Determining real change in conditioned pain modulation: a repeated measures study in healthy volunteers

Harriet I. Kemp , Donna L. Kennedy , Chenxian Wu , Deborah A. Ridout , Andrew S.C. Rice

PII: S1526-5900(19)30847-8
DOI: https://doi.org/10.1016/j.jpain.2019.09.010
Reference: YJPAI 3810

To appear in: *Journal of Pain*

Received date: 12 May 2019
Revised date: 29 August 2019
Accepted date: 19 September 2019

Please cite this article as: Harriet I. Kemp , Donna L. Kennedy , Chenxian Wu , Deborah A. Ridout , Andrew S.C. Rice , Determining real change in conditioned pain modulation: a repeated measures study in healthy volunteers, *Journal of Pain* (2019), doi: https://doi.org/10.1016/j.jpain.2019.09.010

Determining real change in conditioned pain modulation: a repeated measures study in healthy volunteers

Harriet I Kemp[a*]

Donna L Kennedy[a*]

Chenxian Wu[b]

Deborah A Ridout[c]

Andrew SC Rice[a]

* Joint first authors

[a] Pain Research Group, Department of Surgery & Cancer, Imperial College London, United Kingdom

[b] Imperial College London School of Medicine, United Kingdom

[c] Population, Policy and Practice Programme, UCL Great Ormond Street Institute of Child Health, University College London, United Kingdom

Number of pages: 39

Total number of Figures: 6

Total number of Tables: 4

Total number of Supplementary Material: 6

Corresponding author

Dr Donna L Kennedy

Pain Research Group, Department of Surgery & Cancer

Imperial College London

Chelsea & Westminster Hospital Campus

369 Fulham Road

London

SW10 9NH

United Kingdom

Tel: +44 (0)20 3315 8816

d.kennedy@imperial.ac.uk

Highlights

- Reporting in CPM is not standardized and does not consider measurement error
- A distribution-based approach enables the identification of "real" change in CPM
- The proportion of CPM inhibitors and facilitators is paradigm dependent
- Inter-session CPM response is not consistent in healthy subjects
- Standardisation in reporting will underpin emerging clinical utility of CPM

Disclosures

D. L. Kennedy's work was funded by a National Institute for Health Research (NIHR) and Health Education England (HEE) Clinical Doctoral Research Fellowship. H. I. Kemp was funded by a European Commission, NeuroPain FP7 Grant EC (#2013-

602891). D. Ridout has received financial support from the NIHR and HEE for her contribution to this work. Professor A. S. C. Rice received financial support from DOLORisk, a European Union Horizon 2020 research and innovation programme (Grant 633491) for his contribution to this manuscript. Study sponsors provided financial support for the conduct and reporting of this research, they did not contribute to study design; nor collection, analysis or interpretation of data; in the writing of this report; nor in the decision to submit the article for publication.

Conflict of interests

HIK, DLK, CW, DAR – none declared

ASCR - conflicts of interest occurring in last 24 months:

- Orion Pharma funding
- Consultancy and advisory board work for Imperial College Consultants including remunerated work for: Pharmanovo, Galapagos, Toray, Quartet, Lateral, Novartis, Pharmaleads, Orion, Asahi Kasei & Theranexis
- Owner of share options in Spinifex Pharmaceuticals from which personal benefit accrued upon the acquisition of Spinifex by Novartis in July 2015 and from which future milestone payments may occur.

Inventor on patents:

- Rice A.S.C., Vandevoorde S. and Lambert D.M Methods using N-(2-propenyl) hexadecanamide and related amides to relieve pain. WO 2005/079771
- Okuse K. et al Methods of treating pain by inhibition of vgf activity EP13702262.0/ WO2013 110945

Abstract

Conditioned pain modulation (CPM) is a potentially useful biomarker in pain populations; however, a statistically robust interpretation of change scores is required. Currently, reporting of CPM does not consider measurement error. Hence, the magnitude of change representing a "true" CPM effect is unknown. This study determined the standard error of measurement (SEM) and proportion of healthy participants showing a 'true' CPM effect with a standard CPM paradigm.

Fifty healthy volunteers participated in an intersession reliability study using pressure pain threshold (PPT) test stimulus and contact heat, cold water and sham conditioning stimuli. Baseline PPTs were used to calculate SEM and > +/- 2x SEM to determine CPM effect.

SEM for PPT was 0.21 kg/cm^2. An inhibitory CPM effect (>+2SEM) was elicited in 59% of subjects in response to cold stimulus; in 44% to heat. Intrasession and intersession reliability of within-subject CPM response was poor (kappa coefficient <0.36).

Measurement error is important in determining CPM effect and change over time. Even when using reliable test stimuli, and incorporating measures to limit bias and error, CPM intersession reliability was fair and demonstrated a large degree of within-subject variation. Determining "true" change in CPM will underpin future interrogations of intra-individual differences in CPM.

Perspective

This study used a distribution-based statistical approach to identify real change in CPM, based on the standard error of measurement for the test stimulus. Healthy volunteers demonstrate substantial within-subject variation; CPM effect was paradigm dependent at intra-session testing and unstable to the same paradigm at intersession testing.

Key words

Conditioned pain modulation; pressure pain threshold; measurement error; intersession test stability

Introduction

Conditioned pain modulation (CPM) is a psychophysical measure thought to be the human correlate of diffuse noxious inhibitory control (DNIC) (49), initially identified electrophysiologically in rats (22) and suggested as early as 1937 in humans as a phenomenon where 'pain inhibits pain' (9). It is proposed that pain at one site, the 'conditioning stimulus', can modulate (facilitate or inhibit) the experience of pain at a second distant site, the 'test stimulus' (34) via a spino-bulbar-spinal loop (30).

CPM has been suggested as a method for predicting response to treatment in chronic pain patients, as less efficient CPM appeared to correlate with the efficacy of therapies augmenting descending inhibitory systems, such as duloxetine (51). Assessment of CPM prior to intervention has also been suggested as a method for predicting the development of post-surgical pain (38, 50). However, despite consensus work aimed at developing the clinical application of CPM testing (49, 48) our recent review highlighted that methods for determining and reporting CPM effect are not standardised (20).

Consideration is required in the interpretation of CPM change scores and in how participants are classified as demonstrating inhibition or facilitation. At present, if the conditioned test stimulus is rated as less painful than at baseline, regardless of the magnitude of change, generally this is described as CPM inhibition. Conversely, if it is rated as more painful, this is described as facilitation (48). Dichotomizing CPM effect as such suggests participants who are separated by a minute difference in

scores are very different to one another (inhibitors versus facilitators) when in reality they exhibited a similar response (2).

Current practice in reporting of CPM effect usually fails to consider measurement error. All tests perform with some degree of error; therefore, a subject's observed score will be a composite of their 'true score' plus measurement error (25, 33, 35). Measurement error is multifactorial and may be intrinsic to the test stimulus, due to inherent variability in the characteristic being measured or random error, as a result of performance factors such as attention (25,33). For CPM, change in the conditioned test stimulus score must exceed measurement error to be considered a 'true' change or discernible from measurement error.

While there is good evidence for a statistical approach to interpreting change scores in physical measurements (15, 18, 44, 11), this concept is new to psychophysical measures such as CPM . At present, evidence for what constitutes clinically important change in CPM is lacking and we are reliant upon distribution-based approaches for interpreting change scores (41, 27). Most commonly, standard error of measurement (SEM) is used to determine change greater than that due to measurement error, or "real" change in test scores (33, 35, 5, 4, 46). Previously, Locke et al (26) reported SEM to describe CPM effect, interpreting change >1 SEM as meaningful change. However, because SEM is interpreted according to the properties of a Gaussian distribution, there is a 68% chance that a subject's true score falls within ±1 SEM and a 95% chance it falls within ± 2 SEM (32) (or a 5% chance change >±2 SEM is due to measurement error). The use of >±2 SEM is a

statistically robust cut-point (33, 17, 16) and may improve precision in delineating inhibitory and facilitatory CPM effects. Such an approach may aid the identification of CPM paradigms which are most effective, i.e. induce a consistent CPM effect in the greatest proportion of subjects, as necessitated for research and clinical practice.

This study aimed to determine the SEM for a test stimulus and use >±2 SEM to identify change in CPM effect with greater certainty, i.e. change greater than measurement error. It also aimed to identify the CPM effect in healthy volunteers whilst attempting to limit methodological bias, report to standards suggested in a recent systematic review (20), and elucidate inter-session reliability of the test stimulus, conditioning stimulus and CPM effect.

Methods

The REACTION (Reliability and Measurement Error in Conditioned Pain Modulation) study undertook a repeated-measures observational design. REACTION was approved by Imperial College Ethics Committee (IC163176), and healthy volunteers were recruited by advertisement to the general public, and to students and staff of Imperial College London. No financial incentive was offered to participants. Participants signed a written consent form, attended two appointments, 28 days apart and underwent the same test protocol at each appointment. Experiments were conducted at Chelsea & Westminster Campus of Imperial College between March and September 2016.

Participants

Participants were aged 18 years or older and completed a general health questionnaire to screen for any health complaints prior to recruitment. Those with any medical diagnosis (including hypertension), or taking medication other than an oral contraceptive, were excluded. Subjects were also excluded if they reported acute pain in the 48 hours prior to the appointment, if they had taken any analgesic, hypnotic or antidepressant medication in the last 72 hours, if they use illicit drugs and if they were a regular cigarette smoker. This information was collected as the 'baseline screening' reported in the consensus paper published to define a 'healthy volunteer' in psychophysical testing of pain (12).

Participants were asked to refrain from rigorous exercise and drinking caffeine or alcohol for four hours prior to each appointment. Basic demographic information was collected including age, gender, body mass index and ethnicity. Level of education, number of hours of exercise per week, and caffeine intake were also recorded as these have been shown to influence CPM effect. The two sessions were separated by 28 days to limit the influence of the menstrual cycle, although it is recognized that not all subjects would have a 28 days cycle.

Volunteers completed four questionnaires at both appointments. The Hospital Anxiety and Depression Scale (HADS) (52), the Spielberger State-Trait Anxiety Inventory (STAI form Y1 and Y2) (40), the Pain Catastrophizing Scale (42) and the Pain Sensitivity Questionnaire (PSQ) (39) were used to examine the effect of any psychological or personality factors on magnitude and variability in CPM.

Test protocol

Subjects were block randomized to one of three groups, using randomization software. Group allocation determined the order in which subjects received the three different conditioning stimuli (CS) in the test protocol. The reason for allocating subjects to three groups was to attempt to control for order effect of delivery of the conditioning stimulus. Group A received the heat (25x50mm Somedic electrode on the left volar surface of the distal forearm at 46.5°C), then cold (left hand held in a circulating water bath at 12°C), then sham CS (left hand held in circulating water bath at 24°C); group B the cold, then sham, then heat CS; and group C the sham, then heat, then cold CS. All CS were delivered for 90 seconds and the same order was used at both test sessions. The sham stimulus was not intended as a control measure, but to enable the evaluation and comparison of the psychophysical response to a non-noxious stimulus (13).

All participants were tested in the same room, with the same equipment. Two out of three investigators (CW, HIK, DLK) tested all participants and the same investigator had the same role at each appointment for each participant. The temperature of the room was recorded for each appointment. Instructions were read aloud from a script (Supplemental Material 1) to ensure consistent instructions were given to all participants.

The test stimulus was pressure pain threshold (PPT) determined using a pressure algometer (FDN200, Wagner instruments, Greenwich CT, USA), applied at 1 kg/cm^2

per second, on the right forearm at the medial extensor muscle bulk, 2cm distal to the elbow crease. Participants were instructed to say “now” as soon as the usual sensation of pressure changed to an additional sensation of “burning”, “stinging”, “drilling” or “aching” (36). The PPT was repeated three times and the arithmetic mean calculated to give the baseline PPT (as per the German Neuropathic Pain Network [DFNS] Quantitative Sensory Testing protocol (37). The baseline test stimulus was delivered prior to the CS each time. The second TS was delivered in parallel to the CS, starting at 60 seconds after initiation of the CS to allow for the stimulus to become painful and for the intensity to reach a steadier state. It has been suggested that a sequential CPM paradigm, whereby the conditioned test stimulus is evaluated after removal of the conditioning stimulus, may be less influenced by factors such as distraction and therefore a purer measure of the CPM effect (49). However, there is limited evidence for the reliability of sequential CPM paradigms (20), therefore a parallel paradigm was utilized in this study.

Subjects were asked to rate the discomfort of the CS every 10 seconds for 60 seconds and the mean calculated to provide a measure of CS intensity. The word ‘discomfort’ was used instead of ‘pain’ to reduce any anticipatory anxiety and so as not to confuse the participant if the conditioning stimulus was not painful. The second TS was delivered in the same manner as the baseline TS whilst the CS continued. There was a 15-minute break between paradigms to allow for any CPM effect to dissipate (24). Prior to each of the three paradigms, participants were asked to score on a 0 to 100 numerical rating scale how much pain they felt in their right forearm and left hand and arm to ensure they were pain free or that their pain had

returned to a baseline of zero. The outline of the experimental protocol is shown in Figure 1.

Blinding

Each participant was allocated a lead and second investigator that was kept constant at each test session. The lead investigator was not blinded as to group allocation and was responsible for screening, enrolment and delivery of the test instructions and CS. The second investigator was blinded to the group allocation and therefore the order of conditioning stimuli. The second investigator entered the room only to deliver the test stimulus (TS). The subjects' arm was draped therefore covering either the water bath or electrode thermode, so that the second investigator could not identify what CS was being delivered whilst they tested the TS (shown in Figure 2). The water bath remained running during all paradigms throughout the procedure so as not to provide auditory clues as to what CS was being delivered. Subjects were unaware of the test hypothesis and were given similar instructions prior to each stimulus. Subjects with knowledge of CPM (if they worked in pain research or could explain what CPM was) were excluded. Between paradigms at each session, subjects had no exposure to other participants or investigators outside the experiment that might influence or bias the subjects' response.

Statistical Analysis

All subjects were reported; there was no exclusion of 'non-responders'.

i) Measurement error

Standard error of measurement (SEM), a measure of absolute reliability or consistency, was used to identify change in the conditioned test stimulus greater than that indistinguishable from measurement error. The SEM is advocated in reliability studies as it provides an indication of the precision of a score and enables the construction of a confidence interval for scores (35). For clarity, SEM is distinct from standard error of the mean (SE), a measure of precision of the sample mean (1). SEM for the test stimulus (in this study the PPT) was calculated using the repeated baseline PPT measures (each calculated as the mean of the three PPTs as per DFNS protocol (37) delivered prior to each conditioning stimulus at both sessions. Therefore, the six baseline test stimuli, three from day 1 and three from day 28, were included in the following formula (35, 32):

$$SEM = \text{standard deviation of baseline PPT} \times \sqrt{(1-\text{reliability baseline PPT})}$$

A priori, an inhibitory CPM effect was determined to be an increase in PPT greater than 2 SEM, facilitatory CPM a decrease in PPT greater than 2 SEM (32). This allowed each subject to be allocated to one of three groups: inhibitor, facilitator or non-responders (those with a response indistinguishable from measurement error; i.e. a change less than ± 2 SEM). This group allocation was performed for each conditioning stimulus at each test session.

ii) Reliability of the CPM effect

Intraclass correlation coefficients (ICC 2,k) were used to assess reliability of the TS, CS and CPM effect of each different paradigm and reported with 95% confidence

intervals. ICC was chosen as a measure of reliability as it provides information on both the association and the agreement between ratings and to allow for comparison with other published reliability studies.

iii) Effect of order of conditioning stimuli

Between group comparisons (groupings based on order of delivery of conditioning stimulus) of participant characteristics were made using ANOVA or chi-squared tests and data presented as mean (sd). The CPM effect was also compared between groups using an ANOVA to explore if the order of delivery of stimuli influenced CPM effect. Correlation between CPM response to a heat and cold conditioning stimulus was examined using Pearson's correlation.

iv) Factors influencing CPM response

Hierarchical multiple regression, incorporating group allocation, was used to test whether the following factors were associated with CPM effect size on day one, and a separate analysis performed for day 28; gender, age, psychological and personality measures and hours of exercise. Similar regression was used to test the effect of these factors on the size of the difference between the CPM effect on day one and day 28, as a measure of magnitude of variation. Only psychological or personality measures showing a significant bivariate correlation with either noxious stimuli at either test session were included in the regression model.

v) Group comparison based on 2xSEM cut-off

A between-group analysis was performed to compare subject characteristics between those defined as inhibitors, facilitators or non-responders. Comparisons were made using an ANOVA or Fisher's exact test (with Freeman Halton extension). Within subject variability, or the reliability of subjects to remain in the same group allocation (inhibition, facilitation, non-responder) for both noxious thermal conditioning stimuli at each session and for each conditioning stimulus on repeated testing at day 28, was conducted using the kappa co-efficient and reported as κ, p value, 95% confidence interval. The coefficient was interpreted as :<0 as no agreement, 0.01-0.2 as non to slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial and 0.81-1.0 as near perfect agreement (3).

Sample Size Calculation

The study aims included comparison of standard error of measurement, reliability of pain ratings, CPM effect and magnitude of CPM between study groups, therefore the requisite sample size to identify differences between groups was calculated. Sample size was determined with a correlation sample size calculation using the repeated measures ICCs reported in previous studies of CPM (Cathcart et al. (7) (ICC = 0.57); Wilson et al (47) (ICC= 0.39): Lewis et al 2012 (23) (ICC = 0.65); Valencia et al 2013 (43) (ICC = 0.68)). A sample size of 50 (16 per group) was required to achieve a power of 80% and a level of significance of 5% for robust group comparisons.

Results

Fifty participants were recruited to the three test groups. The mean number of days between the two test sessions was 28.2(4.8) days and eight women were using the

oral contraceptive pill. One subject did not return for the second test session as they were not available to test during the 28-day window of the second appointment. Five subjects reported on-going pain at a test site as a result of either the test or conditioning stimuli at the end of at least one of the 15-minute breaks between paradigms (for three subjects this occurred after the heat and prior to the cold CS, in one prior to the heat, following the cold stimulus and in one prior to the sham following the cold stimulus). Pain intensity in those five ranged between 2 and 20 out of 100. The only difference shown to be significant between groups was that Group A was significantly younger than subjects in Group C (Table 1).

CPM effect

There were no significant differences in absolute CPM effect, baseline PPT or score for conditioning stimulus between groups, thus CS order was not shown to be a significant factor (See Supplemental Material 2).

The CPM effect, expressed as an absolute change in PPT in response to the conditioning stimuli, is presented in Figure 3 for both sessions (and as percent change in Supplemental Material 3. Raw PPT pre and during CS are presented in Supplemental Material 4). The mean percent change in PPT during conditioning stimulus on day 1 was 18.2(23.33)% for the heat CS, 26.2(21.49)% for the cold CS and 1.1(13.33)% for the sham stimulus. The largest CPM effect was elicited in response to the cold conditioning stimulus, which was also rated as most painful.

There was inconsistent correlation within individuals for CPM effect in response to heat and cold. On day 1 there was no correlation between subject absolute CPM response to cold and heat (r=0.187, p=0.193), However, on day 28 there was a moderate statistically significant correlation (r-0.431, p=0.002).

Standard error of measurement of the test stimulus

Since SEM is based on a normal distribution, the PPT data were tested for normality. The SEM for the PPT across all baseline PPT tests at both sessions was +/- 0.21kg/cm^2 (using an ICC 0.96). Figure 3 shows 2xSEM (0.42 kg/cm^2) in the context of the results for CPM effect.

Group allocation based on ± 2xSEM

The percentage of subjects who demonstrated inhibition, facilitation or were non-responders to the three CPM paradigms in shown in Figure 4. For each conditioning stimulus, group allocation is contrasted comparing a response greater than 2x the SEM and any magnitude of change in the conditioned test stimulus.

The results of the comparison of demographic factors or questionnaire responses identified between the inhibitor, facilitator or non-responder groups is presented in Supplemental Material 5. On Day 1, significantly more males had an inhibitory response to the cold CS (14 versus 5, p<0.001) but this difference was not identified at the second test session. Although facilitators to the cold CS on day 28 reported a higher mean number of hours exercise, the group size was very small (n=2).

Conditioning stimulus

The mean discomfort score reported for the tonic heat CS was 39.3(22.89) during the first test and 37.1(21.77) on day 28; for the cold water CS was 50.7(23.73) on day one and 50.23 (23.18) on day 29; and for the sham CS was 2.99 (9.20) on day one and 3.2 (7.30) on day 28.

Drop outs

One subject did not attend the second appointment. In this subject, the CPM effect during the first test session to heat was 10.5% and to cold was 6.7% therefore was within the IQR for the total cohort for the heat CS but slightly below the IQR for the cold CS. At one of the testing sessions, two subjects could not tolerate immersion of their hand in the cold water stimulus for 90 seconds on their first attempt. However, on a second attempt after a 10 minute interval, they were able to tolerate the cold conditioning stimulus for 90 seconds as per the test protocol.

Reliability

The intraclass correlations for PPT within and between sessions and for the discomfort scores for conditioning stimuli and CPM effect between sessions are reported in Table 2. Pressure pain threshold showed excellent reliability both within and between sessions, as did the discomfort scores elicited by all three conditioning stimuli. The CPM effect when reported as an absolute change was associated with a slightly higher ICC than when reported as a percentage change. CPM effect in response to a cold CS showed slightly higher reliability than the response to a heat CS.

Within-subject variation

Subjects did not consistently show an inhibitory or facilitatory response to a CS, between sessions as illustrated in Figure 5 or within test sessions, as illustrated in Figure 6. The kappa coefficients comparing inter-session group allocation using the >±2SEM cut-offs for identifying participants as non-responders, inhibitors or facilitators to the study paradigms are shown in Table 3, intrasession kappa coefficients are reported in Table 4. The intersession and intrasession agreement of group allocation was poor to fair.

Factors influencing CPM effect

No significant correlation was identified between the magnitude of CPM effect and how painful the subject reported the conditioning stimulus at either test session. Hierarchical regression revealed no significant effect of group (i.e. conditioning stimulus order) on CPM effect with heat or cold as a conditioning stimulus. The only factors revealed to have an independent significant influence on CPM effect for the heat conditioning stimulus was gender, as females showed a smaller CPM effect on day 1 (B -19.65, p=<0.001) but this was not apparent on day 28 (B 6.00, p=0.88). A higher score on the STAI_Y2 questionnaire, a measure of higher state anxiety, was shown to be associated with increase in CPM effect to the cold stimulus on day 28 only. Age, number of hours of exercise, pain sensitivity and trait anxiety were not significantly associated with CPM effect (see Supplemental Material 6).

Discussion

Determining what magnitude of change constitutes a "true" CPM effect presents a challenge. Consequently, to date CPM has been reported as a continuous measure, enabling between groups comparisons. Importantly, reporting CPM along a continuum has provided evidence for differences in CPM function between chronic pain populations and healthy controls (24) and an association of CPM function and response to pharmacological intervention in patients with painful diabetic neuropathy and painful knee osteoarthritis, respectively (51, 10).

The REACTION study aimed to move from such group level comparisons to exploring CPM effect at the level of the individual. To suggest that subjects with small differences in CPM response have exhibited different effects and are physiologically different is likely unwarranted. However, at present, the classification of CPM effect is not standardized. Meaningful change in scores can be determined using distribution-based methods (27,8). Distribution-based interpretation is based on the statistical distribution of study results; in the present study the 95% confidence interval for standard error of measurement (± 2 SEM) of a commonly used test stimulus in CPM protocols, the pressure pain threshold, in a cohort of healthy volunteers.

CPM effect was determined based on the magnitude of change in the conditioned test stimulus, with change greater than +2SEM interpreted as inhibition and -2SEM as facilitation. Those with change scores <±2SEM were categorized as non-responders to the paradigms under investigation. Categorizing CPM effect as such may aide the identification of paradigms which induce the most stable CPM effect in

the greatest proportion of participants. Further, this will enable the exploration of inter-individual differences in CPM function, broadening understanding of the multiple factors associated with CPM effect and CPM stability over time. Categorizing CPM effect may underpin future predictive studies, whereby CPM is used as a biomarker to identify those at risk of developing chronic pain or those who may benefit from certain therapeutic interventions. Lastly, a statistically robust approach to categorizing CPM effect will improve confidence in the determination that CPM has indeed been “rescued” by an intervention.

The mean percent change CPM response for heat was 18.2% and cold was 26.2%, comparable to the mean percent change CPM response derived in a systematic review of healthy volunteers (34). This review included studies using a range of test and conditioning stimuli and demonstrated that CPM response is conditioning stimulus dependent with more painful stimuli eliciting larger changes. Therefore, our protocol elicits a CPM response comparable to previous healthy volunteer studies and suggests that, if the SEM were considered in such studies, the number of subjects classified as "inhibitors" or "facilitators" would be reduced.

SEM has been reported previously as a measure of response stability for the CPM effect in a patient population (43) and in healthy volunteers (28). However, this approach fails to consider if the magnitude of change in the conditioned test stimulus exceeds measurement error and does not address the question as to whether a CPM effect has been induced. As noted, Locke et al. (26) used ± 1SEM to determine a clinically meaningful CPM effect in healthy volunteers. The investigators

reported 92.8% of participants demonstrated a CPM effect (change in the conditioned test stimulus >1SEM). However, reliability data including SEM were calculated from a sub-population of ten participants rather than the sample of 133 participants, therefore the SEM may not have been representative of measurement error for the sample. And importantly, statistically, the observed score ±1SEM provides a 68% confidence interval for a participant's true score, lacking precision in the interpretation of change scores (32). In the interpretation of change scores, the observed score ±2 SEM provides a 95% confidence interval for a subject's true score (46) affording precision and therefore confidence in the determination of change greater than measurement error, as reported here.

Determination of SEM across test sessions

The REACTION study included two test sessions with a 28-day retest interval. SEM was determined for the test stimulus based on the reliability of three mean scores across test sessions, representative of measurement error over time and therefore real change in repeated measures. For comparison, the SEM for each individual test session (3 PPT from each session) was also calculated; it differed little from the SEM for 6 PPT measures combined for both sessions (Day 1 SEM=0.23; Day 28 SEM=0.26; Combined SEM=0.21). If individualized SEM were used for each test day, a maximum of two further subjects would have been allocated to the non-responder group. This indicates the SEM is relatively stable over time.

It is likely that SEM is population and paradigm specific and therefore should be determined for individual populations in future CPM studies. This requires repeated

measures of the test stimulus to identify stimulus reliability (required to calculate SEM) in a population and experimental condition. Whilst this may seem burdensome, the suggestion of using repeated measures and a range of stimuli within the same study has already been suggested by a consensus agreement (49) as a way of working towards identification of an optimal test paradigm for measuring CPM. This approach need not necessitate repeated test sessions, but the addition of one baseline test stimulus measure incorporated into an experimental protocol.

Non-responders

It is anticipated that categorizing participants as "non-responders" serves two purposes. Firstly, identifying those subjects with minimal or no response may aid the investigation of demographic, physiological or psychosocial factors associated with a limited CPM response. Secondly, should CPM response be used to select patients for inclusion to a clinical trial, use of such robust cut-offs would allow for selection of the most extreme phenotypes and therefore potentially identify those most likely to show a response to any intervention tested.

Facilitators

Our findings highlight that CPM facilitation is a relatively frequent finding (6% of test paradigms), even in a healthy population. It is important that these subjects not be excluded from analysis but that further investigation is performed to identify subject specific characteristics associated with a facilitatory response. Inconsistent reporting of study attrition, including incomplete reporting of results from facilitators in CPM

studies has been highlighted as a risk of bias (33). This may impact on studies of patient populations where a facilitatory response is highlighted as a sign of pathology attributed to the chronic pain condition, rather than potentially being within the range of 'healthy' responses.

It is not clear from the results of this study that individuals can be defined as 'CPM responders' across several different conditioning stimuli as, although there was a correlation in CPM response to two different stimuli on one test day, this was not apparent at the other test session.

Sham paradigm

Responses to the sham stimulus in particular highlight the issue of measurement error. If 'any change in PPT' was taken as a measure of CPM effect, 78% of subjects demonstrated a CPM effect to a non-noxious stimulus, whereas only 24% of participants reported the stimulus elicited pain. In contrast, if a cut-off of +/-2xSEM is used to determine CPM effect, then only 27% of subjects demonstrated a CPM response to a sham conditioning stimulus. It is not clear why nearly a quarter of subjects reported tepid water as painful, even if the pain score given was low. This may indicate an element of expectation or potentially residual sensitization from a previous conditioning stimulus. This response to a sham paradigm may be distinct from a CPM response, instead reflecting the distraction associated with delivering even a non-noxious stimulus and highlights the challenges of using a paradigm where the conditioning and test stimuli are delivered in parallel.

CPM reliability and stability

Our results replicate findings identified in a systematic review (20) which indicate that although the test stimulus and amount of pain reported in response to the conditioning stimulus are very reliable between sessions (ICCs>0.8), the CPM effect itself is less reliable (ICCs 0.11-0.58). This is despite attempts at controlling for known sources of participant variation such as caffeine intake, exercise, state of anxiety, and confirming their status as healthy volunteers.

Methods to control experimental bias such as the use of a scripted protocol, testing in an isolated and replicated environment for each test session, as well as blinding the investigator as to the conditioning stimulus, also did not appear to improve the stability of the CPM response, when compared to ICCs previously reported.

The intersession reliability of CPM has been explored in numerous studies, both with healthy volunteers and in clinical cohorts and employing numerous modalities for administration of painful test and conditioning stimuli. Intersession reliability of the CPM effect has ranged from poor to excellent, demonstrating that CPM reliability is modality and parameter dependent (20). In the present study, intersession reliability was interpreted as fair (ICC = 0.4 – 0.59) for CPM effect using both cold water and contact heat as conditioning stimuli. This is consistent with results reported in previous intersession reliability studies for a PPT- Cold CPM parallel paradigm (29, 19). However, in contrast, Lewis et al (23) reported good intersession reliability (ICC= 0.66 [0.22-0.87]) for the same paradigm with a shorter retest interval of three days and Marcuzzi et al (28) reported poor reliability (ICC= 0.35 [0.16-0.54]) with a retest

interval of four months, suggesting the reliability of the paradigm decreases over time.

It is important to note that 10% of subjects reported persistent pain, albeit of low intensity, as a result of either the test or conditioning stimulus, even after 15 minutes rest. This indicates that although a CPM response may have dissipated (24) a true return to baseline has not been achieved. Therefore future studies involving repeated paradigms should either monitor for persistent pain or incorporate an increased interval time between tests. However, it is likely that the length of time required to ensure return to 'pain-free' baseline is population and test modality specific and may well be different between healthy and patient cohorts.

Standards of Reporting; blinding; data presentation

In this study and the reporting of such, we have attempted to improve rigor and transparency and reduce the risk of bias noted in previously published CPM reliability studies (20). We reported our study according to STROBE guidelines for observational studies to ensure thoroughness and promote standardization in future reporting (45). We have reported a full description of our recruitment strategy, sample characteristics and study participant attrition. We have been novel in introducing investigator blinding to the CPM paradigm. This may be important as, although PPT is a patient reported threshold, administration of the algometer is by the investigator, rather than an automated machine and therefore maybe susceptible to bias. It was anticipated that using a test script would improve consistency and ensure participant blinding to the extent feasible in such a

psychophysical experimental study. We controlled for multiple cofounders including alcohol, caffeine, exercise, smoking, menstrual cycle and pain medication. Test sessions included repeated CPM evaluations, therefore between measures we asked participants to re-evaluate pain, ensuring participants had returned to the baseline testing state prior to subsequent measures. To improve transparency and reduce bias in statistical reporting, we a priori determined the magnitude of CPM effect to be considered inhibition and facilitation (+2 and -2 SEM, respectively). We have presented data to support absolute and relative reliability and importantly have introduced a novel method for presentation of CPM results which aids the interpretation of what is likely to be a true effect and what change may be attributable to measurement error.

Limitations

This study protocol did not include a supra-threshold test stimulus due to limitation of time available with each participant. There is evidence that these stimuli show different reliability to pain thresholds (49) and these should also be investigated in the context of measurement error. The protocol also did not control for habituation as the baseline TS was always measured before the TS under the influence of the conditioning stimulus. Although habituation may reduce the change in PPT observed, this should have been consistent between test paradigms as all were delivered in a similar manner.

Participants were asked to rate the “discomfort” of the thermal conditioning stimuli, rather than the painfulness of the stimuli. This was done to avoid creating confusion

for the participants on exposure to the sham stimulus or the expectation of pain upon stimulus exposure. This terminology was not anticipated to impact on participants' CPM effect as pre-determined intensities for thermal CS were used (13) however this may be a limitation to the robustness of the CPM protocol.

For efficiency, participants completed state depression, anxiety, and pain catastrophizing questionnaires during fifteen-minute breaks between the three test paradigms. However, interspersing such state questionnaires with pain-evoking procedures may have influenced the results of the experimental pain procedures and is therefore a limitation for consideration.

Blinding of the subjects in CPM protocols is difficult and not been previously attempted. While recognized that it is not possible to blind participants to the applied test stimuli, we did ensure participants were "blinded" to the study question, the nature of CPM and to other participants or environmental stimuli that may have influenced participant's responses to the experimental measures. It is possible that having two, rather than the usual one examiner, and that one examiner entering and exiting the room could enhance anxiety or attention towards the stimuli delivered. Draping of the participants arm may also alter visual cues that may affect pain perception (31). This does not appear to be the case however as the magnitude of CPM effect is similar to the mean CPM effect reported by Pud et al in their meta-analysis (34).

This study only determined the SEM for one TS, pressure pain threshold, measured as described in the methods. The SEM will be modality dependent and therefore should be determined for individual test protocols.

Although higher and lower temperature conditioning stimuli (i.e. more painful stimuli) may be associated with a larger CPM response, Granot et al. (13) showed that 12C and 46.5C were sufficient to produce a CPM response. The temperatures in this study were selected to be clinically relevant as pilot work showed that, in particular, lower temperatures to be intolerable. Indeed two subjects in this study could not tolerate 12C at first attempt. In studies where CPM paradigms have been repeated across sessions, the use of lower temperatures, e.g. 4C for the CS does not appear to be increase reliability (20). It can also be argued that using more painful stimuli has the potential to induce more distraction thereby producing a paradigm that measures not only the true CPM effect but also the distractive element.

The observed CPM effect may be a combination of the spino-bulbar-spinal sensory phenomenon as well as non-sensory descending modulation such as attention and anticipation. This study was not designed to elicit the contribution of each of these processes.

Whilst the use of 2xSEM as a cut-off is the most established method of determining change greater than that due to measurement error, other methods of determining an 'important' change in CPM response should be investigated. For example, this could include integrating a subjects' response to sham. However, such methods may

prove difficult in light of the identified stimulus-dependent response and limited reliability of CPM.

Conclusion

This study highlights the importance of measurement error in the interpretation of a change in a test stimulus during exposure to a conditioning stimulus. Interpreting CPM effect as change greater than ±2 standard error of measurement for a test stimulus is a robust distribution-based approach that may improve confidence that observed change is "real" rather than due to measurement error. This distribution-based approach to interpreting CPM effect may assist in the identification of CPM paradigms that induce a CPM effect (inhibitory or facilitatory) in the greatest proportion of participants and afford the greatest degree of within-subject stability over time, prerequisites in the validation of CPM as a robust pain biomarker. Even when using reliable test stimuli, and incorporating measures to limit bias and random error, CPM intersession reliability was fair at the group level and demonstrated a large degree of within-subject variation. Determining "true" change in CPM will underpin future interrogations of intra-individual differences in CPM effect.

Acknowledgments

This report is independent research and the views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Figure Legends

Figure 1. Experimental protocol for Group B as an example

Paradigm order different for Group A and C as order of conditioning stimuli differed. TS= Test stimulus; PPT= pressure pain threshold; CS= Conditioning stimulus; CPM= Conditioned pain modulation; R=right; L=left.

Figure 2. Photograph of CPM test process.

Investigator 2, testing the test stimulus is blinded from identifying which conditioning stimulus was used (written consent given by subject for use of photo).

Figure 3. CPM effect expressed as absolute change in PPT in response to conditioning stimuli for each type of stimulus at both test sessions (21). Results expressed as mean (95%CI). Shaded area represents 2 x standard error of measurement of the PPT therefore the change that could be due to measurement error. CPM= conditioned pain modulation, PPT=Pressure pain threshold.

Figure 4. Proportion of subjects demonstrating a change in absolute PPT in response to three conditioning stimuli (21). Frequencies for those showing any change, and those demonstrating a response >2x standard error of measurement (SEM). PPT=Pressure pain threshold.

Figure 5. Inter-session within-subject variability of absolute CPM effect to thermal conditioning stimuli (21). Comparison of participant responses to hot and cold conditioning stimuli across test sessions. Grey dotted lines represents ± 2 x standard error of measurement of the pressure pain threshold. > +2SEM interpreted as CPM inhibition, >-2SEM as facilitation and < ± 2SEM as non-responder. CPM= conditioned pain modulation.

Figure 6. Intrasession within-subject variability of absolute CPM effect to thermal conditioning stimuli (21). Comparison of participant responses to hot and cold conditioning stimuli across test sessions. Grey dotted lines represents ± 2 x standard error of measurement of the pressure pain threshold. > +2SEM interpreted as CPM inhibition, >-2SEM as facilitation and < ± 2SEM as non-responder. CPM= conditioned pain modulation.

Table Legends

Table 1. Subject demographics and psychological questionnaire scores (21)

BMI=Body mass index; PCS=Pain catastrophizing scale; PSQ= Pain sensitivity questionnaire; STAI_Y1=Spielberger state and trait inventory state questionnaire; STAI_Y2=STAI strait questionnaire; HADS=Hospital anxiety and depression scale

Table 2. Reliability of test measures and CPM effect (21)

Intraclass correlations (ICC 2,k) and 95% confidence intervals for reliability analysis of pressure pain threshold (PPT) both within and between sessions, and pain scores of conditioning stimuli and conditioned pain modulation (CPM) effect between

sessions. ICC of >0.75 represents excellent reliability, 0.6-0.74 good reliability, 0.4-0.59 fair reliability and <0.4 poor reliability (3).

Table 3. Kappa coefficient comparing the allocation of subjects to inhibitor, facilitator or non-responder groups between the two test sessions.

Table 4. Kappa coefficient identifying intra-session reliability in allocation of subjects to inhibitor, facilitator or non-responder groups.

References

1. Altman DG, Bland JM. Standard deviations and standard errors. BMJ 331:903, 2005
2. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 332:1080, 2006
3. Altman DG. Practical statistics for medical research. London: Chapman and Hall; 1991.
4. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 26:217-38, 1998
5. Baumgartner TA. Norm-referenced measurement: Reliability. Safrit MJ, Wood TM, editors. Champaign, IL1989.
6. Biurrun Manresa JA, Fritsche R, Vuilleumier PH, Oehler C, Morch CD, Arendt-Nielsen L. Is the conditioned pain modulation paradigm reliable? A test-retest assessment using the nociceptive withdrawal reflex. PloS One 9:e100241, 2014
7. Cathcart S, Winefield AH, Rolan P, Lushington K. Reliability of temporal summation and diffuse noxious inhibitory control. Pain Res Mang 14:433-8, 2009
8. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. Journal of clinical epidemiology 56:395-407, 2003
9. Duncker K. Some Preliminary Experiments on the Mutual Influence of Pains. Psychologische Forschung. 21:311–26, 1937
10. Edwards RR, Dolman AJ, Martel MO, Finan PH, Lazaridou A, Cornelius M, Wasan AD. Variability in conditioned pain modulation predicts response to NSAID treatment in patients with knee osteoarthritis. BMC musculoskeletal disorders. 17:284, 2016
11. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. Physical therapy 74:777-88, 1994

12. Gierthmuhlen J, Enax-Krumova EK, Attal N, Bouhassira D, Cruccu G, Finnerup NB, Haanpää M, Hansson P, Jensen TS, Freynhagen R, Kennedy JD, Mainka T, Rice AS, Segerdahl M, Sindrup SH, Serra J, Tölle T, Treede RD, Baron R, Maier C. Who is healthy? Aspects to consider when including healthy volunteers in QST-based studies- a consensus statement by the EUROPAIN and NEUROPAIN consortia. Pain. 156:2203-11, 2015.
13. Granot M, Weissman-Fogel I, Crispel Y, Pud D, Granovsky Y, Sprecher E, Yarnitsky D. Determinants of endogenous analgesia magnitude in a diffuse noxious inhibitory control (DNIC) paradigm: do conditioning stimulus painfulness, gender and personality variables matter? Pain 136:142-9, 2008
14. Graven-Nielsen T, Izumi M, Petersen KK, Arendt-Nielsen L. User-independent assessment of conditioning pain modulation by cuff pressure algometry. European journal of pain 21:552-61, 2017
15. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. Physical therapy 86:735-43, 2006
16. Harvill LM. Standard error of measurement. Educational Measurement: Issues and Practice. 10:33-41, 1991
17. Hebert R, Spiegelhalter DJ, Brayne C. Setting the minimal metrically detectable change on disability rating scales. Archives of physical medicine and rehabilitation. 78:1305-8, 1997
18. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 30:1-15, 2000
19. Imai Y, Petersen KK, Morch CD, Arendt Nielsen L. Comparing test-retest reliability and magnitude of conditioned pain modulation using different combinations of test and conditioning stimuli. Somatosens Mot Res. 33:169-77, 2016
20. Kennedy DL, Kemp HI, Ridout D, Yarnitsky D, Rice AS. Reliability of conditioned pain modulation: a systematic review. Pain 157:2410-9, 2016
21. Kennedy DL. Carpal Tunnel Syndrome: An investigation of the impact of neuropathic Pain phenotype on post-operative outcome (CAPS). https://spiral.imperial.ac.uk:8443/handle/10044/1/73694, 2019

22. Le Bars D, Dickenson AH, Besson JM. Diffuse noxious inhibitory controls (DNIC). I. Effects on dorsal horn convergent neurones in the rat. Pain. 6:283-304, 1979
23. 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. 17:98-102, 2012
24. Lewis GN, Rice DA, McNair PJ. Conditioned pain modulation in populations with chronic pain: a systematic review and meta-analysis. Journal of Pain. 13:936-44, 2012
25. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil. 84:719-23, 2005
26. Locke D, Gibson W, Moss P, Munyard K, Mamotte C, Wright A. Analysis of meaningful conditioned pain modulation effect in a pain-free adult population. The Journal of Pain 15:1190-8, 2014
27. Lydick E, Epstein RS. Interpretation of quality of life changes. Qual Life Res. 2:221-6, 1993
28. 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. 158:1217-23, 2017
29. Martel MO, Wasan AD, Edwards RR. Sex differences in the stability of conditioned pain modulation (CPM) among patients with chronic pain. Pain medicine. 14:1757-68, 2013
30. Millan MJ. Descending control of pain. Progress in neurobiology. 66:355-474, 2002
31. Nir RR, Yarnitsky D, Honigman L, Granot M. Cognitive manipulation targeted at decreasing the conditioning pain perception reduces the efficacy of conditioned pain modulation. Pain 153:170-6, 2012
32. Portney LG, Watkins M. Foundations of Clinical Research: Applications to Practice. 3rd ed. Philadelphia, PA: FADavis; 2015.
33. Portney LG, Watkins MP. Foundations of Clinical Research. New Jersey: Prentice Hall Health; 2000.

34. Pud D, Granovsky Y, Yarnitsky D. The methodology of experimentally induced diffuse noxious inhibitory control (DNIC)-like effect in humans. Pain 144:16-9, 2009
35. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clinical rehabilitation. 12:187-99, 1998
36. Rolke R, Andrews K, Magerl W, Treede RD, Pfau D, Klein T, Blunk JA, Geber C, Krumova E, Limbeck C, Magerl W, Maier C, Westermann A, Schuh-Hofer S, Tiede W, Treede RD. The Investigators Brochure: A standardized battery of Quantitative Sensory Testing according to the protocol of the German Research Network on Neuropathic Pain (DFNS) v.2, 2010.
37. Rolke R, Baron R, Maier C, Tölle TR, Treede RD, Beyer A, Binder A, Birbaumer N, Birklein F, Bötefür IC , Braune S , Flor H , Huge V , Klug R , Landwehrmeyer GB , Magerl W, Maihöfner C , Rolko C , Schaub C , Scherens A , Sprenger T, Valet M, Wasserka B. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. Pain. 123:231-43, 2006
38. Sangesland A, Storen C, Vaegter HB. Are preoperative experimental pain assessments correlated with clinical pain outcomes after surgery? A systematic review. Scandinavian journal of pain. 15:44-52, 2017
39. Sellers AB, Ruscheweyh R, Kelley BJ, Ness TJ, Vetter TR. Validation of the English language pain sensitivity questionnaire. Regional anesthesia and pain medicine. 38:508-14, 2013
40. Spielberger C, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. Manual for the State-Trait Anxiety Inventory. Palo Alto, CA Consulting Psychologists Press; 1983.
41. Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. Physical therapy. 76:1109-23, 1996
42. Sullivan MJL, Bishop SR. The Pain Catastrophizing Scale: Development and Validation. Psychological Assessment. 7:524-32, 1995

43. 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 musculoskeletal disorders. 14:182, 2013

44. Vaz S, Falkmer T, Passmore AE, Parsons R, Andreou P. The Case for Using the Repeatability Coefficient When Calculating Test–Retest Reliability. PloS One. 8:e73990, 2013

45. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, Blettner M, Boffetta P, Brenner H, Chêne G, Cooper C, Davey-Smith G, Gagnon F, Greenland P, Greenland S, Infante-Rivard C, Ioannidis J, James A, Jones G, Ledergerber B, Little J, May M, Moher D, Momen H, Morabia A, Morgenstern H, Mulrow CD, Paccaud F, Poole C, Röösli M, Rothenbacher D, Rothman K, Sabin C, Sauerbrei W, Say L, Schlesselman JJ, Sterne J, Syddall H, White I, Wieland S, Williams H, Zou GY. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 370:1453-7, 2007

46. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. Strength Cond Res. 19:231–40, 2005

47. Wilson H, Carvalho B, Granot M, Landau R. Temporal stability of conditioned pain modulation in healthy women over four menstrual cycles at the follicular and luteal phases. Pain. 154:2633-8, 2013

48. Yarnitsky D, Arendt-Nielsen L, Bouhassira D, Edwards RR, Fillingim RB, Granot M, Hansson P, Lautenbacher S, Marchand S, Wilder-Smith O. Recommendations on terminology and practice of psychophysical DNIC testing. European journal of pain. 14:339, 2010

49. Yarnitsky D, Bouhassira D, Drewes AM, Fillingim RB, Granot M, Hansson P, Landau R, Marchand S, Matre D, Nilsen KB, Stubhaug A, Treede RD, Wilder-Smith OH. Recommendations on practice of conditioned pain modulation (CPM) testing. European journal of pain. 19:805-6, 2015

50. Yarnitsky D, Crispel Y, Eisenberg E, Granovsky Y, Ben-Nun A, Sprecher E, Best LA, Granot M. Prediction of chronic post-operative pain: pre-operative DNIC testing identifies patients at risk. Pain. 138:22-8, 2008

51. Yarnitsky D, Granot M, Nahman-Averbuch H, Khamaisi M, Granovsky Y. Conditioned pain modulation predicts duloxetine efficacy in painful diabetic neuropathy. Pain. 153:1193-8, 2012
52. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 67:361-70, 1983

FIGURES
REACTION

Figure 1.

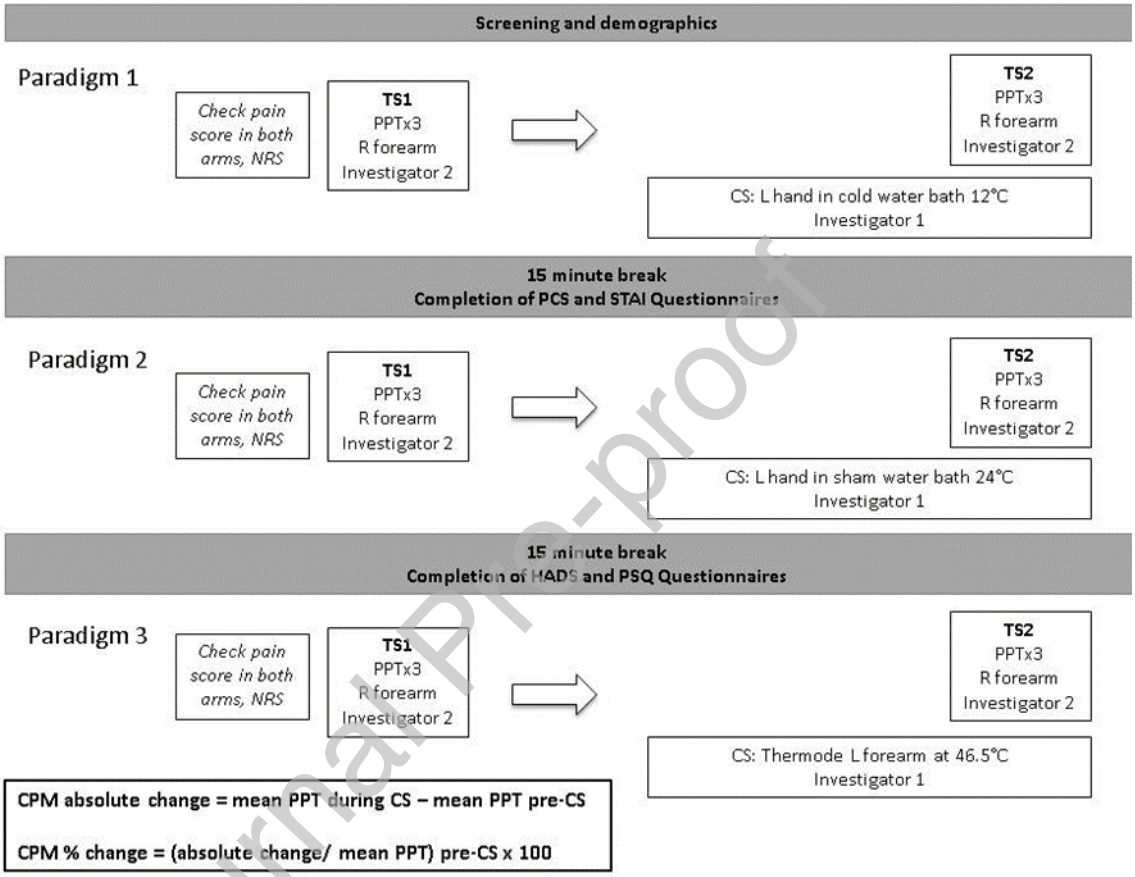


Figure 2.

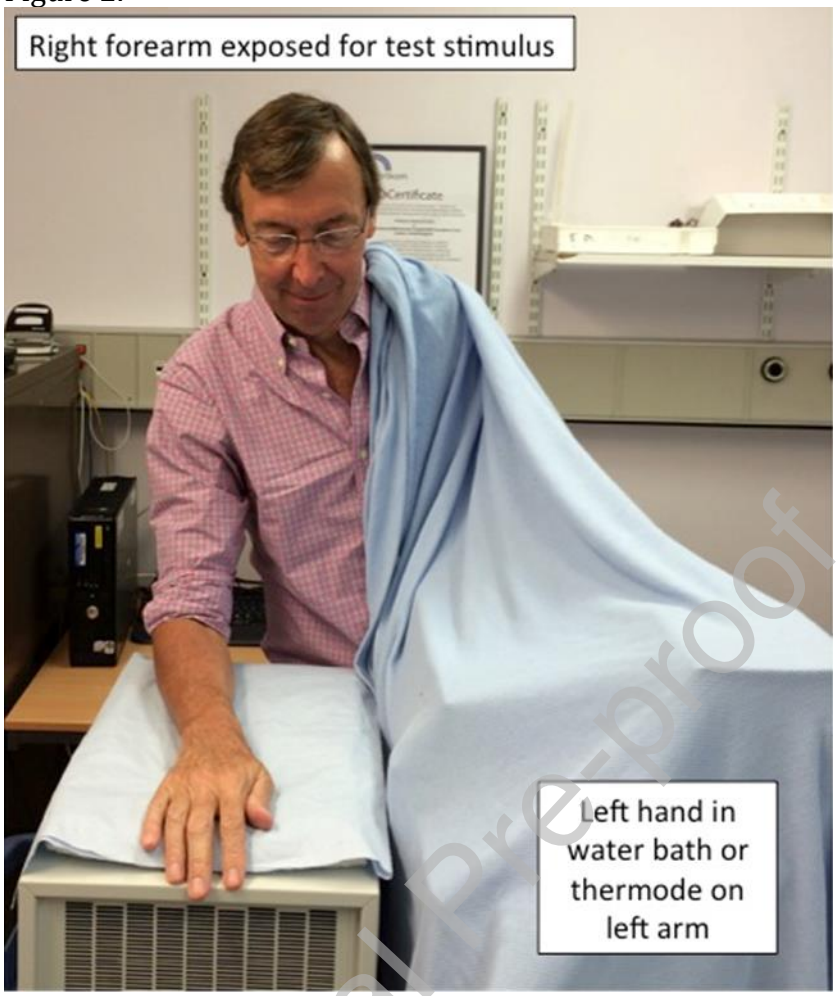


Figure 3.

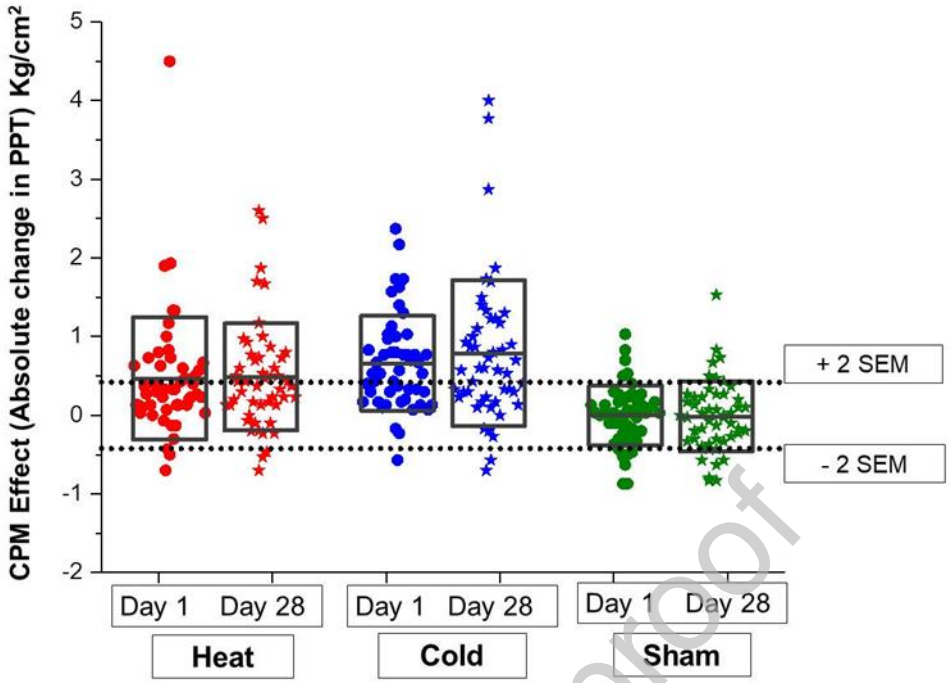


Figure 4.

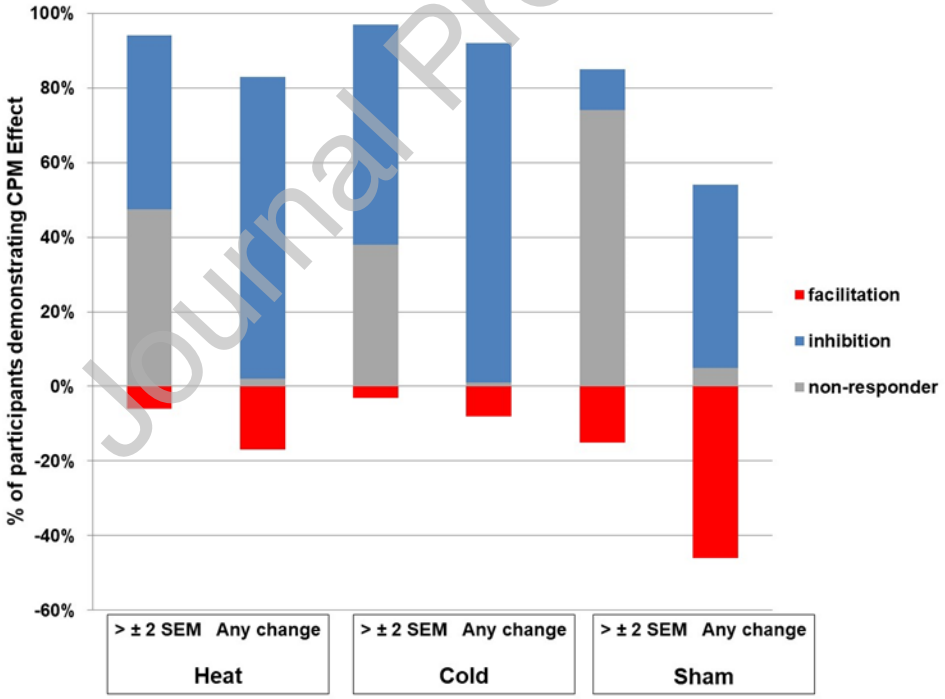


Figure 5.

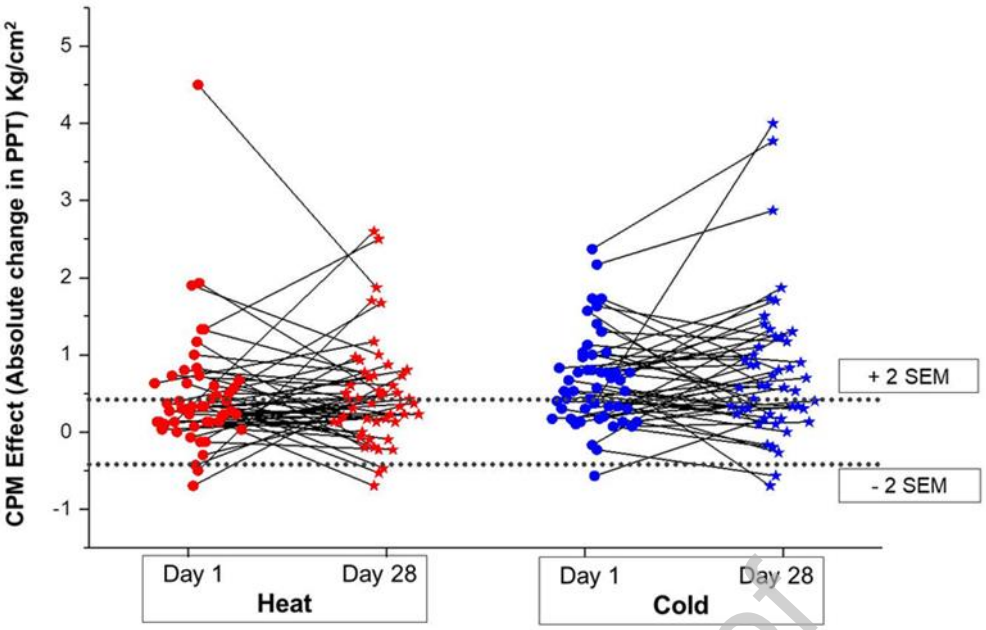
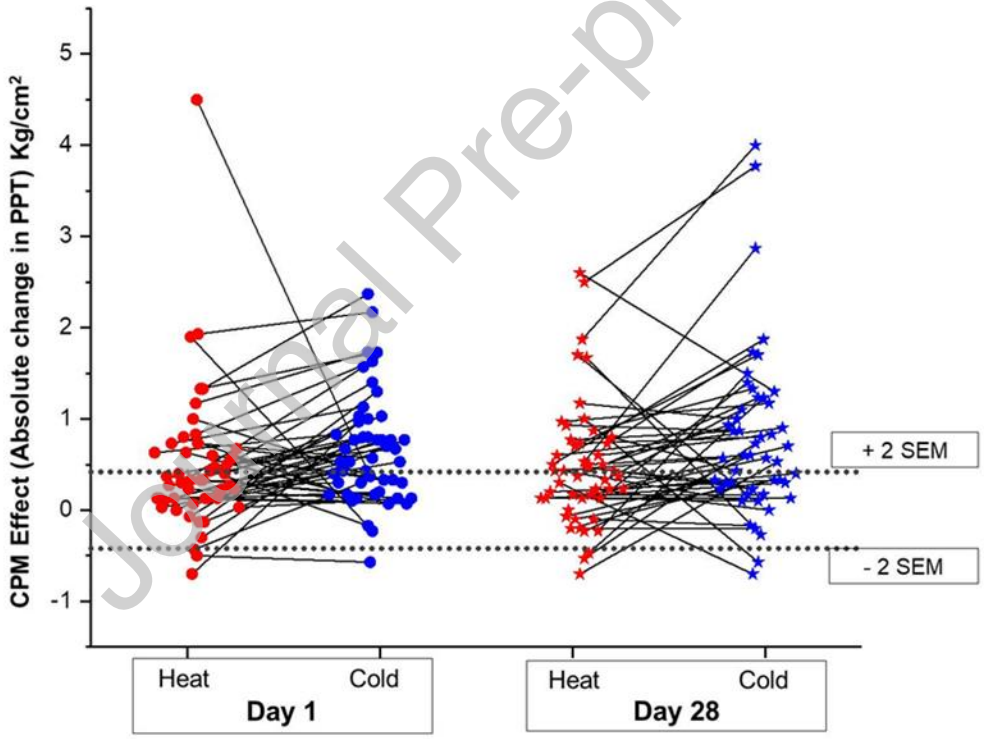


Figure 6.



TABLES

REACTION

Table 1.

	Total cohort (n=50)	**Group A (n=16)**	**Group B (n=17)**	**Group C (n=17)**	**p value for comparison**
Gender male, n (%)	24 (48)	9 (56)	7 (41)	8 (47)	0.68
Age, years	33.9 (11.27)	29.7 (7.98)	32.1 (7.70)	39.8 (14.56)	0.02
BMI (kg/m^2)	24.7 (2.99)	24.4 (3.81)	25.7 (6.55)	23.9 (4.22)	0.54
Ethnicity					0.38
White Caucasian	37 (74)	12 (75)	11 (65)	14 (82)	
Asian	6 (12)	3 (19)	1 (6)	2 (12)	
Black African	1 (2)	0 (0)	1 (6)	0 (0)	
Chinese	5 (10)	1 (6)	3 (17)	1 (6)	
Mixed	1 (2)	0 (0)	1 (6)	0 (0)	
Educational Level					0.75
High school	11 (22)	4 (25)	4 (24)	3 (18)	
Honors Degree	15 (30)	4 (25)	6 (35)	5 (29)	
Masters Degree	19 (38)	7 (44)	4 (24)	8 (47)	
PhD or equivalent	5 (10)	1 (6)	3 (17)	1 (6)	
Hours of exercise per week, hours	3.7 (3.76)	4.1 (2.81)	4.7 (5.15)	2.4 (2.53)	0.17
Average number caffeinated drinks	15.5 (10.60)	15.0 (9.74)	12.7 (10.03)	18.7 (11.64)	0.26

per week					
HADS anxiety score	5.4 (3.04)	5.8 (3.02)	5.1 (3.25)	5.4 (2.98)	0.81
HADS depression score	1.7 (2.27)	1.8 (2.62)	1.8 (2.56)	1.5 (1.66)	0.92
PCS score	12.7 (8.36)	14.8 (10.48)	12.9 (8.31)	10.5 (5.74)	0.35
PSQ score	61.6 (20.80)	61.9 (20.18)	65.7 (19.83)	57.3 (22.64)	0.51
STAI_Y1 score	47.1 (3.76)	47.9 (3.42)	46.5 (3.88)	47.0 (4.07)	0.56
STAI_Y2 score	45.9 (3.76)	47.2 (3.15)	45.0 (12.03)	45.7 (3.29)	0.70

Table 2.

Parameter		**ICC**	**95% confidence interval**
Intrasession PPT	Day 1	0.943	0.909-0.966
	Day 28	0.952	0.924-0.971
Intersession			
PPT	pre heat	0.853	0.688-0.924
	pre cold	0.893	0.782-0.944
	pre sham	0.879	0.775-0.934
Conditioning stimulus discomfort score	heat	0.845	0.728-0.912

	cold	0.847	0.729-0.913
	sham	0.932	0.880-0.961
CPM effect absolute change	heat	0.460	0.040-0.695
	cold	0.582	0.261-0.764
	sham	-0.082	-0.940-0.391
CPM effect % change	heat	0.105	-0.599-0.496
	cold	0.421	-0.036-0.675
	sham	-0.119	-1.00-0.371

Table 3.

	>±2 x SEM group allocation		
	Kappa coefficient	p value	95% confidence interval
Heat	0.143	0.227	-0.02 to 0.48
Cold	0.227	0.082	-0.08 to 0.37
Sham	0.120	0.259	-0.09 to 0.33

Table 4.

	>±2 x SEM group allocation		
	Kappa coefficient	p value	95% confidence interval
Day 1	-0.004	0.976	-0.25 – 0.25
Day 28	0.355	0.004	0.13-0.58