

Review Article

Vladimir Aron*, David Strul, Henrik Bjarke Vaegter, Laurent Pitance, Susan Armijo-Olivo

Reliability and measurement error of exercise-induced hypoalgesia in pain-free adults and adults with musculoskeletal pain: A systematic review

<https://doi.org/10.1515/sjpain-2023-0104>

received September 25, 2023; accepted March 18, 2024

Abstract

Objectives – We systematically reviewed the reliability and measurement error of exercise-induced hypoalgesia (EIH) in pain-free adults and in adults with musculoskeletal (MSK) pain.

Methods – We searched EMBASE, PUBMED, SCOPUS, CINAHL, and PSYCINFO from inception to November 2021 (updated in February 2024). In addition, manual searches of the grey literature were conducted in March 2022, September 2023, and February 2024. The inclusion criteria were as follows: adults – pain-free and with MSK pain – a single bout of exercise (any type) combined with experimental pre-post pain tests, and assessment of the reliability and/or measurement error of EIH. Two independent reviewers selected the studies, assessed their Risk of Bias (RoB) with

the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) RoB tool, and graded the individual results (COSMIN modified Grading of Recommendations Assessment, Development, and Evaluation).

Results – We included five studies involving pain-free individuals ($n = 168$), which were deemed to have an overall “doubtful” RoB. No study including adults with MSK pain was found. The following ranges of parameters of reliability and measurement error of EIH were reported: intraclass correlation coefficients: 0–0.61; kappa: 0.01–0.46; standard error of measurement: 30.1–105 kPa and 10.4–21%; smallest detectable changes: 83.54–291.1 kPa and 28.83–58.21%.

Conclusions – We concluded, with a very low level of certainty, that the reliability and measurement error of EIH is, in pain-free adults, respectively, “insufficient” and “indeterminate.” Future studies should focus on people with MSK pain and could consider using tailored exercises, other test modalities than pressure pain threshold, rater/assessor blinding, and strict control of the sources of variations (e.g., participants’ expectations).

Keywords: exercise-induced hypoalgesia, exercise, humans, psychometrics, pain modulation

* **Corresponding author: Vladimir Aron**, Faculty of Medicine, Institute of Neuroscience (IoNS), Université Catholique de Louvain, Brussels B-1200, Belgium, e-mail: Vladimir.aron@uclouvain.be, tel: +32 494053325

David Strul: Hôpitaux Iris-Sud, Bruxelles, Belgium, e-mail: davidstrul@hotmail.com

Henrik Bjarke Vaegter: Pain Research Group, Pain Center, Department of Anesthesiology and Intensive Care Medicine, University Hospital Odense, Odense, Denmark; Department of Clinical Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark, e-mail: henrik.bjarke.vaegter@rsyd.dk

Laurent Pitance: Institute of Experimental and Clinical Research (IREC), Université Catholique de Louvain, Brussels, Belgium; Oral and Maxillofacial Department, Cliniques Universitaires Saint-Luc, Brussels, Belgium, e-mail: laurent.pitance@uclouvain.be

Susan Armijo-Olivo: Faculty of Economics and Social Sciences, Osnabrück University of Applied Sciences, Osnabrück, Germany; Department of Physical Therapy, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Alberta, Canada, e-mail: susanarmijo@gmail.com

1 Introduction

Exercise-induced hypoalgesia (EIH) refers to an acute reduction in experimental pain (i.e., pain elicited by experimental means) after a single bout of exercise [1], lasting up to 30–45 min [2,3]. Various exercise modes have been used: aerobic [4–6], isotonic [7,8], and isometric [9–11]. Most researchers have used pain thresholds to quantify EIH [12], but some have considered other test modalities such as temporal summation [13] or pain tolerance [5,14]. Studies using pressure and electrical stimuli report EIH more often than those using thermal stimuli [5,15–19]. The location of the stimulus appears to modulate EIH; when stimuli

are applied locally – to a limb involved in the exercise – hypoalgesia seems stronger than when they are applied remotely (i.e., to a limb not, or less, involved in the exercise) [1,4,20]. EIH seems more consistent in healthy people than in those with musculoskeletal (MSK) pain [1,21]. For example, Smith *et al.* [22] showed that, after 30 min of walking, the pressure pain threshold (PPT) at hand, cervical spine, and tibialis anterior muscle increased more in healthy individuals than in patients with whiplash-associated disorders (WAD).

We believe studying EIH could serve different purposes. First, it could provide a model to understand the mechanisms of the acute effects of exercise on human pain processing, which are not clear yet [1]. Second, it could improve the adherence of pain sufferers to exercise interventions. The relationship between exercise-induced pain and EIH seems to differ between pain-free individuals and those with pain conditions [1]. The former consistently report EIH after somewhat painful exercise [20,23], with evidence even suggesting that painful exercise potentiates EIH [24]. In contrast, exercise-induced pain has been associated with impaired EIH in the latter [25,26]. Therefore, in the context of clinical pain, tailoring exercise prescriptions based on individual levels of exercise-induced hypo or hyperalgesia could make sessions more tolerable, increasing adherence to exercise programs [27] and ultimately leading to enhanced clinical outcomes [28]. Third, the lack, or absence, of EIH could be a prognostic factor for clinical pain outcomes [29]. Indeed, the discrepancy in EIH between persons with and without pain led some researchers to hypothesize that an alteration in EIH could predict the prognostic of pain sufferers. Vaegter *et al.* [30] showed that among patients with knee osteoarthritis, those with stronger EIH before a total knee arthroplasty surgery reported less pain 6 months after. However, Woznowski-Vu *et al.* [31] recently showed that EIH had no prognostic value on either pain or disability at 3 months follow-up in adults with back pain. This use of EIH thus warrants further investigation.

A mandatory step to achieve these objectives is to ensure that the hypoalgesic response to exercise is reliable. However, preliminary evidence in pain-free individuals suggests otherwise: studies on EIH using cycling [4,20], walking [5], and isometric [23] exercises combined with PPT and tolerance assessments have reported low-reliability metrics. To our knowledge, there is no published or ongoing systematic review (SR) on this topic, while it would allow us to gather all available evidence, assess its quality, and determine the most reliable ways to assess EIH. Thus, we aim to assess the reliability and measurement error of EIH in pain-free adults and adults with MSK pain, critically

appraise the relevant literature, and formulate recommendations for the assessment of EIH.

2 Methods

Prior to this review, we searched three databases (EPISTEMONIKOS, PROSPERO, and Cochrane Library; from September 12, 2021 to September 19, 2021), which index SR in the health literature. No SR was identified in these databases or by a general web or manual search, and thus, we concluded that the reliability and measurement error of EIH had so far not been systematically reviewed and that no ongoing SR intended to do so. The results of these preliminary searches can be found in Supplementary material: 1. Preliminary searches. We followed the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) guideline for SRs [32] and the reporting guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement; the completed PRISMA 2020 items checklist can be found in Supplementary material: 2. Completed PRISMA 2020 item checklist. The protocol for this SR was registered on PROSPERO on May 10, 2022 (CRD42022325298), before starting the data collection process.

2.1 Eligibility criteria

We used the framework provided by the COSMIN initiative [33] to determine the following eligibility criteria:

- Population: Pain-free adults or adults with MSK pain (acute, chronic, and any MSK location) of any age, gender, and ethnicity. We excluded studies involving individuals younger than 18 years of age, animal models, or studies including adult individuals with medical conditions different from MSK pain or with “secondary” MSK pain (e.g., associated with Parkinson’s disease, due to persistent inflammation or systemic diseases, among others).
- Construct of interest: EIH. It refers to an acute reduction in experimental pain (i.e., pain elicited by experimental means) after a single bout of exercise [1].
- Way of operationalization of the measurement protocol: Studies that used a single bout of exercise (providing mode, intensity, and duration) combined with an experimental pain induction procedure (pain threshold, pain tolerance, or pain ratings of experimentally induced painful stimuli) measured within 2 h upon exercise cessation – we chose this time-frame, also used by Bonello

et al. [34], to target only studies on the effects of acute exercise while excluding those that considered longer lasting effects of exercise. We excluded studies that did not adequately characterize their exercise intervention in terms of mode, intensity, and duration (unless the authors provided the missing information) and studies that assessed pain ratings of clinical pain (i.e., patients with MSK pain reporting a change in their usual pain level after exercise), used repeated exercise sessions in between measurements, did not include a pain-related measure within 2 h upon exercise cessation or combined exercise with another intervention (e.g., manual therapy, neuromuscular electrical stimulation, medication, or supplementation).

- Language: No limits.
- Study type: Due to the nature of our research question, we targeted both observational studies and experimental studies – parallel or cross-over design – from any year. Both phases of the cross-over studies were considered. We excluded narrative reviews, SRs, meta-analyses, commentaries, book chapters, and conference reports. However, the references for these studies were searched.
- Measurement properties of interest: Reliability and measurement error of EIH. Details of these metrics are provided below.

2.2 Information sources and search strategy

The search strategy was designed and conducted by the principal investigator (VA). Team members provided feedback in terms of key words and search strings (peer-reviewed by SA and DS). Five databases were searched using both controlled vocabulary specific to each database, natural language terms related to the different components of the research question and a validated search filter for finding studies on measurement properties designed by Terwee et al. [35]. The details of the search strings for each database can be found in Supplementary material: 3. Search strategy. The following databases were searched initially on November 11, 2021 and updated on February 15 and 16, 2024:

- PUBMED from 1978 to February 2024,
- EMBASE from 1974 to February 2024,
- SCOPUS from 1966 to February 2024,
- PSYCINFO from 1932 to February 2024,
- CINAHL Complete from 1983 to February 2024.

No limits were applied. In total, 4,762 records were exported to a reference manager software (Endnote 20;

endnote.com) and were imported, after duplicates removal, into the Covidence software (covidence.org). After completion of the study selection process, VA conducted manual searches examining the grey literature (March 27, 2022, September 04, 2023, and February 20, 2024) and the reference lists of the included studies (March 18, 2022).

2.3 Selection process

Two independent reviewers (VA and DS) screened – after a pilot phase on 25 references and using SR software (covidence.org) – first, the titles and abstracts of the studies considered for inclusion (first screening round), and then the full text of the studies (second screening round) using the criteria set up above. After each screening round, they discussed their decisions until they reached a consensus. A third reviewer (SA) settled any disagreement.

2.4 Data collection process

The principal investigator (VA) collected – after a pilot phase on three studies with the study team – data from the included studies using a standardized extraction form (Supplementary material: 4. Data extraction form). The data collected were reviewed and discussed with the study team until a consensus was reached. SA settled any disagreement. When information was missing, VA contacted the authors by email. The details of this correspondence can be found in Supplementary material: 5. Correspondence with the authors.

2.5 Data items

We collected data related to the following categories (the full list of the collected variables and outcomes is in Supplementary material: 4. Data extraction form):

- Report information (e.g., authors, year of publication, journal),
- Study information (e.g., study objectives, study question, study design, target population, sample size, eligibility criteria),
- Information regarding exercise (e.g., mode, intensity, duration),
- Test modalities information (i.e., how the authors assessed experimental pain. e.g., stimulus type, test modality, site of test, modulation by exercise),

- EIH determination methods (e.g., scoring of EIH, statistical analysis used, data processing and storage [i.e., “All actions undertaken on the raw data to store it in a usable (electronic) form for later data manipulation”[33]]),
 - Information regarding reliability or measurement error (e.g., type of reliability, repeated component of the procedures, reliability parameters),
 - Information regarding the Risk of Bias (RoB) assessment,
 - Outcomes
- Related to reliability and measurement error: intraclass correlation coefficient (ICC), Pearson’s *r*, Cohen’s kappa, standard error of measurement (SEM).
 - Related to EIH: report of EIH, baseline values for each test modality and location, absolute and relative change values, and numbers of “EIH non-responders/responders” per session.

We defined reliability and measurement error, according to the COSMIN initiative [32], as “The proportion of the total variance in the measurements which is because of ‘true’ differences among patients” and “the systematic and random error of a patient’s score that is not attributed to true changes in the construct to be measured,” respectively. In this definition, “true change” refers to “the average score that would be obtained if the [measurements] were given an infinite number of times” [36]. We collected all the reliability parameters used by the authors of the included studies, but we only reported those that the COSMIN initiative [33] recommends for reliability or measurement error – in the case where ICC and Pearson’s *r* were calculated by the authors, we only reported the ICC. When studies used control and experimental conditions, we collected data related to EIH (e.g., baseline PPT) only from the experimental conditions. A meta-analysis regarding the effect size of EIH was not the aim of this review, and these data were deemed unnecessary.

2.6 Study RoB assessment

We assessed the RoB of the included studies with the COSMIN RoB tool [33], which is recognized as the best current tool to assess the RoB of studies investigating the reliability and measurement error of an outcome measurement instrument. The COSMIN RoB Tool uses a four-point system – “very good,” “adequate,” “doubtful,” or “inadequate” – to rate six common standards for reliability and measurement error – five on design requirements and one on “other flaws” – and additional standards for preferred

statistical methods: three for reliability and two for measurement error (Supplementary material: 6. COSMIN RoB tool). We summarized the RoB for each standard and determined the overall assessment for each study, as recommended by the COSMIN initiative, with the worst-score-counts method: the lowest rating of all standards determined the study RoB. When a study involved multiple test modalities (e.g., PPT and cuff pain tolerance), the RoB was assessed for each of them – we considered that RoB could depend on how each test modality was applied and that within-studies variations were possible. Two independent reviewers (VA and DS) assessed each study and discussed their results until they reached a consensus. A third reviewer (SA) settled any disagreement. We created plots (using R package “robvis” [37]) to report the RoB assessments.

2.7 Data synthesis

We created evidence tables (using R packages “flextable” [38] and “officedown” [39]) to describe the study characteristics, the way of operationalization of EIH, the elements of reliability and measurement error of each individual study, and the outcomes. When studies reported the SEM, we computed the Smallest Detectable Change with a 95% confidence interval (SDC 95; it is a change score that will be observed in only 5% of the individuals that are “truly unchanged” [40]) as $SEM \cdot \sqrt{2} \cdot 1.96$ [41]. When relevant, we organized the data according to the test modality used and its site of application, which we divided into two categories: local (i.e., the stimulus was applied to a limb involved in the exercise) and remote (i.e., the stimulus was applied to a limb not, or less, involved in the exercise). We also created plots (using R package “ggplot2” [42]) to display graphically the values of ICC, SEM, and SDC 95.

We rated the results of the statistics used in each study to assess reliability and measurement error as “+,” “–,” or “?,” respectively “sufficient,” “insufficient,” and “indeterminate,” according to the criteria for good reliability and measurement error from the COSMIN initiative [33] (Supplementary material: 7. COSMIN criteria for good reliability).

We assessed inconsistency in the results (i.e., ratings of “+” or “–” account for less than 75% of all ratings) and compared the designs (e.g., way of operationalizing EIH), populations, and statistical computations of the included studies. Although one of our objectives was to pool reliability parameters from included studies, this was not possible because EIH protocols and exercise parameters greatly differed in each study.

Two independent reviewers (VA and DS) graded the level of evidence of each individual result as “high,”

“moderate,” “low,” or “very low” with the modified Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach published by the COSMIN initiative [32]. A third reviewer (SA) settled any disagreement. GRADE assessments (details about the criteria used for GRADE are provided Supplementary material: 8. COSMIN modified GRADE approach) were based on RoB in included studies, inconsistency in the ratings of the results of the included studies, imprecision (based on the total sample size of the summarized studies), and indirectness (i.e., the population of the summarized studies differed from the target population of this review). We reported the assessment of each outcome in an evidence table.

3 Results

3.1 Study selection

We identified 4,762 records (i.e., the title and/or abstract of a report, which supplies information about a study [43]) through database searching. After duplicate removal, 2,443 records remained, and 2,383 were excluded after screening titles and abstracts. We reviewed 58 full reports; two

potentially eligible reports [44,45] were excluded as only their conference report could be found. The details of the reasons for exclusion after reading the full texts can be found in Supplementary material: 9. Reasons for exclusion of reports. We finally included five studies [4,5,20,23,46]. Then, we searched the references of the included studies and the grey literature, but no new eligible article was found. Figure 1 displays the PRISMA flow diagram of the study selection process.

3.2 Study characteristics

The characteristics of the included studies are described in Tables 1–4.

3.2.1 Study characteristics

Table 1 summarizes the information regarding study design, study setting, funding, sample characteristics, and the manifestation of EIH. Four studies [5,20,23,46] used a cross-over design (but only one [5] was randomized), and one used a within-group pre-post design [4]. All studies [4,5,20,23,46] recruited gender-mixed samples (females/males) of healthy

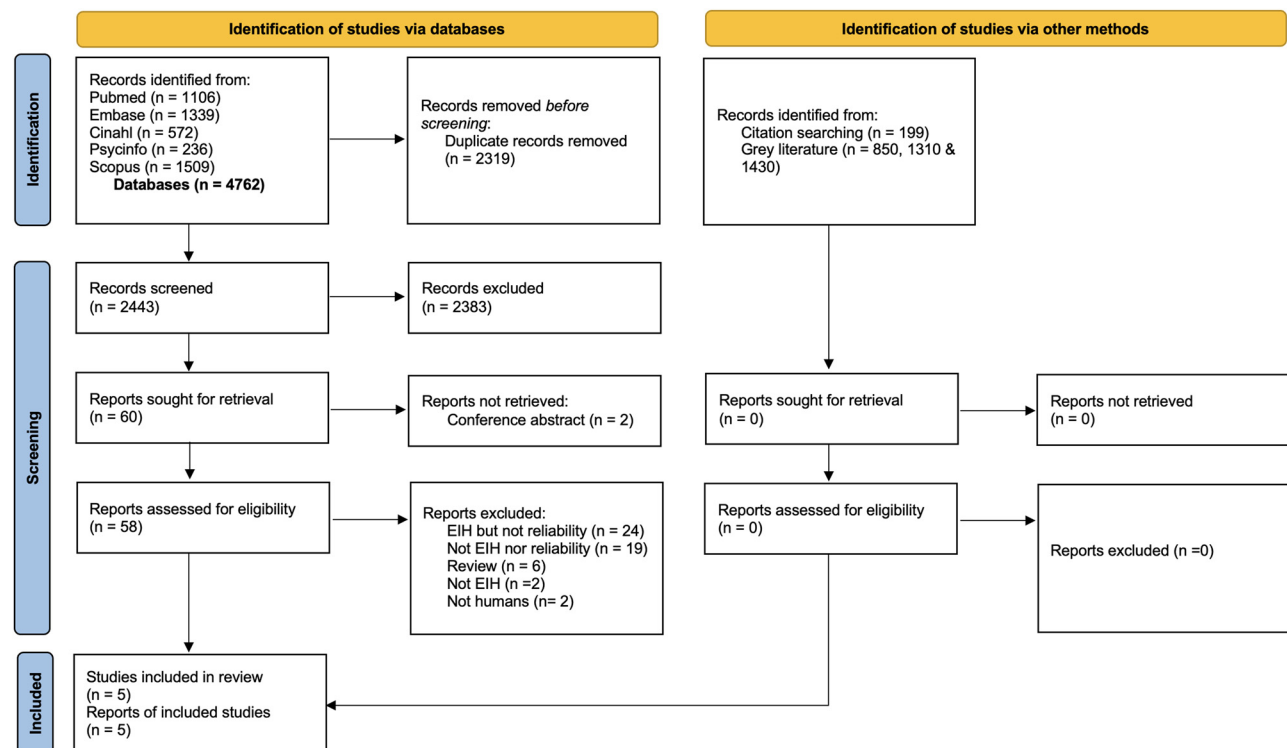


Figure 1: PRISMA flow diagram of included studies (EIH, exercise-induced hypoalgesia).

Table 1: Study characteristics

Reference	Study characteristics			Sample characteristics			Exercise parameters				Control		Test modality		EIH		
	Funding	Setting	Design	n	Age Mean (SD)	Ratio female/male	Condition	Type	Intensity	Duration	Mode	Type	Duration	Type	Site	Session 1	Session 2
Gomolka et al. (2019) [4]	No Funding	Psychology faculty of Ruhr-University of Bochum, Germany	Within-group pre-post	30	24.4 (1.8)	15/15	Healthy	Cycling	85.9% HRmax	15 min	Aerobic	No control		PPT	Biceps femoris	x	x
														PPT	Lower back	–	x
														PPT	Hand	–	–
Hviid et al. (2019) [5]	No Funding	Laboratory at the Pain Center at Odense University Hospital, Odense, Denmark	Randomized cross-over	35	26 (2.9)	18/17	Healthy	Walking	No intensity control	6 min	Aerobic	Rest in a seated position	6 min	PPT	Quadriceps	–	–
														PPT	Upper trapezius	–	–
														cPPT	Calf	–	–
Vaegter et al. (2018) [20,23]	No Funding	Laboratory at the Pain Center at Odense University Hospital, Odense, Denmark	Non-randomized cross-over	34	25.8 (3.4)	13/21	Healthy	Cycling	16RPE	15 min	Aerobic	Rest in a seated position	15 min	PPT	Quadriceps	x	x
														PPT	Upper trapezius	x	x
Vaegter (2019) [46]	No Funding	Laboratory at the Pain Center at Odense University Hospital, Odense, Denmark	Non-randomized cross-over	34	25.3 (3.2)	16/18	Healthy	Cycling	50 (2.5 min); 75 (5 min); 100 (10 min)% LT	15 min	Aerobic	Rest in a seated position	15 min	PPT	Quadriceps	x	x
														PPT	Upper trapezius	–	–
Vaegter (2019) [23]	No Funding	Department of Physiotherapy, University College North Denmark, Aalborg, Denmark	Non-randomized cross-over	35	23.1 (2.2)	17/18	Healthy	Squat	No intensity control	Exhaustion or 3 min	Resistance-Isometric	Rest in a seated position	3 min	PPT	Quadriceps	x	x
														PPT	Upper trapezius	x	x

NA, not applicable; PPT, pressure pain threshold; cPPT, cuff pressure pain threshold; cPTT, cuff pressure pain tolerance; RPE, rate of perceived exertion; LT, lactate threshold; HRmax, maximal heart rate; min, minutes; EIH, exercise-induced hypoalgesia; x, presence of EIH; –, absence of EIH.

Table 2: Way of operationalization of EI_H, data collection, and data processing/storage

Author (Year)	Data collection					Data processing/storage	
	Exercise component			Experimental pain component			
	Equipment (brand)	Procedure	Exercise support	Control	Test modality		
Gomolka et al. (2019) [4]	Stationary ergometer (Corival cpet, the Netherlands), Heart rate monitor (Polar T31, Polar Electro, Finland), Borg's scale 6–20	Subjects warmed up on a stationary ergometer for 2 min at 50 W. Then, the examiner increased the intensity until 85.9% of HRmax was reached. The subjects then pedalled at 70 RPM for 15 min. At the end of the exercise, RPE was assessed with a Borg scale 6–20	NR	No control	PPT	Handheld algometer, probe surface 1 cm ² (Somedic Algometer type II, SBMEDIC Electronics, Sweden)	Subjects said “stop” at the onset of pain, the pressure recording defined the PPT. PPT was assessed at biceps femoris, non-dominant side lower back, and non-dominant hand (prone position; 50 kPa/s; 3; 20 s; randomized and counterbalanced)
Hviid et al. (2019) [5]	Borg's scale 6–20, Cones	The subjects completed two conditions, in a randomized and counterbalanced order 1) 6 min walking supervised by an unidentified trained physiotherapist who provided standardized support 2) control. At the end of the exercise, RPE was assessed with a Borg scale 6–20	Yes	Rest in a seated position, 6 min	PPT	Handheld algometer, probe surface 1 cm ² (Somedic Sales AB, Sweden)	Subjects pressed a button at the onset of pain at non-dominant upper trapezius and dominant quadriceps, the pressure recording defined the PPT (seated position: hands on knees, feet unsupported; 30 kPa/s; 2; 20 s; starting with the quadriceps and then alternated)
					cPPT	VAS, 0–10 cm; 10 cm blood pressure cuff (VBM, Sulz, Germany) connected to an air tank(CPAR, Nocitech, Denmark)	cPPT, cPTT: a cuff around the calf was inflated until 100 kPa while measuring pain with the VAS. cPPT: pressure when pain 1/10. cPTT: pressure when pain 10/10. (seated position; 1 kPa/s; 1; NA; After PPT, cPPT then cPTT)
							PPT values were manually recorded by SG, who transferred them into an SPSS file after the experiment. Data were processed by SG and CT (last author), and securely stored within the Department of Medical Psychology and Medical Sociology of the Ruhr University Bochum PPT, cPPT and cPTT data were initially documented on a physical datasheet (one sheet per session). Cuff data were also stored on cpar/nocitech system PC. After sessions, data were entered into an Excel spreadsheet and stored on secure server at Odense University Hospital

(Continued)

Table 2: Continued

Author (Year)	Data collection					Data processing/storage	
	Exercise component			Experimental pain component			
	Equipment (brand)	Procedure	Exercise support	Control	Test modality		Equipment (brand)
Vaegter et al. (2018) [20,23]	Stationary cycle-ergometer (Ergomedic 928E, Monark Exercise AB, Vansbro, Sweden), Heart rate monitor (Monark Heart Rate Monitor), NRS, Borg's scale 6–20	After a control condition, the subjects cycled for 15 min; starting at 20 W, the resistance was increased of 20 W every min until the subjects reached 16/20 on the Borg's scale. HR was monitored constantly; pain intensity in the legs and RPE were assessed after 2, 3, 6, 9, 12 and 15 min, and maximum intensity (in Watts) was recorded	Yes	Rest in a seated position, 15 min	cPTT PPT	Handheld algometer, probe surface 1 cm ² (Somedic Sales AB, Sweden)	Subjects pressed a button at the onset of pain at non-dominant upper trapezius and dominant quadriceps, the pressure recording defined PPT (NR; 30 kPa/s; 2; 20 s; Excel spreadsheet and stored on secure server at Odense University Hospital)
	Stationary cycle-ergometer (Ergomedic 928E, Monark Exercise AB, Vansbro, Sweden), Heart rate monitor (Monark Heart Rate Monitor), NRS, Borg's scale 6–20	After a control condition, the subjects cycled at 75 RPM for 2.5 min at 50%, 5 min at 75%, and 7.5 min at 100% of the lactate threshold. HR, pain intensity in the legs, and RPE were assessed at 2, 4, 6, 9, 12, and 15 min	Yes	Rest in a seated position, 15 min	PPT	Handheld algometer, probe surface 1 cm ² (Somedic Sales AB, Sweden)	Subjects pressed a button at the onset of pain at non-dominant upper trapezius and dominant quadriceps, the pressure recording defined PPT. (seated position: hands on knees, feet unsupported; 30 kPa/s; 2; 20 s; alternated for each measurement)
Vaegter (2019) [23]	Goniometer, Borg scale 6–20, NRS	After a control condition, the subjects maintained a squat position (back against a wall, feet parallel and shoulder-width apart, hands by their sides, knee joint angle of 100°) for 3 min or until exhaustion. Pain intensity in the legs and RPE were assessed after 1, 2, and 3 min	Yes	Rest in a seated position, 3 min	PPT	Handheld algometer, probe surface 1 cm ² (Somedic Sales AB, Sweden, Horby)	Subjects pressed a button at the onset of pain non-dominant upper trapezius and dominant quadriceps, the pressure recording defined PPT. (seated position: hands on knees, feet unsupported; 30 kPa/s; 3; 20 s; starting with the quadriceps and then alternated)

NA, not applicable; NR, not reported; PPT, pressure pain threshold; cPTT, cuff pressure pain threshold; RPE, rate of perceived exertion; RPM, revolutions per minutes; HR, heart rate; HRmax, maximal heart rate; W, Watts; min, minutes; kPa/s, kilopascal per second; VAS, visual analogic scale; NRS, numeric rating scale.

Table 3: Way of operationalization of EI_H, preparatory actions, and EI_H scoring methods

Author (Year)	Preparatory actions			EI _H scoring method		
	Instructions (adherence control)	Familiarization	Explanation	Procedures	Scoring (class)	Scoring additional (class)
Gomolka et al. (2019) [4]	To refrain from physical exercise, caffeine, and tobacco on the day of testing and the preceding day (NR)	Test modality, blood pressure	Verbal explanation regarding PPT and blood pressure assessment	Exercise component: to adjust seat height (5° knee flexion at the lowest pedal position); to check subjects' comfort. Experimental pain component: NR	Change in PPT: absolute and relative (continuous score)	EI _H responder: absolute change $\geq 1^*SEM$ test modality (dichotomous score)
Hviid et al. (2019) [5]	To refrain from physical exercise on days of participation (Yes, NR)	Test modality	Oral presentation and demonstration of the procedures	Exercise component: NR. Experimental pain component: NR	Change in test modality threshold: absolute and relative (continuous score)	EI _H responder: absolute change $\geq 1^*SEM$ test modality (dichotomous score)
Vaegter et al. (2018) [20,23]	To refrain from physical exercise, coffee, and nicotine on the days of participation (Yes, NR)	Test modality	Verbal introduction to the procedures	Exercise component: to adjust seat height (5° knee flexion at the lowest pedal position). Experimental pain component: NR	Change in PPT: absolute and relative (continuous score)	EI _H responder: absolute change $\geq 1^*SEM$ test modality (dichotomous score)
Vaegter (2019) [46]	To refrain from physical exercises, coffee, and alcohol on the days of participation (Yes, NR)	Cycle ergometer, test modality	Verbal introduction to the procedures	Exercise component: To conduct a sub-maximal exercise test to determine LT; to adjust seat height (5° knee flexion at the lowest pedal position). Experimental pain component: NR	Change in PPT: absolute and relative (continuous score)	EI _H responder: absolute change $\geq 1^*SEM$ test modality (dichotomous score)

(Continued)

Repeated-measures ANOVA 2 × 3 (session 1 vs session 2) × (baseline vs post-exercise vs 15 min post exercise). Bonferroni-corrected *post-hoc* *t*-tests in case of significant main or interaction effects

Repeated-measures ANOVA 2 × 2 (session 1 vs session 2) × (walking vs rest). In case of significant main effects or interactions: *post-hoc* paired *t*-tests with adjustment for multiple comparisons

Mixed-model ANOVA 2 × 2 × 2 (session1 vs session2) × (exercise vs rest) × (quadriceps vs trapezius) × (before vs after) × (male vs female). In case of significant factors or interactions: Bonferroni-corrected *post hoc* *t*-tests

Repeated-measures ANOVA with session (session 2 vs session 3), condition (exercise vs rest), and assessment site (quadriceps vs trapezius) as within-subject factors, and sex (male vs female) as between-subject factor. In case of significant factors or interactions: Bonferroni-corrected *post hoc* *t*-tests

Table 3: Continued

Author (Year)	Preparatory actions			EIH scoring method		
	Instructions (adherence control)	Familiarization	Explanation	Procedures	Scoring (class)	Scoring additional (class)
Vaegter (2019) [23]	To refrain from physical exercises, coffee, and nicotine on the days of participation and between the sessions (Yes, NR)	Test modality	Verbal introduction to the procedures	Exercise component: NR. Experimental pain component: NR	Change in PPT: absolute and relative (continuous score)	EIH responder: absolute change $\geq 1 \times \text{SEM}$ test modality (dichotomous score)
						Mixed-model ANOVA, with session (session 1 vs session 2), condition (exercise vs rest), assessment site (quadriceps vs trapezius), and time (before vs after) as the within-subjects factors and gender as the between subjects factor. In case of significant factors or interactions: Bonferroni-corrected <i>post hoc</i> t-tests

NR, not reported; LT, lactate threshold; PPT, pressure pain threshold; EIH, exercise-induced hypoalgesia; SEM, standard error of measurement; ANOVA, analysis of variance.

subjects (in total, 168 participants were included with a mean age ranging from 23.1 to 26 years) and demonstrated EIH. Three [4,5,46] reported that some combinations of test modality and site did not result in EIH: Gomolka *et al.* [4] assessed PPT at the hand in session 1; Hviid *et al.* [5] used PPT and cuff pressure pain threshold (cPPT) at any location in all sessions; and Vaegter *et al.* [46] assessed PPT at the upper trapezius muscle in all sessions. None of the studies included patients with MSK pain.

3.2.2 Way of operationalization of EIH

Tables 1–3 summarize the information regarding the way of operationalization of EIH. We found methodological similarities across all the included studies [4,5,20,23,46]: EIH was assessed with PPT applied on at least two sites (local and remote); a similar data processing and storage methodology was adopted: the data were first physically recorded then transferred into an electronic spreadsheet and securely stored; subjects were instructed to refrain from physical activity and analgesic medication on the days on participation and were familiarized with the test modalities; EIH was computed from absolute (score test modality_{post-exercise} – score test modality_{baseline}) and relative (absolute change/score test modality_{baseline} $\times 100$) change as a continuous variable; the manifestation of EIH was dichotomized in “responders” and “non-responders” when the absolute change was, respectively, greater and lower than the within-session SEM of the test modality; statistical analysis involved analysis of variance (ANOVA) to test for EIH occurrence (slight variations across studies are detailed in Table 3).

Studies differed regarding their exercise component (see details in Table 2), experimental pain component (see details in Table 2), and preparatory actions (see details in Table 3): three studies [4,20,46] used a cycling exercise, one study [5] used a walking exercise, and one study [23] used an isometric squat exercise; Standardized support during the exercise (i.e., motivation of the participants) was implemented in four studies [5,20,23,46]; one study [4] did not use a control condition, all the others [5,20,23,46] compared exercise with rest; one study [5] used, besides the PPT, the cPPT and the Cuff Pressure Pain Tolerance (cPTT) to induce experimental pain; on the days of participation, subjects were instructed to refrain from caffeine (four studies [4,20,23,46]), nicotine (three studies [4,23,46]), and alcohol (one study [46]). Adjustments of the exercise apparatus (cycle-ergometer settings) were reported in three studies [4,20,46]: one study [46] conducted a sub-maximal test prior to the EIH sessions to determine work-rate at the Lactate Threshold

Table 4: Reliability and measurement error assessment

Author (Year)	Measurement properties	Target population	Rater roles	Repeated component	Variation component	Statistics			
						Reliability		Measurement error	
						Continuous score (Model, Formula)	Dichotomous score	Continuous score (Formula)	Dichotomous score
Gomolka et al. (2019) [4]	Test–retest reliability measurement error	Healthy adults	SG (profession not reported, 2 weeks training), the first author, conducted all the assessments. It is unclear who stored the data, prepared the equipment, and supervised the exercise	The same procedure was repeated on both sessions (except for the familiarization, session 1). The overall focus was to assess the reliability and measurement error of the EIH score	Only the occasions varied ($n = 2$; 1 to 3 weeks interval; same time of day)	ICC (3.1, Agreement)	Unweighted kappa	SEM (consistency)	NA
Hviid et al. (2019) [5]	Test–retest reliability	Healthy adults	The same experienced physiotherapist (3 weeks training) assessed PPT, cPTT, and cPTT on both sessions, stored the data, prepared the equipment, and supervised the exercise and the control conditions	The same procedure was repeated on both sessions. The overall focus was to assess the reliability of the EIH score	Only the occasions varied ($n = 2$; 1 week interval; same time of day)	ICC (3.1, Consistency)	Unweighted kappa	NA	NA
Vaegter et al. (2018) [20,23]	Test–retest reliability measurement error	Healthy adults	The same physiotherapist student (last year of education, 3 weeks training) collected the PPT, stored the data, prepared the equipment, and supervised the exercise and control conditions across sessions	The same procedure was repeated on both sessions (except for the familiarization, session 1). The overall focus was to assess the reliability and measurement error of the EIH score	Only the occasions varied ($n = 2$; 1 week interval; same time of day)	ICC (3.1, Consistency)	Unweighted kappa	SEM (consistency)	NA
Vaegter (2019) [46]	Test–retest reliability	Healthy adults	The same physiotherapist student (last year of education, 3 weeks training) collected the PPT, stored the data, prepared the equipment, and supervised the exercise and control conditions across sessions	The same procedure was repeated on both sessions. The overall focus was to assess the reliability of the EIH score	Only the occasions varied ($n = 2$; 1 week interval; same time of day)	ICC (3.1, Consistency)	Unweighted kappa	NA	NA
Vaegter (2019) [23]	Test–retest reliability Measurement error	Healthy adults	The same physiotherapist student (last year of education, 3 weeks training) collected the PPT, stored the data, prepared the equipment, and supervised the exercise and control conditions across sessions	The same procedure was repeated on both sessions (except for the familiarization, session 1). The overall focus was to assess the reliability and measurement error of the EIH score	Only the occasions varied ($n = 2$; 1 week interval; same time of day)	ICC (3.1, Consistency)	Unweighted kappa	SEM (consistency)	NA

NR, not reported; NA, not applicable; EIH, exercise-induced hypoalgesia; PPT, pressure pain threshold; cPPT, cuff pressure pain threshold; ICC, intraclass correlation coefficient; SEM, standard error of measurement; Model 3.1, two way mixed effects, single rater/measurement, model.

(LT). None of the studies reported preparatory actions related to the equipment used to induce pain.

error was assessed in three studies [4,20,23] using SEM for consistency.

3.2.3 Reliability and measurement error assessment

Table 4 presents information regarding the assessment of reliability and measurement error. All studies [4,5,20,23,46] investigated the test–retest reliability of EIH and targeted a population of healthy human adults. Also, in all of them, the assessor had received training (either 2 weeks [4] or 3 weeks [5,46,20,23]), the same rater conducted the assessment in all sessions, the procedures were repeated on all sessions and only the occasions varied, the sessions took place at the same time of day separated by 1 (all studies [4,5,20,23,46]) to 3 (1 study [4]) weeks, and the reliability of EIH (when computed as a discrete variable) was assessed with unweighted kappa. Two-way mixed effects for single measurement intra-class correlation coefficients (ICC 3.1) for consistency (four studies [5,20,23,46]) and agreement (one study [4]) were used to investigate the reliability of the EIH computed as a continuous variable. Measurement

3.3 RoB in studies

The RoB assessments from studies investigating the reliability and measurement error of EIH are summarized in Figure 2. In all studies [4,5,20,23,46], we considered that subjects were stable between the sessions as only healthy participants were included; thus, we rated R1 and M1 – related to the clinical stability of the population – “very good.” The time interval between sessions ranged from 1 (all studies [4,5,20,23,46]) to 3 weeks (one study [4]); we considered that domains R2 and M2 were “very good.” Four studies [4,20,23,46] stated that measurement sessions were identical – the domains R3 and M3 were judged “very good”; in one study [5], we considered that the authors could have provided more evidence that sessions were identical and we deemed the domains R3 and M3 “adequate.” We considered all the studies [4,5,20,23,46] to be

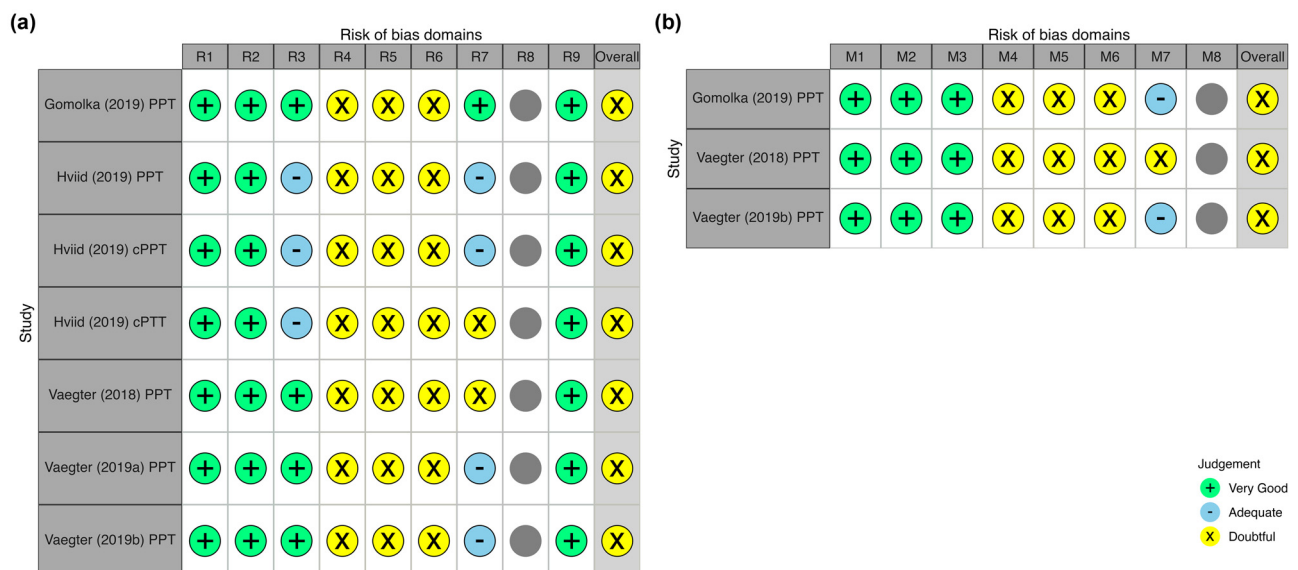


Figure 2: RoB in studies investigating the (a) reliability and (b) measurement error of EIH. Domains: R1/M1, Were patients stable in the time between the repeated measurements on the construct to be measured? R2/M2, Was the time interval between the repeated measurements appropriate? R3/M3, Were the measurement conditions similar for the repeated measurements-except for the condition being evaluated as a source of variation? R4/M4, Did the professional(s) administer the measurement without knowledge of scores or values of other repeated measurement(s) in the same patients? R5/M5, Did the professional(s) assign scores or determine values without knowledge of the scores or values of other repeated measurement(s) in the same patients? R6/M6, Were there any other important flaws in the design or statistical methods of the study? R7, For continuous scores: was an intraclass correlation coefficient (IC) calculated? R8, For ordinal scores: was a (weighted) kappa calculated? R9, For dichotomous/nominal scores: was kappa calculated for each category against the other categories combined? M7, For continuous scores: was the SEM, SDC, limits of agreement or coefficient of variation calculated? M8, For dichotomous/nominal/ordinal scores: Was the percentage specific (e.g. positive and negative) agreement calculated? (cPPT, cuff pressure pain threshold; PPT, pressure pain threshold; cPTT, cuff pressure pain tolerance).

“doubtful” regarding the domains R4, R5, M4, and M5 because the assessor was not blinded [4,5,20,23,46] to the scores of the participants from the previous session when s/he administered the measurement and attributed a score. We considered the domains R6 and M6 as “doubtful” for the following reasons: lack of control of factors potentially affecting exercise performance, exercise metabolism, and/or nociceptive processing, such as menstrual cycle phase [47–49] (all studies [4,5,20,23,46]), pre-exercise nutritional status [50,51] (all studies [4,5,20,23,46]), and consumption of other substances than analgesics (four studies) [5,20,23,46]; lack of control of factors known to affect EIH such as subjects expectations/beliefs [52,53] (all studies [4,5,20,23,46]). We judged the domain R7 as follows: “very good” (one study [4]); the authors used ICC 3.1 agreement to assess the test–retest reliability of EIH as a continuous score; “adequate” (four studies [5,20,23,46]), ICC 3.1 consistency

was used to assess test–retest reliability and the authors provided evidence that no systematic errors occurred (similar EIH display in all sessions, no or minor differences between exercise parameters across sessions); “doubtful” (two studies [5,20]), ICC 3.1 consistency was used to assess test–retest reliability but differences across sessions in the magnitude of EIH were observed – reflecting a lack of agreement between sessions, potentially arising from bias. Domain R9 was considered “very good” in all studies [4,5,20,23,46] – unweighted kappa was used to measure the reliability of EIH as a dichotomous variable. The domain M7 was judged as follows: “adequate” in two studies [4,23], SEM consistency was used to determine measurement error and the authors provided evidence that no systematic errors occurred; “doubtful” in one study [20], SEM consistency was used and the manifestation of EIH varied across sessions. The overall RoB was deemed “doubtful” for all studies [4,5,20,23,46].

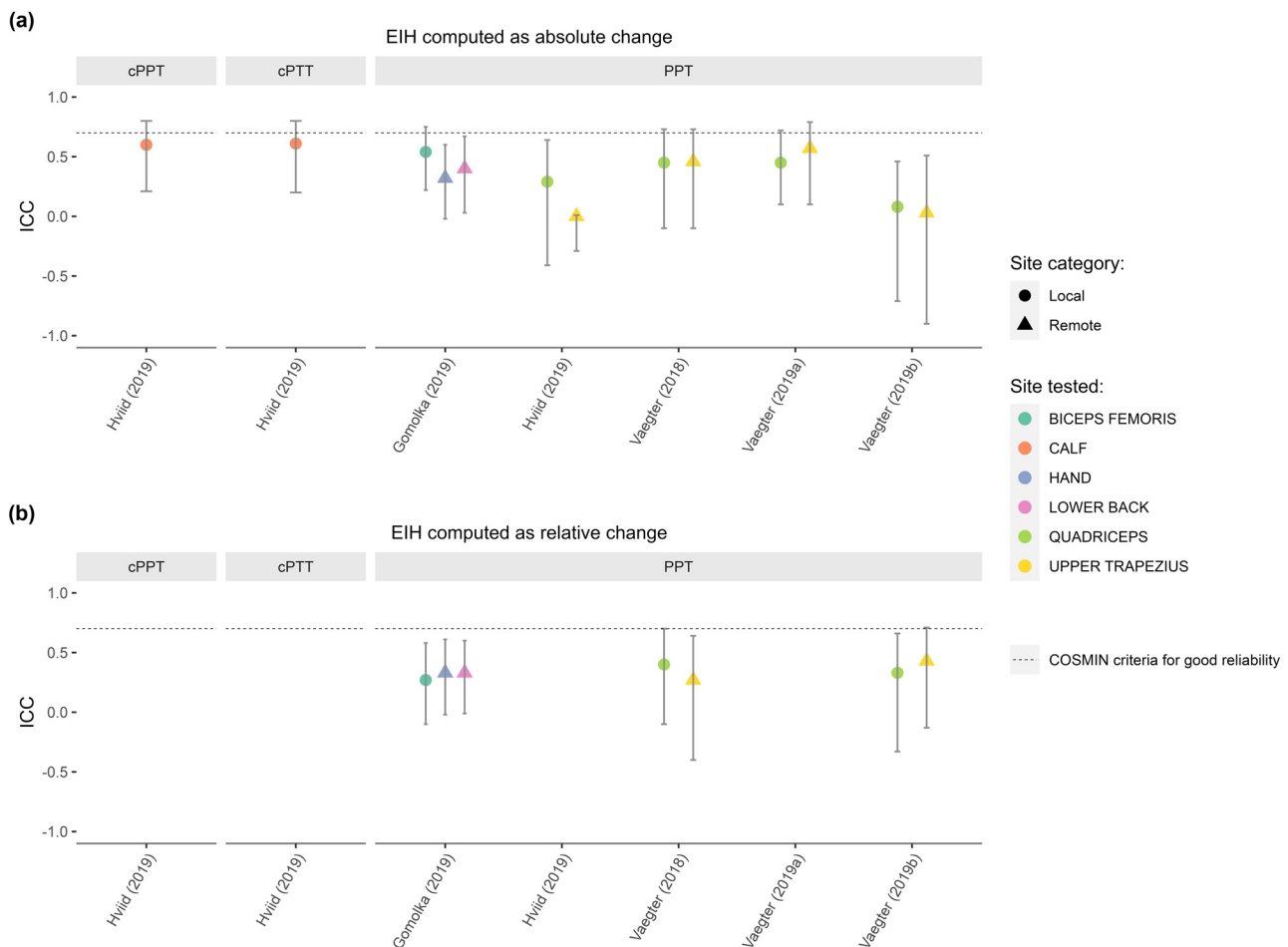


Figure 3: Reliability of EIH. EIH was computed as (a) absolute change and (b) relative change. The ICCs determined in the included studies are reported with their 95% confidence intervals. They vary according to the site of the test and the test modality applied, which was further categorized as local (i.e., the stimulus was applied to a limb involved in the exercise) and remote (i.e., the stimulus was applied to a limb not, or less, involved in the exercise). All values are below the COSMIN criteria for good reliability (ICC > 0.7). (cPPT, cuff pressure pain threshold; PPT, pressure pain threshold; cPTT, cuff pressure pain tolerance).

3.4 Results of individual studies

3.4.1 Reliability and measurement error

Figures 3 and 4, respectively, display the results of reliability and measurement error from the included studies for EIH considered as a continuous score. ICCs for EIH computed as relative and absolute change ranged from 0 to 0.61; kappa for the reliability of EIH considered as a dichotomous score ranged from 0.01 to 0.46; SEM of EIH computed as absolute and relative change, respectively, ranged from 30.1 to 105 kPa and from 10.4 to 21%; SDC 95 of EIH computed as absolute and relative change, respectively, ranged from 83.54 to 291.05 kPa and from 28.83 to 58.21%. Table 5 summarizes this information and the reliability of EIH computed as a dichotomous score. Based on

the COSMIN criteria for good reliability and measurement error, we rated all the ICCs and kappa values as “insufficient” (“–”; ICCs and kappas < 0.7) and the SDC 95 values as “indeterminate” (“?”; minimal important change [MIC] is not defined).

3.4.2 Certainty of evidence

We could not synthesize the individual results because the exercise conditions differed across studies. However, we graded the certainty of evidence for each individual result using a modified GRADE approach (Table 5). The overall RoB assessment of each study, for both reliability and measurement error, was “doubtful,” which corresponds to a “serious” RoB; we downgraded the quality of evidence by

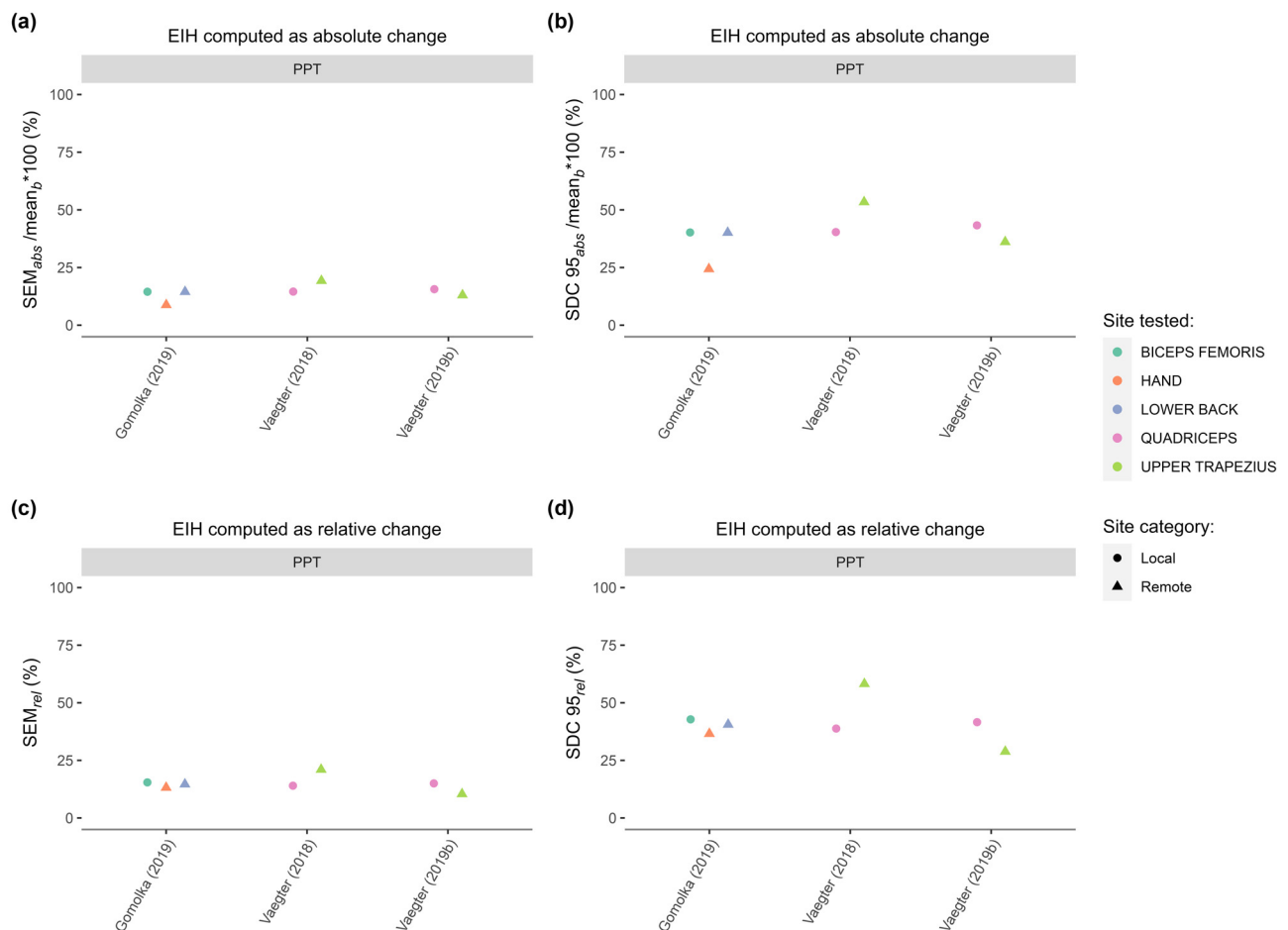


Figure 4: Measurement error of EIH. EIH was computed as (a and b) absolute change and (c and d) relative change. The SEMs of (a) absolute change (SEM_{abs}) and (c) relative change (SEM_{rel}) determined in the included studies are reported. We computed the SDC 95% of (b) absolute change (SDC 95_{abs}) and (d) relative change (SDC 95_{rel}). They vary according to the site of the test and the test modality applied, which was further categorized as local (i.e., the stimulus was applied to a limb involved in the exercise) and remote (i.e., the stimulus was applied to a limb not, or less, involved in the exercise). To compare the SEM_{abs} and SDC 95_{abs} across studies, we divided them by the mean of the mean baseline (mean_b) values across sessions and multiplied by 100. (cPPT, cuff pressure pain threshold; PPT, pressure pain threshold; cPTT, cuff pressure pain tolerance).

Table 5: Results of individual studies, ratings, and GRADE

Author (Year)	Test modality			Reliability				Measurement error									
			Category	EIH continuous		EIH dichotomous		Grade		Absolute change (kPa)		Relative change (%)		Grade			
	Type (units)	Site		ICC absolute change [95% CI]	Rating	ICC relative change [95% CI]	Rating	Kappa [95% CI]	Rating	SEM	SDC 95%	Rating	SEM		SDC 95%		
Gomolka et al. (2019) [4]	PPT (kPa)	Biceps femoris	Local	0.54 [0.22,0.75]	-	0.27 [-0.1,0.58]	-	0.28 [NR,NR]	-	Very low	65.23	180.81	?	15.44	42.80	?	Very low
	PPT (kPa)	Lower back	Remote	0.4 [0.03,0.67]	-	0.33 [-0.01,0.6]	-	0.27 [NR,NR]	-	Very low	64.54	178.90	?	14.63	40.55	?	Very low
	PPT (kPa)	Hand	Remote	0.32 [-0.02,0.6]	-	0.33 [-0.02,0.61]	-	0.32 [NR,NR]	-	Very low	30.14	83.54	?	13.19	36.56	?	Very low
	PPT (kPa)	Quadriceps	Local	0.29 [-0.41,0.64]	-	NR	?	0.07 [-0.24,0.37]	-	Very low	NA	NA	NA	NA	NA	NA	NA
Hviid et al. (2019) [5]	PPT (kPa)	Upper trapezius	Remote	0 [-0.29,0.01]	-	NR	?	0.11 [-0.24,0.47]	-	Very low	NA	NA	NA	NA	NA	NA	NA
	cPPT (kPa)	Calf	Local	0.6 [0.21,0.8]	-	NR	?	0.34 [0.03,0.65]	-	Very low	NA	NA	NA	NA	NA	NA	NA
	cPTT (kPa)	Calf	Local	0.61 [0.2,0.8]	-	NR	?	0.19 [-0.15,0.53]	-	Very low	NA	NA	NA	NA	NA	NA	NA
	PPT (kPa)	Quadriceps	Local	0.45 [-0.1,0.73]	-	0.4 [-0.1,0.7]	-	0.24 [-0.08,0.56]	-	Very low	86.00	238.38	?	14.00	38.81	?	Very low
Vaegter et al. (2018) [20,23]	PPT (kPa)	Upper trapezius	Remote	0.46 [-0.1,0.73]	-	0.27 [-0.4,0.64]	-	0.01 [-0.33,0.34]	-	Very low	65.00	180.17	?	21.00	58.21	?	Very low
	PPT (kPa)	Quadriceps	Local	0.45 [0.1,0.72]	-	NR	?	0.46 [0.17,0.75]	-	Very low	NA	NA	NA	NA	NA	NA	NA
Vaegter (2019) [46]	PPT (kPa)	Upper trapezius	Remote	0.57 [0.1,0.79]	-	NR	?	0.43 [0.11,0.74]	-	Very low	NA	NA	NA	NA	NA	NA	NA
	PPT (kPa)	Quadriceps	Local	0.08 [-0.71,0.46]	-	0.33 [-0.33,0.66]	-	0.08 [-0.25,0.65]	-	Very low	105.00	291.05	?	15.00	41.58	?	Very low
Vaegter (2019) [23]	PPT (kPa)	Upper trapezius	Remote	0.03 [-0.9,0.51]	-	0.43 [-0.13,0.71]	-	0.13 [-0.2,0.46]	-	Very low	53.00	146.91	?	10.40	28.83	?	Very low

NA, not applicable; NR, not reported; EIH, exercise-induced hypoalgesia; ICC, intraclass correlation coefficient; SEM, standard error of measurement; kPa, kilopascal; CI, confidence interval; RoB, Risk of Bias; SDC 95, smallest detectable change with a 95% confidence interval; PPT, pressure pain threshold; cPPT, cuff pressure pain threshold; cPTT, cuff pressure pain tolerance. Rating: rating of the reliability value according to the COSMIN criteria: “-”, insufficient; “?”, indeterminate. Grade: level of the quality of the evidence according to the modified GRADE approach from the COSMIN initiative. The bold letters refer to the rating of reliability and measurement error according to COSMIN.

1 level. Imprecision, which refers to the total sample size, was below 50 in each study; we downgraded the quality of evidence by 2 levels. Indirectness, which refers to differences between the target populations of the studies and the review, was considered absent – all studies [4,5,20,23,46] recruited gender-mixed samples of healthy adults, our population of interest; we did not downgrade the quality of evidence for indirectness. Inconsistency, which relates to the exclusion of heterogeneous results from a pooled result, was considered not applicable – we did not pool any of the results; we did not downgrade the quality of evidence for inconsistency. We downgraded the quality of evidence by three levels in total; thus, the level of certainty of each individual result was considered “very low.” A tabular summary of the modified GRADE results for each study is provided in Supplementary material: 8. COSMIN modified GRADE approach.

4 Discussion

4.1 Main results

To our knowledge, this is the first SR on the reliability and measurement error of EIH. We found five studies [4,5,20,23,46] conducted in pain-free human adults but none in adults with MSK pain. Our results suggest, with a very low level of certainty, that the between sessions test–retest reliability and the measurement error of EIH are, respectively, “insufficient” (i.e., ICCs and kappas < 0.7) and “indeterminate” (unknown MIC) in pain-free human adults.

4.2 General interpretation of the results

4.2.1 Reliability and measurement error

The values of the ICCs of EIH varied across studies, test modalities, and sites and were below the COSMIN criteria for good reliability (ICC > 0.7), suggesting poor reliability of EIH. Overall, they appeared to be the highest (≥ 0.2) when cPPT and cPTT were used to elicit EIH and when the intensity of the exercise was standardized (e.g., using a percentage of maximal HR). The site of the test category (i.e., local vs remote) did not seem to impact the ICC values, suggesting a somewhat smaller involvement in EIH reliability than the test modality and exercise protocol. This is in contrast with the effect of site on the magnitude of EIH, where the hypoalgesic effect of exercise seems to be greater at a limb involved in the exercise [1,4,46].

The COSMIN initiative [33] recommends examining the measurement error parameters in relation to the value of the MIC to determine whether clinically relevant change can be distinguished from measurement error. While the MIC of EIH is currently unknown, we can compare the ranges of the SDC 95 with those of the mean absolute and relative change (used to compute EIH) to estimate the importance of error in the measurements [36]. In the three included studies that assessed measurement error [4,20,23], the ranges of the SDC 95 exceeded the magnitude of EIH. For example, in Gomolka *et al.* [4], the mean of the means of absolute change in PPT assessed at the biceps femoris muscle in the two sessions was 83.04 kPa and the estimated SDC 95 was 180.81 kPa. Thus, to investigate the effect of a variable/intervention on the EIH displayed by an individual with a similar baseline value of absolute change (83 kPa), one would then need to observe more than a twofold change (180 kPa) in EIH to be confident that the observed change is not the mere result of the random fluctuations that 95% of the people could display without “truly” changing [40]. Reliability is, in its formal expression, the ratio of the between subjects’ variability and the total variability (which encompasses the subject’s variability and the measurement error) [36]; therefore, the low estimates of reliability that were observed in the included studies could be explained by a high level of measurement error in EIH.

4.3 Limitations of the evidence included in the review

The evidence included [4,5,20,23,46] in this SR came from a single research group. It is mainly limited by its overall RoB, which was deemed “doubtful” in all studies. Our main concerns were the lack of blindness of the assessor, the lack of control of potential sources of error, and the statistical methods.

4.3.1 Sources of error

Vaegter *et al.* [52] showed that negative pre-exercise information reduces EIH, and none of the included studies [4,5,20,23,46] assessed the participants’ beliefs, expectations, and knowledge regarding exercise, EIH, and pain assessments [53]. There are other factors, which were not controlled in the included studies, that could modulate EIH through an impact on exercise performance, exercise metabolism and/or nociceptive processing: the phase of

menstrual cycle (for aerobic exercise, Hoeger Bement et al. [54] did not find an influence of the phase of the menstrual cycle on EIH after an isometric exercise), the support provided during the exercise, the intake of non-analgesic medications, and the pre-exercise nutritional status. Variations in these elements among participants and across sessions could have contributed to measurement error. A better understanding of the mechanism(s) [1,55,56] that drive EIH will be important to determine the variables to control in future studies.

4.3.2 Statistical methods

ICCs are ratios of the between-subjects variance and the total variance. The total variance includes the between-subjects variance, the residual-variance and the variances due to random and systematic error (i.e., bias) [57]; they are accounted for differently across ICC models and formulas (e.g., the formula for consistency does not include the variance due to bias) [57].

All the included studies [4,5,20,23,46] used ICC 3.1, which implies that the raters were “fixed” (i.e., not random). This choice does not introduce bias in the reliability estimates but limits the generalization of the results to other raters [57]. Two studies [5,20] reported between sessions differences in EIH, potentially caused by systematic errors, and used ICC 3.1 with the formula for consistency thus leading to a risk of overestimation of the ICCs. Yet, it is unlikely that it had a significant impact on their results as their reliability estimates were similar to those of the other studies.

All studies [4,5,20,23,46] assessed the reliability of EIH as a dichotomous score. Another method could have been to use the SDC 95 instead of the SEM – it would have given a more precise level of confidence [40]. However, this approach might have proven unpractical: the SDC 95 of the test modalities that we computed seems to exceed the post-exercise absolute changes reported in the included studies, which would have led to considering all participants as “nonresponders.” For example, the SDC 95 of the PPT at the biceps femoris muscle from Gomolka et al. [4] was 133.05 kPa, while the mean absolute change for the same parameters was 85.5 kPa. Furthermore, neither the SEM nor the SDC 95 is recommended for this use: they are metrics of measurement error, which relates to confidence in the absence of change in people who do not change, rather than confidence in correctly identifying change, which relates to responsiveness [40,41] (i.e., “the ability of an [outcome] instrument to detect change over time in the construct to be measured” [58]). Thus, a preliminary step for a classification approach could be to determine a more adequate criterion for “true

EIH”. A relevant possibility is the MIC, which is a metric of responsiveness and not measurement error and is thus more suitable to identify “responders” than the SDC 95 [40,41]. However, to our knowledge, it has never been investigated for EIH.

4.3.3 Population

The reliability data obtained in pain-free participants cannot be generalized to clinical populations – reliability is a population-dependent construct [36]. The lack of data on the reliability and measurement error of EIH in MSK pain sufferers remains unexplained and stands in stark contrast to the substantive amount of publications comparing EIH in patients with pain-free controls [1,59], with recent studies conducted, among others, in people with chronic low back pain [60], osteoarthritis [61], and WAD [62]. We believe that this knowledge gap stands in the way of effective clinical implementation of EIH: its use to guide exercise prescription or to predict clinical outcomes requires developing ways to induce it reliably in clinical populations and defining cut-off values to classify patients (e.g., the MIC).

4.4 Limitations of the review process

This review has used a comprehensive search strategy and an in-depth data collection process. Because of the heterogeneity in the methods of the included studies (i.e., clinical heterogeneity), we could not quantitatively pool their results. Also, because we only included five studies, we could not provide a statistical or graphical (e.g., funnel plot) assessment of publication bias – we would have lacked statistical power [63]. However, we think that our results are not at high risk of publication bias: our search strategy was not limited to electronic databases; we carefully searched references and grey literature [64], and the reliability estimates of all the included studies were below the criteria for good reliability of the COSMIN initiative [33], whereas publication bias tends to be related to “positive” results [65].

4.5 Implications for future research

We have provided evidence that EIH might be unreliable, which can weaken the conclusions of the numerous studies on the topic [1]. Despite their limitations, our results

demonstrate that the EIH literature must be interpreted with caution, highlighting a knowledge gap regarding EIH reliability (especially in MSK pain sufferers), and can be used to provide recommendations for further research in this field. We suggest that future studies prioritize investigating the reliability and measurement error of EIH in patients with MSK pain and the MIC of EIH in these populations. Furthermore, we recommend that future studies tailor the intensity of the exercise conditions, explore other means of experimental pain induction than PPT, blind the assessors/raters to the scores of the participants, and consider the participants' beliefs and expectations regarding pain and exercise. Also, additional studies are needed to shed light on the sources of variations of EIH and its underlying mechanism(s). For reliability studies, we consider that two-way random-effects ICC models with formula for agreement should be used by researchers – it would allow them to generalize their results to randomly selected raters (and thus to other research teams) and prevent an overestimation of reliability in the presence of bias. Finally, when more studies on the reliability of EIH using a homogenous methodology are published, future SRs on the topic could consider a metaanalysis by subgrouping studies according to the exercise type, exercise intensity, and testing site to improve the precision of the reliability and measurement error estimates – thus increasing the certainty of the results.

5 Conclusions

We conclude, with a very low level of certainty, that EIH is insufficiently reliable and that its measurement error is indeterminate in pain-free human adults. Furthermore, these measurement properties of EIH have never been investigated in patients with MSK pain – a literature gap that should be prioritized by researchers in the field of EIH. To improve their methodological quality, future studies could consider using exercises with standardized intensity (e.g., using a percentage of maximal HR), other experimental pain induction methods than PPT (e.g., cuff pressure pain tolerance), rater blinding, strict control of the sources of variations (e.g., participants' expectations), appropriate computation of ICCs (i.e., agreement), and averages of multiple measurements.

Acknowledgments: This SR slightly deviated from its PROSPERO protocol: we added an author, Prof Henrik Bjarke Vaegter, we used the COSMIN framework to present the eligibility criteria (instead of the PICO format), and the principal investigator collected data in a standardized

extraction form which was then reviewed and discussed with the study team until a consensus was reached (instead of two reviewers collecting data independently).

Research ethics: Not applicable.

Informed consent: Not applicable.

Author contributions: The authors provide their assurance that this work is original, has not been published previously, and is not currently under consideration for publication elsewhere. All authors have read the submission guidelines, and the paper conforms to these guidelines in all respects. All authors have contributed intellectually to the study, meeting the conditions for authorship, and each author has seen and approved the final version of this manuscript.

Competing interests: HBV is an editorial team member of Scandinavian Journal of Pain.

Research funding: VA is supported by a FRS-FNRS FRIA doctoral grant from the French Community of Belgium.

Data availability: The raw data can be obtained on request from the corresponding author.

Supplementary Material: This article contains supplementary material (followed by the link to the article online).

References

- [1] Vaegter HB, Jones MD. Exercise-induced hypoalgesia after acute and regular exercise: experimental and clinical manifestations and possible mechanisms in individuals with and without pain. *Pain Rep.* 2020;5(5):e823.
- [2] Kempainen P, Paalasmaa P, Pertovaara A, Alila A, Johansson G. Dexamethasone attenuates exercise-induced dental analgesia in man. *Brain Res.* 1990;519(1–2):329–32.
- [3] Tomschi F, Lieverkus D, Hilberg T. Exercise-induced hypoalgesia (EIH) in response to different exercise intensities. *Eur J Appl Physiol.* 2022;122(10):2213–22.
- [4] Gomolka S, Vaegter HB, Nijs J, Meeus M, Gajjar H, Hasenbring MI, et al. Assessing endogenous pain inhibition: Test–retest reliability of exercise-induced hypoalgesia in local and remote body parts after aerobic cycling. *Pain Med.* 2019;20(11):2272–82.
- [5] Hviid JCT, Thorlund JB, Vaegter HB. Walking increases pain tolerance in humans an experimental cross-over study. *Scand J Pain.* 2019;19(4):813–22.
- [6] Vaegter HB, Handberg G, Graven-Nielsen T. Similarities between exercise-induced hypoalgesia and conditioned pain modulation in humans. *Pain.* 2014;155(1):158–67.
- [7] Baiamonte BA, Kraemer RR, Chabreck CN, Reynolds ML, McCaleb KM, Shaheen GL, et al. Exercise-induced hypoalgesia: Pain

- tolerance, preference and tolerance for exercise intensity, and physiological correlates following dynamic circuit resistance exercise. *J Sports Sci.* 2017;35(18):1–7.
- [8] Deering R, Pashibin T, Cruz M, Hunter SK, Hoeger Bement M. Fatiguing trunk flexor exercise decreases pain sensitivity in postpartum women. *Front Physiol.* 2019;10:315.
- [9] Alsouhibani A, Vaegter HB, Hoeger Bement M. Systemic exercise-induced hypoalgesia following isometric exercise reduces conditioned pain modulation. *Pain Med.* 2019;20(1):180–90.
- [10] Fingleton C, Smart KM, Doody CM. Exercise-induced hypoalgesia in people with knee osteoarthritis with normal and abnormal conditioned pain modulation. *Clin J Pain.* 2017;33(5):395–404.
- [11] Hansen S, Vaegter HB, Petersen KK. Pretreatment exercise-induced hypoalgesia is associated with change in pain and function after standardized exercise therapy in painful knee osteoarthritis. *Clin J Pain.* 2020;36(1):16–24.
- [12] Naugle KM, Fillingim RB, Riley JL. A meta-analytic review of the hypoalgesic effects of exercise. *J Pain.* 2012;13(12):1139–50.
- [13] Nasri-Heir C, Patil AG, Korczeniewska OA, Zusman T, Khan J, Heir G, et al. The effect of nonstrenuous aerobic exercise in patients with chronic masticatory myalgia. *J Oral Facial Pain Headache.* 2019;33(2):143–52.
- [14] Vaegter HB, Handberg G, Graven-Nielsen T. Hypoalgesia after exercise and the cold pressor test is reduced in chronic musculoskeletal pain patients with high pain sensitivity. *Clin J Pain.* 2016;32(1):58–69.
- [15] Drury DG, Greenwood K, Stuempfle KJ, Koltyn KF. Changes in pain perception in women during and following an exhaustive incremental cycling exercise. *J Sports Sci Med.* 2005;4(3):215–22.
- [16] Jones MD, Nuzzo JL, Taylor JL, Barry BK. Aerobic exercise reduces pressure more than heat pain sensitivity in healthy adults. *Pain Med.* 2019;20(8):1534–46.
- [17] Ring C, Edwards L, Kavussanu M. Effects of isometric exercise on pain are mediated by blood pressure. *Biol Psychol.* 2008;78(1):123–8.
- [18] Ruble SB, Hoffman MD, Shepanski MA, Valic Z, Buckwalter JB, Clifford PS. Thermal pain perception after aerobic exercise. *Arch Phys Med Rehabilitation.* 2005;86(5):1019–23.
- [19] Thomas SD, Carter HH, Jones H, Thijssen DHJ, Low DA. Effects of acute exercise on cutaneous thermal sensation. *Int J Env Res Public Health.* 2020;17(7):E2491.
- [20] Vaegter HB, Dørge DB, Schmidt KS, Jensen AH, Graven-Nielsen T. Test-retest reliability of exercise-induced hypoalgesia after aerobic exercise. *Pain Med.* 2018;19(11):2212–22.
- [21] Weweg MA, Jones MD. Exercise-induced hypoalgesia in healthy individuals and people with chronic musculoskeletal pain: A systematic review and meta-analysis. *J Pain.* 2020;1–11.
- [22] Smith A, Ritchie C, Warren J, Sterling M. Exercise-induced hypoalgesia is impaired in chronic whiplash-associated disorders (WAD) with both aerobic and isometric exercise. *Clin J Pain.* 2020;36(8):601–11.
- [23] Vaegter HB, Lyng KD, Yttereng FW, Christensen MH, Sørensen MB, Graven-Nielsen T. Exercise-induced hypoalgesia after isometric wall squat exercise: A test-retest reliability study. *Pain Med.* 2018;20(1):129–37.
- [24] Ellingson LD, Koltyn KF, Kim JS, Cook DB. Does exercise induce hypoalgesia through conditioned pain modulation? *Psychophysiology.* 2014;51(3):267–76.
- [25] Vaegter HB, Petersen KK, Sjødsholm LV, Schou P, Andersen MB, Graven-Nielsen T. Impaired exercise-induced hypoalgesia in individuals reporting an increase in low back pain during acute exercise. *Eur J Pain.* 2021;25:1053–63.
- [26] Grimby-Ekman A, Ahlstrand C, Gerdle B, Larsson B, Sandén H. Pain intensity and pressure pain thresholds after a light dynamic physical load in patients with chronic neck-shoulder pain. *BMC Musculoskelet Disord.* 2020;21(1):266.
- [27] Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: A systematic review. *Man Ther.* 2010;15(3):220–8.
- [28] Pisters MF, Veenhof C, Schellevis FG, Twisk JWR, Dekker J, De Bakker DH. Exercise adherence improving long-term patient outcome in patients with osteoarthritis of the hip and/or knee. *Arthritis Care Res.* 2010;62(8):1087–94.
- [29] Petersen KK, Vaegter HB, Stubhaug A, Wolff A, Scammell BE, Arendt-Nielsen L, et al. The predictive value of quantitative sensory testing: a systematic review on chronic postoperative pain and the analgesic effect of pharmacological therapies in patients with chronic pain. *Pain.* 2021;162(1):31–44.
- [30] Vaegter HB, Handberg G, Emmeluth C, Graven-Nielsen T. Preoperative hypoalgesia after cold pressor test and aerobic exercise is associated with pain relief 6 months after total knee replacement. *Clin J Pain.* 2017;33(6):475–84.
- [31] Woznowski-Vu A, Uddin Z, Flegg D, Aternali A, Wickens R, Sullivan MJL, et al. Comparing novel and existing measures of sensitivity to physical activity among people with chronic musculoskeletal pain: The importance of tailoring activity to pain. *Clin J Pain.* 2019;35(8):656–67.
- [32] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018;27(5):1147–57.
- [33] Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med Res Methodol.* 2020;20(1):293.
- [34] Bonello C, Girdwood M, De Souza K, Trinder NK, Lewis J, Lazarczuk SL, et al. Does isometric exercise result in exercise induced hypoalgesia in people with local musculoskeletal pain? A systematic review. *Phys Ther Sport.* 2020;49:51–61.
- [35] Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* 2009;18(8):1115–23.
- [36] Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 2015.
- [37] McGuinness L. robvis: An R package and web application for visualizing risk-of-bias assessments; 2019. Disponible sur: <https://github.com/mcguinlu/robvis>.
- [38] Gohel D, Skintzos P. flextable: Functions for Tabular Reporting; 2023. <https://ardata-fr.github.io/flextable-book/https://davidgohel.github.io/flextable/>
- [39] Gohel D, Ross N. officedown: Enhanced « R Markdown » Format for « Word » and PowerPoint [Internet]; 2023. Disponible sur. <https://ardata-fr.github.io/officerverse/https://davidgohel.github.io/officedown/>
- [40] Stratford PW, Riddle DL. When minimal detectable change exceeds a diagnostic test-based threshold change value for an outcome measure: Resolving the conflict. *Phys Ther.* 2012;92(10):1338–47.

- [41] Vet HCW de. *Measurement in medicine*. Cambridge: Cambridge University Press; 2011. p. 350.
- [42] Wickham H. *ggplot2: Elegant graphics for data analysis*. 2nd edn. Cham: Springer International Publishing; Imprint: Springer; 2016. p. 1. (Use R!).
- [43] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
- [44] Lee S, Neogi T, Costello KE, Shah N, Rose MJ, Friscia B, et al. Association of conditioned pain modulation with exercise-induced hypoalgesia in knee osteoarthritis. *Osteoarthritis Cartilage*. 2022;30:S383.
- [45] Kuithan P, Heneghan NR, Rushton A, Falla D. Exercise induced hypoalgesia: stability of measures with functional lumbar spine resistance training. *Physiotherapy*. 2020;107:e49–50.
- [46] Vaegter HB, Bjerregaard LK, Redin MM, Rasmussen SH, Graven-Nielsen T. Hypoalgesia after bicycling at lactate threshold is reliable between sessions. *Eur J Appl Physiol*. 2019;119(1):91–102.
- [47] Oosthuyse T, Bosch AN. The effect of the menstrual cycle on exercise metabolism: implications for exercise performance in eumenorrhoeic women. *Sports Med*. 2010;40(3):207–27.
- [48] McNulty KL, Elliott-Sale KJ, Dolan E, Swinton PA, Ansdell P, Goodall S, et al. The effects of menstrual cycle phase on exercise performance in eumenorrhoeic women: A systematic review and meta-analysis. *Sports Med*. 2020;50(10):1813–27.
- [49] Iacovides S, Avidon I, Baker FC. Does pain vary across the menstrual cycle? A review. *Eur J Pain*. 2015;19(10):1389–405.
- [50] Aird TP, Davies RW, Carson BP. Effects of fasted vs fed-state exercise on performance and post-exercise metabolism: A systematic review and meta-analysis. *Scand J Med Sci Sports*. 2018;28(5):1476–93.
- [51] Vieira AF, Costa RR, Macedo RCO, Coconcelli L, Kruel LFM. Effects of aerobic exercise performed in fasted v. fed state on fat and carbohydrate metabolism in adults: a systematic review and meta-analysis. *Br J Nutr*. 2016;116(7):1153–64.
- [52] Vaegter HB, Thinggaard P, Madsen CH, Hasenbring M, Thorlund JB. Power of words: Influence of preexercise information on hypoalgesia after exercise-randomized controlled trial. *Med Sci Sports Exerc*. 2020;52(11):2373–9.
- [53] Jones MD, Valenzuela T, Booth J, Taylor JL, Barry BK. Explicit education about exercise-induced hypoalgesia influences pain responses to acute exercise in healthy adults: a randomized controlled trial. *J Pain*. 2017;18(11):1409–16.
- [54] Hoeger Bement MK, Rasiarimos RL, DiCapo JM, Lewis A, Keller ML, Harkins AL, et al. The role of the menstrual cycle phase in pain perception before and after an isometric fatiguing contraction. *Eur J Appl Physiol*. 2009;106(1):105–12.
- [55] Da Silva Santos R, Galdino G. Endogenous systems involved in exercise-induced analgesia. *J Physiol Pharmacol*. 2018;69(1):3–13.
- [56] Lesnak JB, Sluka KA. Mechanism of exercise-induced analgesia: what we can learn from physically active animals. *Pain Rep*. 2020;5(5):e850.
- [57] Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation – A discussion and demonstration of basic features. *PLoS ONE*. 2019;14(7):e0219854.
- [58] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539–49.
- [59] Rice D, Nijs J, Kosek E, Wideman T, Hasenbring MI, Koltyn K, et al. Exercise-induced hypoalgesia in pain-free and chronic pain populations: state of the art and future directions. *J Pain*. 2019;20(11):1249–66.
- [60] Patricio P, Mailloux C, Wideman TH, Langevin P, Descarreaux M, Beaulieu LD, et al. Assessment of exercise-induced hypoalgesia in chronic low back pain and potential associations with psychological factors and central sensitization symptoms: A case-control study. *Pain Pract*. 2022;23(3):264–76.
- [61] Bossenger NR, Lewis GN, Rice DA, Shepherd D. The autonomic and nociceptive response to acute exercise is impaired in people with knee osteoarthritis. *Neurobiol Pain*. 2023;13:100118.
- [62] De Kooning M, Coppieters I, Huysmans E, Nijs J, Meeus M, Voogt L, et al. Unravelling impaired hypoalgesia at rest and in response to exercise in patients with chronic whiplash-associated disorders: effects of a single administration of selective serotonin reuptake inhibitor versus selective norepinephrine reuptake inhibitor. *J Clin Med*. 2023;12(15):4977.
- [63] Drucker AM, Fleming P, Chan AW. Research techniques made simple: assessing Risk of Bias in systematic reviews. *J Investig Dermatol*. 2016;136(11):e109-14.
- [64] Dalton JE, Bolen SD, Mascha EJ. Publication bias: The elephant in the review. *Anesth Analg*. 2016;123(4):812–3.
- [65] Dickersin K. Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein HR, Sutton AJ, Borenstein M, editors. *Publication Bias in Meta-Analysis*. Chichester, UK: John Wiley & Sons, Ltd; 2006 [cité 14 juill 2022]. p. 9–33. Disponible sur. <https://onlinelibrary.wiley.com/doi/10.1002/0470870168.ch2>.