Reliability of conditioned pain modulation: a systematic review

Donna L. Kennedy[a],*, Harriet I. Kemp[a], Deborah Ridout[b], David Yarnitsky[c], Andrew S.C. Rice[a]

Abstract

A systematic literature review was undertaken to determine if conditioned pain modulation (CPM) is reliable. Longitudinal, English language observational studies of the repeatability of a CPM test paradigm in adult humans were included. Two independent reviewers assessed the risk of bias in 6 domains; study participation; study attrition; prognostic factor measurement; outcome measurement; confounding and analysis using the Quality in Prognosis Studies (QUIPS) critical assessment tool. Intraclass correlation coefficients (ICCs) less than 0.4 were considered to be poor; 0.4 and 0.59 to be fair; 0.6 and 0.75 good and greater than 0.75 excellent. Ten studies were included in the final review. Meta-analysis was not appropriate because of differences between studies. The intersession reliability of the CPM effect was investigated in 8 studies and reported as good (ICC = 0.6-0.75) in 3 studies and excellent (ICC > 0.75) in subgroups in 2 of those 3. The assessment of risk of bias demonstrated that reporting is not comprehensive for the description of sample demographics, recruitment strategy, and study attrition. The absence of blinding, a lack of control for confounding factors, and lack of standardisation in statistical analysis are common. Conditioned pain modulation is a reliable measure; however, the degree of reliability is heavily dependent on stimulation parameters and study methodology and this warrants consideration for investigators. The validation of CPM as a robust prognostic factor in experimental and clinical pain studies may be facilitated by improvements in the reporting of CPM reliability studies.

Keywords: Conditioned pain modulation (CPM), Diffuse noxious inhibitory control (DNIC), Endogenous pain modulation, Reliability, Systematic review

1. Background

1.1. Conditioned pain modulation

Conditioned pain modulation (CPM) is a psychophysical experimental measure of the endogenous pain inhibitory pathway in humans[17]; the "pain inhibits pain" phenomena.[42] Conditioned pain modulation is believed to represent the human behavioral correlate of diffuse noxious inhibitory control (DNIC),[37] first described in rats.[22] Electrophysiological studies in animals and pharmacological studies in humans have demonstrated that descending influences on spinal nociceptive processing involve the periaqueductal gray, rostral ventromedial medulla and subnucleus reticularis dorsalis, leading to the description of this descending pain modulation pathway as a spino-bulbo-spinal loop.[27]

Conditioned pain modulation paradigms consist of the evaluation of a painful test stimulus followed by a second evaluation

either at the same time as a distant, painful conditioning stimulus (parallel paradigm) or in series after the painful conditioning stimulus has been withdrawn (sequential paradigm).[42] Although pain inhibition is not universal (in some subjects an increase in pain intensity rating is observed) in most subjects the pain intensity experienced with the test stimulus will be reduced during or immediately after exposure to the conditioning stimulus.

Conditioned pain modulation has been investigated extensively in healthy volunteers; however, at present, there are no published normative data for CPM effect and it is unclear what qualifies as a "normal range" effect. In a review of healthy volunteer studies, Pud et al.[35] reported variability in the magnitude of CPM effect was dependent on the CPM paradigm used and that the median CPM effect was 29%. However, this must be interpreted with some caution given the heterogeneity and lack of quality assessment of the included studies. There is good evidence that there is much interindividual difference in the magnitude of CPM related to age, sex, and potentially other as yet unknown variables.[9,10] It has been reported that in some healthy subjects, a CPM effect may be altogether absent,[25] although it is probably more accurate to consider that the spectrum of response may range from significant inhibition to a degree of facilitation dependent on individual variability and CPM paradigm. In healthy volunteer studies, the appreciable variability reported in magnitude and the stability of the CPM effect may be attributable to multiple factors including variation in study characteristics such as study design and testing parameters or variability in sample characteristics as defined by the inclusion and exclusion criteria used to qualify a sample of volunteers as "healthy".[7,11]

At present, there is great interest in the science and conduct of CPM testing as there is a growing body of evidence suggesting that CPM may be an important biomarker of chronic pain and

a predictor of treatment response. However, standardization in the testing of CPM is lacking. A 2014 consensus meeting encouraged investigators to include a second test stimulus or second CPM protocol in study designs for the generation of evidence to enable comparisons, suggested sequential test protocols may be advantageous over parallel protocols for being a purer measure of CPM, and that an upper and lower limb should be default test sites; however, the expert forum concluded that there was insufficient data to support recommendations for the use of a specific CPM protocol,[43] and this has not changed to date. There is evidence to suggest that the magnitude of the CPM effect is dependent on the sensory modality used for delivering the conditioning and test stimuli and the body area tested[29,35] as well as the painfulness of the stimuli;[13] however, at present, there is no gold standard for the testing of CPM. Furthermore, estimating the reliability of CPM, and identifying true change in relation to measurement error, has proven challenging because of heterogeneity in study design and analysis and insufficient reporting.

2. Objectives

To assess the reliability of CPM paradigms in adults, critically appraise the literature against reporting guidelines for prognostic factor research and CPM studies[41,42] and make recommendations for the reporting of future studies.

3. Methods

The protocol for this review was not registered as it does not meet the inclusion criteria of the available web-based repositories. Findings are reported according to the PRISMA guidelines for systematic reviews.[28]

3.1. Literature search

No previously published systematic reviews of the reliability of CPM were located neither in the Cochrane Database of Systematic Reviews nor in a search of the electronic databases MEDLINE, EMBASE, CINAHL, and AMED. The same databases were searched from inception to August 26th, 2015 using the search terms (conditioned pain modulation or diffuse noxious inhibitory control or DNIC or heterotopic noxious conditioning) and (reliability or repeatability or stability) (Appendix A, available online as Supplemental Digital Content at http://links.lww.com/PAIN/A338). Inclusion criteria were full-text English reports of longitudinal observational studies of the repeatability or stability of a CPM test paradigm in adult humans. Two independent reviewers (D.L.K., H.I.K.) screened study titles, abstracts, and where necessary full-text to determine study inclusion (**Fig. 1**). Reference lists of included studies were hand searched for additional eligible studies.

3.2. Data extraction and management

Two review authors independently extracted data using a standardized form (D.L.K., H.I.K.). This included sample size, participant gender and mean age, designation as a healthy volunteer or clinical cohort, test and conditioning stimuli and testing site, testing paradigm (sequential or parallel), retest interval, reliability coefficient for CPM effect, measure of response stability, protocol violations (any deviation from a study protocol that may affect the reliability of the data), and test and conditioning stimulus reliability.

3.3. Risk of bias assessment

The methodological quality and risk of bias of the included studies were assessed by 2 independent raters (D.L.K., H.I.K.) using the Quality in Prognosis Studies (QUIPS) critical assessment tool; a tool specifically developed for use in systematic reviews of prognostic factor studies.[16] The QUIPS appraisal domains are in keeping with the National Institutes for Health (NIH) mandate to improve rigor, transparency, and reproducibility in research.[8,21] For clarity, although published CPM reliability studies do not purport to be prognostic factor studies, it is our intent to initiate and encourage future work toward strengthening the evidence for CPM as a prognostic factor. The QUIPS tool addresses risk of bias in 6 major domains; study participation; study attrition; prognostic factor measurement; outcome measurement; confounding; and statistical analysis and is designed to be operationalized for specific study purposes including specifying key characteristics, omitting irrelevant items, and adding items where required.[17] Criteria in each domain are evaluated, thereby generating an overall rating for each domain as having a "low,"
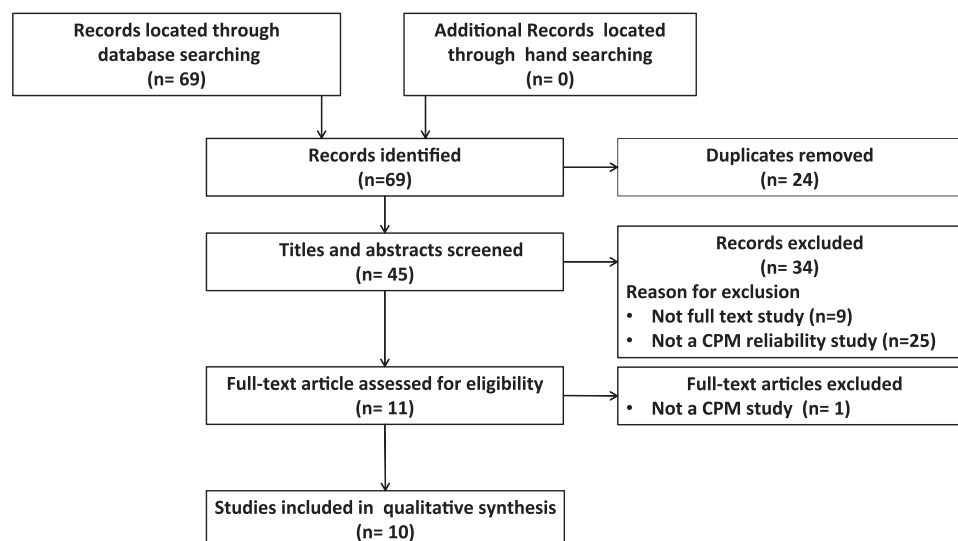
Figure 1. Study flow diagram. CPM, conditioned pain modulation.

"moderate," or "high" risk of bias. For this review, the QUIPS tool was operationalized to be study specific a priori and is reported in Appendix C (available online as Supplemental Digital Content at http://links.lww.com/PAIN/A338). This descriptive approach to quality assessment in systematic reviews is in keeping with current recommendations given the questionable validity and interpretation of existing rating scales.[18]

3.4. Appraisal of reliability data

Reliability data were included in the risk of bias in statistical analysis and interpreted as a measure of the repeatability of a CPM paradigm. Important elements in the statistical analysis of reliability include the reporting of a sample size calculation, an appropriate reliability coefficient and 95% confidence interval for the coefficient and a measure of response stability. Where any of these components were lacking, this was interpreted to increase the risk of bias in statistical analysis and reporting.

Although there is lack of consensus in the appropriate analysis and reporting of reliability for measures which produce continuous data, as does CPM, there is growing evidence to support the use of the intraclass correlation coefficient (ICC) which reflects both the degree of association and agreement among ratings.[34,36,37] Because the ICC is a dimensionless statistic, it is also useful when comparing the repeatability of measures in different units.[5] There are 3 models of ICCs; the choice of model is fundamental in assessing the reliability of clinical or experimental tests and must consider if the use of an instrument or procedure may be generalised to a wider population of random raters, or if performance is user dependent, perhaps reflecting specialist training.

The ICC has been described as a measure of relative reliability as it reflects the degree to which a subject maintains their place in a sample,[1] however reported in isolation, the ICC gives no indication of the magnitude of the disagreement between measures or retests.[36] Response stability, also described as absolute reliability,[1] describes the degree to which a subject's scores will change over repeated tests. A measure of response stability is essential to the practical and clinical interpretation of reliability. Although the ICC provides a dimensionless and easily interpreted point estimate of reliability, a measure of response stability facilitates the comparison of results between reliability studies and enables the judgement of when a change in test score is clinically meaningful rather than possibly attributable to measurement error. Although reliability cannot be interpreted as an all or none concept and acceptable reliability is subjective, there is some consensus that a coefficient less than 0.4 may be interpreted as poor reliability; between 0.4 and 0.59 fair reliability; between 0.6 and 0.75 good reliability; and greater than 0.75 excellent reliability; therefore, the reliability coefficients reported in this review were interpreted as such.[37]

4. Results

Ten studies were selected for inclusion in this review (**Fig. 1**). At screening, excluded records did not pertain to the reliability of CPM or were not full-text papers. One full-text article was excluded and is reported in Appendix B (available online as Supplemental Digital Content at http://links.lww.com/PAIN/A338). No full-text papers examining the reliability of a CPM paradigm were excluded.

4.1. Study characteristics

Summary information for the included studies is reported in **Table 1**. Seven studies investigated CPM in healthy volunteers, 2 studies addressed clinical cohorts, and 1 study included both healthy subjects and a clinical cohort. Eight studies included men and women; 1 study had only men, and 1 study only women. In healthy subject studies, the participants were predominantly below the age of 40, whereas clinical cohort participants were predominantly above the age of 40.

The most commonly investigated test stimulus was pressure pain threshold (PPT) (5 studies), followed by contact heat pain (3 studies). Cold water immersion was the most frequently studied conditioning stimulus (6 studies) followed by hot water immersion (3 studies) and ischemic pain (3 studies). Intersession reliability was investigated in 9 studies with retest intervals varying between 2 and 28 days; intrasession reliability was investigated in 3 studies. The most commonly reported outcome measures were subjective pain threshold (6 studies) and an individualised stimulus intensity required to elicit a predetermined pain intensity (5 studies). Subjective pain intensity rating was measured in 2 studies, a pain-elicited reflex in 2 studies, and subjective pain tolerance in 1 study.

Where reported, study protocol violations and the reliability coefficient for the test and conditioning stimuli are reported in **Table 2**. Protocol violations for the administration of the test and conditioning stimulus include changes to exposure time or intensity of the stimulus from that described a priori and in which case the participant was not excluded from the study. There were no reported study violations in the administration of the test stimuli and 3 reported protocol violations for cold water immersion as a conditioning stimulus.

4.2. Reliability of conditioned pain modulation effect

The intrasession reliability of the CPM effect was investigated in 9 different test–retest measures in 3 studies and was reported as good (ICC = 0.6-0.75) to excellent (ICC >0.75) in 7 of the 9 measures. The intersession reliability of the CPM effect was investigated in 14 different testing paradigms (different test stimuli, outcome measures, and pain intensity) in 8 studies. Investigators in 6 of the 8 studies reported intersession reliability ranging from fair to excellent for a CPM paradigm. Poor intersession reliability was reported for the CPM effect in older adults with chronic pancreatitis and in young women across menstrual cycles (**Table 1**).

4.3. Reliability of test stimuli

Pressure pain threshold was most commonly used as a test stimulus; intrasession reliability was reported as excellent in 2 studies (ICC > 0.75); intersession reliability as good in 2 studies (ICC = 0.60-0.75), and excellent in 1 study. The reliability of contact heat pain was reported in 2 studies. Where a thresholding technique was used to individualise the temperature required to elicit pain at a predetermined intensity, the repeatability of the test stimulus temperature ranged from fair to excellent (ICC = 0.53; ICC = 0.64; and ICC = 0.83). In contrast, the subjective pain rating for the contact heat pain test stimulus ranged from poor to fair (ICC = 0.19; ICC = 0.31; and ICC = 0.4). The reliability of a pain-elicited reflex was reported in 2 studies and ranged from good to excellent (ICC = 0.61; ICC = 0.93) (**Table 2**).

4.4. Reliability of conditioning stimuli

Five studies investigated the intersession reliability of a conditioning stimulus by comparing subjective pain ratings for the stimulus from 2 test sessions. The reliability of pain ratings for immersion in

Table 1

Demographics, CPM paradigm, and reliability results.

Study	Sample size (M/F)	Population, age mean (SD)	Test stimulus, test site	Conditioning stimulus (paradigm)	Retest interval	Reliability coefficient (95% CI)	Response stability
Cathcart et al.[7]	20 (9/11)	HV; male 27 (6.4), female 23 (3.6)	PPT: 1. right middle finger, 2. right trapezius	IP left arm (parallel)	Intrasession	1. Finger ICC = 0.57, 2. shoulder ICC = 0.69 95% CI-NR	CR = 0.35 (±1.69)
Oono et al.[32]	12 (12/0)	HV; 25.6 ± 1.5 (SEM)	PPT, PPTol: 1. right masseter muscle, 2. left forearm, and 3. left tibialis anterior	1. CPT hand, 2. IP upper arm, and 3. pressure pain-head band (parallel)	2 d	NR	Interindividual CV = 41.4%, intraindividual CV = 40.1% for CPM effect with CPT as CS, PPTol at forearm as TS
Lewis et al.[21]	22 (7/15)	HV; 25 (8)	PPT: medial right knee	1. CPT left hand, 2. IP left arm (parallel)	Intrasession; 3 d	Intrasession: CPT ICC = 0.85 (0.62-0.94), IP ICC = 0.75 (0.35-0.90). Intersession: CPT ICC = 0.66 (0.12-0.87), IP ICC = −0.4 (−1.8 to 0.4)	NR
Olesen et al.[31]	62 (38/24)	Painful chronic pancreatitis; 53 (11)	PPT: quadriceps	CPT right hand (sequential)	1 wk	ICC = 0.10 95% CI NR	NR
Martel et al.[24]	55 (35/20)	Chronic back pain; men 48.9 (10.5), women 49.5 (8.9)	PPT: right trapezius	CPT left hand (parallel)	10 d	Overall sample: ICC = 0.59 (0.38-0.74), women: ICC = 0.75 (0.56-0.87), men: ICC = 0.33 (0.12-0.67)	Men ISC = 0.29, women ISC = 0.79, overall ISC = 0.61
Valencia et al.[38]	HV 190 (74/116) patients 134 (87/47)	HV, patients with shoulder pain ; HV 23.02 (6.04), patients 43.83 (17.8)	Contact heat pain (50/100) thenar eminence	CPT contralateral hand (sequential)	HV-intrasession and 1, 3, and 5 d: patients intrasession; presurgery, 3 mo postsurgery	Intrasession, patients, presurgery ICC = 0.54 (0.34-0.68); postop ICC = 0.62 (0.43-0.74). Intrasession HV; ICC = 0.66 (0.55-0.75); ICC = 0.72 (0.62-0.79). Intersession, HVs; female ICC = 0.65 (0.51-0.75); male ICC = 0.82 (0.73-0.88)	Female patients, intrasession presurgery SEM = 5.83; post-surgery SEM = 4.25. Male patients, intrasession presurgery SEM = 7.33; postsurgery SEM = 6.50
Wilson et al.[41]	22 (0/22)	HV; 27 (7)	Contact heat pain (6/10), dominant forearm	HWB (46.5℃) nondominant hand (parallel)	Repeated 8 times over 4 menstrual cycles	ICC = 0.39 (0.23-0.59)	Estimated marginal grand mean ± SE = 1.3 ± 0.3
Biurrun Manresa et al.[4]	34 (34/0)	HV; 27.5 (6.8)	1. NWR threshold at biceps and rectus femoris, 2. electrical pain detection threshold, and 3. pain intensity rating electrical stimulation	CPT contralateral hand (parallel)	Between 1-3 wk (average 11.9 ± 1.9 d)	1. NWR threshold ICC = 0.26 (0-0.55), 2. electrical pain detection threshold ICC = 0.09 (0-0.41), and 3. pain intensity ratings ICC = 0.44 (0.13-0.68).	NWR threshold: Bland–Altman analysis bias = 0.3; LoA = −5.4-6.0; CV (95% CI) = 64.1% (39.1%-81.8%).
Jurth et al.[18]	40 (20/20)	HV; NR	1. NFR biceps femoris (pain 50/100), 2. subjective pain ratings (0-100 NRS)	HWB (parallel)	28 d	CPM effect with NFR ICC = 0.61 (0.36-0.78). Subjective pain ratings for CPM effect ICC = 0.54 (0.26-0.74).	NR
Granovsky et al.[14]	1. 35 (10/25), 2a + b. 30 (15/15)	HV; 1. 26.1 (2.5), 2a + b. 25.9 (2.6)	1. Contact heat pain, 60/100 dominant hand, 2a. 2 thermode + 2b. single test stimulus-contact heat pain, 30/100 nondominant volar forearm	1. HWB dominant hand, 2a + b. contact heat pain dominant upper arm (parallel)	1. 3-7 d, 2a + b. 7 d	CPM effect: 1. ICC = 0.34 (0.03-0.59), 2a. ICC = 0.21 (−0.15 to 0.53), 2b. ICC = 0.59 (0.30-0.78)	NR

CI, confidence interval; CPM, conditioned pain modulation; CPT, cold pressor test; CR, coefficient of repeatability; CS, conditioning stimulus; CV, coefficient of variation; HV, healthy volunteer; HWB, hot water bath; ICC, intraclass correlation coefficient; IP, ischemic pain; ISC, intraindividual stability coefficient; LoA, limits of agreement; MDC, minimal detectable change; M/F, Male/female; NFR, nociceptive flexion reflex; NR, not reported; NWR, nociceptive withdrawal reflex; PPT, pressure pain threshold; PPTol, pressure pain tolerance; SEM, standard error of measurement.

Table 2

Protocol violations in the administration of the test and conditioning stimuli and reliability of test and conditioned stimuli across test sessions.

	TS protocol violations	CS protocol violations	TS test–retest reliability	CS test–retest reliability
Cathcart et al.[7]	PPT-NR	IP-NR	PPT intrasession ICC = 0.82	NR
Oono et al.[32]	PPT, PPTol-NR	CPT 2-4°C, 10 min—most participants did not tolerate on first attempt. IP, mechanical pressure-NR	NR	NR
Lewis et al.[21]	PPT-NR	CPT 12 ± 1°C, 2 min, IP-NR	PPT intrasession ICC = 0.87 (0.60-0.95); PPT intersession ICC = 0.65 (0.05-0.87)	IP NPS intrasession ICC = 0.60 (0.24-0.82), intersession ICC = 0.82 (0.59-0.92). CPT NPS intrasession ICC = 0.94 (0.86-0.98), intersession ICC = 0.80 (0.56-0.92)
Olesen et al.[31]	PPT quadriceps-NR	CPT 2°C, 3 min tolerated for median of 38 s at baseline, 35 s on retest	PPT intersession ICC = 0.79	NR
Martel et al.[24]	PPT trapezius-NR	CPT 4°C, 2 min-NR	PPT intersession ICC = 0.72 (0.56-0.83)	CPT pain ratings ICC = 0.61 (0.41-0.75)
Valencia et al.[38]	Contact heat pain (50/100)-NR	CPT 8°C, 1 min-NR	NR	NR
Wilson et al.[41]	Contact heat pain (6/10)-NR	HWB 46.5°C, 1 min-NR	Temperature °C intersession ICC = 0.83 (0.72-0.91), VNPS intersession ICC = 0.40 (0.24-0.60)	HWB VNPS ICC = 0.79 (0.68-0.89)
Biurrun Manresa et al.[4]	1. NWR threshold-NR; unable to elicit NWR in 5 subjects (13%), 2. Electrical pain detection threshold-NR. 3. pain rating- electrical stimulation-NR	CPT <2°C, 2 min or until reaching 7/10 on VAS—4 of 34 (12%) of subjects did not tolerate continuously	Intersession: 1. NWR ICC = 0.93 (0.87-0.97), 2. electrical pain detection threshold ICC = 0.67 (0.43-0.82), 3. pain intensity ratings ICC = 0.85 (0.71-0.92)	NR
Jurth et al.[18]	NFR-NR	HWB 46.5°C, 200 s-NR, 1 subject excluded	NR	NR
Granovsky et al.[14]	1. Contact heat Pain[60]-NR 2a,b. Contact heat Pain[30]-NR	1. HWB 46.5°C, 1 minute-NR; 2a,b. contact heat, TS + 0.5°C-NR	Intersession: 1. bath-thermode contact heat °C ICC = 0.53; contact heat NPS ICC = 0.31, 2a. 2 thermode contact heat pain °C ICC = 0.64; mean NPS ICC = 0.19, 2b. single test stimulus contact heat pain test stimulus NPS ICC = 0.15	1. Bath-thermode HWB NPS ICC = 0.76, 2a. 2 thermode contact heat pain VAS ICC = 0.16, 2b. single test stimulus not reported

CPT, cold pressor test; CS, conditioning stimulus; HWB, hot water bath; ICC, intraclass correlation coefficient; IP, ischemic pain; NFR, nociceptive flexion reflex; NPS, numerical pain scale; NWR, nociceptive withdrawal reflex; PPT, pressure pain threshold; PPTol, pressure pain tolerance; TS, test stimulus; VAS, visual analogue scale; VNPS, verbal numeric pain scale.

a hot water bath ranges from fair to excellent (ICC = 0.54; ICC = 0.76; ICC = 0.79); for immersion in cold water good to excellent (ICC = 0.61; ICC = 0.80), and for ischemic pain excellent (ICC = 0.82). Poor reliability (ICC = 0.16) was reported for contact heat pain (Pain[30] + 0.5°C) as a conditioning stimulus (**Table 2**).

4.5. Risk of bias in included studies

Results for the assessment of risk of bias are reported in **Table 3**. A moderate to high risk of bias for study participation and study attrition was found. The risk of bias for prognostic factor measurement was moderate as reporting of investigator or participant blinding was lacking. Risk of bias in study confounding ranged from low to high; for outcome measurement was assessed as low and for risk of bias in statistical analysis; and reporting was moderate to high.

5. Discussion

5.1. Summary of results

The aim of this review was to determine if CPM is reliable. This review incorporated 9 studies reporting 23 test–retest measures of various CPM test paradigms in heterogeneous populations, and therefore meta-analysis of results was not appropriate.

However, 78% of reported reliability coefficients for intrasession reliability were interpreted as good (ICC = 0.6-0.75) or excellent (ICC > 0.75). Intersession reliability was reported in 8 studies, and reliability coefficients were interpreted as good or excellent in 50% of studies. The reliability of a CPM paradigm is dependent on test and conditioning stimulus, stimulation parameters, test sites, and study population.

5.2. Reporting and risk of bias

In this review, there was a moderate to high risk of bias for both study participation and study attrition. A recently published consensus paper defines the characteristics of healthy subjects in quantitative sensory testing studies.[11] For the reader to ascertain susceptibility to bias, we suggest in future studies the source of the target population, the sampling frame and methods of recruitment, the place or places and dates of recruitment, study inclusion and exclusion criteria, the numbers recruited to the numbers enrolled, and baseline characteristics of the study sample are reported. In addition to facilitating the assessment of risk of bias, more thorough description of a study sample aids the generalization of results to other populations (**Table 3**).

The aim in rating risk of attrition bias is determining the possibility that the prognostic factor, in this case CPM effect, is

Table 3

Risk of bias in CPM reliability studies (Hayden et al.[16], Hayden et al.[17]).

Study	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Confounding	Statistical analysis and reporting
Cathcart et al.[6]	Moderate	Moderate	Moderate	Low	Low	High
Oono et al.[33]	High	Moderate	Moderate	Low	Moderate	High
Lewis et al.[21]	High	Moderate	Moderate	Low	Low	Moderate
Olesen et al.[32]	Moderate	Moderate	High	Low	High	High
Martel et al.[26]	Moderate	Moderate	Moderate	Low	Moderate	Moderate
Valencia et al.[38]	Moderate	Moderate	High	Low	Moderate	Moderate
Wilson et al.[41]	Low	Low	Moderate	Low	High	Moderate
Biurrun Manresa et al.[3]	Moderate	Moderate	Moderate	Low	High	Low
Jurth et al.[19]	High	Moderate	Moderate	Low	Low	High
Granovsky et al.[14]	High	Moderate	Moderate	Low	Low	Moderate

different for those who complete versus those who do not complete the study. Generally, a moderate risk of attrition bias was found. Study dropouts were not consistently reported, nor was information provided on key characteristics of those who dropped out of the studies which would have enabled an appraisal of whether those who dropped out differed systematically from those who continued in the study.

The risk of bias for prognostic factor measurement was generally moderate; reporting of investigator or participant blinding was lacking. Although assessor blinding is challenging in measures such as CPM, future investigations might consider how this can be addressed. For most studies, it is unclear what information the participants received regarding the experiment which may have influenced their response or created expectation, or what their exposure was between intersession measures. Additionally, there was lack of detail regarding the standardization of test instructions between participants, and in a number of studies the conditioning stimulus was not consistent for all participants.

Risk of bias in study confounding ranged from low to high. In healthy volunteer studies, common exclusions included pain conditions, pain medication, and psychiatric history. However, it was common that baseline and retest measures of health and pain were not used, making the assumption that participants were indeed pain free at retest. Although it is difficult to interpret the effect of confounding on reliability (**Fig. 2**), it would seem that there may be an association. In studies of intersession reliability, there seems to be a trend with lower risk of bias in confounding associated with greater reliability. This would suggest that in studies with lower risk of bias, important factors that may influence the CPM effect were controlled for between sessions, thereby improving repeatability.

The risk of bias in statistical analysis and reporting was rated as moderate to high. The publication dates of the studies included in this review range from 2009 to 2015 and although the reporting of statistical methods has improved with subsequent publications, it is important that improvements continue to be made in this area. As noted previously, the precision of a reliability coefficient is dependent of an appropriate sample size and at present, sample size calculations are generally lacking in CPM reliability studies. And although the model of ICC used for statistical analysis should be reported, this has been consistently under-reported.

It is clear that reducing risk of bias in the conduct and reporting of CPM reliability studies is essential to improve transparency and

make gains towards the identification of robust, reliable CPM paradigms. At present, a moderate to high risk of bias for prognostic factor measurement may be introducing random error into testing, and thereby reducing reliability. As noted above, the same may be said for risk of bias in confounding, with lack of control for important participant-related variables subsequently reducing retest reliability. In contrast, risk of bias for study participation, study attrition and analysis, and reporting may be unintentionally over inflating reliability estimates. It is only with improved rigour in study design and reporting that we can move toward standardisation in testing.

5.3. Reliability of test and conditioning stimulus

Although the test and conditioning stimulus must be noxious, the methods and parameters for delivering these stimuli vary. If a test or conditioning stimulus is overly painful, it is possible that it may not be tolerated by all participants, and therefore the stimulus is not applied uniformly to the sample. There is evidence to suggest that the repeatability of the various test and conditioning stimuli vary across sessions, and this lack of repeatability of the components of the CPM paradigm may reduce the repeatability for the sum of the paradigm (**Table 2**).

For the studies included in this review, there were no reports of participants not tolerating the test stimulus (PPT, contact heat, nociceptive withdrawal, or flexion reflexes) as specified in the study protocols, therefore creating a protocol violation. As the test stimuli described are phasic, this brief exposure to a noxious stimulus seems well tolerated. In comparison, the conditioning stimuli reported (ischemic pain, cold pressor test [CPT], contact heat, hot water bath, and contact heat) are tonic, vary in intensity and exposure, and in how well they are tolerated by participants. Using ischaemic pain[33] and contact heat[14] as conditioning stimuli, there were no reported participant withdrawals, ie, all participants tolerated the stimulus for the period specified in the protocol. In contrast, participant tolerance to immersion in the CPT and hot water bath appear to be time and temperature dependent. This suggests that CPT temperatures of between 8° and 12°C and for up to 2 minutes and hot water bath immersion at 46.5°C for 1 minute are sufficient to induce inhibition and are well tolerated by participants, ensuring that the conditioning stimulus is consistent for all participants and thereby perhaps improving repeatability. This is consistent with the findings of Granot et al.[13] regarding the intensity of heat and cold pain necessary to induce
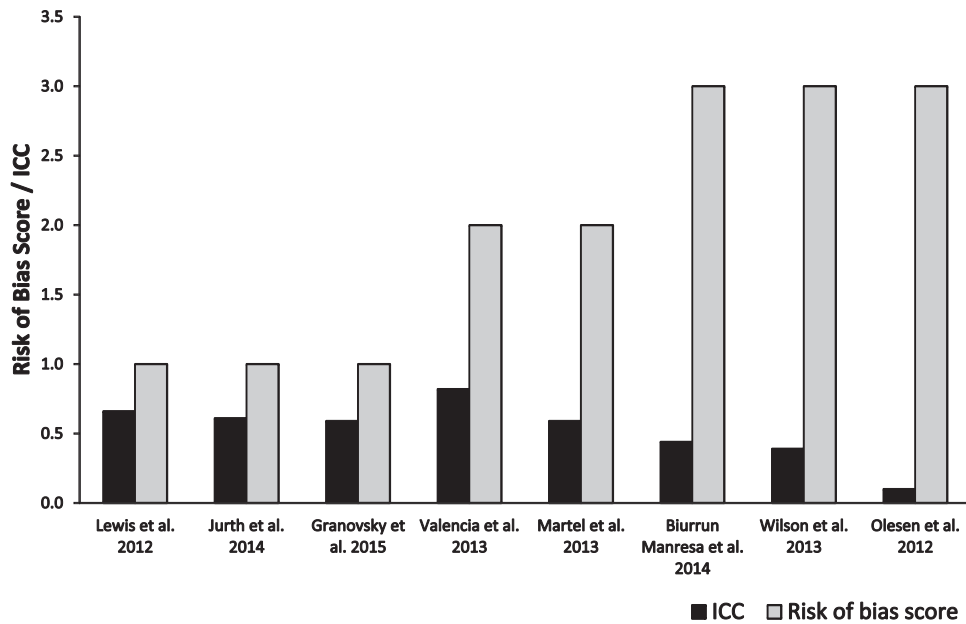
Figure 2. Risk of bias in study confounding and reliability. The ICC is the highest reported reliability coefficient for CPM effect. For risk of bias score, 1 = low risk; 2 = moderate risk; and 3 = high risk. CPM, conditioned pain modulation; ICC, intraclass correlation coefficient.

CPM. These findings have important implications for the investigation of CPM paradigms in populations with chronic, painful conditions; if a stimulus is not well tolerated by a sample of healthy volunteers, it is perhaps even less likely to be tolerated by patients who are in pain.

5.4. The reliability of parallel vs sequential paradigms

Two studies, Olesen et al.[32] and Valencia et al.[39] investigated sequential CPM paradigms with reliability reported as poor, and good to excellent, respectively. The remainder investigated parallel paradigms with intersession reliability ranging from poor to good; therefore, it is impossible to conclude from the available evidence if there is greater reliability for one paradigm over another (**Table 1**).

5.5. Timing of intrasession assessments

For the 3 studies that investigated intrasession reliability, the wash-out period between intrasession assessments included 2 minutes, 15 minutes, and 60 minutes,[6,23,38] respectively. With a 2-minute wash-out reliability ranged from fair to good, for 15 minutes good to excellent, and for 60 minutes fair to good; therefore, it is difficult to discern the impact of wash-out time on intrasession reliability from this review (**Table 1**).

5.6. Nonresponders

An important consideration in the clinical or experimental utility of a CPM paradigm is whether or not the paradigm induces a CPM effect and, if so, in what proportion of subjects. Although the reporting of absolute and percentage change in CPM effect speaks to the magnitude of change, that is, the reduction in pain ratings or increase in threshold of the test stimulus after exposure to the conditioning stimulus, this approach does not consider the measurement error inherent in the test stimulus and may be misleading. Locke et al.[25] has described the calculation of

a meaningful CPM effect as a percentage change from baseline (increase in pain threshold or decrease in pain ratings) greater than the inherent measurement error. In this review, judging from the reported value and standard deviation for the CPM effect, it is clear that there are differences in the response to the various CPM paradigms with some participants demonstrating inhibition of pain and others demonstrating facilitation. Although some investigators have described "non-responders," this reporting is not standardized and requires improvement for transparency. Although the consideration of measurement error in the calculation of a potentially clinically meaningful effect is new to CPM studies, it is statistically robust and widely used for the interpretation of change scores.[34,36] This approach may aid the interpretation of results across studies.

5.7. Important findings regarding conditioned pain modulation test design

After exposure to a CPM conditioning stimulus, it is unclear how long pain inhibition persists. Although it may be stimulus dependent, pain inhibition secondary to cold water immersion continues 10 minutes after removal of the conditioning stimulus but has resolved at 15 minutes.[24] The time for resolution of inhibition has important implications for intrasession reliability studies and studies investigating multiple pain measures.

Cold water immersion was the most frequently reported conditioning stimulus in this review; however, stimulus parameters vary. Olesen et al.[32] used cold water immersion at 2°C for 3 minutes as a conditioning stimulus and reported that most patients were unable to remain in the conditioning stimulus for 3 minutes because of the intensity of pain, suggesting these may be inappropriate parameters for patients with a painful condition. In this study, the reliability of the CPM effect was poor (ICC = 0.10) possibly because of random error introduced by systematic differences in exposure to the conditioning stimulus.

The choice of outcome measure or response has important implications for CPM reliability. Static measures of PPT, or the

point where stimulation just becomes painful, demonstrate good to excellent reliability and in contrast, when statically measuring pressure pain tolerance, or the point when the painfulness of stimulation just becomes intolerable, retest reliability is poor to fair.[32] Similarly, a difference is seen in the outcome or response measure to contact heat with the individualised temperature of the contact heat pain test stimulus demonstrating fair to excellent reliability, whereas the pain ratings for exposure to contact heat range from poor to fair.

There is evidence for sex differences in CPM effect. Martel et al.[26] investigated CPM in patients with back pain, assessing the influence of demographics including age, sex, medication use, pain severity, and psychological factors including catastrophising and negative affect. They reported sex differences for the magnitude and stability of the CPM effect; however, regarding demographic and psychological variables, there was no significant association with CPM magnitude or stability and sex. This was supported by Valencia et al.[39] in an investigation of the influence of shoulder pain intensity and sex on CPM stability in patients with presurgical and postsurgical shoulder pain and in healthy volunteers with exercise-induced shoulder pain. They found while the reliability of CPM was not related to shoulder pain intensity in either group, the reliability of the CPM effect differed between sexes with female patients and male healthy volunteers demonstrating greater reliability.

Objective measures such as pain-elicited reflexes are appealing as test stimuli for their potential to decrease subjectivity and random error and therefore to improve reliability. Biurrun Manresa et al.[3] and Jurth et al.[19] investigated the intersession reliability of CPM in healthy volunteers using the nociceptive withdrawal or flexion reflex as an objective, reliable measure of spinal nociceptive processing[4] as a test stimulus. Biurrun Manresa et al.[4] reported excellent reliability for the repeatability of the pain-elicited reflex test stimulus, whereas the reliability of the cold water immersion–induced CPM effect was poor. In contrast, Jurth et al.[19] reported good reliability for the hot water–induced CPM effect. These results suggest the pain-elicited reflex may be a reliable test stimulus, and the difference in the reliability of the CPM effect in the 2 studies may be secondary to the parameters of the conditioning stimulus. The pain-elicited reflex may be found to increase the objectivity and reliability of the CPM paradigm and warrants further investigation in other populations and in combination with other noxious conditioning stimuli.

As standardization in the testing of CPM is lacking, it is important to consider novel test paradigms. Granovsky et al.[14] investigated the reliability of CPM in healthy volunteers using a protocol which was novel for introducing the second test stimulus before rather than after the introduction of the conditioned stimulus. The intersession reliability was reported as fair (ICC = 0.59); however, it is possible that in using a predetermined value for tonic heat pain as a conditioning stimulus habituation to temperature may occur, with the intensity of the conditioning stimulus dropping below that necessary to induce CPM in some subjects.[13] Although the single-test stimulus paradigm is enticing for the reduction in testing time, further reliability studies including an investigation of response stability are warranted.

Although work is required to standardise the evaluation and interpretation of CPM as an experimental and clinical measure, it is apparent that CPM has great potential as a clinically important measure or biomarker. In a systematic review and meta-analysis, Lewis et al.[24] appraised the risk of bias and synthesised the evidence from 30 studies comparing CPM between chronic pain populations and control groups. They reported that nearly 70% of comparisons revealed a statistically significant reduction in CPM in patients with chronic pain and an acceptable level of bias in included studies, providing good evidence that patients with chronic pain conditions have a significantly reduced CPM effect as compared with healthy individuals. In surgical populations, it has been reported that patients with less efficient CPM are at greater risk of developing chronic postoperative pain[40,44] and that CPM may be predictive of subsequent pain relief (Wilder–Smith, personal communication). In pharmacological studies, it has been demonstrated that in patients with painful diabetic neuropathy, CPM predicts the analgesic effectiveness of duloxetine[45] and tapentadol (Niesters et al, personal communication) and can be activated by tapentadol.[30]

Although it seems that CPM is often deficient in patients with chronic pain conditions, it is unclear to what degree deficient endogenous pain modulation may be a cause or an effect of the chronic pain condition. Emerging evidence suggests that deficient CPM may be the result of a chronic pain condition, whether that pain be neuropathic or nociceptive in nature, and that when pain is alleviated, CPM is restored. This restoration or rescue of CPM has been demonstrated with the pharmacological treatment of pain[30,44] and after joint replacement surgery in patients with painful hip osteoarthritis[20] and painful knee osteoarthritis.[15]

Questions persist as well as to the nature of CPM as a stable trait or a transient state and as to how CPM is influenced by environment and context. Although it is known from animal studies that DNIC in the rat can function independently of cortical control, it is unclear in humans how the descending modulation of pain may be cognitively confounded.[2] It may be that patients with chronic pain have difficulty disengaging from their pain toward a distracting stimulus, or that psychological factors such as anxiety or hypervigilance interfere with the pain inhibition response.[2] It has been demonstrated in humans that cognitive manipulation can effect CPM; pain inhibition under CPM seems to depend on the perceived level of the conditioned stimulus pain rather than solely on its physical intensity.[31] Additionally, in humans, there is evidence to support an association of mood and affect with CPM. In a double-blind placebo-controlled randomized trial of intranasal oxytocin, Goodin et al.[12] demonstrated that oxytocin augmented CPM and reduced negative mood and anxiety.

There is evidence to suggest much potential for CPM to serve as a useful prognostic factor and predictor of response to therapeutic intervention in patients with chronic and neuropathic pain. As such, the evaluation of CPM may aid clinical decision making, assist in informing patients about possible outcomes, be used to identify risk groups for stratified management, and be a potentially modifiable target.[17] However, for a measure such as CPM to be a clinically useful prognostic factor, it must produce consistent results with minimal measurement error, ie, it must be reliable. Estimating the reliability of CPM presents a challenge because just as there has been much heterogeneity in the investigations of CPM testing paradigms, variability in the analysis and reporting of the reliability of CPM has been equally heterogeneous.

5.8. Review limitations

No meta-analysis was performed; therefore our findings regarding the reliability of CPM amount to a qualitative synthesis of the evidence. Additionally, our findings are limited by the quality of reporting in the included studies. Although we attempted to control for the induction of reviewer bias by relying on double screening of studies, data extraction, and assessment of risk of bias, the risk of reviewer bias is nonetheless a consideration.

6. Conclusions

There is evidence to suggest that CPM is a reliable measure; however, the degree of reliability is dependent on stimulation parameters, study methodology, and the population of interest. The validation of CPM as a robust prognostic factor in experimental and clinical pain studies will be facilitated by improvements in the reporting of CPM reliability studies.

6.1. Recommendations for future research

It has been recommended that the CPM effect should be reported as both the absolute change and the percent change (when appropriate for the level of measurement) in the perceived test stimulus induced by the conditioning stimulus and a measure of variability should be included.[42] Recommendations for future reliability studies include due consideration of how the results for a sample of participants may be generalized to a population of interest. Gierthmuhlen et al.[11] has described important data collection domains for healthy volunteer quantitative sensory testing studies which may be equally pertinent for dynamic measures such as CPM, including but not limited to sociodemographic data, medical history and current health status, pain coping strategies, psychological factors, history of alcohol and drug abuse, smoking and use of recreational drugs, current medication, depression and anxiety scores, the frequency of any pain episodes during the last 3 to 6 months, and self-reported sleep measurements. Consideration should be given to blinding of both the investigator and the participants of CPM studies, standardization of test instructions, and as to how the test environment and exposure to investigators and other study participants may bias performance or results. The intensity and exposure time for the conditioning stimulus should be of a magnitude that the stimulus is uniform for all participants. Attempts to control for known confounders should be made, with an accounting of confounders at both baseline and retest. Lastly, improvements in the statistical design and analysis of CPM reliability studies are essential if progress is to be made toward standardization in CPM testing and reporting. The inclusion of a sample size calculation, an appropriate reliability coefficient and 95% confidence interval, and a measure of response stability will aid the interpretation of results and the comparison between studies. Thorough data reporting including measures of central tendency and variability for ratings for test stimulus, conditioning stimulus, conditioned test stimulus and CPM magnitude, the number of responders and nonresponders and how this was established, the intrasession or intersession reliability for the test and conditioning stimulus, and where appropriate the absolute and percentage change for the CPM effect will aid comparison of testing paradigms across studies and substantiate the repeatability and inherent variability of the CPM paradigm.

Conflict of interest statement

The authors have no conflicts of interest to declare.

Acknowledgements

Appendix A. Supplemental Digital Content

Supplemental Digital Content associated with this article can be found online at http://links.lww.com/PAIN/A338.

Supplemental media

Video content accompanying this article can be found at http://links.lww.com/PAIN/A330.

References

[1] Baumgartner TA. Norm-referenced measurement: Reliability. MJ Safrit, TM Wood (Eds.) Champaign, IL: Human Kinetics, 1989.

[2] Bingel U, Tracey I. Imaging CNS modulation of pain in humans. Physiology (Bethesda) 2008;23:371–80.

[3] Biurrun Manresa JA, Fritsche R, Vuilleumier PH, Oehler C, Morch CD, Arendt-Nielsen L, Andersen OK, Curatolo M. Is the conditioned pain modulation paradigm reliable? A test-retest assessment using the nociceptive withdrawal reflex. PloS One 2014;9:e100241.

[4] Biurrun Manresa JA, Neziri AY, Curatolo M, Arendt-Nielsen L, Andersen OK. Test-retest reliability of the nociceptive withdrawal reflex and electrical pain thresholds after single and repeated stimulation in patients with chronic low back pain. Eur J Appl Physiol 2011;111:83–92.

[5] Bruton A, Conway JH, Holgate ST. Reliability: what is it and how is it measured? Physiotherapy 2000;86:94–9.

[6] Cathcart S, Winefield AH, Rolan P, Lushington K. Reliability of temporal summation and diffuse noxious inhibitory control. Pain Res Manag 2009; 14:433–8.

[7] Coghill RC, Yarnitsky D. Healthy and normal? The need for clear reporting and flexible criteria for defining control participants in quantitative sensory testing studies. PAIN 2015;156:2117–8.

[8] Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. Nature 2014;505:612–13.

[9] Edwards RR, Ness TJ, Weigent DA, Fillingim RB. Individual differences in diffuse noxious inhibitory controls (DNIC): association with clinical variables. PAIN 2003;106:427–37.

[10] Ge HY, Madeleine P, Arendt-Nielsen L. Sex differences in temporal characteristics of descending inhibitory control: an evaluation using repeated bilateral experimental induction of muscle pain. PAIN 2004;110:72–8.

[11] Gierthmuhlen J, Enax-Krumova EK, Attal N, Bouhassira D, Cruccu G, Finnerup NB, Haanpaa M, Hansson P, Jensen TS, Freynhagen R, Kennedy JD, Mainka T, Rice A, Segerdahl M, Sindrup SH, Serra J, Tolle 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 2015;156:2203–11.

[12] Goodin BR, Anderson AJB, Freeman EL, Bulls HW, Robbins MT, Ness TJ. Intranasal oxytocin administration is associated with enhanced endogenous pain inhibition and reduced negative mood states. Clin J Pain 2015;31:757–67.

[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 2008;136:142–9.

[14] Granovsky Y, Miller-Barmak A, Goldstein O, Sprecher E, Yarnitsky D. CPM test-retest reliability: "standard" vs "single test-stimulus" protocols. Pain Med 2016;17:521–29.

[15] Graven-Nielsen T, Wodehouse T, Langford RM, Arendt-Nielsen L, Kidd BL. Normalization of widespread hyperesthesia and facilitated spatial summation of deep-tissue pain in knee osteoarthritis patients after knee replacement. Arthritis Rheum 2012;64:2907–16.

[16] Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med 2006;144:427–37.

[17] Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med 2013; 158:280–6.

[18] Higgins JPT, Green S (eds). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

[19] Jurth C, Rehberg B, von Dincklage F. Reliability of subjective pain ratings and nociceptive flexion reflex responses as measures of conditioned pain modulation. Pain Res Manag 2014;19:93–6.

[20] Kosek E, Ordeberg G. Abnormalities of somatosensory perception in patients with painful osteoarthritis normalize following successful treatment. Eur J Pain 2000;4:229–38.

[21] Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, Crystal RG, Darnell RB, Ferrante RJ, Fillit H, Finkelstein R, Fisher M, Gendelman HE, Golub RM, Goudreau JL, Gross RA, Gubitz AK, Hesterlee SE, Howells DW, Huguenard J, Kelner K, Koroshetz W, Krainc D, Lazic SE, Levine MS, Macleod MR, McCall JM, Moxley RT III, Narasimhan K, Noble LJ, Perrin S, Porter JD, Steward O, Unger E, Utz U, Silberberg SD. A call for transparent reporting to optimize the predictive value of preclinical research. Nature 2012;490:187–91.

[22] Le Bars D, Dickenson AH, Besson JM. Diffuse noxious inhibitory controls (DNIC). I. Effects on dorsal horn convergent neurones in the rat. PAIN 1979;6:283–304.

[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 2012;17:98–102.

[24] 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.

[25] 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. J pain 2014;15:1190–8.

[26] Martel MO, Wasan AD, Edwards RR. Sex differences in the stability of conditioned pain modulation (CPM) among patients with chronic pain. Pain Med 2013;14:1757–68.

[27] Millan MJ. Descending control of pain. Prog Neurobiol 2002;66: 355–474.

[28] Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009;62:1006–12.

[29] Nahman-Averbuch H, Yarnitsky D, Granovsky Y, Gerber E, Dagul P, Granot M. The role of stimulation parameters on the conditioned pain modulation response. Scand J Pain 2013;4:10–14.

[30] Niesters M, Proto PL, Aarts L, Sarton EY, Drewes AM, Dahan A. Tapentadol potentiates descending pain inhibition in chronic pain patients with diabetic polyneuropathy. Br J Anaesth 2014;113: 148–56.

[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 2012;153:170–6.

[32] Olesen SS, van Goor H, Bouwense SA, Wilder-Smith OH, Drewes AM. Reliability of static and dynamic quantitative sensory testing in patients with painful chronic pancreatitis. Reg Anesth Pain Med 2012;37:530–6.

[33] Oono Y, Hongling N, Lima Matos R, Wanga 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.

[34] Portney LG, Watkins MP. Foundations of Clinical Research. New Jersey: Prentice Hall Health, 2000.

[35] Pud D, Granovsky Y, Yarnitsky D. The methodology of experimentally induced diffuse noxious inhibitory control (DNIC)-like effect in humans. PAIN 2009;144:16–19.

[36] Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil 1998;12: 187–99.

[37] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.

[38] Valencia C, Fillingim RB, Bishop M, Wu SS, Wright TW, Moser M, Farmer K, George SZ. Investigation of central pain processing in post-operative shoulder pain and disability. Clin J Pain 2014;30:775–86.

[39] Valencia C, Kindler LL, Fillingim RB, George SZ. Stability of conditioned pain modulation in two musculoskeletal pain models: investigating the influence of shoulder pain intensity and gender. BMC Musculoskelet Disord 2013;14:182.

[40] Wilder-Smith OH, Schreyer T, Scheffer GJ, Arendt-Nielsen L. Patients with chronic pain after abdominal surgery show less preoperative endogenous pain inhibition and more postoperative hyperalgesia: a pilot study. J Pain Palliat Care Pharmacother 2010;24:119–28.

[41] 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 2013;154:2633–38.

[42] 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. Eur J Pain 2010;14:339.

[43] 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. Eur J Pain 2015;19:805–6.

[44] 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.

[45] 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.