessment sites, which indicates that healthy participants presenting higher PPT values over the neck region also present higher PPT over the upper back or lower leg, and *vice versa*. However, no correlation was found between PPT and cuff-PPTol-7 or pressure-induced referred pain. This lack of correlation may be explained by the different stimulation intensities of cuff-PPTol-7 (high intensity) and sustained pressure (intensity slightly above the pain threshold) stimuli compared to PPT. In contrast to handheld pressure algometry assessing a single point, cuff algometry evaluates a larger area by applying a tourniquet around the extremity [37]. This has been suggested to remove the variability of skin sensitivity and instead assess diffuse pain sensitivity in deeper tissues [44]. The lack of correlation between the size of the area of pain and PPT has been previously observed [28], and may be due to the high variability in pain referral patterns compared to PPT.

## 4. Clinical implications and future research

Going from bench to bedside is a natural step of basic pain research [45]. Within pain science, assessing the pain-sensory profile in a research context has allowed

researchers to classify patients, establish prognosis, and predict response to specific treatments [11]. However, translating basic research to clinical practice is a delicate process that requires valid and reliable procedures [45]. This study showed that PPT, PPTol, and pressure-induced referred pain are procedures with high levels of correlation and agreement between intra-individual measurements, indicating that they can be replicated over time with a small degree of error [31]. In other words, these procedures are stable, and their changes should not be attributed to measurement error but ideally would largely reflect a change in the MSK condition. Therefore, these procedures may be suitable for monitoring the pain-sensory profile of patients over a long-term period, *e.g.*, to assess the recovery process over time or evaluate the impact of a treatment. Nevertheless, it is still essential to investigate the relationship between the pain-sensory profile and clinical pain, as to date, it is unclear to what extent psychophysical measures are related to the overall pain experience [46,47]. In any case, assessing the pain-sensory profile could be considered part of a holistic patient evaluation, supporting clinical decision-making.

This study showed that the MDC of PPT varies between the different body sites, and was lower in the neck than in the upper back and leg. This is in line with previous findings, which reported that PPT assessment, due to its high specificity but moderate sensitivity, is a more valuable procedure to confirm changes than discard changes [35]. In general, this study showed an increase in the MDCs of psychophysical tests over time, indicating that the longer the time between assessments, the greater the difference is needed for it to be considered a true change. However, more studies are needed to confirm these results.

Lastly, the strong positive correlation between all PPT sites (*i.e.*, the splenius capitis, upper trapezius, soleus, and infraspinatus muscles) indicates that in healthy individuals, a higher PPT at one site is related to higher recordings at other sites over time, which is in line with previous studies in acute pain populations [20]. These findings suggest that a less time-consuming assessment protocol, including one local and one distal site to the painful region, may be sufficient to evaluate the evolution of the pain-sensory profile over time.

## 5. Strengths and limitations

The main strength of this study is the performance of a systematic assessment protocol to reduce the impact of confounders such as the misunderstanding of what a "pain threshold" means, controlling room conditions,

and being consistent with the time of day of assessments, to avoid the influence of circadian rhythms [22]. Another strength of this study is that psychosocial factors were analyzed as covariates and controlled throughout the 6-month assessment period.

Also, simultaneously presenting short-term and long-term reliability values within the same cohort, makes it easier to observe temporal changes in MDCs. Consequently, future studies should consider this factor when calculating sample sizes, especially when including a follow-up measurement [33].

This study was limited by the relatively small sample size in the original study [20], which did not adequately account for the power needed in the current reliability study. However, this fact mainly affected establishing the reliability of the CPM response, as this has previously been shown prone to be highly affected by variability between individuals [18,20]. Nevertheless, for the rest of the psychophysical tests (*i.e.*, PPT, cuff-PPTol, referred pain), considering previous ICC values of long-term reliability being between 0.65 and 0.90 [18], 18 participants would have been sufficient to reach a power of 80% with an alfa error of 0.05 for a single measurement two-way mixed model [48]. It is highly recommended that clinicians and researchers consider the confidence intervals of reliability values when interpreting results rather than reducing the interpretation to the final value of the ICCs [49]. This applies particularly to the CPM response which had large confidence intervals and relatively poor test-retest reliability, particularly at 6 months (**Table 2**). Another limitation that hampers the generalizability to clinical practice is that the current study was based on data from a healthy population which may not be comparable to the clients commonly seen in clinical practice although it can still serve to provide reference values to compare clinical data against. In summary, employing the methods described in this study in clinical research or practice should account for the limitations outlined above.

In conclusion, assessing the pain-sensory profile of healthy participants over a 6-month period found reliable results when using PPT and PPTol, although the MDCs increased slightly over time. In contrast, the CPM paradigm was shown to be unstable over time, indicating that it may not be suitable for assessing changes over time. This study showed a high correlation between PPT sites, suggesting that a brief assessment battery comprising two PPT points, together with other psychophysical tests, could be sufficient to obtain useful information on the pain-sensory profile.

## DATA AVAILABILITY

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## ACKNOWLEDGMENTS

## CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

## FUNDING

## AUTHOR CONTRIBUTIONS

Pablo Bellosta-López: Study conception, methodology, writing/manuscript preparation; Víctor Doménech-García: Study conception, methodology, writing/manuscript preparation; Thorvaldur Skuli Palsson: Writing/manuscript preparation, methodology; Pablo Herrero: Methodology; Steffan Wittrup Mcphee Christensen: Study conception, Writing/manuscript preparation.

## ORCID

Pablo Bellosta-López, https://orcid.org/0000-0003-4750-9077
Víctor Doménech-García, https://orcid.org/0000-0002-7007-0368
Thorvaldur Skuli Palsson, https://orcid.org/0000-0002-4418-0133
Pablo Herrero, https://orcid.org/0000-0002-9201-0120
Steffan Wittrup Mcphee Christensen, https://orcid.org/0000-0001-9068-0641

## REFERENCES

1. McRae M, Hancock MJ. Adults attending private physiotherapy practices seek diagnosis, pain relief, improved function, education and prevention: a survey. J Physiother 2017; 63: 250-6.
2. Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2021; 396: 2006-17. Erratum in: Lancet 2021; 397: 198.
3. McMahon SB, Dargan P, Lanas A, Wiffen P. The burden of musculoskeletal pain and the role of topical non-steroidal anti-inflammatory drugs (NSAIDs) in its treatment. Ten underpinning statements from a global pain faculty. Curr Med Res Opin 2021; 37: 287-92.
4. Mills SEE, Nicolson KP, Smith BH. Chronic pain: a review of its epidemiology and associated factors in population-based studies. Br J Anaesth 2019; 123: e273-83.
5. Roy JS, Bouyer LJ, Langevin P, Mercier C. Beyond the joint: the role of central nervous system reorganizations in chronic musculoskeletal disorders. J Orthop Sports Phys Ther 2017; 47: 817-21.
6. Edwards RR, Dworkin RH, Sullivan MD, Turk DC, Wasan AD. The role of psychosocial processes in the development and maintenance of chronic pain. J Pain 2016; 17(9 Suppl): T70-92.
7. Rabey M, Slater H, O'Sullivan P, Beales D, Smith A. Somatosensory nociceptive characteristics differentiate subgroups in people with chronic low back pain: a cluster analysis. Pain 2015; 156: 1874-84.
8. Lluch E, Nijs J, Courtney CA, Rebbeck T, Wylde V, Baert I, et al. Clinical descriptors for the recognition of central sensitization pain in patients with knee osteoarthritis. Disabil Rehabil 2018; 40: 2836-45.
9. Daenen L, Nijs J, Roussel N, Wouters K, Van Loo M, Cras P. Dysfunctional pain inhibition in patients with chronic whiplash-associated disorders: an experimental study. Clin Rheumatol 2013; 32: 23-31.
10. Beales D, Mitchell T, Moloney N, Rabey M, Ng W, Rebbeck T. Masterclass: a pragmatic approach to pain sensitivity in people with musculoskeletal disorders and implications for clinical management for

musculoskeletal clinicians. Musculoskelet Sci Pract 2021; 51: 102221.

11. Georgopoulos V, Akin-Akinyosoye K, Zhang W, Mc-Williams DF, Hendrick P, Walsh DA. Quantitative sensory testing and predicting outcomes for musculoskeletal pain, disability, and negative affect: a systematic review and meta-analysis. Pain 2019; 160: 1920-32.

12. Pavlaković G, Petzke F. The role of quantitative sensory testing in the evaluation of musculoskeletal pain conditions. Curr Rheumatol Rep 2010; 12: 455-61.

13. Graven-Nielsen T, Arendt-Nielsen L. Assessment of mechanisms in localized and widespread musculoskeletal pain. Nat Rev Rheumatol 2010; 6: 599-606.

14. Kanner R. Chapter 1 - definitions. In: Pain management secrets. 3rd ed. Edited by Argoff CE, McCleane G. Mosby. 2009, pp 9-14.

15. Nir RR, Yarnitsky D. Conditioned pain modulation. Curr Opin Support Palliat Care 2015; 9: 131-7.

16. Doménech-García V, Palsson TS, Boudreau SA, Bellosta-López P, Herrero P, Graven-Nielsen T. Healthy pain-free individuals with a history of distal radius fracture demonstrate an expanded distribution of experimental referred pain toward the wrist. Pain Med 2020; 21: 2850-62.

17. Zhu GC, Böttger K, Slater H, Cook C, Farrell SF, Hailey L, et al. Concurrent validity of a low-cost and time-efficient clinical sensory test battery to evaluate somatosensory dysfunction. Eur J Pain 2019; 23: 1826-38.

18. Marcuzzi A, Wrigley PJ, Dean CM, Adams R, Hush JM. The long-term reliability of static and dynamic quantitative sensory testing in healthy individuals. Pain 2017; 158: 1217-23.

19. Nothnagel H, Puta C, Lehmann T, Baumbach P, Menard MB, Gabriel B, et al. How stable are quantitative sensory testing measurements over time? Report on 10-week reliability and agreement of results in healthy volunteers. J Pain Res 2017; 10: 2067-78.

20. Christensen SWM, Bellosta-López P, Doménech-García V, Herrero P, Palsson TS. Changes in pain sensitivity and conditioned pain modulation during recovery from whiplash-associated disorders. Clin J Pain 2021; 37: 730-9.

21. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011; 64: 96-106.

22. Tanaka Y, Shigetoh H, Sato G, Fujii R, Imai R, Osumi M, et al. Classification of circadian pain rhythms and pain characteristics in chronic pain patients: an observational study. Medicine (Baltimore) 2021; 100: e26500.

23. Rejas J, Ribera MV, Ruiz M, Masrramón X. Psychometric properties of the MOS (Medical Outcomes Study) Sleep Scale in patients with neuropathic pain. Eur J Pain 2007; 11: 329-40.

24. Azocar F, Areán P, Miranda J, Muñoz RF. Differential item functioning in a Spanish translation of the Beck Depression Inventory. J Clin Psychol 2001; 57: 355-65.

25. García Campayo J, Rodero B, Alda M, Sobradiel N, Montero J, Moreno S. [Validation of the Spanish version of the Pain Catastrophizing Scale in fibromyalgia]. Med Clin (Barc) 2008; 131: 487-92. Spanish.

26. 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: e0160866.

27. Yarnitsky D, Bouhassira D, Drewes AM, Fillingim RB, Granot M, Hansson P, et al. Recommendations on practice of conditioned pain modulation (CPM) testing. Eur J Pain 2015; 19: 805-6.

28. Doménech-García V, Palsson TS, Herrero P, Graven-Nielsen T. Pressure-induced referred pain is expanded by persistent soreness. Pain 2016; 157: 1164-72.

29. Doménech-García V, Skuli Palsson T, Boudreau SA, Herrero P, Graven-Nielsen T. Pressure-induced referred pain areas are more expansive in individuals with a recovered fracture. Pain 2018; 159: 1972-9.

30. Boudreau SA, Badsberg S, Christensen SW, Egsgaard LL. Digital pain drawings: assessing touch-screen technology and 3D body schemas. Clin J Pain 2016; 32: 139-45.

31. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016; 15: 155-63. Erratum in: J Chiropr Med 2017; 16: 346.

32. Furlan L, Sterr A. The applicability of standard error of measurement and minimal detectable change to motor learning research-a behavioral study. Front Hum Neurosci 2018; 12: 95.

33. 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: 38-41. Erratum in: Scand J Pain 2016; 13: 17.

34. Black CD, Pickowitz KE. Day-to-day reliability of pressure pain threshold and pain ratings in college-

aged men. Int J Rehabil Res 2015; 38: 213-8.

35. Walton DM, Levesque L, Payne M, Schick J. Clinical pressure pain threshold testing in neck pain: comparing protocols, responsiveness, and association with psychological variables. Phys Ther 2014; 94: 827-37.

36. Kvistgaard Olsen J, Fener DK, Waehrens EE, Wulf Christensen A, Jespersen A, Danneskiold-Samsøe B, et al. Reliability of pain measurements using computerized cuff algometry: a DoloCuff reliability and agreement study. Pain Pract 2017; 17: 708-17.

37. Graven-Nielsen T, Vaegter HB, Finocchietti S, Handberg G, Arendt-Nielsen L. Assessment of musculoskeletal pain sensitivity and temporal summation by cuff pressure algometry: a reliability study. Pain 2015; 156: 2193-202.

38. Arroyo-Fernandez R, Bravo-Esteban E, Domenech-Garcia V, Ferri-Morales A. Pressure-induced referred pain as a biomarker of pain sensitivity in fibromyalgia. Pain Physician 2020; 23: E353-62.

39. Kennedy DL, Kemp HI, Ridout D, Yarnitsky D, Rice ASC. Reliability of conditioned pain modulation: a systematic review. Pain 2016; 157: 2410-9.

40. Lewis GN, Heales L, Rice DA, Rome K, McNair PJ. Reliability of the conditioned pain modulation paradigm to assess endogenous inhibitory pain pathways. Pain Res Manag 2012; 17: 98-102.

41. Valencia C, Kindler LL, Fillingim RB, George SZ. Stability of conditioned pain modulation in two musculoskeletal pain models: investigating the influence of

shoulder pain intensity and gender. BMC Musculoskelet Disord 2013; 14: 182.

42. Roldan CJ, Abdi S. Quantitative sensory testing in pain management. Pain Manag 2015; 5: 483-91.

43. Mason KJ, O'Neill TW, Lunt M, Jones AKP, McBeth J. Psychosocial factors partially mediate the relationship between mechanical hyperalgesia and self-reported pain. Scand J Pain 2018; 18: 59-69.

44. Polianskis R, Graven-Nielsen T, Arendt-Nielsen L. Spatial and temporal aspects of deep tissue pain assessed by cuff algometry. Pain 2002; 100: 19-26.

45. Jensen B. Chronic pain assessment from bench to bedside: lessons along the translation continuum. Transl Behav Med 2016; 6: 596-604.

46. Palsson TS, Christensen SWM, De Martino E, Graven-Nielsen T. Pain and disability in low back pain can be reduced despite no significant improvements in mechanistic pain biomarkers. Clin J Pain 2021; 37: 330-8.

47. McPhee ME, Graven-Nielsen T. Recurrent low back pain patients demonstrate facilitated pronociceptive mechanisms when in pain, and impaired antinociceptive mechanisms with and without pain. Pain 2019; 160: 2866-76.

48. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med 1998; 17: 101-10.

49. Kamper SJ. Showing confidence (intervals). Braz J Phys Ther 2019; 23: 277-8.