medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

HOW TO MAKE CALIBRATION LESS PAINFUL – A PROPOSITION OF AN AUTOMATIC, RELIABLE AND TIME-EFFICIENT PROCEDURE.

Karolina Świder[1], Ricardo Bruña[2,3], Stephan Moratti[1]

Short title: HOW TO MAKE CALIBRATION LESS PAINFUL

1 Department of Experimental Psychology, Psychology Faculty, Universidad Complutense de Madrid
2 Department of Radiology, Universidad Complutense de Madrid; IdISSC, Spain
3 Centre for Cognitive and Computational Neuroscience (C3N), Universidad Complutense de Madrid, Spain

The corresponding author: Karolina Swider, tel. +34623106406, email: kswider@ucm.es

The manuscript category: original article

Funding Sources: This work was supported by H2020 EU-funded Marie Skłodowska-Curie Individual Fellowship (MSCA-IF), Grant agreement ID: 896262.

Conflicts of interest: None declared

Key words: experimental pain, calibration, electric stimuli, linear model, gender

SIGNIFICNCE: We propose an automatic calibration method that ensures stimuli-to-pain ratings linearity and provides means to quantify its correctness, allowing the choice of an optimal stimulation value. We propose metrics on convergence, R^2, and maximum stimuli current as calibration rejection criteria. Finally, for the first time we indicate the sex differences in sensation threshold and pain thresholds for intraepidermal electrical stimulation via WASP electrodes.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Abstract

Background: The description of the calibration in pain studies is scarce. This study provides an in-depth analysis and proposes an automatic pain-threshold calibration that quantifies its correctness, minimizes its duration and invasiveness, and considers validation of linearity between stimuli intensity and Numerical Rating Scale (NRS) values.

Methods: 70 healthy volunteers (37 males) underwent a staircase calibration method adjusted for A-delta specific intraepidermal electrical stimulation. The sensation (*t*) and pain thresholds (*T*) intensities were calculated using: the traditional threshold method (TM), a linear regression model (LRM), and two variants of a proposed truncated LRM (tLRM). The tLRM analyses data up to an optimal point, chosen according to the maximum goodness-of-fit value (R^2). The optimal value of R^2 and the convergence based on the regression line gradient formed the base for rejection criteria of failed procedure. Then, we compared the different methods for thresholds assessment as well as mid-painful stimulation (*MP*), equivalent to $1.5*T$ or NRS equal to 8, using repeated measures ANOVA.

Results: Four participants were rejected according to our criteria. tLRM methods exhibit higher R^2 than LRM with 36% less stimuli application. Compared to other methods, tLRM result in significantly higher *t* and lower *T* and *MP* intensities. Last, significantly lower *t* and *T* were found for female compared to men, regardless of the calibration method.

Conclusions: The proposed method of calibration reduces the number of stimuli delivered and can be easily adapted to other simulation modalities. However, future investigation is recommended using adaptive LRM fitting and verification of reliability.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

1. Introduction

Pain perception is a subjective experience, influenced by a combination of psychosocial, affective, cognitive, and biological factors (Bushnell et al., 2013; Tracy, 2017). Despite of attempts of finding alternative methods (Gruss et al., 2019; Lautenbacher et al., 2022; Lundeberg et al., 2001; Wang et al., 2022), self-reports continue to be the most reliable method for evaluation of pain. In laboratory and clinical settings, e.g., pain studies where experimental pain models are used to test analgesic compounds, standardized stimuli of different modalities are required to produce controlled pain sensation (Olesen et al., 2012). Such studies require a calibration procedure to map stimuli intensity to individual perception using one-dimensional scales, that are not free from certain limitations and weaknesses (Wagemakers et al., 2019). There are concerns about the doubtful linearity of the pain scores, hampering the use of mathematical calculations (Lazaridou et al., 2019), and about forcing translations of complex non-linear perception onto a linear scale (Berger and Baria, 2022). The reach of these limitations is inconclusive, as some studies show that numerical rating scales (NRS) behave as a ratio scale (Van Der Heide et al., 2009; Oliveira et al., 2014), while others contradict it (Hartrick et al., 2003; Price et al., 1994).

We propose an automatic calibration method, with metrics for quantifying its accuracy and reliability. We compare the *T* and *t*, both computed via classical approaches as the threshold-based method (TM) and via a family of linear regression methods (LRM), with the objective of choosing the method that minimizes the number of painful stimuli applicated during the calibration stage. While LRM is based on all data, the proposed truncated LRM (tLRM) discards painful stimuli according to a set of conditions to insure higher degree of linearity. It is a measure to avoid a decrease in pain-related response as a reaction to repetitive painful stimulation (i.e., habituation) (De Paepe et al., 2019). Increasing intensity of pain stimulation in TM and LRM might lead to stimuli habituation, and therefore may introduce non-linearities to the NRS ratings, resulting in an overestimation of the *T* intensity. For this reason, we predict that *T* values calculated using TM and LRM will be higher than ones derived from the proposed tLRM. Furthermore, in laboratory studies it is common to set the stimulation intensity using a function of *T* (Bąbel et al., 2017; Colloca et al., 2010) or taking scale scores higher than *T* (Rütgen et al., 2015; Świder et al., 2017). Here we will present a comparison of mid-painful stimulation calculated using our linear regression family and discuss the results.

Finally, sex differences in pain sensitivity are not always consistent across experimental pain modalities (Hashmi and Davis, 2014). Despite this ambiguity, the prevailing opinion is that women are characterized by higher sensitivity in detecting all modalities of perceptual stimuli including pain and touch (Fillingim and King, 2009). We will compare sex differences in both thresholds for Adela-specific electrodes, something that, to the best of our knowledge, has not been done before.

2. Methods

2.1. Participants

In total, 70 healthy volunteers (37 males; age mean±SD: 25.47±5.34) participated in the calibration phase of a magnetoencephalography (MEG) study, which description and results are beyond the scope of the present work. From this original sample, 4 participants were excluded based on the results of the calibration stage (see “Data analysis” section below), and therefore the final sample for this work consists of 66 individuals (35 males; age between 18 and 36, mean±SD: 25.62±5.34).

Participants were recruited by an advertisement on the project website (https://neuroconmsca.wordpress.com/) from students and staff of the Complutense University of Madrid (UCM), the Centre for Biomedical Technology in Madrid (CTB), and the Basque

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Center on Cognition, Brain and language (BCBL) in San Sebastian. With respect to the main MEG pain study, there were specific exclusion criteria which include: age below 18 and above 36 years, pregnancy, chronic diseases including chronic pain or migraine, recurrent pain, neurological or psychiatric diseases, heart disease, repeated unconsciousness, external and internal tissue damage, use of any type of medication or drug (psychoactive medication/substances such as antidepressants, antiepileptics, antipsychotics or illegal drugs), family history of epilepsy/photic epilepsy episode, claustrophobia, left-handedness, implantation of metal elements (e.g. endoprostheses, implants, metallic staples) and active implant (e.g.: pacemaker, neurostimulator, insulin pump, ossicle prosthesis), metal wire behind teeth and tattoos. Additionally, only participants who scored less than 8 on *Hospital Anxiety and Depression Scale* (HADS) (Zigmond and Snaith, 1983) and did not undergo a magnetic resonance study for 48 hours before the experiment could sign up to the study. All participants were encouraged to ask questions, were informed that they could resign form participation in any moment of the study and were guaranteed to receive 10 euros as compensation. All volunteers signed a written informed consent form.

2.2. Stimulation hardware

Fig. 1 shows the main setup for the procedure.

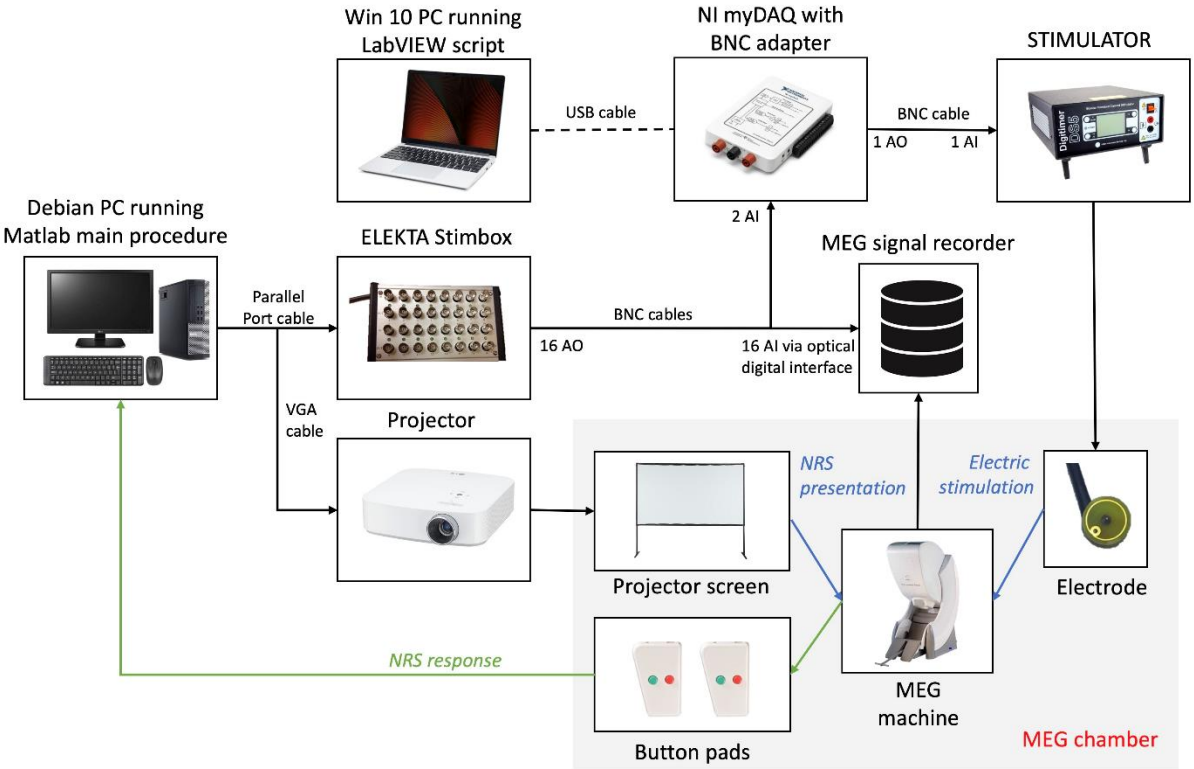

Figure 1. Laboratory set up used. Presented set up was used at CTB. Note that for the 24 participants that participated in the study at BCBL, Vipxxx system (https://vpixx.com/) for stimulation was used instead of StimBox, and Windows computer was used instead of Debian 10 system.

The stimulation was controlled by a PC running on Debian Linux using a script written in Matlab (Mathworks©), using the Psychophysics Toolbox extensions (Brainard, 1997). This setup allowed to synchronize the computer, the MEG system and the stimulation device. Painful stimulation was controlled via the Elekta Stimulus Trigger Interface (STI102) StimBox, which is equipped with sixteen binary input/output channels (BNC sockets) which generate a set of sixteen 5V analogue signals. We used two channels to control the stimulation

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

strength. The complete stimulation hardware setup included: (1) PC with Windows 10 and LabVIEW, (2) NI myDAQ with Florida Research Instruments Inc. myDAQ BNC adapter for x10 oscilloscope probes with connectors, (3) Digitimer DS5 Bipolar Constant Current Stimulator, (4) WASP electrodes with connection cables, and (5) 3 BNC cables. The complete setup is shown in the Fig. 1 and further described below.

2.2.1. Electrodes

We used WASP electrodes (Brainbox, United Kingdom), which selectively activate Adelta nociceptors, with very small stimulation intensity, approximately 0.01–2.0 mA (Seidel et al., 2015). WASP electrodes have two gold-plated solder pads to connect wires and a platinum pin at the top side of the electrode. The platinum pin is mounted to project approximately 0.5 mm from the surface of the board ±0.25mm. By pressing the electrode, the needle tip (0.2 mm) is inserted into the skin, adjacent to the free nerve endings of the thin myelinated fibres in the epidermis and superficial part of the dermis (Lefaucheur et al., 2012). The insertion of the electrode pin electrode causes no bleeding nor visible damage. The stimuli strength is set according to temporal summation of a long continuous duration (e.g. up to 500 ms) (Rütgen et al., 2015) or pulse train (Inui and Kakigi, 2012) of fixed intensity. For this reason, WASP electrodes have a lower probability of generating adverse secondary effects like skin irritation. To adapt WASP electrodes for the MEG environment, a specific nonmagnetic cable was attached (Kupferschaltlitze versilbert MX 250v 7 drähtig) to the electrode by its manufacturer (Brainbox team). This cable was then connected to the 10 m extender cable that was connected to the stimulator's output socket. During calibration, two WASP electrodes were attached to the dorsum of the left hand (Seidel et al., 2015) with a spacing of 8 cm, but only one electrode was active and used to delivered 300-ms stimulations of ascending and descending intensities.

2.2.2. Stimulator

Although there is a myriad of medical grade electric stimulator types, we used the Digitimer DS5 Bipolar Constant Current Stimulator, as it is possible to control its output via analogue voltage input. The stimulator was placed outside of the MEG chamber to avoid electromagnetic interferences. A sufficiently long BNC cable was used to connect its voltage input socket to the NI myDAQ.

2.2.3. NI myDAQ

The stimulation PC used in this work, combined with the StimBox of the MEG system, was able to provide TTL (transistor-to-transistor logic) signals of 5V. However, the DS5 stimulator requires an analogue input for controlling the stimulation intensity. We used a NI myDAQ Data Acquisition Device (National Instruments, Austin, TX, USA) in combination with a myDAQ BNC adapter (Florida Research Instruments Inc., Cocoa Beach, FL, USA) as a drive, accepting these TTL values and generating a varying voltage output. The driver accepted two TTL signals, connected through its BNC inputs, for setup, and generated an analogue signal, connected to its BNC output, to control the intensity level of the DS5. The correct functioning of this setup was checked using a digital voltmeter measuring the analogue output of the myDAQ.

2.2.4. PC with Windows 10 and LabVIEW

NI myDAQ was controlled via USB by a PC running a LabVIEW (National Instruments, Austin, TX, USA) script. LabVIEW is a system-design platform and development environment for visual programming. Execution flow is determined by the structure of graphical block diagrams. The script was running continuously, listening to the input channels at 1 kHz sampling rate and generating the requested output, if any. Table 1 describes the action of the script given the input it received.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Table 1. Input signal and its corresponding action performed by the LabVIEW script.

Input	**Signal duration**	**Action**
Signal coming from ai0 channel only	0,5 sec	Increase the output voltage by 1mV (equivalent to increasing stimulation current by 0.01 mA)
	1,0 sec	Increase the output voltage by 2mV (equivalent to increasing stimulation current by 0.02 mA)
	1,5 sec	Increase the output voltage by 5mV (equivalent to increasing stimulation current by 0.05 mA)
Signal coming from ai1 channel only	0,5 sec	Decrease the output voltage by 1mV (equivalent to decreasing stimulation current by 0.01 mA)
	1,0 sec	Decrease the output voltage by 2mV (equivalent to decreasing stimulation current by 0.02 mA)
	1,5 sec	Decrease the output voltage by 5mV (equivalent to decreasing stimulation current by 0.05 mA)
Signal coming simultaneously from both ai0 and ai1 channels	0,3 sec	Send electrical stimulation of the previously selected intensity and 300 ms of duration.

2.3. Procedure

The participants were asked to enter the magnetically shielded room, take a seat in the MEG chair, facing a projector screen 50 cm away, and have their hands placed on a board in front of them. The experimenter briefed them that the objective of a calibration procedure was to individually determine the stimulation intensities that would be used during the main experiment, and that at that moment no MEG measurements were taken. They were informed that only one of the electrodes placed in their hand would be used during the calibration procedure, that they would receive multiple increasing and decreasing intensities of electrical stimuli starting from 0mA, and that they would need to score their sensation according to an 11-point NRS using two 2-button response pads (Current Design, USA) (see Fig. 1).

During this briefing, the operator prepared the skin of the participant (the impedance of the skin was reduced by cleaning the skin with a cloth soaked in alcohol which eliminated lipids and dead cells form the skin surface), and attached the WASP electrodes to the dorsum of the left hand (Seidel et al., 2015). After that, the participants were left alone in the MEG chamber, and assured that the experimenter could see them and was in verbal contact during all calibration procedure. The procedure began with the NRS practice exercise. Thereby, the participants learned how to use 2-Button Response Pads to choose the correct number on the NRS. First, they were provided with written description of the exercise on the projector screen and then they were given 4 practice trials during which they had to choose numbers (2, 5, 7, 4) on the scale. If they made a mistake, the trial was repeated.

After the exercise, the participants underwent the calibration procedure which was completely automatic – no intervention from the experimenter was needed. The procedure applied electrical stimuli according to the timeline presented in Fig. 2a, and followed the well-known staircase method, shown in detail in Fig. 2b.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

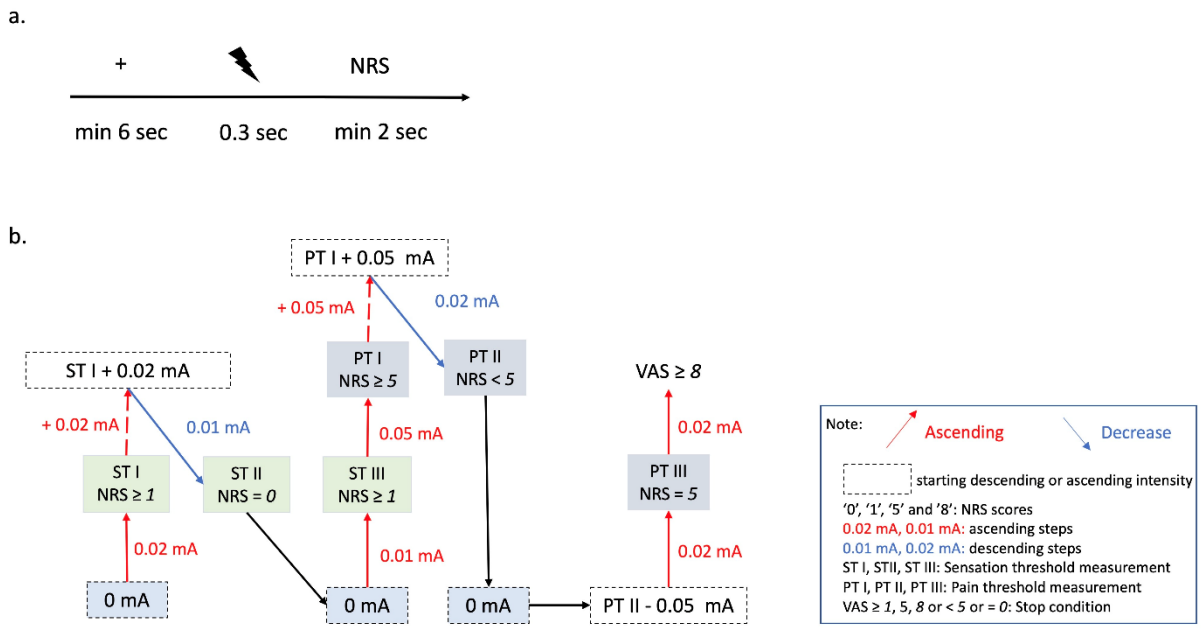

Figure 2. Calibration procedure. A. calibration timeline. B. Staircase procedure used in the study. Sensation (t) and pain threshold (T) intensities were established by the staircase procedure, taking an average of the three readings (ST I, STII and STIII for t and PT I, PTII and PTIII for T) corresponding to VAS rating 1 and 5, respectively.

First, the participant is presented with a fixation cross for a minimum of 6 seconds, with the higher boundary determined by the NRS response duration. Then, the electric stimulation, with the current level of intensity, was applied, indicated in Fig. 2a by a lightning bold icon. Thereafter, the participant was asked to rank the level of pain of the stimulus, if any. Finally, the current level of intensity of the stimulation was modified up or down, depending on the participant response. To reduce the probability of attenuation, and to reduce the number of steps required to achieve the t and T values, we used three different ascending/descending steps (0.01, 0.02, and 0.05, see Fig. 2b). As a result, the interstimulus interval (ISI) was kept at minimum 8 seconds, to further minimalize the probability of habituation.

After each stimulus, the participants were scoring their sensations on the 11-point NRS ranging from 0 – '*No electrical sensation*', 1 – *'I start to feel something'*, 5 – *'It starts to be painful'* to 10 – '*The strongest painful sensation imaginable'*, using the response pads. The numbers 1, 5 and 10 on the scale indicated the sensation threshold (t), the pain threshold (T), and the maximum pain tolerance, respectively. This scale resembles the one used in the studies by Świder and colleagues (Świder et al., 2017) and by Romero and collaborators (Roa Romero et al., 2013). The maximal stimulation that the participants received was the stimulation corresponding to an NRS of 8. Fig. 3 presents an example of the output of the data collected during the calibration procedure. The graph was presented to the participant and served as a visual cue to check if the data exhibited a habituation pattern.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

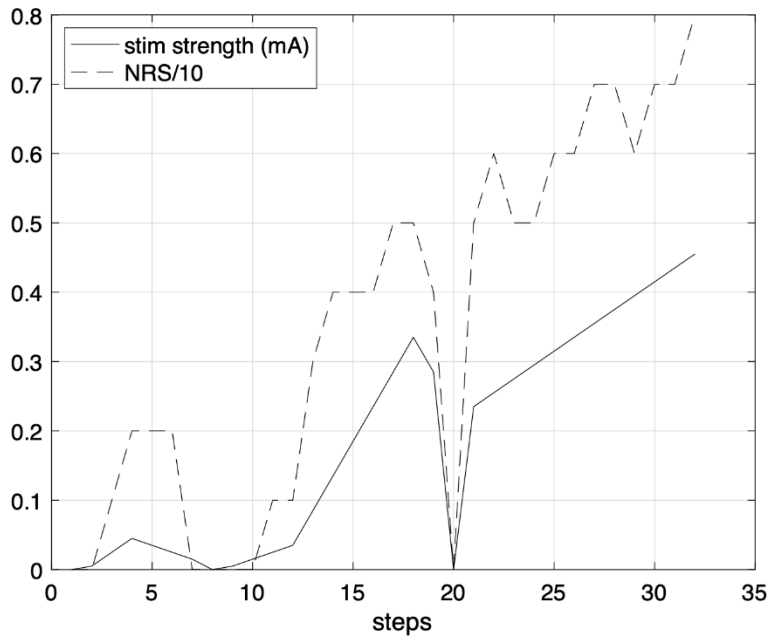

Figure 3. An example output provided at the end of the calibration procedure. Note: x-axis represents the number of steps and y-axis is the simulation current in mA (blue line) and the response NRS divided by 10 (red line).

In a post-test, 4 min after completing calibration, to checking if the correct stimuli intensities were calculated for each participant, three intensities of different strength were delivered in ascending order. Participants were asked to verbally rate stimuli intensity using the same NRS scale. The stimuli intensities were calculated based on LRM and equation: 3*t* mA for tactile, 1.5*T*+0.02 mA for low-painful (LP) and LP*1.5 mA for mid-painful (MP).

The calibration procedure design was approved by the Ethics Committee of the Universidad Complutense de Madrid (UCM), the Universidad Politécnica de Madrid (UPM), and the Basque Centre on Cognition, Brain and Language (BCBL), and followed the Helsinki Declaration and national and European Union regulations as part of a the MSCA project. None of the participants showed any signs of tissue damage nor reported genuinely adverse experiences as a result of undergoing the calibration procedure.

2.4. Calibration methods

We calculated the *t* and *T* values using four methods: threshold method (TM), linear regression method (LRM) applied to the whole set of data, and two modifications of LRM, the truncated linear regression method (tLRM).

2.4.1. Threshold method (TM)

The threshold method calculates *t* by taking an average of three stimulation intensities (ST I, ST II, ST III) rated as 1 on the NRS (see Fig 2b), a procedure used in previous pain studies (Bąbel et al., 2017; Świder and Bąbel, 2013). The first rating (ST I) was achieved in the first ascending curve (with a step 0.02 mA), selecting the first stimulation current with an NRS equal to 1. After surpassing ST I by 0.02mA, the second rating (ST II) was taken in the first descending curve, selecting the stimulation current just before the first NRS equal to 0. The third rating (ST III) was achieved in the second ascending curve, selecting again the first stimulation current corresponding to NRS equal to 1 (see Fig 2b).

An analogous calculation was performed for *T*, by taking the average of three stimulation currents (PT I, PT II, PT III) scored as 5 on the NRS (see Fig 2b). The first rating (PT I) was achieved in the second ascending curve, selecting the stimulation current

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

corresponding to an NRS equal to 5. After surpassing PT I by 0,05 mA, the second rating (PT II) was taken in the second descending curve, selecting the stimulation current just before an NRS equal to 4. The third rating (PT III) was achieved in the third ascending curve, that started from PT II minus 0.5 mA, selecting the stimulation current corresponding to an NRS equal to 5 (see Fig 2b).

2.4.2. Linear Regression Method (LRM)

We used a linear regression to model the participant response (the NRS) as a function of the stimulation intensity. We used the Matlab function *fitlm*, which fits a specified model, in our case a linear regression model, to variables in the dataset, and returns the linear model in the form of the equation $NRS = m \cdot I + c$, where I is the stimulation current, m is the gradient of the relation between I and NRS, and c is the intercept. We also obtained R^2 as a statistical measure that determines the proportion of variance in the dependent variable explained by the independent variable. In other words, it is a measure of how well the regression model explains the data and ranges between 0 (0% of data variability is explained by the model) to 1 (100% of data variability is explained by the model). It is calculated according to the following formula

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)}{\sum_i (y_i - \bar{y})},$$

where y_i is the observed dependent value, f_i is the predicted dependent value, and $\bar{y}$ is the mean of the observed dependent values.

2.4.3. Truncated Linear Regression Method (tLRM)

In this work we propose two novel calibration methods, namely tLRMm and tLRMmc, obtained by modifying the LRM method to use only a specific part of the calibration data. The difference between both methods is that tLRMm assumes that the intercept is zero (this is, an input current of 0 has a corresponding NRS of 0), while tLRMmc does not. As previously, we used the Matlab function *fitlm* to create the regression model.

In both, the linear model is obtained by applying the linear regression model not once, like in the case of LRM at the end of the calibration procedure, but iteratively when a new data point is collected, starting from the second stimulation. As a result, m, c, and R^2 are obtained as a vector of size $n-1$, where n is the total number of stimulations. Later, we truncate the vector based on the position of maximum R^2, given that calibration was run for sufficiently many steps (lower limit of optimization range). It has been empirically observed that once the participant reached the second zero current stimulation (not including the initial zero current stimulation at step 0, this is, the end of the second descending curve), i, the model delivers reliable predictions. Additionally, to reduce the number of painful stimuli, we need to stablish a maximum number of steps (upper limit of optimization range), which also restricts the procedure from running too long. For this reason, the highest pain sensation NRS rating reached during the tLRM methods will be approximately 5. In this work, the optimization range was defined between $i-r$ and $i+r$ (in our case r was chosen as 3), and the linear model where the maximum R^2 was found was returned as the optimal. Figure 4 shows an example of how the procedure chooses the truncation point.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

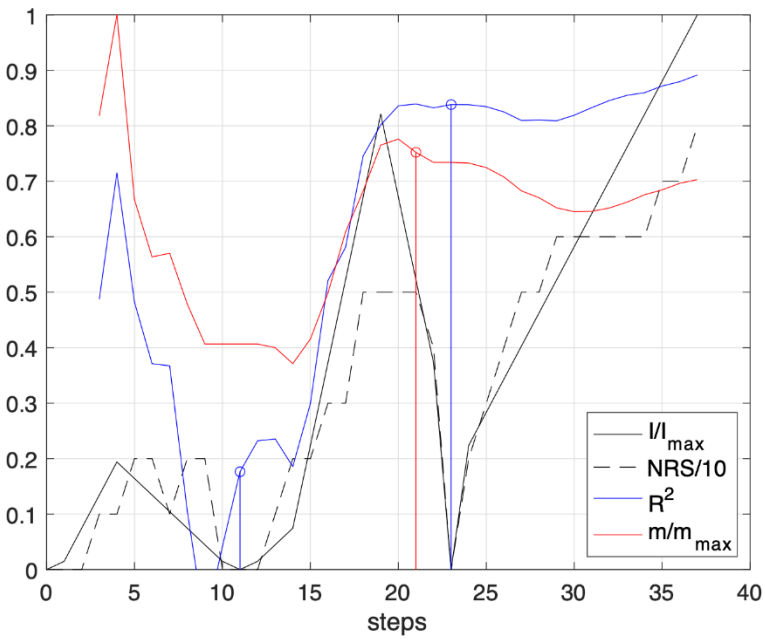

Figure 4. An example of tLRM output. The x axis represents calibration steps, starting from 0. The y axis corresponds to: stimuli strength, I, normalized by its maximum value (black line); corresponding NRS rating divided by 10 (black dashed line); evolution of the coefficient R2 for the linear regression; and linear regression equation gradient, m, normalized by its maximum value (red line). Blue vertical lines represent points where zero current stimulation was applied, and the red vertical line represents the optimal truncation point.

3. Data analysis

3.1. Rejection criteria

To ensure that only good-quality data are analysed, we quantitatively checked which participants "failed" the calibration procedure, which is seldomly performed in pain research studies. This verification was performed only for the tLRM, as the evaluation of these methods is the object of this study.

We propose three rejection criteria with their corresponding thresholds. The first one is the coefficient of determination, R^2. A high R^2 value indicates that the data are linear and that they fit the regression model for predicting the stimulation currents I corresponding to their *NRS*. A recent study on thermal stimulation (Amir et al., 2022) established that a successful calibration procedure, R^2 should be at least 0.4. The second criterium is related to the convergence of the parameters. If the calibration procedure is stopped at a point where the regression model parameters are invariant of collecting more data, we can say that the calibration procedure has reached the final solution. As a measure of convergence, we propose to modify the classical Cauchy's convergence criterion based on the gradient m:

$$\sum_{i=o}^{o+r-1} \frac{m_{i+1} - m_i}{m_o} < \epsilon$$

where o subscript symbolizes the optimal truncation step and ϵ is a limit which we proposed to set to 0.25. Finally, and to both prevent participants from receiving multiple pain stimulations of very high intensity and prevent habituation effects during the main phase of the experiment, we set a limit to the stimulation current (i.e., the third criterium). In the case of our experiment, we chose an upper limit to the pain sensation (NRS=5) as 0.635mA. Any participants with pain sensation T above this level was rejected. This value is based on our previous pilot study, which indicated that stimulation above 1mA exceeds safety limits.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

3.2. Statistical analysis

We investigated if there are differences in the goodness of fit for the three linear regression models (LRM, tLRMm and tLRMmc). The R^2 values distributions were negatively skewed (-1.41, -1.24 and -1.31 respectively) and did not satisfy the Anderson-Darling goodness-of-fit hypothesis test for normality ($p<0.005$). To mitigate the skewness and normalize the data, a set of 40 different power transformations was tested, which concluded that power 4.69 yielded an optimal result (skewness of 0.24, -0.30, and -0.35, respectively and Anderson-Darling goodness-of-fit hypothesis p=0.054, 0.040, and 0.055, respectively). Next, we performed a mixed effects repeated measure ANOVA for the R^2 values with *Linear Regression Type* (LRM, tLRMm, tLRMmc) as a within- and *Gender* as between-subject factor (females and males).

In order to indicate which calibration method requires a smaller number of stimuli application, we used a nonparametric Wilcoxon signed rank test on the number of stimuli required by LRM vs tLRM (note that, by definition, the number of stimuli required by tLRMm and tLRMmc is equal).

Afterward, we investigated if there were differences between the four tested calibration methods. We performed a mixed effects repeated measure ANOVA for the calculated values of *T* and *t* with *Threshold Type* (sensation and pain threshold) and *Gender* (females and males) as between- and *Calibration Type* (TM, LRM, tLRMm, tLRMmc,) and *Threshold Type* (sensation and pain threshold) as within-subject factors.

Next, we compared the intensity of mid-pain stimulation calculated for each of the calibration method. For TM we based mid-pain stimulation intensity on the equation $1.5*T$ used in previous studies (Bąbel et al., 2018; Colloca et al., 2010). For linear models, the intensity leading to an NRS of 8 was used. Then, we performed a repeated measures ANOVA for mid-pain intensity (mA) with *Calibration Type* (TM, LRM, tLRMm, tLRMmc) as within- and *Gender* (females and males) as between-subject factor.

If necessary, Greenhouse-Geisser correction was applied for violations of sphericity, and in the case of interaction effects, the repeated measure ANOVA was followed by multiple comparison tests, using Bonferroni correction for multiple comparisons, in order to indicate the meaning of the effect.

Data analysis was performed using MATLAB (9.9.0.1524771 (R2020b) Update 2) and statistical analysis were performed using IBM SPSS (Version 26).

4. Results

4.1. Rejected participants

According to the rejection criteria, we determined that 4 participants (5.7%) need to be excluded from further analysis due to a low R^2 value (below 0.4, see Fig. 5a and Fig. 5d), high stimulation currents above 0.635mA for NRS of 5 (see Fig. 5b), or lack of convergence (see Fig. 5c and Fig. 5d).

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

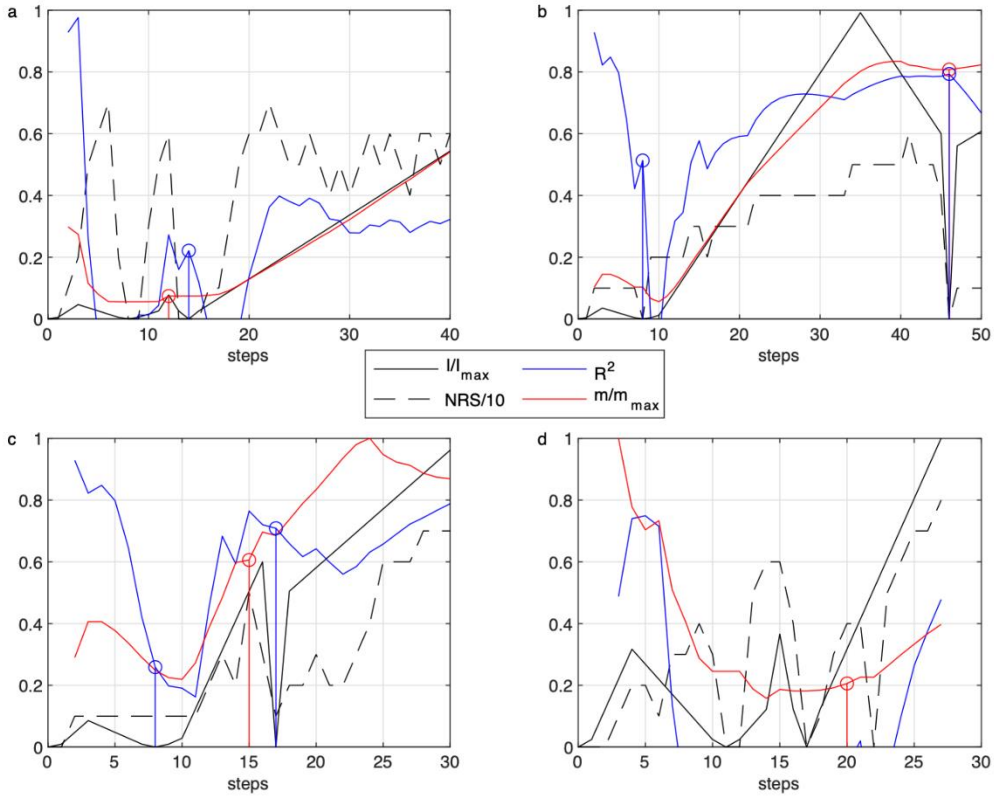

Figure 5. Participants rejected from further analysis (for plot description refer to

Figure 3). Rejection reasons: a - low R2 and high number of delivered stimuli; b - high stimuli current and high number of delivered stimuli; c - lack of convergence and mistake in 0 mA stimulus scoring; d - lack of convergence and mistake in 0mA stimuli scoring.

4.2. Statistical results

For LRM and tLRM, the R^2 values were above 0.75. The relation between stimuli intensity and NRS ratings is, therefore, linear, what is in line with our hypothesis.

Results of repeated measure ANOVA for R^2 revealed a statistically significant main effect of the *Linear Regression Type* ($F_{(2,65)}$= 17.25, p<.001, η_p^2=.21) and no effects of *Gender* ($F_{(1,64)}$=.06, p=.81, η_p^2=.16) nor interaction effect *Linear Regression Type * Gender* ($F_{(1,64)}$= 2.10, p=.15, η_p^2=.02). The pairwise comparisons of *Linear Regression Type* indicated statistically significant differences between LRM vs. tLRMm, LRM vs. tLRMmc and tLRMm vs. tLRMmc. The results indicate that the tLRM models (tLRMmc reported higher R^2 than tLRMm, as would be expected due to a higher number of degrees of freedom) ensure the highest degree of linearity, as was hypothesized (see Table 2).

Table 2. Descriptive statistics of investigated variables for all four calibration models: TM, tLRMm, tLRMmc and LRM.

Calibration method	Gender	t [a] (Mean±SD)	T [b] (Mean±SD)	Mid-painful stimulation[c] (Mean±SD)	Steps number [d] (Mean±SD)	R^{2e}	N[f]
TM	Female Male	0.02±0.01 0.03±0.02	0.26±0.32 0.37±0.15	0.40±0.23 0.56±0.22	37.68±2.64 46.56±2.49	N/A	31 35

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

	Total	0.02±0.02	0.32±0.16	0.48 ±0.24	42.38±1.88		66
LRM	Female	0.04±0.03	0.23±0.16	0.38±0.23	37.68±2.64	0.80±0.09	31
	Male	0.07±0.05	0.35±0.16	0.56±0.22	46.56±2.49	0.75±0.15	35
	Total	0.05±0.05	0.29±0.17	0.47±0.23	42.38±1.88	0.78±0.13	66
tLRMm	Female	0.04±0.02	0.19±0.10	0.30±0.14	23.55±0.75	0.81±0.12	31
	Male	0.05±0.02	0.27±0.11	0.43±0.16	26.69±1.1.7	0.81±0.13	35
	Total	0.05±0.02	0.23±0.11	0.36±0.16	25.21±0.73	0.81±0.12	66
tLRMmc	Female	0.04±0.02	0.19±0.10	0.30±0.14	23.55±0.75	0.83±0.11	31
	Male	0.05±0.03	0.27±0.11	0.43±0.16	26.69±1.1.7	0.83±0.13	35
	Total	0.04±0.03	0.23±0.11	0.37±0.16	25.21±0.73	0.83±0.12	66

Note: [a] sensation threshold; [c] pain threshold; [c] Pain stimulation corresponding to 1.5*T for TM and NRS of 8 for all linear models; [d] number of steps/stimuli of each calibration [e] coefficient of determination; [f] Participant number.

Wilcoxon signed rank test for the number of stimuli required by LRM vs tLRM revealed a statistically significant difference ($p<0.001$). tLRM take on average 17.2 steps less than LRM, what is in line with our hypothesis. Descriptive statistics of the number of steps for each calibration methods are presented in Table 2.

The results of repeated measures ANOVA for individual threshold values of *T* and *t* are presented in Table 3. We found statistically significant main effects of *Calibration Type*, *Threshold Type* and *Gender*, as well as interaction effects of *Calibration Type* Threshold Type* and *Threshold Type*Gender*.

Table 3. The results of repeated measures ANOVA for pain and sensory thresholds.

Main and interactions effects	**F**	***df***	***P***	η_p^2
Calibration Type	19.06	3,64	**<0.001**	0.23
Threshold Type	326,74	1,64	**<0.001**	0.83
Gender	11.34	1,64	**0.001**	0.15
Calibration Type Threshold Type*	37.42	3,64	**<0.001**	0.37
*Calibration Type*Gender*	1.5	3,64	0.21	0.02
*Threshold Type*Gender*	9.89	1,64	**<0.01**	0.13
Calibration Type Threshold Type*Gender*	1.51	3,64	0.23	0.02
Within-factor multiple comparison tests for T values *Calibration Type* Threshold Type*	**F**	***df***	***P***	η_p^2
LRM vs TM			**<0.001**	
tLRMm vs TM			**<0.001**	
tLRMmc vs TM			**<0.001**	
tLRMm vs tLRMmc			0.10	
tLRMm vs LRM			**0.001**	
tLRMmc vs LRM			**<0.01**	
Within-factor multiple comparison tests for t values *Calibration Type* Threshold Type*	**F**	***df***	***P***	η_p^2
LRM vs TM			**<0.001**	
tLRMm vs TM			**<0.001**	
tLRMmc vs TM			**<0.001**	

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

tLRMm vs tLRMmc			0.42	
tLRMm vs LRM			0.81	
tLRMmc vs LRM			0.14	
Between-factor multiple comparison *Threshold Type*Gender*	**F**	***df***	***P***	η_p^2
T female vs males	10.90	1,64	<**0.01**	0.15
t females vs males	8.07	1,64	<**0.01**	0.11

Note: TM: Threshold Method; tLRMm: Truncated Linear Regression Method that assumes that the coefficient *c* (the intercept) is zero; tLRMmc: Truncated Linear Regression Method where the coefficient *c* is not equal to zero; T: pain threshold; t sensation threshold.

The within-factor multiple comparison tests for the interaction *Calibration Type*Threshold Type* for the *T* values showed statistically significant difference between LRM vs. TM, tLRMmc vs. TM, tLRMm vs. TM, tLRMm vs. LRM, tLRMmc vs. LRM (see Table 3 and Fig. 6), what is in line with our hypothesis that *T* values calculated for longer lasting TM and LRM may be higher than ones calculated based on the tLRM. Analogous exploratory analysis performed for *t* values revealed that there was a statistically significant effect in the comparison between TM and the three types of LRM (see Table. 3 and 4 and Fig. 6). It is interesting to note that for four participants the LRM method predicted negative stimuli intensity for the *t* threshold.

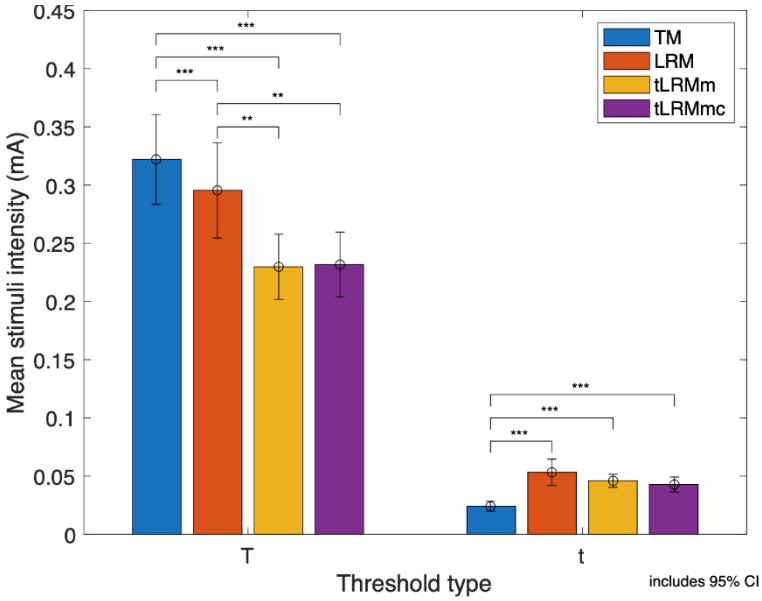

Figure 6. T and t values in mA for each of the four tested calibration methods. Note: ** and *** corresponds to p<0.01 and p<0.001 respectively.

The pairwise comparison between the tLRMm and tLRMmc for *t* and *T* found no differences, what is in line with our hypothesis.

The within-factor multiple comparison tests showed sex differences in the interaction *Threshold Type*Gender*. For both *T* and *t*, we found statistically significant differences between both genders, what is in line with our hypothesis of existence of gender difference in *t* and *T*. The results are presented in Tab. 3 and Fig. 7.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

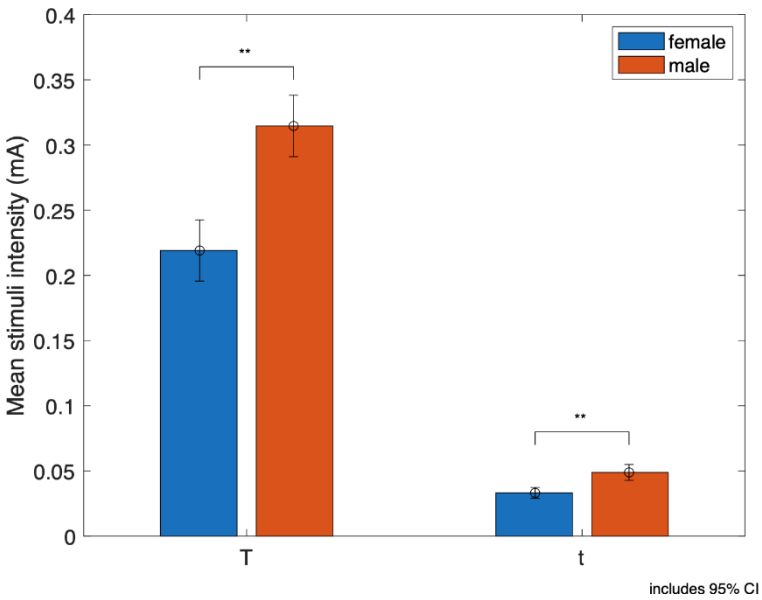

Figure 7. T and t values for males and females. For females, lower intensity for both t and T compared to male participants. Note: ** corresponds to p<0.01.

The ANOVA analysis for mid-painful stimulation indicated statistically significant main effects of *Calibration Type* ($F_{(3,64)}$=19.38, p<.001, η_p^2=.23) and *Gender* ($F_{(1,64)}$= 10.63, p>.01, η_p^2=.14), but no interaction effect *Calibration Type* Gender* ($F_{(1,64)}$=0.93, p> .05, η_p^2=.01) (see Fig. 8 and Table 2). The pairwise comparisons indicate a statistically significant difference (p<.05) between TH vs. LRM, TH vs. LRMm, TH vs. LRMmc, LRM vs. LRMm, and LRM vs. LRMmc (see Fig. 8 and Table 2). These results are in line with our hypothesis that TM and LRM may overestimate the intensity of the pain stimuli.ç

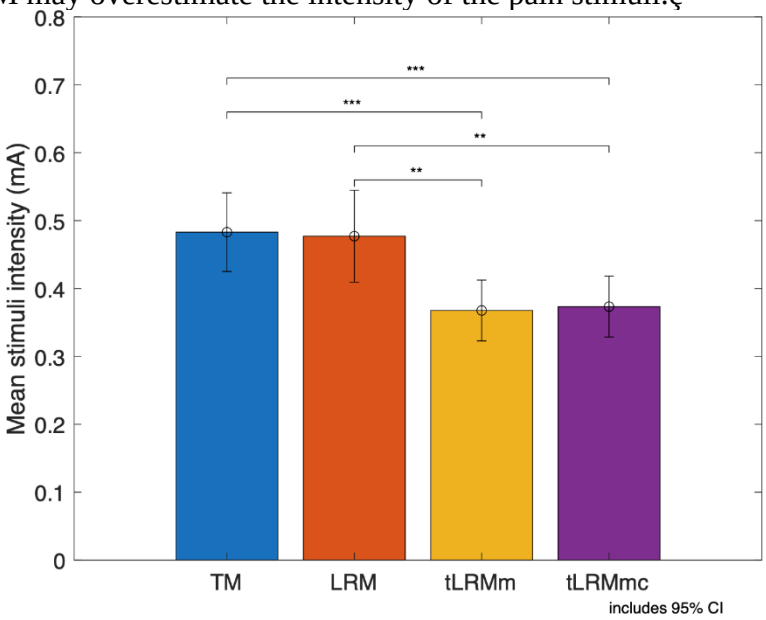

Figure 8. Mid-painful stimulation estimated for each of four calibration methods. Note: ** and *** corresponds to p<0.01 and p<0.001 respectively. For TM the equation 1.5T was used in order to set mid-painful stimulation.

The post-test revealed that, for most of the participants, mid-painful stimulation was indicated for NRS scores between 6-9, low-painful between 3-6, and tactile between 2-3.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

5. Discussion and conclusions

In the present study we aimed to develop an alternative calibration method to the traditional threshold-based method (TM), taking advantage of the linear regression model. In line with our hypothesis, the collected NRS scores exhibit high levels of linearity with the stimuli intensity, which is demonstrated by high values of R^2 for LRM and both tLRM. This result is coherent with previous studies where NRS scores are linearly correlated with stimuli intensity (Van Der Heide et al., 2009; Oliveira et al., 2014). The lack of consistency in the previous studies might be associated with the investigated laboratory stimulation type or with the specific clinical pain type.

Furthermore, we have proven that tLRM show higher R^2 values, indicating that considering only an optimal set of data gives better results than a complete one, likely due to habituation effects in the latter. In the study we used a 11-point NRS where both non-painful and painful stimuli are rated (Colloca et al., 2010; Roa Romero et al., 2013; Świder et al., 2017) which, together with tLRM, allows for performing estimations on stimuli intensity and NRS ratings. Finally, truncating data at an optimal point has another advantage, namely decreasing the number of delivered stimuli. On average this meant that the tLRM required 17.2 less stimuli which shortens the procedure, is less painful for the participant and has a potential for avoiding stimuli habituation.

Since data linearity has been proven, the R^2 parameter can be used as a measure of quantifying the correctness and reliability of the calibration procedure. This allows for an automatic (no human labour required) and instant (immediately after calibration end) procedure for participants rejection according to the three proposed criteria. As a result, the experimenter can avoid perpetuating errors in the further stages of the experiment, which can save time and financial resources (some investigation techniques, such as MEG, have an elevated cost per participant).

One of the main objectives of the study was to use non-painful stimuli to make predictions about painful stimuli (tLRM), which could allow to minimize the method runtime and the number of delivered stimuli, especially the painful ones. It still remains unclear, and more data needs to be collected to determine, which of the evaluated methods (TM, LRM, tLRMm and tLRMmc) is more reliable. On one hand, LRM and tLRM yield statistically similar results for the value of t, which are higher than for the TM method. This means that either the linear regression models overestimate the required stimulus, or the TM underestimates it. We suspect that the latter is more probable, due to a higher sensitivity to stimuli at the beginning of the procedure, when t is predicted by TM. On the other hand, only tLRMm and tLMRmc yield statistically similar results for T values, but their mean is lower than for TM and LRM. This is in line with our hypothesis that, in general, TM and LRM would indicate higher stimuli intensity as they require more stimuli application and therefore participants can suffer from habituation. Habituation can be minimized by using a carefully chosen ISI, which in our study was at least 8s. In studies using analogous electrodes (intraepidermal electrical stimulation, IES) the rest time between three stimulation sets was 1 min (Tanaka et al., 2021) or ISI between 2.5 and 3.5s (Poulsen et al., 2020). Taking this into account, the ISI used in our study can be considered adequate, but in future studies we suggest including breaks between ascending/descending curves, to further avoid these effects. Nevertheless, the higher degree of linearity in tLRM is a strong indication that non-painful stimulation calibration can be extrapolated to provide prediction of painful stimulation. To verify it, more evidence should be collected, for example by applying a set of random stimuli, after a completed calibration procedure and sufficient resting, to validate the correctness of the linear model.

Next, since tLRMm and tLRMmc yield statistically similar results for t, T, and mid-painful stimuli, we propose to use the tLRMm method, as it is described by only one parameter, m, and forces the condition that for no stimulation the NRS is 0. This, in turn, makes

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

comparisons between participants and experiments easier, and avoids not realistic situations where negative stimulation is predicted for low NRS, as was observed for a few cases of LRM.

Regarding gender differences, our results indicate that females require a lower stimulation intensity to reach both the sensation and painful thresholds. This may suggest that women have a higher sensory acuity than men, at least in response to the electric stimulation via WASP electrodes. This conclusion is not consistent across other studies (Racine et al., 2012a, 2012b); however, the predominant view is that women have a greater pain responsiveness than men for most pain modalities (Sorge and Totsch, 2017). This is not unique only to pain, as there is evidence that women are often more perceptive than men across multiple sensory domains (Hashmi and Davis, 2014). Women were shown to have a greater detection and discrimination sensitivity than men to tactile (Boles and Givens, 2011), olfactory (Bontempi et al., 2021), and visual stimuli (Shaqiri et al., 2018). Our results are also in line with the clinical reality which has shown that gender differences exists with respect to pain tolerance and thresholds, and that there is a higher prevalence of chronic pain conditions for females (Meints et al., 2018; Sorge and Totsch, 2017; Templeton, 2020). Finally, gender differences in pain perception and modulation exists at molecular, cellular, and system level (Presto et al., 2022).

In relation to the above, the calibration procedure was made as automatic as possible in order to improve standardization and avoid experimenter biases (Engskov et al., 2021). Moreover, to control the effect of experimenter gender during the experiment, both a woman and a man were present in the laboratory. The literature shows a significant ambiguity in that topic, and it is possible that pain modality may play a crucial role. For instance, previous studies indicate higher thresholds for electric pain determined in the presence of a female investigator regarding participant gender (Engskov et al., 2021). On the other hand, another study, where pleasure stimuli was used, shows that men showed higher average pain thresholds when tested by female experimenter (Gijsbers and Nicholson, 2005). Interestingly, heat pain thresholds were not shown to be significantly influenced by experimenter gender (McDougall et al., 2021). For this reason, in future studies it is recommended to investigate the role of experimenter gender on pain thresholds produced via WASP electrodes.

The last analysis involved a comparison between predicted levels for mid-painful stimulus. There is no statistically significant difference between TM based on 1.5*T and LRM. At the same time, both are different to the intensity levels predicted by tLRM, which are significantly lower. We hypothesize that higher values for TM and LRM may be a consequence of habituation. However, we cannot exclude that the tLRM mid-painful stimulation could be underestimated, as the post test for LRM, where low- and mid-painful stimulations were delivered, indicated that stimulation was predicted correctly. A possible way of mitigating this effect could be using an adaptive staircase method (Atlas et al., 2010).

Concluding, the proposed method, including its quantification of correctness and the rejection criteria, can be easily adapted to other stimulation methods and hence extended to any neurophysiological studies on pain. According to multi-modal approach in human pain research (Neziri et al., 2011), this method allows for a standardized comparison between different stimulation modalities, facilitating translation of their result to a clinical setting. We established procedures for the application of the stimulus as well as an specific automatic calibration of electric stimulation delivered via WASP electrodes, as the replicability of the laboratory setup seems to be crucial for the further comparison of study results. Furthermore, linear regression satisfies the definition of the IASP, which recommends that pain threshold should be at the level at which 50% of stimuli would be recognized as painful instead of the least stimulus intensity at which pain is perceived (Raja et al., 2020).

Finally, our method is not free from limitations and the need for further refinements. To be sure that calibration leads to adequate stimuli evaluation, a random strength stimulation

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

verification of reliability should be performed *a posteriori*. Additionally, this could be used to check if the calibration results are consistent over time, as was done before for thermal stimulation (Amir et al., 2022). Reliability tests are important for future experimental studies on WASP electrodes and studies where pain biomarkers or signatures are evaluated (Davis et al., 2020; Pleil et al., 2018). Furthermore, the intra-epidermal electrodes, such as WASP, have been found to be sensitive to their positioning relative to the location of the nerve fibre (Poulsen et al., 2020), which could have influenced the calibration. Finally, more research is needed to determine the factors influencing the sex-based differences in perception of *t* and *T* stimuli, such as possible differences of electrode attachment, or gender physiological, chemical, and biophysical skin differences (Rahrovan et al., 2018).

6. Acknowledgements

We would like to thank Gabriel Delojo Pradas for his enthusiasm and practical support in conducting the study.

7. Author contributions

KS: conceptualization, data collection and analysis, ms preparation; RB: conceptualization, ms revision; SM: ms revision. All authors discussed the results and commented on the manuscript.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

8. Reference

Amir, C., Rose-McCandlish, M., Weger, R., Dildine, T.C., Mischkowski, D., Necka, E.A., Lee, I. seon, Wager, T.D., Pine, D.S., Atlas, L.Y. (2022). Test-Retest Reliability of an Adaptive Thermal Pain Calibration Procedure in Healthy Volunteers. *J Pain* 23, 1543–1555.
Atlas, L.Y., Bolger, N., Lindquist, M.A., Wager, T.D. (2010). Brain Mediators of Predictive Cue Effects on Perceived Pain. *J Neurosci* 30, 12964–12977.
Bąbel, P., Adamczyk, W., Świder, K., Bajcar, E.A., Kicman, P., Lisińska, N. (2018). How Classical Conditioning Shapes Placebo Analgesia: Hidden versus Open Conditioning. *Pain Med* 19, 1156–1169.
Bąbel, P., Bajcar, E.A., Adamczyk, W., Kicman, P., Lisińska, N., Świder, K., Colloca, L. (2017). Classical conditioning without verbal suggestions elicits placebo analgesia and nocebo hyperalgesia. *PLoS One* 12, 1–17.
Berger, S.E., Baria, A.T. (2022). Assessing Pain Research: A Narrative Review of Emerging Pain Methods, Their Technosocial Implications, and Opportunities for Multidisciplinary Approaches. *Front Pain Res* 3, 1–26.
Boles, D.B., Givens, S.M. (2011). Laterality and sex differences in tactile detection and two-point thresholds modified by body surface area and body fat ratio. *Somatosens Mot Res* 28, 102–109.
Bontempi, C., Jacquot, L., Brand, G. (2021). Sex Differences in Odor Hedonic Perception: An Overview. *Front Neurosci* 15.
Brainard, D.H. (1997). The Psychophsycis Toolbox. *Spat Vis* 10, 433–436.
Bushnell, M.C., Čeko, M., Low, L.A. (2013). Cognitive and emotional control of pain and its disruption in chronic pain. *Nat Rev Neurosci* 14, 502–511.
Colloca, L., Petrovic, P., Wager, T.D., Ingvar, M., Benedetti, F. (2010). How the number of learning trials affects placebo and nocebo responses. *Pain* 151, 430–439.
Davis, K.D., Aghaeepour, N., Ahn, A.H., Angst, M.S., Borsook, D., Brenton, A., Burczynski, M.E., Crean, C., Edwards, R., Gaudilliere, B., Hergenroeder, G.W., Iadarola, M.J., Iyengar, S., Jiang, Y., Kong, J.T., Mackey, S., Saab, C.Y., Sang, C.N., Scholz, J., Segerdahl, M., Tracey, I., Veasley, C., Wang, J., Wager, T.D., Wasan, A.D., Pelleymounter, M.A. (2020). Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. *Nat Rev Neurol* 16, 381–400.
Engskov, A.S., Lejbman, I., Åkeson, J. (2021). Randomized cross-over evaluation of investigator gender on pain thresholds in healthy volunteers. *GMS Ger Med Sci* 19, 1–8.
Fillingim, R., King, C. (2009). Sex, gender, and pain: a review of recent clinical and experimental findings. *J Pain* 10, 447–485.
Gijsbers, K., Nicholson, F. (2005). Experimental Pain Thresholds Influenced by Sex of Experimenter. *Percept Mot Skills* 101, 803–807.
Gruss, S., Geiger, M., Werner, P., Wilhelm, O., Traue, H.C., Al-Hamadi, A., Walter, S. (2019). Multi-modal signals for analyzing pain responses to thermal and electrical stimuli. *J Vis Exp* 2019, 1–12.
Hartrick, C.T., Kovan, J.P., Shapiro, S. (2003). The Numeric Rating Scale for Clinical Pain Measurement: A Ratio Measure? *Pain Pract* 3, 310–316.
Hashmi, J.A., Davis, K.D. (2014). Deconstructing sex differences in pain sensitivity. *Pain* 155, 10–13.
Van Der Heide, E.M., Buitenweg, J.R., Marani, E., Rutten, W.L.C. (2009). Single pulse and pulse train modulation of cutaneous electrical stimulation: A comparison of methods. *J Clin Neurophysiol* 26, 54–60.
Inui, K., Kakigi, R. (2012). Pain perception in humans: Use of intraepidermal electrical stimulation. *J Neurol Neurosurg Psychiatry* 83, 551–556.
Lautenbacher, S., Hassan, T., Seuss, D., Loy, F.W., Garbas, J.U., Schmid, U., Kunz, M.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

(2022). Automatic Coding of Facial Expressions of Pain: Are We There Yet? *Pain Res Manag* 2022.
Lazaridou, A., Elbaridi, N., Edwards, R., Berde, C.B.B. (2019). Pain assessment. In *Cambridge Handbook of Psychology, Health and Medicine: Third Edition*, (Elsevier), pp. 39–46.
Lefaucheur, J.P., Ahdab, R., Ayache, S.S., Lefaucheur-Ménard, I., Rouie, D., Tebbal, D., Neves, D.O., Ciampi de Andrade, D. (2012). Pain-related evoked potentials: A comparative study between electrical stimulation using a concentric planar electrode and laser stimulation using a CO 2 laser. *Neurophysiol Clin* 42, 199–206.
Lundeberg, T., Lund, I., Dahlin, L., Borg, E., Gustafsson, C., Sandin, L., Rosén, A., Jan, K., Eriksson, S. V. (2001). Reliability and Responsiveness of Three Different Pain Assessments. *J Rehabil Med* 279–283.
McDougall, J.F., Bailey, N.G.N., Banga, R., Linde, L.D., Kramer, J.L.K. (2021). The Influence of Examiner Gender on Responses to Tonic Heat Pain Assessments: A Preliminary Investigation. *Front Pain Res* 2, 1–10.
Meints, S.M., Wang, V., Edwards, R.R. (2018). Sex and Race Differences in Pain Sensitization among Patients with Chronic Low Back Pain. *J Pain* 19, 1461–1470.
Neziri, A.Y., Curatolo, M., Nüesch, E., Scaramozzino, P., Andersen, O.K., Arendt-Nielsen, L., Jüni, P. (2011). Factor analysis of responses to thermal, electrical, and mechanical painful stimuli supports the importance of multi-modal pain assessment. *Pain* 152, 1146–1155.
Olesen, A.E., Andresen, T., Staahl, C., Drewes, A.M. (2012). Human experimental pain models for assessing the therapeutic efficacy of analgesic drugs. *Pharmacol Rev* 64, 722–779.
Oliveira, I.M., Machado, A.R.P., Pereira, A.A., Andrade, A.O. (2014). Estimated features from surface EMG of the lower limb correlate with the subjective sensation of pain. *Psychol Neurosci* 7, 355–361.
De Paepe, A.L., Amanda, A.C., Crombez, G. (2019). Habituation to pain: A motivationalethological perspective. *Pain* 160, 1693–1697.
Pleil, J.D., Wallace, M.A.G., Stiegel, M.A., Funk, W.E. (2018). Human biomarker interpretation: the importance of intra-class correlation coefficients (ICC) and their calculations based on mixed models, ANOVA, and variance estimates. *J Toxicol Environ Heal - Part B Crit Rev* 21, 161–180.
Poulsen, A.H., Tigerholm, J., Meijs, S., Andersen, O.K., Mørch, C.D. (2020). Comparison of existing electrode designs for preferential activation of cutaneous nociceptors. *J Neural Eng* 17.
Presto, P., Mazzitelli, M., Junell, R., Griffin, Z., Neugebauer, V. (2022). Sex differences in pain along the neuraxis. *Neuropharmacology* 210, 109030.
Price, D.D., Bush, F.M., Long, S., Harkins, S.W. (1994). A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. *Pain* 56, 217–226.
Racine, M., Tousignant-Laflamme, Y., Kloda, L.A., Dion, D., Dupuis, G., Choinière, M. (2012a). A systematic literature review of 10 years of research on sex/gender and experimental pain perception - part 1: are there really differences between women and men? *Pain* 153, 602–618.
Racine, M., Tousignant-Laflamme, Y., Kloda, L.A., Dion, D., Dupuis, G., Choinire, M. (2012b). A systematic literature review of 10 years of research on sex/gender and pain perception - Part 2: Do biopsychosocial factors alter pain sensitivity differently in women and men? *Pain* 153, 619–635.
Rahrovan, S., Fanian, F., Mehryan, P., Humbert, P., Firooz, A. (2018). Male versus female skin: What dermatologists and cosmeticians should know. *Int J Women's Dermatology* 4,

medRxiv preprint doi: https://doi.org/10.1101/2022.10.03.22280662; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

122–130.
Raja, S.N., Carr, D.B., Cohen, M., Finnerup, N.B., Flor, H., Gibson, S., Keefe, F.J., Mogil, J.S., Ringkamp, M., Sluka, K.A., Song, X.J., Stevens, B., Sullivan, M.D., Tutelman, P.R., Ushida, T., Vader, K. (2020). The Revised IASP definition of pain: concepts, challenges, and compromises. *Pain* 161, 1976–1982.
Roa Romero, Y., Straube, T., Nitsch, A., Miltner, W.H.R., Weiss, T. (2013). Interaction between stimulus intensity and perceptual load in the attentional control of pain. *Pain* 154, 135–140.
Rütgen, M., Seidel, E.-M., Silani, G., Riečanský, I., Hummer, A., Windischberger, C., Petrovic, P., Lamm, C. (2015). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proc Natl Acad Sci* 112, E5638–E5646.
Seidel, E.M., Pfabigan, D.M., Hahn, A., Sladky, R., Grahl, A., Paul, K., Kraus, C., Küblböck, M., Kranz, G.S., Hummer, A., Lanzenberger, R., Windischberger, C., Lamm, C. (2015). Uncertainty during pain anticipation: The adaptive value of preparatory processes. *Hum Brain Mapp* 36, 744–755.
Shaqiri, A., Roinishvili, M., Grzeczkowski, L., Chkonia, E., Pilz, K., Mohr, C., Brand, A., Kunchulia, M., Herzog, M.H. (2018). Sex-related differences in vision are heterogeneous. *Sci Rep* 8, 1–10.
Sorge, R.E., Totsch, S.K. (2017). Sex Differences in Pain. *J Neurosci Res* 95, 1271–1281.
Świder, K., Bąbel, P. (2013). The effect of the sex of a model on nocebo hyperalgesia induced by social observational learning. *Pain* 154, 1312–1317.
Świder, K., Wronka, E., Oosterman, J.M., van Rijn, C.M., Jongsma, M.L.A. (2017). Influence of transient spatial attention on the P3 component and perception of painful and non-painful electric stimuli in crossed and uncrossed hands positions. *PLoS One* 12, 1–22.
Tanaka, S., Gomez-Tames, J., Wasaka, T., Inui, K., Ueno, S., Hirata, A. (2021). Electrical Characterisation of Aδ-Fibres Based on Human in vivo Electrostimulation Threshold. *Front Neurosci* 14, 1–9.
Templeton, K.J. (2020). Sex and Gender Issues in Pain Management. *J Bone Joint Surg Am* 102, 32–35.
Tracy, L.M. (2017). Psychosocial factors and their influence on the experience of pain. *Pain Reports* 2, 2–3.
Wagemakers, S.H., Van Der Velden, J.M., Gerlich, A.S., Hindriks-Keegstra, A.W., Van Dijk, J.F.M., Verhoeff, J.J.C. (2019). A systematic review of devices and techniques that objectively measure patients' pain. *Pain Physician* 22, 1–13.
Wang, L., Guo, Y., Dalip, B., Xiao, Y., Urman, R.D., Lin, Y. (2022). An experimental study of objective pain measurement using pupillary response based on genetic algorithm and artificial neural network. *Appl Intell* 52, 1145–1156.
Zigmond, A.S., Snaith, R.P. (1983). The hospital anxiety and depression scale. *Acta Psychiatr Scand* 67, 361–370.