Assessment of CPM reliability

quantification of the within-subject reliability of 10 different protocols

Vaegter, Henrik Bjarke; Petersen, Kristian Kjær; Mørch, Carsten Dahl; Imai, Yosuke; Arendt-Nielsen, Lars

Published in:
Scandinavian Journal of Pain

DOI (link to publication from Publisher):
10.1515/sjpain-2018-0087

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Original experimental

Henrik Bjarke Vaegter*, Kristian Kjær Petersen, Carsten Dahl Mørch, Yosuke Imai and Lars Arendt-Nielsen

Assessment of CPM reliability: quantification of the within-subject reliability of 10 different protocols

https://doi.org/10.1515/sjpain-2018-0087
Received May 16, 2018; revised June 22, 2018; accepted June 26, 2018; previously published online July 14, 2018

Abstract

Background and aims: Conditioned Pain Modulation (CPM) is a well-established phenomenon and several protocols have shown acceptable between-subject reliability [based on intraclass correlation coefficient (ICC) values] in pain-free controls. Recently, it was recommended that future CPM test-retest reliability studies should explicitly report CPM reliability based on CPM responders and non-responders (within-subject reliability) based on measurement error of the test stimulus. Identification of reliable CPM paradigms based on responders and non-responders may be a step towards using CPM as a mechanistic marker in diagnosis and individualized pain management regimes. The primary aim of this paper is to investigate the frequency of CPM responders/non-responders, and to quantify the agreements in the classification of responders/non-responders between 2 different days for 10 different CPM protocols.

Methods: Data from a previous study investigating reliability of CPM protocols in healthy subjects was used. In 26 healthy men, the test-stimuli used on both days were: Pain thresholds to electrical stimulation, heat stimulation, manual algometry, and computer-controlled cuff algometry as well as pain tolerance to cuff algometry. Two different conditioning stimuli (CS; cold water immersion and a computer-controlled tourniquet) were used in a randomized and counterbalanced order in both sessions. CPM responders were defined as a larger increase in the test stimulus response during the CS than the standard error of measurement (SEM) for the test-stimuli between repeated baseline tests without CS.

Results: Frequency of responders and non-responders showed large variations across protocols. Across the studied CPM protocols, a large proportion (from 11.5 to 73.1%) of subjects was classified as CPM non-responders when the test stimuli standard error of measurements (SEM) was considered as classifier. The combination of manual pressure algometry and cold water immersion induced a CPM effect in most participants on both days ($n = 16$). However, agreement in the classification of CPM responders versus non-responders between days was only significant when assessed with computer-controlled pressure pain threshold as test-stimulus and tourniquet cuff as CS ($\kappa = 0.36$ [95% CI, 0.04–0.68], $p = 0.037$).

Conclusions and implications: Agreements in classification of CPM responders/non-responders using SEM as classifier between days were generally poor suggesting considerable intra-individual variation in CPM. The most reliable paradigm was computer-controlled pressure pain threshold as test-stimulus and tourniquet cuff as conditioning stimulus. However while this CPM protocol had the greatest degree of agreement of classification of CPM responders and non-responders across days, this protocol also failed to induce a CPM response in more than half of the sample. In contrast, the commonly used combination of manual pressure algometry and cold water immersion induced a CPM effect in most participants however it was inconsistent in doing so. Further exploration of the two paradigms and classification of responders and non-responders in a larger heterogeneous sample also including women would further inform the clinical usefulness of these CPM protocols. Future research in this area may

***Corresponding author: Henrik Bjarke Vaegter,** PhD, Pain Research Group, Pain Center South, University Hospital Odense, Heden 7-9, Indgang 200, DK – 5000, Odense C, Denmark,
Phone: +45 65413869, Fax: +45 65415064,
E-mail: henrik.bjarke.vaegter@rsyd.dk; and Institute of Clinical Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark
Kristian Kjær Petersen, Carsten Dahl Mørch and Lars Arendt-Nielsen: SMI, Department of Health Science and Technology, School of Medicine, Aalborg University, Aalborg, Denmark
Yosuke Imai: Clinical Development-Center, Asahi Kasei Pharma Corporation, Tokyo, Japan

be an important step towards using CPM as a mechanistic marker in diagnosis and in developing individualized pain management regimes.

Keywords: conditioned pain modulation; pain sensitivity; pain threshold; pain tolerance; reliability; test-retest; cuff algometry.

1 Introduction

In chronic pain patients, several studies have demonstrated impaired conditioned pain modulation (CPM) compared with pain-free controls [1]. Moreover, the CPM response has demonstrated some promise in predicting future pain status [2], the efficacy of analgesics [3], and non-pharmacological treatment [4] contributing to clinical pain profile and management. Assessment of CPM is frequently demonstrated as a change in a test stimulus response (e.g. increase in pressure pain thresholds or decrease in heat pain ratings) by a painful conditioning stimulus (e.g. cold water) applied contralaterally to the test stimulus [5], although various modes of test and conditioning stimuli have been used. In general, CPM is a well-established phenomenon and several protocols have shown acceptable between-subject reliability (based on intraclass correlation coefficient (ICC) values) in pain-free controls [6, 7]. ICC values is a measure of how much of the total variability (within- and between-subject) is explained by between-subject variability, thus reflecting the protocols' ability to differentiate CPM responses between subjects.

Despite a massive increase in the number of studies investigating the magnitude and between-subject reliability of different CPM paradigms in humans, the authors of a recent systematic review investigating test-retest reliability of CPM [6] concluded that the vast majority of studies did not consider measurement error of the test stimulus (e.g. standard error of measurement, SEM) as some of the change in test stimulus during the conditioning stimulus may be due to measurement error (e.g. habituation). Thus, it was recommended that future CPM test-retest reliability studies should explicitly report CPM reliability based on responders and non-responders (within-subject reliability) based on measurement error of the test stimulus.

A previous experimental cross-over study investigated the between-days reliability of eight different CPM protocols in 26 healthy subjects and reported poor to fair between-subject test-retest reliability across 2 different days for all protocols (all ICCs <0.54) [5]. Although the coefficient of variations for the different test stimuli used were reported as a measure of intra-individual variation between two repeated test stimuli without conditioning stimulus, no information on frequency of CPM responders and non-responders considering these intra-individual variations were reported. As identification of CPM paradigms with adequate within-subject reliability may be a further step towards individualized pain medicine [8], these data provide a good opportunity to describes the frequency of CPM responders versus non-responders across these 10 different CPM protocols considering the standard error of measurement (SEM) for the test stimuli.

Thus, the primary aim of this explorative analysis was to investigate the frequency of CPM responders/non-responders, and to quantify the agreements in the classification of CPM responders versus non-responders using SEM as classifier between 2 different days for 10 different CPM protocols.

2 Methods

This explorative analysis was performed using data on pain sensitivity assessed at 2 different days with four different test modalities before and during two different conditioning stimuli included in a previous experimental cross-over study [5]. Assessments were performed in 26 healthy men (mean age of 25.3 ± 5.6) years (range: 18–42) who provided written informed consent and the experimental study was conducted in accordance with the Declaration of Helsinki, and approved by the local ethical committee (N-20150055).

2.1 Procedure

A complete description of the methodology has been published previously [5]. In short, subjects participated in 2 identical test days at approximately the same time of the day and separated by 1–3 weeks. On each day, pain thresholds to electrical stimulation (Digitimer DS5; Digitimer, Welwyn Garden City, UK), heat stimulation (ATS thermode; Medoc Advanced Medical Systems, Ramat-Yishai, Israel), manual algometry (Somedic Sales AB, Horby, Sweden), and computer-controlled cuff algometry (Nocitech, Aalborg, Denmark) as well as the pain tolerance to computer-controlled cuff algometry (Nocitech, Aalborg, Denmark) were assessed on the dominant side of the body in randomized and counterbalanced order. Assessment of pain sensitivity to the different modalities was assessed four times (two baseline assessments and two assessments during conditioning pain stimuli) each day, before

and during two different conditioning stimuli, respectively. Conditioning pain stimuli were induced by cold water immersion and a computer-controlled tourniquet on the non-dominant side of the body in randomized and counterbalanced order. During the conditioning stimuli, subjects used a 10 cm electronic visual analog scale (VAS) to rate the pain intensity due to the conditioning stimulus. Zero and 10 cm extremes on the VAS were defined as "no pain" and as "maximal pain", respectively. The conditioning stimuli were separated by 45 min to avoid carry-over effects. Previous studies have consistently demonstrated short-lasting hypoalgesic effects of CPM of less than 15 min [9, 10].

2.2 Classification of CPM responders versus non-responders

Subjects who had an increase in the test stimulus response during conditioning which was larger than the SEM of each of the test stimuli for the four baseline measurements (two in each session) was classified as CPM responders and subjects who did not have an increase in the test stimulus response during conditioning larger than the SEM of the test stimulus was classified as CPM non-responders.

2.3 Statistical analyses

Data were analyzed using SPSS Statistics, version 24 (IBM, Armonk, NY, USA).

2.3.1 Baseline pain sensitivity assessments, SEM for the different pain tests and pain intensity during CS

Mean and standard deviations (SD) were calculated for the four baseline pain test measurements, and the absolute difference between two baseline measurements within the same day were investigated with 95% confidence intervals (CI). Repeated measures analysis of variances (RM-ANOVAs) between the four baseline pain test stimuli was performed to investigate potential systematic bias across time. Then, the group SEM for each pain test stimuli were estimated as the square root of the mean square error term in the RM-ANOVA output from the SPSS Statistics [11]. Similarly, mean and SD were calculated for the pain intensity reported during cold water immersion and tourniquet cuff as conditioning stimuli, and the absolute difference between days were investigated with 95% CI. Intraclass correlation coefficients (ICCs) based on a single rating, absolute agreement, 2-way mixed effect model were used reflecting the ability of the different pain tests to differentiate values between individuals. An ICC above 0.75 was taken as excellent reliability, 0.40–0.75 was fair to good reliability, and less than 0.40 defined poor reliability [12].

2.3.2 CPM responders versus non-responders

The numbers of CPM responders versus non-responders at day 1 and day 2 were calculated for the different CPM protocols based on SEM as classifier as described. Cohen's kappa coefficient was used to investigate agreement between which subjects were classified as CPM responders versus non-responders on the 2 days for each of the different CPM protocols. *p*-Values less than 0.05 were considered significant. Kappa values of 0.81–1.0 was interpreted as almost perfect agreement, 0.61–0.80 as substantial agreement, 0.41–0.60 as moderate agreement, 0.21–0.40 as fair agreement, and 0.0–0.20 as poor agreement [13]. In addition, and based on the recommendations for future research on CPM [6], the effect of the conditioning stimuli on the test stimuli was calculated as the absolute change (Day 1 minus Day 2) and relative change [(Day 1 minus Day 2) divided by Day 1]*100% in test stimuli during conditioning compared with before conditioning, and SEM for the CPM responses between days were also calculated.

3 Results

3.1 Baseline pain sensitivity assessments, SEM for the different pain tests and pain intensity during CS

Baseline assessments during the 2 days with the five different pain test stimuli (Table 1) showed no systematic errors in day 1 or day 2, as indicated by the 95% CI of the mean differences, where zero lies within the intervals. However, significant differences in heat pain thresholds and pressure pain threshold assessed with cuff algometry across the four measurements were noted. The test-retest reliability of electrical pain threshold, PPT, cPPT and cPTT were excellent with ICCs above 0.80. The test-retest reliability of heat pain threshold was fair to good.

Based on the SEM as classifier, the minimal differences in pain thresholds a subject needed to be considered a CPM responder were 0.58 mA for electrical stimulation,

Table 1: Absolute and relative within-day test-retest reliability for five different pain test stimuli.

Test stimulus	Day 1 Test 1 Mean + SD	Day 1 Test 2 Mean + SD	Absolute difference Day 1 Mean + SD (95% CI)	Day 2 Test 1 Mean + SD	Day 2 Test 2 Mean + SD	Absolute difference Day 2 Mean + SD (95% CI)	*p*-Value between all 4 tests	ICC (95% CI)	Standard error of measurement
Electrical pain threshold (mA)	4.13 ± 2.40 mA	3.84 ± 2.19 mA	0.29 ± 0.77 mA (−0.02 to 0.60)	4.25 ± 2.29	4.23 ± 2.22	0.02 ± 0.68 (−0.26 to 0.29)	0.53	0.93 (0.88–0.97)	0.58 mA
Heat pain threshold (°C)	43.33 ± 2.1 °C	43.20 ± 2.36 °C	0.13 ± 1.38 °C (−0.43 to 0.68)	44.03 ± 1.64	43.93 ± 1.77	0.09 ± 1.34 (−0.45 to 0.64)	0.008	0.70 (0.54–0.84)	1.03 °C
PPT (kPa)	454 ± 145 kPa	470 ± 138 kPa	−16 ± 89 kPa (−52 to 20)	495 ± 184 kPa	480 ± 175 kPa	15 ± 53 kPa (−6 to 36)	0.183	0.82 (0.71–0.91)	68 kPa
cPPT (kPa)	23.68 ± 7.87 kPa	23.37 ± 8.17 kPa	0.31 ± 5.71 kPa (−1.99 to 2.62)	24.99 ± 8.85 kPa	26.83 ± 9.46 kPa	−1.84 ± 5.46 kPa (−4.05 to 0.36)	0.012	0.76 (0.61–0.87)	4.07 kPa
cPTT (kPa)	51.38 ± 14.07 kPa	49.42 ± 13.99 kPa	1.97 ± 8.17 kPa (−1.33 to 5.26)	51.28 ± 14.72 kPa	51.65 ± 16.33 kPa	−0.37 ± 5.56 kPa (−2.61 to 1.87)	0.510	0.84 (0.74–0.92)	5.91 kPa

PPT = pressure pain threshold assessed with manual algometry; cPPT = pressure pain threshold assessed with computer-controlled cuff-algometry; cPTT = pressure pain tolerance assessed with computer-controlled cuff-algometry. *p*-Value based on repeated-measures ANOVA with the four time points as repeated measure. ICC based on two-way mixed-effects model with absolute agreement, single measures.

1.03 °C for heat, 68 kPa for manual algometry, and 4.07 kPa for computer-controlled cuff algometry, respectively. The minimal differences in pain tolerance a subject needed to be considered a CPM responder was 5.91 kPa for computer-controlled cuff algometry.

For the conditioning stimuli, cold water immersion was rated as significantly more painful on the electronic VAS compared with tourniquet cuff across days and pain test stimuli, but no significant differences between days were found for either of the CS (Table 2). It was noted that the ICC values of pain intensity ratings (VAS scores) between days during tourniquet cuff were somewhat higher compared with cold water immersion.

3.2 CPM responders versus non-responders

3.2.1 Cold water immersion as conditioning stimulus

As illustrated in Table 3, the numbers of CPM responders versus non-responders at day 1 and day 2 for the different pain test stimuli showed large variation in number of subjects classified as CPM responders on both days ranging from five subjects (19.2%) subjects when computer-controlled pressure pain tolerance was used as test stimulus to 16 subjects (61.5%) subjects when manual pressure pain threshold was used as test stimulus. Across the five different test stimuli 11.5% to 46.2% were classified as CPM non-responders. There was no significant agreement in the classification of CPM responders between day 1 and day 2 for any of the test stimuli in combination with cold water immersion ($\kappa < 0.31$, $p > 0.1$).

3.2.2 Tourniquet cuff as conditioning stimulus

The numbers of CPM responders versus non-responders classified at day 1 and day 2 for the different pain test stimuli when tourniquet cuff was used as conditioning stimulus (Table 4) also showed some variation in number of subjects classified as CPM responders on both days ranging from two subjects (0.1%) subjects when electrical or heat pain thresholds or computer-controlled pressure pain tolerance were used as test stimuli to six subjects (23.1%) subjects when manual pressure pain threshold was used as test stimulus. Across the five different test stimuli 30.8% to 73.1% were classified as CPM non-responders. Examination of the agreement in classification of CPM responders versus non-responders between days showed a significant agreement only for CPM assessed with computer-controlled pressure pain threshold as test stimulus

Table 2: Absolute and relative between-day test-retest reliability for pain intensity rated during the conditioning stimuli (cold pressor test and tourniquet cuff) in combination with five different pain test stimuli.

Conditioned stimulus	Test stimulus	Pain intensity (VAS: 0–10 cm) Day 1 Mean + SD	Pain intensity (VAS: 0–10 cm) Day 2 Mean + SD	Absolute difference in VAS Mean + SD (95% CI)	*p*-Value (between day 1 and day 2)	ICC (95% CI)
Cold pressor test	Electrical	7.8 ± 1.3[a]	7.8 ± 1.4[a]	0.001 ± 0.97 (−0.38 to 0.40)	0.97	0.77 (0.54–0.89)
	Heat	7.9 ± 1.9[a]	8.0 ± 1.5[a]	−0.13 ± 1.52 (−0.71 to 0.51)	0.74	0.60 (0.28–0.80)
	PPT	7.6 ± 1.6[a]	7.6 ± 1.4[a]	0.02 ± 1.43 (−0.53 to 0.57)	0.96	0.59 (0.27–0.80)
	cPPT	7.5 ± 1.6[a]	7.4 ± 1.8[a]	0.12 ± 1.60 (−0.54 to 0.77)	0.72	0.55 (0.20–0.77)
	cPTT	7.5 ± 1.6[a]	7.4 ± 1.8[a]	0.12 ± 1.60 (−0.54 to 0.77)	0.72	0.55 (0.20–0.77)
Tourniquet cuff	Electrical	6.0 ± 2.1	6.0 ± 2.0	−0.01 ± 1.60 (−0.66 to 0.64)	0.97	0.71 (0.44–0.86)
	Heat	5.8 ± 2.0	5.8 ± 2.0	−0.05 ± 1.48 (−0.65 to 0.54)	0.85	0.74 (0.50–0.88)
	PPT	6.0 ± 2.2	6.0 ± 2.1	−0.05 ± 1.7 (−0.75 to 0.65)	0.88	0.69 (0.41–0.85)
	cPPT	6.0 ± 1.9	5.9 ± 2.0	0.12 ± 1.22 (−0.37 to 0.61)	0.62	0.81 (0.62–0.91)
	cPTT	6.0 ± 1.9	5.9 ± 2.0	0.12 ± 1.22 (−0.37 to 0.61)	0.62	0.81 (0.62–0.91)

Pain intensity was rated on a 10 cm electronic visual analog scale (VAS) with 0 and 10 cm extremes on the VAS defined as "no pain" and as "maximal pain", respectively. PPT = pressure pain threshold assessed with manual algometry; cPPT = pressure pain threshold assessed with computer-controlled cuff-algometry; cPTT = pressure pain tolerance assessed with computer-controlled cuff-algometry; VAS = Visual Analog Scale. ([a]$p < 0.001$) Significant difference between pain intensity ratings due to cold pressor test and tourniquet cuff.

[$\kappa = 0.36$ (95% CI: 0.04 to 0.68), $p = 0.037$], indicating fair agreement in classification of CPM responders versus non-responders with SEM as classifier between 2 different days when assessed with this CPM protocol.

On a group level, no significant differences were found in magnitude of the CPM responses between days for any of the CPM protocols [Table 5; $F(1,25) < 3.09$, $p \geq 0.09$].

4 Discussion

This study investigated the frequency of CPM responders/non-responders, and quantified the agreements in the classification of responders/non-responders between 2 different days for 10 different CPM protocols. Across the studied CPM protocols, a large and variable proportion (from 11.5–73.1%) of subjects was classified as CPM non-responders when the test stimuli standard error of measurements (SEM) was considered as classifier. The number of subjects who had an increase in the test stimulus response during conditioning which was larger than the SEM of the test stimulus showed large variations between the different test and conditioning stimuli combinations. Agreements in the classification between the 2 days were in general poor across the protocols. Only when CPM was assessed with computer-controlled pressure pain threshold as test stimulus and tourniquet cuff as the conditioning stimulus a significant and fair agreement in the classification of CPM responders and non-responders between days was noted. In addition, this CPM protocol had the lowest intra-individual variation in CPM responses between the 2 days (3.5%).

It is noteworthy, that the commonly used CPM protocol using manual pressure pain threshold in combination with cold water immersion [6], which had the highest percentage of CPM responders on both days (61.5% on both days), and a relatively low intra-individual variation in CPM response between days (13%) showed a poor and non-significant agreement in classification of CPM responders versus non-responders between days. This could be influenced by the small sample size and the small number of CPM non-responders ($n = 3$) in both days limiting interpretation of the results. Moreover, the relatively low ICC values for the pain intensity experienced during the cold water immersion could influence the reliability of the subsequent CPM response.

This explorative analysis is the first to investigate the numbers of CPM responders versus non-responders using test stimulus SEM as classifier between different days for 10 different CPM protocols. In combination with the previously described between-days test-retest reliability ICC values for the CPM protocols under study [5], these results suggest that considerable intra-individual difference in CPM between days exists. The discrepancies in individual CPM responses between days can be caused by adaptation or expectation effects [14]. Moreover, the variability could also be caused by natural physiological variability, variability from the assessment techniques, or variability introduced by the experimenter, although this would potentially be reduced with standardized computer-controlled techniques.

Table 3: Crosstabulations of the CPM responders and non-responders at day 1 and day 2 assessed with cold pressor test in combination with five different pain test stimuli.

	CPM responders day 2	
	Yes	**No**
A. Electrical pain threshold as test stimulus (EPT)		
Cohen's kappa coefficient:		
[$\kappa = 0.08$ (95% CI, −0.31 to 0.46), $p = 0.70$]		
% agreement (yes or no on both days): 53.8%		
CPM responders day 1		
Yes	7	6
No	6	7
B. Heat pain threshold as test stimulus (HPT)		
Cohen's kappa coefficient:		
[$\kappa = 0.31$ (95% CI, −0.05 to 0.66), $p = 0.11$]		
% agreement (yes or no on both days): 65.4%		
CPM responders day 1		
Yes	10	6
No	3	7
C. Manual pressure pain threshold as test stimulus (PPT)		
Cohen's kappa coefficient:		
[$\kappa = 0.28$ (95% CI, −0.13 to 0.69), $p = 0.15$]		
% agreement (yes or no on both days): 73.1%		
CPM responders day 1		
Yes	16	4
No	3	3
D. Computer-controlled pressure pain threshold as test stimulus (cPPT)		
Cohen's kappa coefficient:		
[$\kappa = 0.09$ (95% CI, −0.27 to 0.46), $p = 0.62$]		
% agreement (yes or no on both days): 53.8%		
CPM responders day 1		
Yes	6	8
No	4	8
E. Computer-controlled pressure pain tolerance as test stimulus (cPTT)		
Cohen's kappa coefficient:		
[$\kappa = 0.26$ (95% CI, −0.11 to 0.63), $p = 0.17$]		
% agreement (yes or no on both days): 65.4%		
CPM responders day 1		
Yes	5	3
No	6	12

Responders and non-responders are classified based on the standard error of measurement (SEM) for two repetitive test stimulus assessments without conditioning. Responders are defined as an increase in the test stimulus during the conditioning stimulus larger than the test stimulus before the conditioning stimulus plus 1 SEM. PPT = pressure pain threshold assessed with manual algometry; cPPT = pressure pain threshold assessed with computer-controlled cuff-algometry; cPTT = pressure pain tolerance assessed with computer-controlled cuff-algometry.

On a group level, no significant differences in magnitude of CPM responses between day 1 and day 2 were found suggesting no systematic group bias in CPM between days.

In patients with chronic pain, several studies have demonstrated impaired CPM compared with asymptomatic controls [1], however the frequency and agreement of CPM responders considering measurement error of the test stimuli between different days in patients with pain is currently unknown.

Strict standardization procedures and reduced bias induced by the person assessing pain sensitivity e.g. via computer-controlled assessment methodology may reduce the intra-individual variation of CPM protocols.

4.1 Implications

These results clearly indicate that choice of protocol used to assess CPM has an influence on the number of subjects who demonstrate hypoalgesia and on how reliable this response is when tested on different days. These findings may suggest that different pain inhibitory systems are

Table 4: Crosstabulations of the CPM responders and non-responders at day 1 and day 2 assessed with tourniquet cuff in combination with five different pain test stimuli.

	CPM responders day 2	
	Yes	**No**
A. Electrical pain threshold as test stimulus (EPT)		
Cohen's kappa coefficient:		
[$\kappa = 0.33$ (95% CI, −0.13 to 0.79), $p = 0.09$]		
% agreement (yes or no on both days): 80.8%		
CPM responders day 1		
Yes	2	2
No	3	19
B. Heat pain threshold as test stimulus (HPT)		
Cohen's kappa coefficient:		
[$\kappa = 0.235$ (95% CI, −0.14 to 0.61), $p = 0.15$]		
% agreement (yes or no on both days): 73.1%		
CPM responders day 1		
Yes	2	6
No	1	17
C. Manual pressure pain threshold as test stimulus (PPT)		
Cohen's kappa coefficient:		
[$\kappa = 0.08$ (95% CI, −0.30 to 0.46), $p = 0.69$]		
% agreement (yes or no on both days): 53.8%		
CPM responders day 1		
Yes	6	5
No	7	8
D. Computer-controlled pressure pain threshold as test stimulus (cPPT)		
Cohen's kappa coefficient:		
[$\kappa = 0.358$ (95% CI, 0.04 to 0.68), $p = 0.037$]		
% agreement (yes or no on both days): 69.2%		
CPM responders day 1		
Yes	5	7
No	1	13
E. Computer-controlled pressure pain tolerance as test stimulus (cPTT)		
Cohen's kappa coefficient:		
[$\kappa = 0.20$ (95% CI, −0.23 to 0.62), $p = 0.32$]		
% agreement (yes or no on both days): 73.1%		
CPM responders day 1		
Yes	2	4
No	3	17

Responders and non-responders are classified based on the standard error of measurement (SEM) for two repetitive test stimulus assessments without conditioning. Responders are defined as an increase in the test stimulus during the conditioning stimulus larger than the test stimulus before the conditioning stimulus plus 1 SEM. PPT = pressure pain threshold assessed with manual algometry; cPPT = pressure pain threshold assessed with computer-controlled cuff-algometry; cPTT = pressure pain tolerance assessed with computer-controlled cuff-algometry.

activated depending on the combination of test stimuli (e.g. cutaneous of deeper tissue stimulation) and conditioning stimuli which may trigger different mechanisms. Further research are warranted in this area to better understand potential consequences and treatment implications.

5 Conclusion

For most CPM protocols, a large proportion of subjects were classified as CPM non-responders when the test stimuli standard error of measurement was considered as classifier. Considerable intra-individual difference exists between different days, and the agreements in classification of CPM responders versus non-responders were in general poor. The most reliable paradigm was computer-controlled pressure pain threshold as test stimulus and tourniquet cuff as conditioning stimulus. However while computer controlled pressure pain threshold test stimulus and tourniquet cuff had the greatest degree of agreement of classification of CPM responders and non-responders across days, this protocol also failed to induce a CPM

Table 5: Absolute and relative between-day test-retest reliability for CPM assessed with cold pressor test and tourniquet cuff in combination with five different pain test stimuli.

Conditioned stimulus	Test stimulus	CPM response Day 1 Mean + SD (95% CI)	CPM response Day 2 Mean + SD (95% CI)	Absolute difference in CPM response Mean + SD (95% CI)	Relative difference in CPM response Mean + SD (95% CI)	*p*-Value	Standard error of measurement
Cold pressor test	Electrical	0.83 ± 0.99 (0.43–1.23)	0.73 ± 0.72 (0.44–1.02)	0.1 ± 1.16 (–0.38 to 0.57)	57.98 ± 369.00 (–91.06 to 207.03)	0.68	0.82
	Heat	1.55 ± 1.26 (1.04–2.06)	1.11 ± 0.92 (0.74–1.48)	0.44 ± 1.48 (–0.15 to 1.04)	132.92 ± 474.04 (–58.55 to 324.39)	0.14	1.04
	PPT	153 ± 141 (96–210)	123 ± 101 (83–164)	30 ± 124 (–21 to 79)	13 ± 284 (–102 to 128)	0.24	88
	cPPT	5.42 ± 7.22 (2.50–8.33)	4.12 ± 6.53 (1.48–6.75)	1.30 ± 7.25 (–1.63 to 4.23)	135.23 ± 246.61 (35.62–234.83)	0.37	5.13
	cPTT	2.70 ± 8.04 (–0.55 to 5.95)	5.08 ± 5.91 (2.69–7.46)	–2.38 ± 7.00 (–5.20 to 0.45)	–149.69 ± 774.68 (–462.59 to 163.21)	0.10	4.95
Tourniquet cuff	Electrical	0.026 ± 0.84 (–0.31 to 0.37)	0.25 ± 0.61 (0.001–0.49)	–0.22 ± 0.97 (–0.61 to 0.17)	–51.71 ± 493.87 (–251.19 to 147.77)	0.26	0.69
	Heat	0.41 ± 1.29 (–0.11 to 0.93)	–0.01 ± 1.19 (–0.49 to 0.47)	0.42 ± 1.27 (–0.09 to 0.93)	10.37 ± 133.18 (–43.42 to 64.16)	0.10	0.89
	PPT	76 ± 111 (31–121)	75 ± 80 (43–108)	1 ± 135 (–54 to 55)	118 ± 313 (–8 to 245)	0.98	95
	cPPT	3.24 ± 5.58 (0.99–5.49)	1.59 ± 4.35 (–0.19 to 3.31)	1.68 ± 4.86 (–0.29 to 3.64)	3.52 ± 254.92 (–99.44 to 106.49)	0.09	3.44
	cPTT	2.08 ± 5.23 (–0.03 to 4.20)	1.59 ± 6.41 (–1.0 to 4.18)	0.49 ± 7.68 (–2.61 to 3.59)	111.10 ± 465.87 (–77.06 to 299.27)	0.75	5.43

CPM calculated as absolute difference between test stimuli with and without conditioning stimuli; PPT = pressure pain threshold assessed with manual algometry; cPPT = pressure pain threshold assessed with computer-controlled cuff-algometry; cPTT = pressure pain tolerance assessed with computer-controlled cuff-algometry.

response in more than half the sample. In contrast, the combination of PPT and cold water immersion was more effective in inducing a CPM effect however it was inconsistent in doing so. Further exploration of the two paradigms and classification of responders and non-responders in a larger heterogeneous sample also including women would further inform the clinical usefulness of these CPM protocols. Future research in this area may be an important step towards using CPM as a mechanistic marker in diagnosis and in developing individualized pain management regimes.

Authors' statements

Research funding: This work was supported by the Innovation Fund Denmark (136-2014-5), The Shionogi Science Program, and TaNeDS Europe.

Conflict of interest: Nocitech is partly owned by the Aalborg University. There are no actual or potential conflicts of interest for any of the authors.

Informed consent: All participants provided written informed consent.

Ethical Approval: The experimental study was conducted in accordance with the Declaration of Helsinki, and approved by the local ethical committee (N-20150055).

References

[1] Lewis GN, Rice DA, McNair PJ. Conditioned pain modulation in populations with chronic pain: a systematic review and meta-analysis. J Pain 2012;13:936–44.

[2] 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 2008;138:22–8.

[3] Yarnitsky D, Granot M, Nahman-Averbuch H, Khamaisi M, Granovsky Y. Conditioned pain modulation predicts duloxetine efficacy in painful diabetic neuropathy. Pain 2012;153:1193–8.

[4] Fingleton C, Smart K, Doody C. Exercise-induced hypoalgesia in people with knee osteoarthritis with normal and abnormal conditioned pain modulation. Clin J Pain 2017;33:395–404.

[5] 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 2016;33:1–9.

[6] Kennedy DL, Kemp HI, Ridout D, Yarnitsky D, Rice AS. Reliability of conditioned pain modulation: a systematic review. Pain 2016;157:2410–9.

[7] Oono Y, Nie H, Matos RL, Wang K, Arendt-Nielsen L. The inter- and intra-individual variance in descending pain modulation evoked by different conditioning stimuli in healthy men. Scand J Pain 2011;2:162–9.

[8] Yarnitsky D, Granot M, Granovsky Y. Pain modulation profile and pain therapy: between pro- and antinociception. Pain 2014;155:663–5.
[9] Vaegter HB, Handberg G, Graven-Nielsen T. Similarities between exercise-induced hypoalgesia and conditioned pain modulation in humans. Pain 2014;155:158–67.
[10] Vaegter HB, Handberg G, Jorgensen MN, Kinly A, Graven-Nielsen T. Aerobic exercise and cold pressor test induce hypoalgesia in active and inactive men and women. Pain Med 2015;16:923–33.
[11] Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005;19:231–40.
[12] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.
[13] Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[14] Goffaux P, Redmond WJ, Rainville P, Marchand S. Descending analgesia – when the spine echoes what the brain expects. Pain 2007;130:137–43.