Title page

1) Title: Psychophysical or spinal reflex measures when assessing conditioned pain modulation?

2) Running head: Comparing CPM effects with different test-stimuli

3) Authors: M. U. Lie[1,2], E. Petriu[3,4], D. Matre[5], P. Hansson[6,7], O. K. Andersen[8], J. A. Zwart[1,2,4], K. B. Nilsen[1,4,9]

4) Affiliations:

1. Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital, Oslo, Norway
2. Faculty of Medicine, University of Oslo, Oslo, Norway
3. Faculty of Health Science, Oslo Metropolitan University, Oslo, Norway
4. Department of Neurology, Oslo University Hospital, Oslo, Norway
5. Department of Work Psychology and Physiology, National Institute of Occupational Health, Oslo, Norway
6. Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
7. Department of Pain Management and Research, Division of Emergencies and Critical Care, Oslo University Hospital, Oslo, Norway
8. Center for Neuroplasticity and Pain (CNAP), SMI®, Department of Health Science and Technology, Aalborg University, Aalborg, Denmark
9. Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway

5) Corresponding author:

Marie Udnesseter Lie
FORMI, Oslo University Hospital
Pb. 4956 Nydalen
0424 Oslo
Norway
Tel: +47 22117768
E-mail: maulie@ous-hf.no

6) Manuscript category: original article

7) Funding: The present study received funding from the Oslo University Hospital and Oslo Metropolitan University. Author Ole Kæseler Andersen received funding from the Danish National Research Foundation (DNRF121).

8) Conflict of interests: The authors have no conflict of interests to declare.

9) Significant statement: The large difference in CPM effect between the two protocols suggests that the CPM effect relates to pain perception rather than nociception on the spinal level. Due to poor absolute intra-rater reliability, we recommend caution and further research before using any of the investigated CPM protocols in clinical decision making on an individual level.

Introduction

Assessment of endogenous pain modulatory function may carry a potential for stratification of treatment and follow-up of pain patients. One such measure is conditioned pain modulation (CPM), which assesses an individual's inherent ability to alter the central nervous system processing set up by a nociceptive stimulus (termed test-stimulus) in the presence of another nociceptive stimulus (termed conditioning stimulus) (Yarnitsky et al., 2010). CPM has been shown to be altered in several chronic pain conditions (Lewis et al., 2012) and deficits may predict development of post-operative pain (Wilder-Smith et al., 2010; Yarnitsky et al., 2008) and treatment response (Nahman-Averbuch, Dayan, et al., 2016; Yarnitsky et al., 2012). There is, however, a large variation in applied CPM methodology, which limits the generalization of conclusions for application in daily clinical practice (Pud et al., 2009). Thus, there is a need for standardized and reliable methods to measure CPM (Yarnitsky et al., 2015).

CPM is usually assessed with psychophysical outcome measures, i.e., pain intensity ratings of supra-threshold stimuli or pain threshold assessment (Kennedy et al., 2016; Pud et al., 2009), clearly involving subjective interpretation of the stimulus-induced percept. A systematic review suggests that CPM is a reliable measure, but the degree of reliability heavily depends on methodology (Kennedy et al., 2016). In a previous study, we reported large variability when using a protocol involving a thermal test-stimulus (Lie et al., 2017). Assessing CPM with standardized spinal reflex measures such as the nociceptive withdrawal reflex (NWR) elicited by electrical stimulations, may potentially be more reliable since such a measure may be less influenced by cognitive processes than psychophysical measures (Sandrini et al., 2005). One must however keep in mind that the withdrawal to the electrical stimulus is a reflex and not dependent of pain perception. Although not a painful outcome measure, it is commonly used as test-stimulus in CPM studies (Pud et al., 2009). The reliability of neuronal activity induced by an electrical test-stimulus has been investigated (Biurrun Manresa et al., 2014; Jurth et al., 2014), but not compared with more commonly used psychophysical stimuli such as thermal test-stimuli. Therefore, the aim of the present study was to compare the CPM effect and test-retest reliability between a CPM protocol using a thermal test-stimulus and a psychophysical outcome with a CPM protocol with an electrical test-stimulus and a spinal reflex outcome.

Methods

2.1 Study design

This was an experimental crossover study comparing two CPM protocols with different test-stimuli (thermal vs. electrical) and different outcome measures (psychophysical vs. spinal reflex) but the same conditioning stimulus. A pretest was performed to familiarize subjects with the stimulations and pain intensity rating procedures, before the baseline test-stimulus was applied according to the protocol. After a 5-minute break, the test-stimulus in parallel with a conditioning stimulus was applied. A 30-minute break followed to eliminate carryover effects before the other protocol was carried out with the same procedure contralaterally (Fig. 1). The experiment was repeated with a minimum interval of 7 days. The second session was identical to the first session in regards to randomization.

A computerized block-randomization for the order of protocol and the test side was conducted prior to the experiments. The subjects were informed of the testing procedure, but were not told whether the conditioning stimulus would influence the test-stimulus and were thus blinded for the study hypothesis. Subjects were also blinded for readouts from the stimulation instruments. A female experimenter (E.P) carried out all experiments. Instructions, placement of instruments, room temperature (21°C–23°C), and the experimenter's clothes were standardized.

A written informed consent was obtained prior to participation. The study was approved by the Regional committee for medical and health research ethics (project no. 2010/2927) and conducted in accordance with the Declaration of Helsinki. Subjects received a gift certificate of NOK 500 for participation.

2.2 Subjects

Men and women self-reported to be healthy and aged 18-45 years were recruited by advertisement at local hospitals and colleges/universities in Oslo, Norway. Exclusion criteria were: chronic pain, somatic or psychiatric disease, headache for more than two days a month, hypertension (>140/90, assessed prior to the experiment after a 5 minute rest), pregnancy (self-reported), breastfeeding, acquaintance with the experimenter and regular use of medication (including non-prescription pain killers) except oral contraceptives. Subjects were requested not to work nightshifts 48 hours before the experiment, not to consume alcohol or pain killers 24 hours before the experiment, or caffeine or tobacco the last hour before the experiment.

A priori power analysis based on previous studies from our laboratory (Lie et al., 2017; Nilsen et al., 2014) showed that 20 subjects were needed to detect a 10% difference in the absolute CPM effect in a paired Student t-test between the two protocols with a standard deviation of 1.5 cm on a 10 cm visual analog scale (VAS, left end: 'no pain', right end: 'worst pain imaginable'), assuming a two-sided significance level of 5% and 80% power.

1.3 Test-protocol

2.3.1 Psychophysical outcome

Test-stimulus was contact heat stimulation induced by a 30×30 mm Peltier thermode (Medoc, Ramat Yishai, Israel) (baseline temperature: 32°C, increase rate: 2°C/second, decrease rate: 8°C/second) applied on the proximal volar aspect of the forearm with a constant temperature for 120 seconds (Fig. 2a). The subjects continuously rated the pain intensity of the test-stimulus on a computerized 10 cm horizontal VAS by scrolling the wheel on a computer mouse. The stimulation site of the test-stimulus alone and the test-stimulus in parallel with the conditioning stimulus was not overlapping to avoid sensitization or habituation. The temperature of the test-stimulus was aimed to reflect pain intensity equal to approximately 6 cm on 10 cm VAS. In order to find this temperature the following procedure was followed: An average of three tests of heat pain tolerance threshold tested with the methods of limits (baseline: 32°C, increase rate: 1°C/second) minus 2°C was calculated. The estimated temperature was tested with a 30 seconds heat stimulus positioned on the volar aspect of the opposite forearm. If the first 20 seconds was rated outside 4-9/10 cm VAS the temperature was adjusted accordingly.

2.3.2 Spinal reflex outcome

Subjects were lying at a medical plinth with the back rest inclined 135 degrees relative to the horizontal level, and a pillow under the knees assuring knee flexion of 45 degrees. At stimulation sites existing hair was removed and the skin was lightly abraded and cleaned with sterilizing alcohol.

Electrocutaneous stimulation was applied through surface Ag/AgCl-electrode (30x22 mm, type Neuroline 720, Ambu A/S Denmark) placed on the medial aspect on the arch of the foot, and a large surface electrode (5x10 cm, Axelgaard, USA) placed on the dorsum of the foot just proximal to the toes (Fig. 2b). This ensured that the stimulus was perceived in the arch of the foot. The electrodes were repositioned if subjects felt radiating sensation into the toes or on the dorsum of the foot. Recording electrodes were placed on the ipsilateral tibialis anterior muscle by three surface Ag/AgCl-electrodes (30x22 mm, type Neuroline 720, Ambu A/S Denmark, one reference electrode) with an inter-electrode distance of 2 cm. The skin was cleaned and abraded again if high impedance (>5kOhm) occurred.

Trains of five 1 ms rectangular pulses (felt as a single stimulus) was delivered at 200 Hz with a 4 ms interpulse interval with Dolosys Paintracker (Dolosys GmbH, Berlin, Germany). Spinal reflexes measures may be difficult to standardize in clinical settings, and we wanted to use a commercial device which could be easy to implement in forthcoming clinical studies. Dolosys Paintracker is a commercial device which is easy to transport and set up and is therefore beneficial to use as a bedside-/point-of-care-test compared to other devices currently used to induce and measure

electrical stimuli. The device is specifically developed to determine reflex thresholds continuously over a longer period of time.

The intensity of the electrical stimulus was the current needed to evoke a reflex threshold with inter-stimulus intervals randomized between 8-12 seconds to minimize stimulus predictability. The amplitude of the electromyographic (EMG) reflex responses to the electrical stimulations was converted to a peak z-score defined as the baseline-adjusted maximum divided by standard deviation of the EMG amplitudes before stimulation. The NWR threshold was defined as a peak z-score of≥12 in the post-stimulus interval of 70-150 ms (France et al., 2009). Electrical stimulations started at 1mA and increased with a rate of 0.5 or 1mA until threshold was detected (minimum 8 values were needed for threshold calculation. After threshold detection, repeated stimulations were given for 120 seconds, resulting in a total of 10 electrical stimulations due to the inter-stimulus interval. Each stimulus was adjusted to be as close to a peak z-score of 12, e.g. if the previous stimulus elicited a large response, the intensity of the next stimulus was decreased. If this resulted in a threshold below a peak z-score of 12, the next stimulus was increased. Values of the 3 previous stimulations were used to determine if the intensity changed by 0.5mA or 1.0 mA, which ensures precise threshold determination with the smallest possible steps (Instructions for use Paintracker, Dolosys GmbH). Subjects were told to relax their leg as much as possible, and were reminded to relax if muscle contractions in the leg (high EMG noise) were present between stimulations.

The overall level of unpleasantness and pain intensity of the electrical stimulations were rated verbally on a 0-10 numerical rating scale (NRS) (0 = 'no pain' / 'no unpleasantness', 10 = 'worst pain imaginable' / 'worst unpleasantness imaginable') after the test-stimulus was terminated.

2.4 Conditioning stimulus

A 7°C circulating water bath (LAUDA Alpha RA8, LAUDA-Brinkman LP., USA) was used as conditioning stimulus in both protocols at the hand contralateral to the test-stimulus side (Fig. 2c). With water up to the wrist, the hand was held wide open and steady for 120 seconds or until the pain forced the subject to withdraw the hand from the water bath. After 120 seconds, subjects were asked to rate the overall pain intensity of the conditioning stimulus on a 0-10 NRS (0 = 'no pain', 10 = 'worst pain imaginable').

2.5 Data analysis

The CPM effect was defined as the difference in average pain intensity or NWR threshold between the test-stimulus alone and the test-stimulus in parallel with the conditioning stimulus. The CPM effect was also calculated as a percent change (CPM effect/test-stimulus alone x 100). The percent change in CPM effect was used when comparing the two protocols due to different parameters used to calculate the CPM effect. Additionally, subjects were categorized as CPM responders or non-responders. Subjects with decreased pain ratings during conditioning stimulus were defined as CPM responders in the protocol with the thermal test-stimulus, while subjects with increased reflex threshold were defined as CPM responders in the protocol with the electrical test-stimulus.

Statistical analyses were conducted using SPSS Statistics v. 21 (IBM, Armonk, NY). Findings with P-values ≤0.05 were regarded as significant. The distribution of data was assessed in preliminary analyses by a Shapiro–Wilk test and inspection of descriptive statistics, histograms, boxplots, and Q-Q plots. These analyses did not indicate any extreme values or distributions that would affect the planned parametric analysis.

To determine whether a CPM effect was present, pain ratings or NWR threshold during the test-stimulus alone were compared with pain ratings or NWR thresholds during the test-stimulus in parallel with conditioning stimulus in paired sample Student t tests. Differences in CPM effect between the two protocols were estimated with repeated-measures analysis of variance (RM ANOVA), with session (levels: first session vs second session) and protocol (levels: thermal protocol vs electrical protocol) as factors.

Intraclass correlation coefficients with a 2-way random-effect model ($ICC_{2,1}$) and absolute agreement definition for single measures were used to assess relative reliability (0.4: poor reliability; 0.4–0.59: fair reliability; 0.6–0.75: good reliability; 0.75: excellent reliability (Shrout et al., 1979).

Bland-Altman plot and its related limits of agreement were used to assess the absolute reliability. Bias was calculated as the mean difference between the two sessions by subtracting the mean CPM effect in the first session from the second session, and then evaluated with a 1-sample Student's T test. 95% limits of agreement was calculated as mean difference ± 1.96 x SDdiff (SDdiff = SD of the mean difference).

Results

Twenty-eight subjects were included in the study. One subject did not participate in the second session for unknown reasons. One subject was excluded when previous participation in a similar study was revealed and one subject was excluded due to missing data because of technical issues. Thus, a total of 25 (11 females) were included in the analysis. Sample characteristics are presented in table 1.

3.1 CPM effect

The mean CPM effect for the thermal protocol was -2.2 cm, representing a -46.0% decrease between pain ratings during test-stimulus alone and pain ratings during test-stimulus in parallel with the conditioning stimulus ($p<0.001$) (Fig. 3a). The mean CPM effect for the electrical protocol was 0.4 mA, representing a 4.5% increase between the NWR threshold during test-stimulus alone and NWR threshold during test-stimulus in parallel with conditioning stimulus ($p=0.216$) (Fig. 3b). The difference in CPM effect between the two protocols was significant ($p<0.001$) (Fig. 4) with a partial eta^2 effect size of 0.7. No significant differences in CPM effect was found between sessions ($p=0.618$), and no interactions between sessions and protocols ($p=0.949$). Post hoc analysis (RM ANOVA adjusted for changes in thresholds) showed, in contrast to the NWR thresholds, a significant CPM effect when using pain ratings (-32.5% decrease, $p=0.002$, partial eta^2 effect size 0.4) or unpleasantness ratings (-26.1% decrease, $p<0.001$, partial eta^2 effect size 0.5) of the electrical test-stimulus, comparing ratings during test-stimulus alone with ratings during test-stimulus in parallel with the conditioning stimulus. A mean baseline noise of 0.6 µV was found with no significant difference between test-stimulus alone and during test-stimulus in parallel with conditioning stimulus, indicating low baseline muscle activity in both conditions.

3.2 Reliability

Detailed reliability values are shown in Table 2 and 3. The ICC values of the CPM effect in both protocols were in the 0.40 – 0.59 range, which suggests fair relative reliability. Regarding absolute reliability, no bias was observed as there was no significant difference in mean difference between sessions in the protocol with thermal test-stimulus ($p = 0.631$) or the protocol with electrical test-stimulus ($p = 0.616$). Large limits of agreement were observed for the CPM effect in both protocols, which indicates large intra-individual differences between sessions (Fig. 5).

Discussion

Our data showed significantly larger CPM effect using a protocol with a psychophysical outcome from using a thermal test-stimulus compared to a spinal reflex outcome using an electrical test-stimulus, where the latter protocol failed to detect a CPM effect. Fair relative reliability was observed for the CPM effect in both protocols. The absolute reliability indices in both protocols displayed good agreement in the mean CPM effect between the two sessions. However, high intra-individual variability was observed for both protocols.

4.1 CPM effect

The large difference in CPM effect between the two protocols (41.5 %) indicate that the perceptual pain experience from a thermal test-stimulus is more prone to modulation during the conditioning stimulation than the nociceptive withdrawal reflex assessed by an EMG response to an electrical test-stimulus. This is somewhat consistent with previous studies. Studies using 120 seconds heat test-

stimulus report a CPM effect between -29 – -47% (Lie et al., 2017; Matre et al., 2016; Nilsen et al., 2014; Potvin et al., 2008; Tousignant-Laflamme et al., 2008), while studies using electrical test-stimulus giving rise to a NWR, report a CPM effect between 11.5 – 40% (Biurrun Manresa et al., 2014; Bouhassira et al., 2003; Jurth et al., 2014; Sandrini et al., 2006). The somewhat higher CPM effect in other studies using an electrical test-stimulus in comparison to the result of the present study may be due to different testing sites. The reflex in the present study was elicited from the plantar surface of the foot and the response was measured from the anterior tibial muscle. The comparable studies stimulated the sural nerve trunk and recorded from the biceps femoris muscle. It is argued that sural nerve stimulation often is found intolerable resulting in a large number of failed tests, and that the currently employed set-up is less dependent on exact electrode positioning and demonstrates better test-retest reliability than sural nerve stimulation (Bouhassira et al., 2003; Jensen et al., 2015). Another difference, which could contribute to differences found in the CPM effect between the present study and the comparable studies, is that they did not track the reflex threshold over a longer period of time (120 seconds).

In addition to a larger CPM effect, a larger proportion of CPM responders were observed using the protocol with thermal test-stimulus compared to the protocol with electrical test-stimulus. A possible explanation for lower CPM effect and fewer CPM responders when using electrical test-stimulus compared to thermal test-stimulus could be related to differences in the intensity of the test-stimulus between the two protocols in regards to pain intensity, pain quality, and the duration of the stimulus. The NWR threshold has been reported to be correlated with the subjective pain threshold (Sandrini et al., 2005). If this is the case, it is possible that a floor effect for the CPM effect for the electrical test-stimulus is present. The thermal test-stimulus was aimed to reflect a pain intensity of 6/10 on a VAS to prevent floor- or ceiling effects. One could argue that a supra-threshold, e.g., a NWR threshold x 1.5 instead of the NWR threshold may have resulted in a larger CPM effect in the protocol with electrical test-stimulus and also have more methodological similarity to the protocol with thermal test-stimulus. However, earlier studies have suggested that the NWR threshold is sufficient to detect a change in test-stimulus evoked by the conditioning stimulus CPM effect and importantly, is more reliable than supra-threshold stimulation (Biurrun Manresa et al., 2014; Jurth et al., 2014). The NWR is commonly considered a proxy for nociception, due to its longer latency and higher threshold than the tactile reflex which first appears after an electrical stimulation (Willer, 1977). Still, it is still a possibility that the motor response may be contaminated by innocuous somatosensory processes, such as startle reactions and voluntary movements (although we attempted to reduce such influence by familiarization during pre-tests) or modulated by other types of descending control, e.g. emotions (Rhudy et al., 2008) or attention/distraction (Bjerre et al., 2011).

The difference between the outcomes of the protocols may also be a result of different sites of stimulation, which can give rise to activity in different pain modulatory pathways. Two upper limbs are used in the protocol with the thermal test-stimulus, which may possibly reflect a segmental spinal inhibitory effect (although not necessarily limited to that). A combination of a lower limb and an upper limb is used in the protocol with electric test-stimulus, which may be more influenced by an ascending-descending modulatory activity. However, a recent study (Graven-Nielsen et al. (2017) did not find any differences in CPM effect between upper and lower limb stimulation sites when using the same test-stimulus at different locations.

The large difference in CPM effect between the two protocols in our study raises questions as to the mechanisms of CPM. Larger CPM effect when the pain percept component is evaluated compared to when reflex processes are measured, suggests that CPM depends more on cognitive/evaluative aspects of the pain percept than on nociception. This theory is supported by our post-hoc analysis where a significant CPM effect was observed when using pain ratings (32.5%) or unpleasantness ratings (26.1%) of the electrical test-stimulus. This result is in contrast to the traditional theory of a more limited neural system interaction, i.e., diffuse noxious inhibitory controls based on animal research, which is considered to rest on a spinal-supraspinal-spinal feedback loop. However, CPM in humans has shown to be highly influenced by supraspinal processes (Nahman-Averbuch, Nir, et al., 2016). Whether the modulation of pain perception found in the present study is influenced by previous pain experiences, expectations, mood, attention or other modulatory influences from the central nervous system have not been embraced in the present study protocol and needs to be addressed in future research.

A 7° cold water bath was chosen to induce pain ratings close to tolerance to ensure maximal CPM effect for all subjects, since conditioning stimulus with temperatures inducing higher pain intensity have shown to increase the CPM effect compared to temperatures inducing lower pain intensity or non-painful temperatures (Granot et al., 2008; Tousignant-Laflamme et al., 2008; Willer et al., 1989). However, it is desirable that the temperature and duration is tolerable enough to complete the conditioning stimulus according to protocol.

The two protocols have many methodological differences that may affect the CPM effect and make comparison of the outcome of the two protocols difficult. First, the two protocols differ with respect to stimulation parameters such as type of stimulus, duration, stimulus intensity as well as pain intensity. Secondly, when increasing electrical stimulation intensity from 0, there is a range where stimulation is perceived as non-noxious. This means that the scales properties are not directly comparable. When the CPM effect is reported as a percent change for both methods, this may enhance the difference when comparing the CPM effect of the two protocols. However, both thermal and electrical protocols are commonly used to assess CPM and although it is difficult to find measures that are 100% comparable, the comparison of different protocols is important to find a golden standard protocol for CPM assessment.

4.2 Test-retest reliability

Fair relative reliability was found in both protocols. In other studies using thermal test-stimuli, ICC values between 0.21 – 0.62 have been found (Gehling et al., 2016; Granovsky et al., 2015; Imai et al., 2016; Valencia et al., 2013; Wilson et al., 2013). A recent systematic review concludes that differences in reliability heavily depend on stimulation parameters. However, in the present study, the protocol with thermal test-stimulus was identical to a protocol used in a previous study conducted at our laboratory (Lie et al., 2017)which reported good relative reliability (ICC value 0.60). The difference in ICC values between our present (0.40) and our previous study highlights the variation in results despite identical protocols. It also emphasizes the limitations of ICC values as a measure of test-retest reliability. ICC strongly depends on the sample's heterogeneity; ICC values are lower in a homogenous group than in a heterogenous group although the difference in the outcome between sessions are the same in both groups (Atkinson et al., 1998). High ICC values will also occur when subjects maintain their position in the sample across repeated measurements, even though the measurement (i.e., CPM effect) may have changed from session to session. Using ICC alone may lead to false conclusions regarding repeatability and it is therefore recommended to also include measures of absolute reliability in test-retest reliability studies (Atkinson et al., 1998; de Vet et al., 2011; Kennedy et al., 2016). The relative reliability observed in studies using electrical test-stimulus is also conflicting; values between 0.26 (Biurrun Manresa et al., 2014) and 0.61 (Jurth et al., 2014) have been reported. A possible explanation for the poor reliability in our study may be different placement of the electrodes from session to session, even though we tried to prevent this by standardized localization of the stimulation sites. In addition, the two sessions were not conducted at the same time during the day. Time of the day may to a minor degree influence the CPM effect (Aviram et al., 2015).

In both protocols, the bias between sessions was close to zero, suggesting absence of learning effects etc. However, large intra-individual variability was observed in both protocols, which indicate that neither of the protocols evokes a reliable CPM effect in healthy adults on an individual level. When it comes to comparing which of the two methods that is most reliable, the different outcome measured challenge the interpretation of the analysis. The level of absolute reliability depends solely on what is acceptable for practical use (Lexell et al., 2005). Considering the average CPM effect of 0.4 mA using the electrical test-stimulus, the wide range of limits of agreement (-3.4 – 3.8 mA) seems to constitute a genuine reliability problem. Levels of minimal detectable change when using NRS or VAS at 1-2 NRS points or 1-2 cm VAS is often considered acceptable. In the present, study 7 subjects (28%) showed a CPM difference between sessions of more than 2 cm on the VAS when using the thermal test-stimulus. Considering such a high proportion of subjects with high variability between tests, the implementation of CPM tests employed in the present study is of limited value in clinical practice for stratification or prognostic purposes. However, whether CPM is a fluctuating parameter in healthy controls and a more stable parameter in patients suffering from pain conditions, should be

addressed by future research before dismissing the applicability of the testing paradigm in clinical decision making.

Conclusion

The present study demonstrated a variable but fairly pronounced inhibitory CPM effect when the outcome measure is a psychophysical assessment of a thermal test-stimulus. Employing a spinal reflex outcome set up by a point-of-care device with electrical test-stimulus failed to demonstrate a CPM effect. Put together these results raise questions about the mechanisms involved in CPM testing. Fair relative reliability was observed for the CPM effect in both protocols, and poor absolute reliability was found in both protocols due to high intra-individual variability. One should be cautious to extrapolate the results from healthy adults to patients, and the large variability observed in our study calls for extended research in the clinical population before finally concluding on the applicability of CPM methodology in clinical decision making on an individual level.

Acknowledgements

The authors are most grateful for invaluable assistance from Ingrid Fjeldheim Bånerud and Monica Wigemyr for coordinating the experiments and Linda Margareth Pedersen for feedback on the project.

Author contributions

All authors contributed to the conception and design or analysis and interpretation of data as well as making intellectual contributions to the manuscript's content. All authors discussed the results and commented on the manuscript.

References

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med, 26*(4), 217-238.

Aviram, J., Shochat, T., & Pud, D. (2015). Pain perception in healthy young men is modified by time-of-day and is modality dependent. *Pain Med, 16*(6), 1137-1144. doi:10.1111/pme.12665

Biurrun Manresa, J. A., Fritsche, R., Vuilleumier, P. H., Oehler, C., Morch, C. D., Arendt-Nielsen, L., . . . Curatolo, M. (2014). Is the conditioned pain modulation paradigm reliable? A test-retest assessment using the nociceptive withdrawal reflex. *PLoS One, 9*(6), e100241. doi:10.1371/journal.pone.0100241

Bjerre, L., Andersen, A. T., Hagelskjaer, M. T., Ge, N., Morch, C. D., & Andersen, O. K. (2011). Dynamic tuning of human withdrawal reflex receptive fields during cognitive attention and distraction tasks. *Eur J Pain, 15*(8), 816-821. doi:10.1016/j.ejpain.2011.01.015

Bouhassira, D., Danziger, N., Attal, N., & Guirimand, F. (2003). Comparison of the pain suppressive effects of clinical and experimental painful conditioning stimuli. *Brain, 126*(Pt 5), 1068-1078.

de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine*. USA: Cambridge University Press.

France, C. R., Rhudy, J. L., & McGlone, S. (2009). Using normalized EMG to define the nociceptive flexion reflex (NFR) threshold: further evaluation of standardized NFR scoring criteria. *Pain, 145*(1-2), 211-218. doi:10.1016/j.pain.2009.06.022

Gehling, J., Mainka, T., Vollert, J., Pogatzki-Zahn, E. M., Maier, C., & Enax-Krumova, E. K. (2016). Short-term test-retest-reliability of conditioned pain modulation using the cold-heat-pain method in healthy subjects and its correlation to parameters of standardized quantitative sensory testing. *BMC Neurol, 16*, 125. doi:10.1186/s12883-016-0650-z

Granot, M., Weissman-Fogel, I., Crispel, Y., Pud, D., Granovsky, Y., Sprecher, E., & Yarnitsky, D. (2008). Determinants of endogenous analgesia magnitude in a diffuse noxious inhibitory control (DNIC) paradigm: do conditioning stimulus painfulness, gender and personality variables matter? *Pain, 136*(1-2), 142-149. doi:10.1016/j.pain.2007.06.029

Granovsky, Y., Miller-Barmak, A., Goldstein, O., Sprecher, E., & Yarnitsky, D. (2015). CPM Test-Retest Reliability: "Standard" vs "Single Test-Stimulus" Protocols. *Pain Med*. doi:10.1111/pme.12868

Graven-Nielsen, T., Izumi, M., Petersen, K. K., & Arendt-Nielsen, L. (2017). User-independent assessment of conditioning pain modulation by cuff pressure algometry. *Eur J Pain, 21*(3), 552-561. doi:10.1002/ejp.958

Imai, Y., Petersen, K. K., Morch, C. D., & Arendt Nielsen, L. (2016). Comparing test-retest reliability and magnitude of conditioned pain modulation using different combinations of test and conditioning stimuli. *Somatosens Mot Res*, 1-9. doi:10.1080/08990220.2016.1229178

Jensen, M. B., Biurrun Manresa, J., & Andersen, O. K. (2015). Reliable estimation of nociceptive withdrawal reflex thresholds. *J Neurosci Methods, 253*, 110-115. doi:10.1016/j.jneumeth.2015.06.014

Jurth, C., Rehberg, B., & von Dincklage, F. (2014). Reliability of subjective pain ratings and nociceptive flexion reflex responses as measures of conditioned pain modulation. *Pain Res Manag, 19*(2), 93-96.

Kennedy, D. L., Kemp, H. I., Ridout, D., Yarnitsky, D., & Rice, A. S. (2016). Reliability of conditioned pain modulation: a systematic review. *Pain, 157*(11), 2410-2419. doi:10.1097/j.pain.0000000000000689

Lewis, G. N., Rice, D. A., & McNair, P. J. (2012). Conditioned pain modulation in populations with chronic pain: a systematic review and meta-analysis. *J Pain, 13*(10), 936-944. doi:10.1016/j.jpain.2012.07.005

Lexell, J. E., & Downham, D. Y. (2005). How to assess the reliability of measurements in rehabilitation. *American Journal of Physical Medicine & Rehabilitation, 84*(9), 719-723. doi:10.1097/01.phm.0000176452.17771.20

Lie, M. U., Matre, D., Hansson, P., Stubhaug, A., Zwart, J. A., & Nilsen, K. B. (2017). A tonic heat test stimulus yields a larger and more reliable conditioned pain modulation effect compared to a phasic heat test stimulus. *Pain Reports, 2*(6), e626. doi:10.1097/PR9.0000000000000626

Matre, D., Andersen, M. R., Knardahl, S., & Nilsen, K. B. (2016). Conditioned pain modulation is not decreased after partial sleep restriction. *Eur J Pain, 20*(3), 408-416. doi:10.1002/ejp.741

Nahman-Averbuch, H., Dayan, L., Sprecher, E., Hochberg, U., Brill, S., Yarnitsky, D., & Jacob, G. (2016). Pain Modulation and Autonomic Function: The Effect of Clonidine. *Pain Med, 17*, 1292–1301. doi:10.1093/pm/pnv102

Nahman-Averbuch, H., Nir, R. R., Sprecher, E., & Yarnitsky, D. (2016). Psychological Factors and Conditioned Pain Modulation: A Meta-Analysis. *Clin J Pain, 32*(6), 541-554. doi:10.1097/AJP.0000000000000296

Nilsen, K. B., Olsen, I. C., Solem, A. N., & Matre, D. (2014). A large conditioned pain modulation response is not related to a large blood pressure response: a study in healthy men. *Eur J Pain, 18*(9), 1271-1279. doi:10.1002/j.1532-2149.2014.486.x

Potvin, S., Stip, E., Tempier, A., Pampoulova, T., Bentaleb, L. A., Lalonde, P., . . . Marchand, S. (2008). Pain perception in schizophrenia: no changes in diffuse noxious inhibitory controls (DNIC) but a lack of pain sensitization. *J Psychiatr Res, 42*(12), 1010-1016. doi:10.1016/j.jpsychires.2007.11.001

Pud, D., Granovsky, Y., & Yarnitsky, D. (2009). The methodology of experimentally induced diffuse noxious inhibitory control (DNIC)-like effect in humans. *Pain, 144*(1-2), 16-19. doi:10.1016/j.pain.2009.02.015

Rhudy, J. L., Williams, A. E., McCabe, K. M., Russell, J. L., & Maynard, L. J. (2008). Emotional control of nociceptive reactions (ECON): do affective valence and arousal play a role? *Pain, 136*(3), 250-261. doi:10.1016/j.pain.2007.06.031

Sandrini, G., Rossi, P., Milanov, I., Serrao, M., Cecchini, A. P., & Nappi, G. (2006). Abnormal modulatory influence of diffuse noxious inhibitory controls in migraine and chronic tension-type headache patients. *Cephalalgia, 26*(7), 782-789. doi:10.1111/j.1468-2982.2006.01130.x

Sandrini, G., Serrao, M., Rossi, P., Romaniello, A., Cruccu, G., & Willer, J. C. (2005). The lower limb flexion reflex in humans. *Prog Neurobiol, 77*(6), 353-395. doi:10.1016/j.pneurobio.2005.11.003

Shrout, P. E., & Fleiss, J. L. (1979). Intracalss Correlations: Uses in Assesing Rater Reliability. *Psychological Bulletin, 86*(2), 420-428.

Tousignant-Laflamme, Y., Page, S., Goffaux, P., & Marchand, S. (2008). An experimental model to measure excitatory and inhibitory pain mechanisms in humans. *Brain Res, 1230*, 73-79. doi:10.1016/j.brainres.2008.06.120

Valencia, C., Kindler, L. L., Fillingim, R. B., & George, S. Z. (2013). Stability of conditioned pain modulation in two musculoskeletal pain models: investigating the influence of shoulder pain intensity and gender. *BMC Musculoskelet Disord, 14*, 182. doi:10.1186/1471-2474-14-182

Wilder-Smith, O. H., Schreyer, T., Scheffer, G. J., & Arendt-Nielsen, L. (2010). 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, 24*(2), 119-128. doi:10.3109/15360281003706069

Willer, J. C. (1977). Comparative study of perceived pain and nociceptive flexion reflex in man. *Pain, 3*(1), 69-80.

Willer, J. C., De Broucker, T., & Le Bars, D. (1989). Encoding of nociceptive thermal stimuli by diffuse noxious inhibitory controls in humans. *J Neurophysiol, 62*(5), 1028-1038.

Wilson, H., Carvalho, B., Granot, M., & Landau, R. (2013). Temporal stability of conditioned pain modulation in healthy women over four menstrual cycles at the follicular and luteal phases. *Pain, 154*(12), 2633-2638. doi:10.1016/j.pain.2013.06.038

Yarnitsky, D., Arendt-Nielsen, L., Bouhassira, D., Edwards, R. R., Fillingim, R. B., Granot, M., . . . Wilder-Smith, O. (2010). Recommendations on terminology and practice of psychophysical DNIC testing. *Eur J Pain, 14*(4), 339. doi:10.1016/j.ejpain.2010.02.004

Yarnitsky, D., Bouhassira, D., Drewes, A. M., Fillingim, R. B., Granot, M., Hansson, P., . . . Wilder-Smith, O. H. (2015). Recommendations on practice of conditioned pain modulation (CPM) testing. *Eur J Pain, 19*(6), 805-806. doi:10.1002/ejp.605

Yarnitsky, D., Crispel, Y., Eisenberg, E., Granovsky, Y., Ben-Nun, A., Sprecher, E., . . . Granot, M. (2008). Prediction of chronic post-operative pain: pre-operative DNIC testing identifies patients at risk. *Pain, 138*(1), 22-28. doi:10.1016/j.pain.2007.10.033

Yarnitsky, D., Granot, M., Nahman-Averbuch, H., Khamaisi, M., & Granovsky, Y. (2012). Conditioned pain modulation predicts duloxetine efficacy in painful diabetic neuropathy. *Pain, 153*(6), 1193-1198. doi:10.1016/j.pain.2012.02.021