Are changes in nociceptive withdrawal reflex magnitude a viable central sensitization proxy? Implications of a replication attempt

A. Guekos a,b,* A.C. Grata c, M. Hubli d,e, M. Schubert d,e, P. Schweinhardt a,d

a Integrative Spinal Research, Department of Chiropractic Medicine, Balgrist University Hospital, University of Zurich, Zurich, Switzerland
b Decision Neuroscience Lab, Institute of Human Movement Sciences and Sport, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland
c Department of Cell and Developmental Biology, Feinberg School of Medicine, Northwestern University, Chicago IL, United States
d Faculty of Medicine, University of Zurich, Zurich, Switzerland
e Spinal Cord Injury Center, Balgrist University Hospital, University of Zurich, Zurich, Switzerland

Article history:
Accepted 21 September 2022
Available online 1 October 2022

Keywords:
Nociceptive withdrawal reflex
Central sensitization
Response rate
Reliability

Highlights
- A previously observed increase in nociceptive withdrawal reflex (NWR) magnitude during application of painful heat was not replicated.
- Strong parameter sensitivity in central sensitization and NWR studies warrants precise reporting of study methods to guarantee replicability.
- Stimulating the medial plantar nerve and recording from the tibialis anterior muscle yielded response rates of up to 79% for repetitive suprathreshold stimulations.

Abstract
Objective: The nociceptive withdrawal reflex (NWR) has been proposed to read-out central sensitization (CS). Replicating a published study, it was assessed if the NWR magnitude reflects sensitization by painful heat. Additionally, NWR response rates were compared for two stimulation, the sural nerve at the lateral malleolus (SU) and the medial plantar nerve on the foot sole (MP), and three recording sites, biceps femoris (BF), rectus femoris (RF), and tibialis anterior (TA) muscles.

Methods: 16 subjects underwent one experiment with six blocks of eight transcutaneous electrical stimulations to elicit the NWR while surface electromyography was collected. Tonic heat was concurrently applied in the same dermatome. Temperatures rose from 32°C in the first to 46°C in the last block following the previously published protocol.

Results: Tonic heat did not influence NWR magnitude. The highest NWR response rate was obtained for MP–TA combination (79%). Regarding elicitation in all three muscles, SU stimulation outperformed MP (59% vs 57%).

Conclusions: The replication failed. NWR magnitude as a CS proxy in healthy subjects needs continued investigation. With respect to response rates, MP–TA proved efficient, whereas SU stimulation seemed preferable for multiple muscle recordings.

Significance: Unclear methodological descriptions in the original study affected CS and NWR replication. The NWR magnitude changes induced by CS may closely depend on the different stimulation methods used.

© 2022 International Federation of Clinical Neurophysiology. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The nociceptive withdrawal reflex (NWR) is a polysynaptic spinal response to noxious stimulation. Activation of nociceptive Aδ and C fibres, and of mechanosensitive Aβ fibres initiates the reflex pathway, prompting an involuntary muscle response of flex-
ors and extensors, withdrawing the affected limb (Pierrot-Deseilligny and Burke, 2012). It is used to study nociceptive processing on the spinal level in healthy human participants and in chronic pain patients (Sandrini et al., 2005; Sklarevski and Ramadan, 2002; Manresa et al., 2011; Willer, 1977; Downman, 1991; Wallwork et al., 2017).

As a spinal read-out, the NWR is an interesting tool to investigate central sensitization (CS) (Manresa et al., 2010; Manresa et al., 2013; Arendt-Nielsen et al., 2018; Sandrini et al., 2005). CS is defined by the International Association for the Study of Pain (IASP) as an increased responsiveness of nociceptive neurons in the central nervous system to their normal or subthreshold afferent input (Loeser and Treede, 2008). It is considered to be pathophysiologically important in chronic pain (Arendt-Nielsen et al., 2018; Woolf, 2011). Because direct neuronal recordings are not possible in humans (Arendt-Nielsen et al., 2018; Woolf, 2011), CS induction - also called reflex hyperalgesia or dynamic mechanical allodynia (Quesada et al., 2021; Woolf, 2011) - can be influenced by supraspinal facilitation or inhibition (Bjerre et al., 2011; Terry et al., 2016; Willer et al., 1979). Thus, it is crucial to choose a measurement for the NWR with high test–retest reliability in spite of potential supraspinal influences.

Most studies using the NWR after experimental CS induction focus on effects on the NWR threshold \( I_0 \) (Manresa et al., 2014; Linde et al., 2021; Leone et al., 2021). It is one of the most studied characteristics of the NWR because changes in \( I_0 \) potentially reflect altered nociceptive processing, especially in clinical populations (Sandrini et al., 1993; Arendt-Nielsen et al., 1994; Edwards et al., 2007; Manresa et al., 2011; Linde et al., 2021). However, conflicting results regarding within-subject test–retest reliability of \( I_0 \) have been reported for studies with healthy subjects (Manresa et al., 2014; Linde et al., 2021; Neziri et al., 2010; French et al., 2005). Therefore, \( I_0 \) might not be the best choice to assess the effects of CS induction. Instead, the NWR magnitude - also called reflex size - might be a viable alternative because it has been observed to remain stable in test–retest studies (Jurt et al., 2014). In addition, it has been observed to reflect CS induction (Manresa et al., 2010; Ellrich and Treede, 1998). In particular, one study reported effects on NWR magnitude using noxious heat, which can be interpreted as reflecting successful CS induction (Ellrich and Treede, 1998). In contrast, a recently published study using capsaicin in combination with noxious heat for CS induction reported no effect on the NWR magnitude (Leone et al., 2021). In both studies, the NWR magnitude had been quantified by the area under the curve (AUC) of the sEMG recordings. To help clarify the controversial results, the primary objective of the present study was to replicate the previously published results (Ellrich and Treede, 1998), where a thermally induced sensitized state had changed the NWR magnitude.

In that study, the NWR had been elicited by transcutaneous electrical stimulations to the medial plantar nerve (MP) on the foot sole proximal to the big toe and recorded from the ipsilateral tibialis anterior muscle (TA) by surface electromyography (sEMG) (Ellrich and Treede, 1998). Stimulation to a distal nerve and sEMG recording from one of the muscles involved in the withdrawal is a practical approach to using the NWR in experimental studies. Because of its simplicity, it can be integrated into designs involving outcome measures from different modalities (Rhudy et al., 2020). More advanced approaches to studying the NWR are also used, such as muscle synergy analyses (Jure et al., 2019). However, their integration into a multimodal design is more challenging than just combining one stimulation and one recording site. Apart from MP–TA (Meinck et al., 1985), another typical combination uses the sural nerve (SU) dorsal to the lateral malleolus as the stimulation and the biceps femoris muscle (BF) as the recording site (Hugon, 1973; Willer, 1977), although other combinations are used as well, some of which alternatively stimulate free nerve endings instead of trunks (Sandrini et al., 2005; Jensen et al., 2015; Jure et al., 2020).

For MP–TA, \( I_0 \) exhibits higher between sessions test–retest reliability than for SU–BF (Jensen et al., 2015). However, a high between-sessions reliability of \( I_0 \) does not guarantee successful repetitive elicitations within experimental sessions, when the NWR is attempted to be elicited repeatedly at suprathreshold currents (Willer, 1977; Sklarevski and Ramadan, 2002; Sandrini et al., 2005). Conversely, a stimulation and recording site combination might be reliable for elicitations but not for \( I_0 \) determination. In contrast to \( I_0 \), the NWR response rate has not been extensively studied, except for the so-called "habituation" effect (Dimitrijević et al., 1972; Arendt-Nielsen et al., 1994; Sandrini et al., 2005). But, habituation observations only focus on the declining NWR amplitudes over time of successful elicitations and not on the number of actually elicited reflexes. Therefore, the secondary objective of the present study was to determine the stimulation and recording site combination yielding the highest NWR response rate, i.e. the highest percentage of NWR elicitations.

To address both research objectives, the experimental paradigm from the published study was not only replicated using the MP–TA combination as in the original design, wherein the NWR was repeatedly elicited at MP and noxious heat applied to the same dermatome resulted in a NWR magnitude increase at the TA (Ellrich and Treede, 1998). But, it was also replicated for the SU–BF combination. In order to compare more than just two combinations, sEMG recordings were in both cases simultaneously collected from three muscles, TA, BF, and the rectus femoris muscle (RF), leading to a total of six comparable combinations.

Based on the results of the original study (Ellrich and Treede, 1998), it was expected that the NWR magnitude would increase for MP–TA and SU–BF between the baseline temperature and the two highest temperatures. For the secondary objective, it was hypothesized, because of the reported high \( I_0 \) reliability for MP–TA (Jensen et al., 2015), that (i) MP–TA would show the highest NWR response rate and (ii) stimulations at MP would trigger a higher NWR response rate in all three muscles compared to SU.

2. Methods

2.1. General study design

The study followed a within-subject experimental design, closely following a previously published study (Ellrich and Treede, 1998). The original design was reproduced as faithfully as possible using all the published information. Changes were only made if deemed necessary in order to avoid biasing the results. All changes are discussed in detail in the supplementary material A. Study approval was obtained from the ethics board of the Canton of Zurich. All participants underwent a single experimental session that consisted of two identical testing procedures, one for each stimulation site combination as in the original design, wherein the NWR was repeatedly elicited at MP and noxious heat applied to the same dermatome resulted in a NWR magnitude increase at the TA (Ellrich and Treede, 1998). But, it was also replicated for the SU–BF combination. In order to compare more than just two combinations, sEMG recordings were in both cases simultaneously collected from three muscles, TA, BF, and the rectus femoris muscle (RF), leading to a total of six comparable combinations.

Based on the results of the original study (Ellrich and Treede, 1998), it was expected that the NWR magnitude would increase for MP–TA and SU–BF between the baseline temperature and the two highest temperatures. For the secondary objective, it was hypothesized, because of the reported high \( I_0 \) reliability for MP–TA (Jensen et al., 2015), that (i) MP–TA would show the highest NWR response rate and (ii) stimulations at MP would trigger a higher NWR response rate in all three muscles compared to SU.

2.2. Participants

Participants were healthy, pain-free adults who gave written informed consent prior to inclusion according to the Declaration of Helsinki. Exclusion criteria were the presence of any low back pain on more than two consecutive days within the last year, a serious disease, e.g. heart failure or clinical depression, any pain...
condition, or inability to tolerate the applied electrical stimulations or heat stimuli.

2.3. Sample size

The previously reported results corresponded to effect sizes (calculated as Cohen’s d (Cohen, 1988)) of 0.7, 1, and 2.4, respectively (Ellrich and Treede, 1998). Using an effect size of 0.85, a significance level of 0.05, and a power of 0.8 the necessary sample size was determined to be 13. It was decided to recruit 17 participants to account for potential dropouts or unusable data.

2.4. Testing procedure

2.4.1. Participant setup

For the duration of the experiment, participants were in a semi-supine position on a test bed, with the back rest inclined at 60° and their legs comfortably positioned. To ensure muscle relaxation in the test leg, a rolled-up towel was placed underneath the popliteal fossa. All tests were performed on the right leg. During the tests, participants were asked to read a newspaper article in order to be distracted from the test stimulations as a distraction-related facilitation of the NWR had been shown (Bjerre et al., 2011; Terry et al., 2016).

2.4.2. Electrical simulation and sEMG recordings

Transcutaneous constant current stimulation and ipsilateral sEMG recordings were performed using a Dantec Keypoint G4 EMG/EP Workstation (Natus Medical Inc, San Carlos CA, US). The recording time window extended from 120 ms pre- to 380 ms post-stimulation.

Fig. 1 shows a schematic of the stimulation and recording setup during the experiment. On each of the three recording sites, two Ambu® (Ambu A/S, Copenhagen, DK) Blue Sensor NF-50 1.5 mm surface electrodes were placed, one on the muscle belly and one on the nearest reference point, i.e. near the distal tendon insertion for BF and RF and on the shin for TA. Before attaching the electrodes, the skin was cleaned with ethanol and exfoliated with Nuprep® gel (Weaver and Company, Aurora CO, USA) to achieve impedances < 5 kΩ. For SU stimulation, two Ambu® Neuroline 700 2 mm surface electrodes were placed 2 cm apart, whereas for MP stimulation one Ambu® Neuroline 700 2 mm surface electrode was placed as the cathode on the foot sole proximal to the medial side of the ball of the foot and one ValuTrode® (Axelgaard Manufacturing Co., Lystrup, DK) VTX 5 cm surface electrode as the anode on the dorsum of the foot.

Every electrical stimulation consisted of a train of five rectangular stimuli of 1 ms duration delivered at 200 Hz. For each of the two stimulation sites, eight blocks of 90s were performed with pauses of 300s between blocks (Fig. 2). During the 90s, electrical stimulations were applied. The first two blocks served to determine I₀ in either the TA (for MP stimulation) or the BF (for SU stimulation). The TA and BF muscles were chosen for the TA (for MP stimulation) or the BF (for SU stimulation). The I₀ was defined as the higher of the single stimulation and 3-stimulation thresholds because it has been observed that this definition yields high reliability of I₀ determination (Terry et al., 2001; Terry et al., 2016; Rhudy et al., 2020). The single stimulation threshold is the lowest current to reliably elicit a reflex with a single electrical stimulation, as defined above. The 3-stimulation threshold refers to the third such stimulation in a triplet of identical electrical stimulations delivered at 2 Hz. It is the lowest current to reliably elicit a reflex with the third stimulation of such a triplet (Terry et al., 2001; Terry et al., 2016; Rhudy et al., 2020). In both, a single ascending staircase (Rhudy and France, 2011) was used with increments of 1 mA starting at 1 mA until the first reflex was reached and continued for two more increments for validation. The order of these first two blocks, i.e. starting either with the determination of the single or the 3-stimulation threshold, was counterbalanced across subjects. The subsequent six test blocks employed eight single stimulations with a random interstimulation interval of 5 to 15s to avoid predictability, which has been shown to modulate the NWR response (Jure et al., 2020). The eight stimulations were delivered at eight different currents, starting from I₀-4 mA to I₀+1 mA in steps of 1 mA followed by one stimulation at 1.5I₀ and one stimulation at 2I₀, always in that order.

Participants were instructed to rate each stimulation on a numerical scale ranging from 0 (no sensation) to 200 (maximum tolerable pain), with 100 corresponding to the pain threshold.

2.4.3. Heat stimuli

The electrical stimulations were applied during the application of 90s tonic heat stimuli using a 3 cm x 3 cm computer controlled Peltier thermode (TSA II, Medoc, St. Ramat Yishai, IL). Fig. 2 shows the heat stimuli procedure following the previously published design (Ellrich and Treede, 1998). The respective target tempera-
tures of the six test blocks were 32, 36, 39, 42, 45, and 46 °C, always in that order. During the two \( I_0 \) determination blocks the temperature was kept at 32 °C baseline temperature.

The thermode was attached proximal to the site of electrical stimulation, i.e. on the lateral lower leg, proximal to the ankle for SU and on the sole of the foot for MP stimulation respectively (Fig. 1). These positions were chosen following the previously published design (Ellrich and Treede, 1998) and to ensure that the tonic heat stimulus would be applied in the same dermatome as the electrical stimulations, L5 and S1 respectively (Downs and Laporte, 2011). The thermode remained attached until all blocks for the first site had been performed and was then switched to the other site. For each of the six target temperatures ascending and descending ramps of 8 °C/s were used. The temperature returned to baseline (32 °C) for the 300s pause in between blocks. After every block, participants were asked whether they considered the applied temperature as painful or not painful.

### 2.5. Data analysis

The sEMG signals were sampled at 48 kHz and downsampled to 6 kHz, rectified, band-pass filtered from 10 Hz to 500 Hz and amplified up to 125 times. Between –120 ms pre- and 380 ms post-stimulation, traces for all applied stimulations were automatically saved into separate text files (.txt) with stimulation number logged for every stimulation. All suprathreshold stimulations, i.e. applying currents \( \geq I_0 \), were further analysed, i.e. 2304 of 4608 (= 8 stimuli · 6 blocks · 3 muscles · 2 stimulation sites · 16 subjects). The time window of interest for NWR occurrence extended from 60 to 150 ms post-stimulation. The lower bound of 60 ms was chosen because it corresponds to the reported earliest onset for the NWR defining activation of \( A_0 \) fibres in similar studies (Andersen, 2007). The upper bound was set at 150 ms to avoid contamination of the reflex response by voluntary muscle activation (Downan, 1992).

A reflex was computationally judged to be present if the NWR mean z-score \( Z \) was \( \geq 1.4 \). \( Z \) was defined as the difference between the mean of the 60 to 150 ms post-stimulation amplitude and the mean of the \( -110 \) to \( -20 \) ms pre-stimulation amplitude divided by the standard deviation of the \( -110 \) to \( -20 \) ms pre-stimulation amplitude. The \( Z \) criterion is a well validated approach to automatically detect the NWR (France et al., 2009; Rhudy and France, 2007). However, it is susceptible to contamination of the sEMG by spontaneous muscle activation. Visual inspection of the traces is necessary to eliminate false classifications (Rhudy and France, 2007; France et al., 2009) although alternative methods have been proposed (Jensen et al., 2013). Thus, all sEMG traces were also visually inspected by the two investigators performing the tests (AG and ACG) to eliminate any potential false negative or false positive automatic identifications. When either of the investigators disagreed with the \( Z \) criterion the traces were given to two expert raters (MH and MS) who had not been not involved in the data collection. The independent assessments of the two expert raters were compared. In case of disagreements, the respective traces were given to an external expert rater to act as an arbiter. In total, 113 (4.9%) out of the analysed 2304 traces were identified as either false negative (110 i.e. 4.8%) or false positive (3, i.e. 0.1%).

Statistical analyses were performed in R 3.4.4 (R Foundation for Statistical Computing, Vienna, Austria) using the tidyverse collection of packages as well as the packages pwr 1.3.0, fmsb 0.7.0, nlme 3.1.131, lme4 1.1.21, relaimpo 2.2.3, rjags 4.10 and runjags 2.0.4.6 relying on JAGS 4.3.0. Demographic information of the study participants as well as the data from the traces to be used for further analysis were stored in a comma separated values (csv) file.

As all subsequent analysis focussed on reflexes with suprathreshold current stimulations it was investigated whether there was a potential association of the \( I_0 \) with the demographic characteristics of the study participants, namely with age, height, sex, handedness, time of measurement, and stimulation order (SU or MP first). For continuous variables, Spearman’s rank correlation tests, and for categorical variables, Kruskal–Wallis rank sum tests were performed. Because reflex responses between the two stimulation sites were compared, it was also tested whether there was an association between the two respective \( I_0 \) for the two stimulation sites using Spearman’s rank correlation tests. \( P < 0.05 \) was considered significant for all analyses.

For the primary objective, determining and analysing a potential influence of temperature on the NWR, the magnitude of all obtained reflexes was quantified using the NWR Cohen’s-\( d \) (D). D was defined as the difference between the mean of the 60 to 150 ms post-stimulation amplitude and the mean of the \(-110 \) to \(-20 \) ms pre-stimulation amplitude divided by the pooled standard deviation of the \(-110 \) to \(-20 \) ms pre-stimulation amplitude. D is a robust way of quantifying the magnitude of the NWR with a very high test–retest reliability (Rhudy et al., 2020; Terry et al., 2016; France et al., 2009; Rhudy et al., 2009; Rhudy et al., 2008; Rhudy and France, 2007). For comparability with the results from the original study (Ellrich and Treede, 1998), the area under the sEMG curve (AUC) was also calculated as an alternative quantification metric.

Fig. 2. Heat stimuli procedure. Schematic of the experimental blocks. The first two blocks are used to determine the threshold \( I_0 \) of the nociceptive withdrawal reflex (NWR). During the following six blocks heat stimuli of increasing temperatures are applied while electrical stimulations w.r.t. \( I_0 \) are applied.
Bayesian normal-normal models were used to assess the effect size for the temperature influence on D for all muscle responses. Models without and with a simple hierarchical structure were applied to account for individual differences on the participant level.

Additionally, a classical linear model with temperature as the sole predictor was trained. In a separate mixed effects model subject ID was entered as a random effect.

Because the original design can be interpreted as a split-model design, albeit not stated as such in (Ellrich and Treede, 1998), a separate mixed effects model was trained with temperature as whole-plot factor, suprathreshold current as split-plot factor, subject ID as block factor, and a random effect per combination of participant and temperature.

Following the analysis in the original study (Ellrich and Treede, 1998), paired samples t-tests for AUC at the two highest temperatures (i.e. 45 and 46 °C) compared to baseline were performed. For comparability with the results from the original study, this analysis was restricted to the MP–TA combination for electrical stimulations that were rated as not painful (i.e. < 100) and to subjects who had considered the heat stimuli of 45 and 46 °C as painful.

The second outcome measure was the presence of a NWR (binary variable) which was analysed for the two stimulation (MP, SU) and the three recording sites (BF, RF, TA). Bayesian beta-binomial models were used to determine the highest density intervals (HDI) for elicitation of all six stimulation and recording site combinations separately as well as for the respective MP and SU elicited reflexes combined. These models allowed to determine robust ranges of NWR response rates using a particular stimulation and recording site combination. The effectiveness of one combination compared to another was then calculated with the help of the so-called “rate ratio” (RR) (Fleiss et al., 2003), i.e. the ratio of the two calculated ranges of the combinations in question, with effectiveness defined as 1-RR.

In addition, the data were analysed using a classical (i.e. frequentist) statistical approach. For every stimulation and recording site combination, the NWR response rate, i.e. the percentage of elicited NWR at suprathreshold currents was calculated. The six combinations (MP–BF, MP–RF, MP–TA, SU–BF, SU–RF, SU–TA) were compared using \( \chi^2 \) tests of independence for between subjects and McNemar’s tests for within subjects analysis. The analysis was conducted between the two stimulation sites for all three muscles together (1 group-wise comparison) as well as for every muscle separately (9 pair-wise comparisons). Additionally, comparisons of muscle responses for the same stimulation site were performed (2 - 3 pair-wise comparisons).

The four subthreshold stimulations served to indirectly test for a potential decrease in reflex threshold \( I_0 \) over the course of the experiment. To that end, the NWR response rates for all subthreshold stimulations were determined for the MP–TA and SU–BF combinations at each of the six applied temperatures to test for a potential increase of the response rates with increasing temperature. Because \( I_0 \) had been obtained for MP–TA and SU–BF respectively increased response rates for these combinations would indicate a lowered reflex threshold.

### 3. Results

#### 3.1. Participant characteristics

In total, 17 participants were recruited. One had to be excluded because of not tolerating the two highest temperatures used. This led to a study sample of 16 participants of equal sex distribution and a mean age of 26.8 ± 4.7 years.

No association between the stimulation-site-specific \( I_0 \) and the demographic characteristics of the study participants was found (Table 1). Therefore, they were not included as covariates in the further analysis. The \( I_0 \) at MP (8.6 ± 4.9 [mA]) and SU (11.4 ± 6.1 [mA]) were positively correlated (Spearman’s \( r = 0.65 \) with \( p = 0.006 \)).

#### 3.2. No observable influence of noxious heat stimuli on the NWR

Fig. 3 shows the sEMG recordings of two representative subjects for the MP–TA and SU–BF combinations following 1.5 \( I_0 \) stimulation for all six temperature blocks, with the NWR clearly visible within the 60 to 150 ms post-stimulation time window.

The original study claimed a facilitation of the electrically elicited reflex response by painful heat. A significant increase of the NWR magnitude, quantified as AUC, was reported for stimulations in the last two experimental blocks, i.e. when a noxious tonic heat stimulus of 45 or 46 °C was respectively applied. At 46 °C, the increase of AUC had been significant in all subjects who had been tested with stimulation currents corresponding to about 1.2 \( I_0 \) and 3 \( I_0 \) respectively. However, the increase at 45 °C had been significant only for stimulations corresponding to about 3 \( I_0 \) (Ellrich and Treede, 1998).

In the present study, all participants reported that they considered the two highest temperatures (45 and 46 °C) as painful, thus fulfilling the prerequisite of the original study. It was assessed whether a tonic heat stimulus, applied in the same dermatome as the concurrent NWR elicitation, had on effect on the NWR magnitude, as reported in the original study (Ellrich and Treede, 1998), when comparing between the two highest, i.e. painful, temperatures (45 and 46 °C) and the baseline (32 °C). To that end, a Bayesian normal-normal model without hierarchical structure was first used. The model was run for both stimulation sites, for all three recording sites and suprathreshold stimulation currents together and separately.

For MP stimulation, comparing the magnitudes between 32 and 46 °C as well as between 32 and 45 °C yielded median effect sizes of 0.01 (HDI\(_{95} = [-0.19,0.21]\)) and −0.06 (HDI\(_{95} = [-0.26,0.13]\)). For SU stimulation, the median effect sizes were −0.03 (HDI\(_{95} = [-0.24,0.16]\)) and 0.09 (HDI\(_{95} = [-0.11,0.29]\)), respectively.

| Table 1 | Characteristics of the study participants (n = 16). |
|-----------------|-------------------------------|
| **Continuous Variables** | **Association to \( I_0 \) at MP:** \( p (\rho) \) | **Association to \( I_0 \) at SU:** \( p (\rho) \) |
| **Age (years)** | 26.8 ± 4.7 | 0.76 (0.08) | 0.19 (0.34) |
| **Height (cm)** | 174.4 ± 7.2 | 0.48 (0.19) | 0.90 (0.03) |
| **Binary Variables** | **Distribution** | **Association to \( I_0 \) at MP:** \( p (\chi^2) \) | **Association to \( I_0 \) at SU:** \( p (\chi^2) \) |
| **Sex (m:f)** | 1:1 | 0.92 (0.01) | 0.92 (0.01) |
| **Handedness** | 13:3 | 0.50 (0.46) | 0.54 (0.37) |
| **Day Time** | 1:1 | 0.13 (2.33) | 0.48 (0.50) |
| **Start Site** | 1:1 | 0.22 (1.47) | 0.14 (2.18) |

Abbreviations: \( \sigma \) = standard deviation; \( I_0 \) = threshold (current of nociceptive withdrawal reflex; MP = medial plantar nerve electrical stimulation site; SU = sural nerve electrical stimulation site; m = male; f = female; AM = ante meridiem; PM = post meridiem; \( p = p \)-value; \( r = \) rank correlation coefficient; \( \chi^2 = \chi^2 \) statistic; \( n \) = number of study participants.

Associations were calculated using the Spearman’s rank correlation test (continuous variables) and the Kruskal–Wallis rank sum test (binary variables).
For the MP–TA pair, for which effect sizes from 0.7 to 2.4 had been reported in (Ellrich and Treede, 1998), the median effect sizes were 0.01 (HDI95=[-0.34,0.35]) and 0.04 (HDI95=[-0.39,0.30]).

For the SU–BF pair, median effect sizes of 0.07 (HDI 95=[-0.41,0.27]) and 0.12 (HDI95=[-0.23,0.46]) were obtained.

No significantly greater effect sizes resulted from the other stimulation and recording site combinations nor from separately assessing the data at the different suprathreshold stimulation currents (results not shown).

Second, hierarchical Bayesian normal-normal models were applied to account for individual differences between the participants. The resulting median effect sizes were greater but did still not come close to the reported ones. In the case of MP–TA, the present median effect sizes were 0.02 (HDI95=[-0.80,0.90]) and 0.05 (HDI95=[-0.87,0.79]), respectively. And for SU–BF, the median effect sizes were -0.11 (HDI95=[-0.92,0.71]) and 0.19 (HDI95=[-0.65,1.06]). Again, separate assessments for the other stimulation and recording site combinations displayed the same behaviour (results not shown).

Fig. 4 shows the variation of the reflex magnitude in the BF, RF, and TA muscles after suprathreshold stimulations as quantified by D in the presence of a tonic heat stimulus in the same dermatome. The temperature of the heat stimulus increased from 32°C (baseline) in the first to 46°C in the last block following the published design (Ellrich and Treede, 1998). As determined in the analysis, no clear pattern emerged with increasing temperature in any of the three recorded muscles, neither for MP nor for SU stimulation. This did not change when the four stimulation currents (I0, I0+1 mA, 1.5I0, 2I0) were considered separately (see supplementary material B).

Using a classical statistical approach, neither a linear regression model with temperature as the sole predictor nor a mixed effects model had any explanatory value (see supplementary material B). Similarly, the split-plot model analysis did not reveal any significant influence of the whole-plot factor temperature on the reflex magnitude (see supplementary material B).

Following the original study (Ellrich and Treede, 1998), the NWR magnitude for the MP–TA combination was also quantified by the area under the curve (AUC) of the sEMG recordings for the 50 to 100 ms post-stimulation time window for all stimulations corresponding to about 1.2I0, i.e. for the stimulation at I0+1 mA, for 10 out of 16 participants. In order to comply with the original design (Ellrich and Treede, 1998), two participants were excluded from this analysis because they had consistently rated the stimulations at I0 + 1 mA as painful, i.e. >100 (see supplementary material D for details on reported pain ratings). Paired samples t-tests were performed for the AUC at either of the two highest temperatures (i.e. 45 and 46°C) and at 36°C. In contrast to the published results (Ellrich and Treede, 1998), the AUC did not continuously increase with temperature but rather fluctuate (Fig. 5) and was not significantly different neither between 46 and 36 °C nor between 45 and 36 °C (p = 0.086 and p = 0.105, respectively). The analysis was restricted to AUC of sEMG curves from actual reflexes, i.e. fulfilling the Z criterion (interval mean z score ≥1.4 (France et al., 2009; Rhudy and France, 2007)). However, the original article did not state whether a scoring criterion was used or not. If the AUC of all sEMG curves were considered the p-values greatly increased, both between 46 and 36 °C as well as between 45 and 36 °C (p = 0.289 and p = 0.273, respectively). If the same tests were performed for D the differences between 46 and 36 °C as well as...
between 45 and 36 °C were not significant either (p = 0.075 and p = 0.067, respectively). Additionally, the values for D were lower at the higher temperatures thus contradicting the results from the original study (Fig. 5). Again, the differences became more insignificant (p = 0.568 and p = 0.4485 respectively) if all values of D were considered.

3.3. NWR response rate

Fig. 6 depicts NWR response rates for all six stimulation and recording site combinations when suprathreshold stimulation currents were used. Overall, MP–TA, SU–BF, and SU–RF showed the highest NWR response rates of 79.4%, 66.9%, and 61.2% respectively. Fig. 7 shows the NWR response rate per recording and stimu-
ulation site combination w.r.t. the stimulation current that was applied to elicit the reflex, i.e. the change of reflex responses w.r.t. stimulation current. The steepest increase was observed for MP–BF and MP–RF. The detailed numbers for every combination are provided in the supplementary material C.3. The numbers of reflex responses, i.e. elicited NWR, per stimulation and recording combination for all suprathreshold currents are shown in Table 2. These data were entered into Bayesian beta-binomial models.

When all three muscles and all four stimulation currents were considered together, the 95% HDI (HDI95) for the NWR response rate ranged from 0.56 to 0.62 (median = 0.59) for SU stimulation and from 0.54 to 0.60 (median = 0.57) for MP stimulation. This means that 59% of the time reflexes were elicited with SU stimulation, and 57% with MP stimulation. However, the HDI95 for the calculated greater effectiveness extended from 0.03 to 0.10. Because 0 still lay within the HDI95, the SU stimulation did statistically not yield greater effectiveness to elicit a NWR than MP stimulation.

Comparing the overall two best stimulation and recording site combinations (MP–TA and SU–BF), the effectiveness was greater for MP–TA (HDI95 = [0.75, 0.83], median = 0.79) than for SU–BF (HDI95 = [0.62, 0.71], median = 0.67). The latter did not show greater effectiveness than the third placed SU–RF (HDI95 = [0.56, 0.66], median = 0.61).

The HDI95 for all six pairs, as well as the results of the classical statistical approach are presented in the supplementary material C. By and large, the classical approach yielded the same results.

The NWR response rates for subthreshold stimulations did not increase with temperature, neither for MP–TA nor for SU–BF (see supplementary material C.2).

### 4. Discussion

First, this study aimed to replicate the results of (Ellrich and Treede, 1998). The NWR had been shown to reflect ongoing sensi-

![Fig. 7. Change of nociceptive withdrawal reflex (NWR) response rates w.r.t. stimulation current for every stimulation and recording site combination. Top row (left to right): MP–BF, MP–RF, MP–TA. Bottom row (left to right): SU–BF, SU–RF, SU–TA. Boxplots indicate the 25 to 75% interquartile range and median (thick horizontal line), with whiskers extending to 10 and 90% respectively. Abbreviations: MP = medial plantar nerve electrical stimulation site; SU = sural nerve electrical stimulation site; BF = biceps femoris muscle; RF = rectus femoris muscle; TA = tibialis anterior muscle; \(I_0\) = reflex threshold current.](image)

### Table 2

Numbers of elicited NWR (response rate) per stimulation site, muscle, and current.

|       | MP           |       | SU           |       |
|-------|--------------|-------|--------------|-------|
|       | BF           | RF    | TA           | BF    | RF    | TA    |
| \(I_0\) | 22 (22.9%)   | 18 (18.8%) | 65 (67.7%)  | 43 (44.8%) | 40 (41.7%) | 38 (39.6%) |
| \(I_0 + 1\) mA | 31 (32.3%)   | 28 (29.2%) | 68 (70.8%)  | 58 (60.4%) | 54 (56.2%) | 42 (43.8%) |
| 1.5 \(I_0\) | 58 (60.4%)   | 83 (86.5%) | 89 (92.7%)  | 78 (81.2%) | 74 (77.1%) | 54 (56.3%) |
| 2 \(I_0\) | 75 (78.1%)   | 68 (70.8%) |             |       |       |       |

Abbreviations: MP = medial plantar nerve electrical stimulation site; SU = sural nerve electrical stimulation site; BF = biceps femoris muscle; RF = rectus femoris muscle; TA = tibialis anterior muscle; \(I_0\) = reflex threshold current; NWR = nociceptive withdrawal reflex; response rate = percentage of elicited NWR.

Maximum possible elicited NWR per stimulation and recording site combination and current \(> I_0\): 96 (6 blocks × 16 participants).
tization processes, which resembled thermally induced CS in the same dermatome where the reflex was elicited. In the present study, this finding could not be replicated. Although the documented design of (Ellrich and Treede, 1998) was reproduced as closely as possible, no systematic influence of the thermal CS induction paradigm on the NWR magnitude was observed. Because the potential CS induction was not verified by other proxies, such as secondary hyperalgesia, it cannot be said with absolute certainty whether CS was really induced. Yet, the absence of a modulation of the NWR (see 3.2) as well as of the pain ratings of the electrical stimulations (see supplementary material D), indicate that, in the present study, CS was probably not induced. Thus, it remains an open question if NWR magnitude changes can serve as a reliable CS proxy in studies using painful heat to induce CS in healthy participants as has been shown for CS induction using electrical stimulation (Manresa et al., 2010).

Second, this study investigated the NWR response rate after repetitive electrical stimulation. Six pairs of stimulation and recording sites were tested. Three of these showed response rates of 61% or higher. MP–TA was the combination that yielded the highest NWR response rate (79.4%), followed by SU–BF (66.9%) and SU–RF (61.2%). The difference between MP–TA and SU–BF as well as between MP–TA and SU–RF was significant. It is therefore concluded that MP–TA is the best choice for obtaining robust NWR response rates in an experiment including repetitive elicitation. This information is crucial because the NWR response rate determines the minimum number of stimulations that are necessary for meaningful results.

4.1. The reflex magnitude may depend on experimental and analytical parameters

In the present study, no significant increase of the NWR magnitude was observed when a noxious heat stimulus was applied in the same dermatome as the concurrent electrical stimulation to elicit the reflex. This is in contrast to the in the original study (Ellrich and Treede, 1998) that reported an increase in NWR magnitude, which would indicate a successful induction of CS through the applied heat pain paradigm. There are several potential reasons for this discrepancy: (1) As indicated above, it is unclear whether or even unlikely that CS was induced in the present study. Although the exact same induction paradigm as in the original study was used, it is conceivable that, for some unknown reason, the participants of the present study sensitized less. (2) A Type I error in the original study or a Type II error in the present study. (3) Slight adaptations of the number and rate of the electrical stimulations. (4) Details of the statistical analysis.

Regarding (2), the results of the original study were obtained from a cohort of 11 participants, of which just 5 were used for all published analyses. These small sample sizes might have biased the results, hampering the present replication attempt that relied on 16 participants.

Regarding (3), the number and rate of the electrical stimulations to elicit the reflex can have a direct impact on the results. In the original study, 18 electrical stimulations were applied at 0.2 Hz (Ellrich and Treede, 1998). This frequency is close to the reported 0.3 Hz threshold for temporal summation of pain (TSP) which is often used to study CS with the help of the NWR (Price, 1972; Terry et al., 2001; Terry et al., 2016; Rhudy et al., 2020; Arendt-Nielsen et al., 1994; Serra et al., 2004; Sandrini et al., 2005). TSP, a phenomenon similar to wind-up in animal studies, is an increase in perceived pain over time during repetitive noxious stimulations of the same intensity (Price, 1972; Price et al., 1977).

TSP has been reported to cause earlier onsets of the NWR (Arendt-Nielsen et al., 1994; Terry et al., 2001), which is normally observed in the 60 to 150 ms post-stimulation interval. This would explain why in the original study reflex responses were reported for the 50 to 100 ms post-stimulation time window (Ellrich and Treede, 1998). However, a shortening of onset latencies for the activation of Aβ and C fibres is not considered characteristic for CS processes as longer or unchanged latencies have been reported in CS studies (Baek et al., 2016; Liang et al., 2016; De Schoenmacker et al., 2021). It seems therefore unlikely that the original study observed CS and more likely that they observed TSP. In the present study, a potential influence of TSP was avoided by reducing the number of electrical stimulations to eight, delivered at random interstimulation intervals of 5 to 15 s (see supplementary material A for all changes to the original design).

Alternatively, it should be noted that, using stimulations on the foot sole, onset latencies as early as 60 ms post-stimulation are reported for the activation of nociceptive Aβ fibres that defines the NWR. In this context, the 60 ms post-stimulation moment has been characterized as a natural separation point between non-nociceptive and nociceptive components visible in the sEMG signal (Andersen, 2007). Therefore, a time window of 50 to 100 ms post-stimulation could theoretically contain the response of activated Aβ fibres, i.e. the NWR, even in the absence of TSP. Conversely, it seems unlikely that the observed reflex responses in the original study were predominantly non-nociceptive, although they were interpreted as such in the original article, because the published traces show no activity before 60 ms (Ellrich and Treede, 1998).

Regarding (4), the NWR magnitude was quantified in the original study using the average AUC of the obtained reflexes (Ellrich and Treede, 1998). However, the AUC might not be the best criterion to quantify and analyse the reflex magnitude. Because AUC has been reported to be significantly worse for reflex identification than Z or D (Rhudy and France, 2007; France et al., 2009) it is conceivable that quantification and analysis using AUC misrepresents the actual muscle response activity. This would explain the striking difference in the present replication study between the evolution of the AUC from the lowest to the highest temperature in comparison to D, with AUC increasing and D decreasing (Fig. 5). For future studies investigating the viability of the NWR magnitude as a CS proxy, it is therefore recommended not to use AUC as a quantification because it could distort the results. Relying on a more robust metric for quantification, such as D, might improve reproducibility of findings relating to subtle differences in NWR response.

4.2. One muscle drives the withdrawal action for each stimulation site

The withdrawal action to electrical stimulation was always most reliable in one of the three muscles of which responses were recorded. For MP stimulation it was the TA, and for SU stimulation the BF muscle. This pattern is in line with other published results that muscle activation depends on the stimulation site, indicative of withdrawal strategies (Massé-Alarie et al., 2019), and justifies the use of these specific two combinations. It can be excluded that it was due to the specifics of the sEMG recordings because the muscle with the most reliable response changed between stimulation sites although the recording specifics remained identical. The importance of one particular muscle suggests that the withdrawal action is typically driven by this specific muscle indicating that the NWR is organized in a hierarchical pattern on the spinal level. This interpretation is in line with proposed hypotheses regarding hierarchical muscle activation strategies (Tresch and Jarc, 2009; Pierrot-Deselligny and Burke, 2012).

The purpose of such a pattern would be to prevent any potential harm to the limb by moving the innervation sites (i.e. the sites of nociceptive input in a naturalistic situation) of the electrically stimulated nerve away. The muscle that dominates the withdrawal action should be the one most effectively removing the limb. For a
stimulus on the foot sole (as in the case of an electrical MP stimulation), a contraction of the TA muscle results in dorsiflexion and inversion of the foot (Junjea and Hubbard, 2021), thereby ensuring removal of the foot sole from the stimulus. Likewise, in the case of SU stimulation at the ankle, contracting the BF muscle achieves the same result; the knee is flexed with the ankle region and the lateral side of the foot being immediately removed from their current position.

The other muscles that are involved in the withdrawal would only assist the dominating muscle in moving the affected limb away from the nociceptive input. In the case of MP stimulation, the involvement of the two other muscles that were observed (BF and RF) was limited. Their contribution was only detected in approximately half of all elicited reflex responses, probably because the contraction of the dominating TA is sufficient to move the foot away. The muscles of the thigh, i.e. BF and RF, might in the instance of a nociceptive stimulus on the sole of the foot only subserve stabilization. Thus, their activation would be expected to be more variable. Interestingly, the contribution of BF and RF increased with increasing stimulation currents (see supplementary material C), presumably because higher currents signal greater potential harm. Greater harm in turn warrants greater supraspinal control, which has been suggested to be increased for proximal muscles such as BF and RF (Jure et al., 2019). Therefore, the whole limb is moved with the help of these two muscles.

In the case of SU stimulation, the dominating BF was most of the time assisted by the RF in the withdrawal. The NWR response rates of these two muscles were not statistically different for suprathreshold stimulations. This could be due to the complementary nature of their movements regarding knee and hip. Specifically, the BF acts as a hip extensor and a knee flexor while the RF is a hip flexor and a knee extensor (Darby and Cramer, 2014). Thus, in the withdrawal, the dominating action of the BF is probably knee flexion and of the RF hip flexion to effectively withdraw the lower limb from nociceptive input. In contrast, the TA is significantly less activated (see supplementary material C), probably because dorsi-flexion of the foot does not aid to move the lateral side of the foot, i.e. the innervation territory of the sural nerve, away from the putative nociceptive source.

The above considerations are limited by focussing on individual muscle responses to electrical stimulation of nerve trunks. For an in-depth assessment of the underlying NWR pattern involving multiple muscle activations other approaches, such as muscle synergy analyses (Jure et al., 2019), and using alternative stimulation and recording site combinations would be more suitable. However, those kind of approaches would have been beyond the scope of the present study, which was primarily a replication attempt.

4.3. The MP–TA combination is not always the best choice

The MP–TA stimulation and recording site combination showed significantly higher NWR response rates than the SU–BF and SU–RF combinations. In addition, participants typically consider suprathreshold stimulations at MP less painful than at SU (Jensen et al., 2015) as can be observed in the collected pain ratings (see supplementary material D). Still, MP–TA might not always be the combination of choice for NWR studies, for two main reasons.

First, stimulating the MP nerve on the foot sole is not straightforward. The ideal position for the stimulation electrode is proximal to the medial side of the ball of the foot, i.e. close to where the arch of the foot ends and the ball begins. This region is often covered by calloused skin, the extent of which may depend on foot anatomy, walking patterns, and skin properties. Because of its composition and thickness, calloused skin influences the electrical conductivity and thus the nociceptive input. The skin would have to be peeled off before an experiment in order to achieve comparability between subjects. This would result in a lengthy preparatory procedure and potentially cause sensitization of the testing area. In contrast, stimulating the SU nerve does not require any preparation except for identifying the SU by manual palpation before placing the stimulation electrodes over its retromalleolar pathway.

Second, the MP is a mixed sensory and motor nerve, whereas the SU is a purely sensory nerve. Electrical stimulation of a mixed nerve can cause noticeable motor responses, such as toe twitching in case of a MP stimulation. These motor responses may confuse the participant and prompt voluntary muscle responses, which in turn may contaminate the EMG recordings.

Additionally, the difference in NWR response rates (79.4% for MP–TA compared to 66.9% for SU–BF), albeit significant, is not big enough to be the sole decisive factor when choosing the stimulation and recording site combination in a clinical study. Depending on the research question, it might be advisable to choose SU–BF instead of MP–TA, e.g., if a mixed nerve stimulation is to be avoided. If multiple muscle responses are to be recorded simultaneously, the choice of stimulation site should guarantee high NWR response rates in all recorded muscles. In the present study, SU stimulation yielded better results than MP when all three muscles, BF, RF, and TA, were considered.

Using the SU–RF combination might be a viable alternative to both MP–TA and SU–BF. If only the RF reflex responses are recorded, the NWR threshold \( I_0 \) should be determined at this muscle. That might yield even higher NWR response rates than in the present study, where for all SU stimulations the \( I_0 \) of the BF was used.

4.4. Clear methods’ sections support replicability

The present study attempted to replicate results that had been published over 20 years ago (Ellrich and Treede, 1998), based on the published, i.e. publicly available information. However, in certain instances, such as threshold determination, exact currents used for electrical stimulation, or number of experimental runs, this information proved to be unclear or incomplete. Hence, some potential changes to the original design were inevitable (see Methods and supplementary material A). All changes were made after carefully considering how they might affect the results of the replication attempt. Overall, the changes have presumably not affected the outcome of the replication as a whole although a small influence cannot be fully excluded.

For replication studies, it is of paramount importance that the methods published in the original study are precise and unequivocal. Otherwise, results of previous studies cannot be completely understood and evaluated by different researchers at a later date. And, it would become virtually impossible to scrutinize historic results, that may form part of the canon literature in a certain field, against the backdrop of new knowledge that might have emerged over time. Thus, a just comparison could not be made. In the present case, the understanding of CS and of the NWR has greatly advanced over the past two decades (Quesada et al., 2021; Arendt-Nielsen et al., 2018; Sandrini et al., 2005; Andersen, 2007; Jure et al., 2020). Because of these advances, it was worthwhile to re-examine the original results (Ellrich and Treede, 1998) in order not just to confirm or dismiss them but to also put them into context with the new knowledge in the field of CS and NWR studies. Unfortunately, such an assessment was not fully possible due to the aforementioned incompleteness of the published methodology.

The results from the present replication attempt would discourage the use of the published paradigm for CS studies although a final verdict cannot be given because small changes to the original design had to be made. Small changes can still impact the results of potential CS proxies, if the expected modulation could be very sub-
tle, such as in the case of a modulation of the NWR magnitude. Therefore, methods in studies on CS involving the NWR magnitude as an outcome measure need to be clearly and precisely reported.

5. Conclusion

Tonic heat stimuli that were concurrently applied in the same dermatome as the electrical stimulations to elicit the NWR did not significantly alter the NWR magnitude in the present study despite closely reproducing a previous design that reflected CS (Ellrich and Treede, 1998). The viability of NWR magnitude changes as a CS proxy after induction with painful heat should only be assessed if the induction can be independently confirmed, e.g., by the presence of secondary hyperalgesia. For future replicability, methods of CS studies involving the NWR need to be reported in detailed and precise fashion.

The NWR can be reliably elicited in a repeated measures experiment. Presently, the MP–TA stimulation and recording site combination yielded the highest response rates for the NWR. If simultaneous sEMG read-outs in more than one muscle are used, the SU stimulation site appears to be the superior option.

Funding

Primary funding (as part of a project grant to PS): Swiss National Science Foundation, Bern, Switzerland.

Additional funding for equipment and materials: Clinical Research Priority Program PAIN, University of Zurich, Zurich, Switzerland; AO Foundation, Chur, Switzerland; ETH Foundation, Zurich, Switzerland.

Authors contributions

AG, ACG, MH, MS, and PS contributed to the study design, literature research and writing of the manuscript; AG and ACG carried out the experiments and data collection; AG and PS carried out the statistical analysis and interpretation; PS obtained the ethical approval for the study.

Conflicts of interest

None of the authors have potential conflicts of interest to be disclosed.

Acknowledgements

The authors would like to thank Dr. Jan Rosner for his valuable input and support.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.clinph.2022.09.011.

References

Andersen OK. Studies of the organization of the human nociceptive withdrawal reflex: Focus on sensory convergence and stimulation site dependency. Acta Physiol 2007;189:1–35.

Andersen OK, Felsby S, Nicolaisen L, Bjerre P, Jensen TS, Arendt-Nielsen L. The effect of Ketamine on stimulation of primary and secondary hyperalgesic areas induced by capsaicin–a double-blind, placebo-controlled, human experimental study. Pain 1996;66(1):51–62.

Arendt-Nielsen L, Bremann J, Sindrup S, Bak P. Electrophysiological and psychophysical quantification of temporal summation in the human nociceptive system. Eur J Appl Physiol Occup Physiol 1994;68(3):266–73.

Arendt-Nielsen L, Morlion B, Perrot S, Dahan A, Dickenson A, Kress H, Wells C, Bouhassira D, Mohr Drewes A. Assessment and manifestation of central sensitisation across different chronic pain conditions. Eur J Pain 2018;22(2):216–41.

Baek S-H, Seek HY, Koo YS, Kim B-J. Lengthened cutaneous silent period in fibromyalgia suggesting central sensitization as a pathogenesis. Pain Res One 2016;11(2):e140248.

Bjerre L, Andersen AT, Hagelskjær MT, Ge N, Mørch CD, Andersen OK. Dynamic tuning of human withdrawal reflex receptive fields during cognitive attention and distraction tasks. Eur J Pain 2011;15(8):816–21.

Cohen J. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates; 1988.

Curatolo M, Muller M, Ashraf A, Neziri AY, Streiterberg K, Andersen OK, Arendt-Nielsen L. Pain hypersensitivity and spinal nociceptive hypersensitivity in chronic pain: prevalence and associated factors. Pain 2015;156(11):2373–82.

Darby S, Cramer C. General anatomy of the spine, spinal cord, and ans. St. Louis, Mosby; 2014.

De Schonmaker I, Berry C, Bloun J-S, Rosner J, Hübli M, Jutzelter CR, Kramer JL. An intensity matched comparison of laser-and contact heat evoked potentials. Sci Rep 2021;11(1):1–12.

Dimitrijević MR, Faganel J, Gregoš M, Nathan P, Trontelj J. Habituation: effects of regular and stochastic stimulation. J Neurol Neurosurg Psychiatry 1972;35(2):234–42.

Downman R. Spinal and supraspinal correlates of nociception in man. Pain 1991;45(1):269–81.

Downman R. Possible startle response contamination of the spinal nociceptive withdrawal reflex. Pain 1992;49(2):187–97.

Downs MB, Laporte C. Conflicting dermatome maps: educational and clinical implications. J Orthop Sports Phys Ther 2011;41(6):427–34.

Edwards L, Ring C, France CR, Alabi M, McIntyre D, Carroll D, Martin U. Nociceptive flexion reflex thresholds and pain during rest and computer game play in patients with hypertension and individuals at risk for hypertension. Biol Psychol 2007;76(1–2):72–82.

Ellrich J, Treede R-D. Convergence of nociceptive and non-nociceptive inputs onto spinal reflex pathways to the tibialis anterior muscle in humans. Acta Physiol Scand 1998;163(4):391–401.

Fleiss JL, Levin R, Paik MC. Statistical methods for rates and proportions. John wiley & sons; 2003.

France CR, Rhuday PJ, McGlone S. Using normalized EMG to define the nociceptive flexion reflex (NFR) threshold: further evaluation of standardized NFR scoring criteria. Pain 2009;145(1–2):211–8.

French DJ, France CR, France JL, Arnott LF. The influence of acute anxiety on assessment of nociceptive flexion reflex thresholds in healthy young adults. Pain 2005;114(3):358–63.

Genschel U, Becker C, Schließende Statistik: Grundlegende Methoden. Springer-Verlag; 2006.

Hugon M. Exteroceptive reflexes to stimulation of the sural nerve in normal man. Human Reflexes, Pathophysiology of Motor Systems, Methodology of Human Reflexes 1973;vol. 3. Karger Publishers; 1973. p. 713–29.

Jensen MB, Manresa JAB, Frahm KS, Andersen OK. Analysis of muscle fiber conduction velocity enables reliable detection of surface EMG crosstalk during detection of nociceptive withdrawal reflexes. BMC Neurosci 2013;14(1):1–11.

Jensen MB, Manresa JR, Andersen OK. Reliable estimation of nociceptive withdrawal reflex thresholds. J Neurosci Methods 2015;253:110–5.

Jenessa P, Hubbard JB. Anatomy, bony pelvis and lower limb, tibialis Anterior Muscles. StatPearls [Internet] 2021.

Jure FA, Arguissain FG, Biurrun Manresa JA, Graven-Nielsen T, Andersen OK. Conditioned pain modulation. Pain Res Manag 2014;19(2):93–6.

Juneja P, Hubbard JB. Anatomy, bony pelvis and lower limb, tibialis Anterior Muscles. StatPearls [Internet] 2021.

Kohler JD, Treede R-D. The Kyoto protocol of IASP basic pain terminology. Pain 2005;114(3):358–63.

Kostrzewa RM, Photopoulo J, Dohrenwend BS. Research Design and Data Analysis in Social Science: An Empirical Approach. New York: McGraw-Hill; 1986.

Kostrzewa RM, Photopoulo J, Dohrenwend BS. Research Design and Data Analysis in Social Science: An Empirical Approach. New York: McGraw-Hill; 1986.

Kostrzewa RM, Photopoulo J, Dohrenwend BS. Research Design and Data Analysis in Social Science: An Empirical Approach. New York: McGraw-Hill; 1986.

Kostrzewa RM, Photopoulo J, Dohrenwend BS. Research Design and Data Analysis in Social Science: An Empirical Approach. New York: McGraw-Hill; 1986.
Manresa JAB, Neziri AY, Curatolo M, Arendt-Nielsen L, Andersen OK. Test–retest reliability of the nociceptive withdrawal reflex and electrical pain thresholds after single and repeated stimulation in patients with chronic low back pain. Eur J Appl Physiol 2011;111(1):83–89.

Manresa JAB, Nguyen GP, Curatolo M, Moeslund TB, Andersen OK. Probabilistic model for individual assessment of central hyperexcitability using the nociceptive withdrawal reflex: a biomarker for chronic low back and neck pain. BMC Neurosci 2013;14(1):1–12.

Massé-Alarie H, Salomoni SE, Hodges PW. The nociceptive withdrawal reflex of the trunk is organized with unique muscle receptive fields and motor strategies. Eur J Neurosci 2019;50(2):1932–47.

Meinck H-M, Küster S, Benecke R, Conrad B. The flexor reflex-influence of stimulus parameters on the reflex response. Electroencephalogr Clin Neurophysiol 1985;61(4):287–98.

Neziri AY, Andersen OK, Petersen-Felix S, Radanov B, Dickenson AH, Scaramozzino P, Arendt-Nielsen L, Curatolo M. The nociceptive withdrawal reflex: normative values of thresholds and reflex receptive fields. Eur J Pain 2010;14(2):134–41.

Pierrot-Deseilligny E, Burke D. The circuitry of the human spinal cord: its role in motor control and movement disorders. Cambridge University Press; 2012.

Price DD. Characteristics of second pain and flexion reflexes indicative of prolonged central summation. Exp Neurol 1972;37(2):371–87.

Price DD, Hu JW, Dubner R, Gracely RH. Peripheral suppression of first pain and central summation of second pain evoked by noxious heat pulses. Pain 1977;3(1):57–68.

Quesada C, Kostenko A, Ho I, Leone C, Nochi Z, Stouffs A, Wittayer M, Caspani O, Finnerup NB, Mouraux A, et al. Human surrogate models of central sensitization: A critical review and practical guide. Eur J Pain 2021;25(7):1339–49.

Rhudy JL, France CR. Reliability of the nociceptive flexion reflex (NFR) threshold in human participants: a comparison of different scoring criteria. Pain 2007;128(3):244–53.

Rhudy JL, France CR. Reliability and validity of a brief method to assess nociceptive reflex response to pain: further validation of the nociceptive flexion reflex (NFR) as a measure of nociception using multilevel modeling. Psychophysiology 2009;46(3):939–48.

Rhudy JL, Green BA, Arnaud RC, France CR. Taxometric analysis of biceps femoris EMG following electrocutaneous stimulation over the sural nerve: Determining the latent structure of the nociceptive flexion reflex (NFR). Int J Psychophysiol 2008;69(1):18–26.

Rhudy JL, Lannon EW, Kuhn BL, Palit S, Payne MF, Sturycz CA, Hellman N, Güereca YM, Toledo TA, Huber F, et al. Assessing peripheral fibers, pain sensitivity, central sensitization, and descending inhibition in Native Americans: main findings from the Oklahoma Study of Native American Pain Risk (OK-SNAP). Pain 2020;161(2):388.

Sandrini G, Arrigo A, Bono G, Nappi G. The nociceptive flexion reflex as a tool for exploring pain control systems in headache and other pain syndromes. Cephalalgia 1993;13(1):21–7.

Sandrini G, Serrao M, Rossi P, Romaniello A, Cruccu G, Willer JC. The lower limb flexion reflex in humans. Prog Neurobiol 2005;77(6):353–95.

Serrao M, Rossi P, Sandrini G, Parisi L, Amabile GA, Nappi G, Pierelli F. Effects of diffuse noxious inhibitory controls on temporal summation of the RIII reflex in humans. Pain 2004;112(3):353–60.

Sklarevski V, Ramadan N. The nociceptive flexion reflex in humans. Pain 2002;96(3).

Terry EL, France CR, Bartley EJ, DelVentura JL, Kerr KL, Vincent AL, Rhudy JL. Standardizing procedures to study sensitization of human spinal nociceptive processes: comparing parameters for temporal summation of the nociceptive flexion reflex (TS-NFR). Int J Psychophysiol 2011;81(3):263–74.

Terry EL, Thompson KA, Rhudy JL. Does pain catastrophizing contribute to threat-evoked amplification of pain and spinal nociception? Pain 2016;157(2):456–65.

Treich MC, Jarc A. The case for and against muscle synergies. Curr Opin Neurol 2009;19(6):601–7.

Wallwork SB, Grabber L, O’Connell NE, Catley MJ, Moseley GL. Defensive reflexes in people with pain—a biomarker of the need to protect? A meta-analytical systematic review. Rev Neurosci 2017;28(4):381–96.

Willer JC. Comparative study of perceived pain and nociceptive flexion reflex in man. Pain 1977;3(1):69–80.

Willer JC, Bourbeau F, Albe-Fessard D. Supraspinal influences on nociceptive flexion reflex and pain sensation in man. Brain Res 1979;179(1):61–8.

Woolf CJ. Central sensitization: implications for the diagnosis and treatment of pain. Pain 2011;152(3):S2–S15.