To calibrate or not to calibrate? A methodological dilemma in experimental pain research

Waclaw M Adamczyk  
*University of Lübeck*

Tibor M Szikszay  
*University of Lübeck*

Hadas Nahman-Averbuch  
*Washington University School of Medicine in St. Louis*

Jacek Skalski  
*Jerzy Kukuczka Academy of Physical Education*

Jakub Nastaj  
*Jerzy Kukuczka Academy of Physical Education*

*See next page for additional authors*

Follow this and additional works at: https://digitalcommons.wustl.edu/oa_4

Part of the Medicine and Health Sciences Commons

Please let us know how this document benefits you.

**Recommended Citation**

Adamczyk, Waclaw M; Szikszay, Tibor M; Nahman-Averbuch, Hadas; Skalski, Jacek; Nastaj, Jakub; Gouverneur, Philip; and Luedtke, Kerstin, "To calibrate or not to calibrate? A methodological dilemma in experimental pain research." *Journal of Pain.* 23, 11. 1823 - 1832. (2022).  
https://digitalcommons.wustl.edu/oa_4/1256

This Open Access Publication is brought to you for free and open access by the Open Access Publications at Digital Commons@Becker. It has been accepted for inclusion in 2020-Current year OA Pubs by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
To Calibrate or not to Calibrate? A Methodological Dilemma in Experimental Pain Research

Waclaw M. Adamczyk,*,† Tibor M. Szikszay,† Hadas Nahman-Averbuch,§ Jacek Skalski,∗ Jakub Nastaj,* Philip Gouverneur,§ and Kerstin Luedtke†,*

†Laboratory of Pain Research, Institute of Physiotherapy and Health Sciences, The Jerzy Kukuczka Academy of Physical Education, Katowice, Poland
†Institute of Health Sciences, Department of Physiotherapy, Pain & Exercise Research Luebeck (P.E.R.L.), University of Lübeck, Lübeck, Germany
§Center of Brain, Behavior and Metabolism (CBBM), University of Lübeck, Lübeck, Germany
§Washington University Pain Center, Department of Anesthesiology, Washington University School of Medicine, St. Louis, Missouri
*Institute of Medical Informatics, University of Lübeck, Lübeck, Germany

Abstract: To calibrate or not to calibrate? This question is raised by almost everyone designing an experimental pain study with supra-threshold stimulation. The dilemma is whether to individualize stimulus intensity to the pain threshold / supra-threshold pain level of each participant or whether to provide the noxious stimulus at a fixed intensity so that everyone receives the identical input. Each approach has unique pros and cons which need to be considered to i) accurately design an experiment, ii) enhance statistical inference in the given data and, iii) reduce bias and the influence of confounding factors in the individual study e.g., body composition, differences in energy absorption and previous experience. Individualization requires calibration, a procedure already irritating the nociceptive system but allowing to match the pain level across individuals. It leads to a higher variability of the stimulus intensity, thereby influencing the encoding of “noxiousness” by the central nervous system. Results might be less influenced by statistical phenomena such as ceiling/floor effects and the approach does not seem to rise ethical concerns. On the other hand, applying a fixed (standardized) intensity reduces the problem of intensity encoding leading to a large between-subjects variability in pain responses. Fixed stimulation intensities do not require pre-exposure. It can be proposed that one method is not preferable over another, however the choice depends on the study aim and the desired level of external validity. This paper discusses considerations for choosing the optimal approach for experimental pain studies and provides recommendations for different study designs.

Perspective: To calibrate pain or not? This dilemma is related to almost every experimental pain research. The decision is a trade-off between statistical power and greater control of stimulus encoding. The article decomposes both approaches and presents the pros and cons of either approach supported by data and simulation experiment.

© 2022 The Author(s). Published by Elsevier Inc. on behalf of United States Association for the Study of Pain, Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Key words: Calibration, Experimental Pain, Fixed Stimulus, Individualized Intensity, Scaling, offset analgesia.

Disclosures: The work was prepared thanks to funding from Polish National Science Center (2020/37/B/HS6/04196). The authors declare no conflict of interest.

Address reprint requests to Dr. Waclaw M. Adamczyk, Laboratory of Pain Research, Institute of Physiotherapy and Health Sciences, The Jerzy Kukuczka Academy of Physical Education, ul. Mikolowska 72A, 40-065 Katowice, Poland. E-mail: w.adamczyk@awf.katowice.pl
In basic pain science, the application of noxious stimuli is fundamental to evoke pain experimentally. This allows to study pain mechanisms and support translational research, filling the gap between studies on animal models and clinical populations. A common dilemma is whether to apply stimuli of fixed or individually calibrated intensity. In some fields, both approaches are equally common, however, the rationale for the choice of one or the other is often tacitly reported in the methods section of experimental pain studies. Mostly, authors do not give reasons for choosing the particular method (likely due to the routine practice in their lab), and if they do, it is justified by only providing the purpose of the study. For example, Fust et al.26 explain the choice, noting that in their study they calibrated the intensity because they were primarily interested in “pain intensity not stimulus intensity per se”. In the study of Pedersen et al.50 authors were interested in reproducing intensity not stimulus intensity per se”. In some fields, both approaches are equally common, however, the rationale for the choice of one or the other is often tacitly reported in the methods section of experimental pain studies.63,68,82

The dilemma has been also addressed in two fMRI studies in which both approaches were contrasted. For example, results from a study by van den Bosch et al.6 pointed out that both intensities (calibrated vs. fixed) may activate similar brain regions but also that calibrated intensity was linked to activity detected in a cluster of the white matter within the corpus callosum. As the authors discussed, this can be the result of movement artifacts having higher pain levels in the calibrated compared to the fixed condition (Appendix 1). On the other hand, successful pain ‘match’ between these two conditions has been performed in a later study by Quiton et al.56 Authors have found that the type of stimulus, i.e., calibrated (pain = 50/100) vs. fixed affected the reliability of the BOLD signal in some of the regions of interests (ROI). For example, the insula, which is a core structure involved in pain perception, showed reliable bilateral activity in response to noxious heat of 48°C compared to calibrated intensity.64 Another structure which is part of the “neurological pain signature”, the SII,56,77 was in contrast, reliably activated with a calibrated intensity. The authors also investigated the spatial patterns of the BOLD signal associated with both stimulus types and found only partial overlap. As described as a coefficient between 0 and 1, the activity in the insula weakly overlapped in its posterior part (ipsilateral), but strong spatial agreement was found in the anterior insula contralaterally.56,65

Another study which is (indirectly) relevant for the discussed dilemma is the one by Nir et al.58 Here, a conditioned pain modulation paradigm was used in which thermal pain was applied to one hand before and during exposure of the contralateral hand into a hot water (conditioning stimulus). The temperature was fixed (45.5°C) but the perception of this conditioning stimulus was, however, manipulated in the experiment using a nocebo manipulation. Thus, in one group, participants perceived the conditioning stimulus as ‘more intense’ because of a nocebo cream application that was supposed to increase the level of pain. Results showed that not the actual thermal noxious intensity, but the perceived pain intensity caused by the conditioning stimulus influenced the CPM effect. This might suggest that the use of a calibrated paradigm is accurate to maintain the same pain level and reduce the impact of the conditioning stimulus variability on the CPM response.

Furthermore, fixed intensity has the potential to elucidate neural mechanisms of individual differences in pain at baseline was e.g., 25.5 vs. 8.1 in the LSP group and 18.1 vs. 7.5 in the HSP group. Moreover, the CPM assessment with fixed temperature elicited a lower inhibitory effect in HSP participants difference not significant. Since the fixed stimulus elicited only mild pain scores at baseline in the LSP group, the inhibitory effect of the CPM paradigm could be potentially limited by the mild pain ratings at baseline, resulting in skewed data and a possible floor effect for the LPS group. Thus, the results of the CPM effect in participants with low pain sensitivity might not reflect their full inhibitory capabilities and using a more intense stimulus might have revealed a greater inhibitory effect.
Fixed Stimulus Intensity

In this approach, all participants receive the same parameters of stimulation to activate nociceptors through nerve endings (e.g., using heat) or to directly activate nociceptive fibers (electrical stimulation). Although using a fixed intensity is often used as a stand-alone methodological decision, reducing the number of stimuli applied, it is sometimes coupled with a familiarization/training session. The aim of the training session is to teach subjects how to provide pain ratings, to reduce the feeling of novelty or to learn how to use an experimental equipment i.e., the computerized Visual Analogue Scale.

With or without familiarization, fixed intensity leads to pain rating variability, that can arise from combinations of multiple biological, psychological, and social factors. Of note, everyone has his/her own unique history of painful experiences affecting subsequent painful events, and thus perception. Indeed, the fact that prior exposure shapes the pain experience has been shown using procedures of classical conditioning and observational learning of pain. For example, in a typical classical conditioning experiment, volunteers underwent a procedure in which they learned that one colour e.g., blue was linked to low pain and another colour e.g., orange preceded high pain. Babel et al. showed that participants continued to feel more intense pain after exposure to orange compared to blue colour, even though the stimulus intensity was equal after the initial conditioning phase.

An identical stimulus can be perceived and expressed by a myriad of possibilities. Fillingim reported for example, that a stimulus of 48°C applied via a thermode can provoke perceived pain intensities varying between 4 and 100 with a mean pain rating at the level of 71.8 on a 0 to 100 (most intense pain imaginable) scale. This behavioral effect is also reflected by a profound variability of brain activations. For instance, Coghill et al. showed that the anterior cingulate cortex (ACC), primary sensory cortex (SI) and prefrontal cortex (PFC) exhibited more robust activity in response to a fixed stimulus in highly sensitive compared to less sensitive individuals. Interestingly, not only different people perceive the same stimulus differently, but there is also a significant within-subject variability, with the most profound example referring to the same stimulus being rated as painful or not within the same session, depending on the trial.

However, the fixed stimulus intensity approach — although frequently used and straightforward — has some important implications (Table 1) as well as advantages and disadvantages (Appendix 2):

i. The variability of pain ratings can raise ethical concerns as some individuals might be exposed to

| Table 1. Comparison of two different approaches |
|--------------------------|------------------------------------------|--------------------------|
| **ITEM** | **CALIBRATION PROCEDURE** | **FIXED STIMULUS INTENSITY** |
| Construct | Equal pain intensity | Equal stimulus intensity |
| Variability | Large in terms of input (nociception), Less variability in pain perception | Large in terms of input, Less variability in terms of input |
| Statistical inference | Unlikely to be affected by ceiling and/or floor effects | Prone to floor/ceiling effects |
| Neural activity | Different intensities may distinguish stimulus encoding at the spinal and thalamus level | Same stimulus may produce equal afferent activation patterns until thalamus level |
| Feasibility | Time-consuming technique | Time-saving technique |
| Translation | Studies on humans less comparable to studies on other species | Studies in humans comparable to studies with animal models |
| Sensitization | Depending on the calibration protocol, potential to irritate the nociceptive system | Lack of pre-exposure in the experiment setup |
| Study design | Can be preferred in studies targeted at subjective outcomes (e.g., pain ratings) | Can be preferred in studies targeted at physiological outcomes (e.g., fMRI) |
| Peripheral factors | Calibration reduce influences resulted from body composition, e.g., fat tissue or epidermal thickness | Affected by body composition, e.g., epidermal thickness and innervation density linked with sensitivity |
| Study design | Desirable in follow-up studies: control for pain sensitivity over-time | Beneficial in assessment of inter-individual pain differences |
| Generalizability | Generalization only to population of a given pain intensity level | Generalization only to population having the same nociceptive focal |
| Ethical | An adapted and tolerable stimulus is always applied (see IASP note) | For some individuals, the applied stimulus can be unbearable or even harmful |
| Pain measurement | In principle affected by lesser reliability | In principle more reliable due to greater variance in pain |

Abbreviations: fMRI, Functional Magnetic resonance Imaging; IASP, International Association for Studying Pain.
individual stimulus that they cannot or can barely tolerate. This variability seems to be dictated by the type of pain modality used in the particular study: Coefficient of Variation (CoV, a ratio of the standard deviation to the mean value) of 16% in pain response was reported using contact heat,23 32% using noxious cold-water stimulation58 and 47% using electrical stimuli.79

ii. The approach is prone to statistical effects such as ceiling and/or floor effects23 with some volunteers rating the stimulus using extreme values.19 This might be problematic in studies that assess a change in pain levels in response to an intervention or a manipulation, as extremely high or low ratings at baseline might limit the visibility of the intervention effect.

iii. A fixed approach limits the generalizability of the results within a domain. As the aim of basic science research in humans is to fill the gap between animal and clinical research, fixed intensity mimics populations of patients who have similar/comparable (in its magnitude) nociceptive input(s), e.g., size of the injury. The same stimulus e.g., 48°C can serve as a proxy to infer from populations having the same size of injury/nociception and thus, might not be generalized to patients with less intense or more extreme injury.

iv. Results from studies with fixed stimulus intensities may be more suitable for assessing the relationship between pain and autonomic (physiological) responses. That approach reduces correlations with different inputs that are a feature of the calibrated approach.43 It has been shown, for example that trodermal activity is stimulus rather than pain-dependent.47 This might have an advantage in fMRI studies as intensity must not be regressed during analysis.

v. Interestingly, this method can have profound effects on the reliability of pain measurement, with measurements in the fixed approach having larger reliability compared to the calibrated approach. This aspect has been demonstrated mathematically, in simulations,80 and real data.56

Individual Stimulus Intensity

The process of individualization of the pain intensity requires a pre-exposure to the stimulus during a so-called calibration procedure. Calibrations differ depending on the psychophysical testing procedures. Two common approaches are the method of limits and the method of levels (ramp and hold). In the former, the intensity is constantly increased (or decreased) while volunteers respond verbally or actively (e.g., by pressing a button) when a pain threshold or pre-specified pain intensity (e.g., 50/100) is reached,53 while in the latter, volunteers are exposed to series of stimuli of gradually increased and/or decreased intensity (e.g., temperature); volunteers decide posthoc (after the stimulus) whether the stimulus elicited the target sensation (pain threshold, or e.g., pain of 50/100) or provide their ratings in real-time.

When compared, the method of levels produced lower thresholds compared to the method of limits.15 Some calibration protocols rely on a random sequence of stimuli. As a result, a stimulus-response function is plotted, and a given value read out from the function. Not only different calibration protocols exist, also, within-protocol diversities have been reported (see14,44,46). In that sense, the method of levels could be performed with different baseline temperatures, increase rates, stimulus durations, steps, and intervals. For instance, heat pain thresholds have been determined with rate of 2, 1 or 0.5°C/s. Some participants might require a longer calibration to identify the pre-specified pain intensity level and will be exposed to a larger number of noxious stimuli compared to other participants. This variability might lead to sensitization and could impact the results of the study.

The outcome of the procedure is a unique individual intensity of noxious stimuli which is supposed to evoke similar pain levels across participants. The main advantage of this approach is that participants are exposed to stimuli causing -in theory- equal pain intensities, thereby reducing the possibility of floor and/or ceiling effects43 and at the same time ethical concerns: according to IASP “In any pain research, stimuli should never exceed a subject’s tolerance limit and subjects should be able to escape or terminate a painful stimulus at will”. If pain intensity is the primary outcome this approach is likely to be preferable as it seems to be less prone to random noise, as noise might affect one of the tested calibration intensities but is not likely to impact all of them. In case of occurrence of unexpected sensitization or habituation, the original pain level can be restored using recalibration (see e.g.,26). Furthermore, this approach bridges the inter-individual differences related to body-composition. For example, it has been shown that the body composition (different regions) and, e.g., BMI influence sensory thresholds (but see also18), and thus, pain sensitivity.15,55,61 Interestingly, in humans, heat perception thresholds are generally expressed as the temperature of the stimulation device, although thresholds for thermoreceptor activation have often been described at the receptor level.19 However, the temperature of the stimulated skin region depends
on several factors, including the initial temperature of the skin, the diffusion capacity between the skin and the stimulation surface, the ability of the skin to distribute heat throughout the tissue, and the depth at which the thermoreceptors are located. Thus, in experimental heat stimulation, the temperatures at the skin surface and at the level of the thermoreceptors may differ significantly. This can be remedied by mathematical (psychophysical) models, a further possible individualization of the stimulation intensity, here related to the parameters of intensity and duration, depending on the receptor depth or the thickness of the epidermis.

Calibration, or re-calibration can have an advantage over fixed intensities (Appendix 2) in studies with multiple observations (e.g., longitudinal), as significant intra-individual differences exist when pain is assessed over days or weeks or even within minutes. Similarly, the effect of inter-individual factors that modulate pain perception is reduced in such an approach. In contrast, individualization hampers the reliability of the measurement. In response to a recent fMRI study focusing on reliability aspects of pain-related brain activity and pain reports, Woo & Wager proved with simulations that the reliability coefficient (ICC) is lower if the sample is characterized by a homogenous level of pain, typical for calibration studies. It is not possible to individualize self-reported pain in non-humans’ species because the way they potentially can do this is markedly different.

Calibration, depending on the protocol employed, irritates the nociceptive system which is encompassed by the subsequent assessment.

Calibration only partly reduces variance in pain ratings. Even though one calibrates the pain, it still varies among individuals and some investigators use a range of acceptable pain ratings such as 5-6 or 40-60. Thus, variability relates to not only peripheral input, but also pain ratings per se. This can be confirmed by our own data (see Fig 1) and all other reports with individual approaches. Furthermore, calibration is not 100% precise. Determining the pain level using a calibration is not a prerequisite for that level in the later parts of the experiment (see examples of unsuccessful calibrations here).

Different intensities used in different individuals can hamper the results’ interpretation. This is important as together with the increase in the intensity of stimulation, different fibers are recruited. Activation of e.g. A-delta fibers characterizes different sensory quality and latency: when sensation can be described as “pricking” or “stabbing” this is an indication for A-delta fiber activity. In turn, C fibers-mediated pain could be distinguishing by a quality
of “dull” or “throbbing”, “cramping”, “aching”. In general, the increase in the intensity of stimulation (mA) leads to overlapping activity of fibers with large and small diameters: smaller fibers (e.g., A-delta) have higher activation thresholds. This variability of sensory input provokes a perception of different quality. Although this phenomenon has been investigated in depth in electrical stimuli, it does not mean that such a problem is irrelevant to other modalities, including heat. This problem seems to be more relevant for the calibration approach in which utilized intensity differs substantially.

Contrasts in the Two Approaches

Analytically, the two approaches differ significantly. As mentioned above, variance in the intensity used is only one side of the coin. One could hypothesize that the variance in pain ratings is lower when calibrating the intensity to e.g., an intensity of 50/100. Indeed, the differences can be visually observed in both approaches in Fig 1. Two approaches were compared in terms of pain ratings collected via an offset analgesia (OA) paradigm (Appendix 3). It is assumed that this paradigm reflects the efficiency in descending pain inhibition and relies on the application of stimuli consisting of three temperatures (T1, T2, T3). The intensity of the noxious stimulus during the T1 interval is equal to the intensity in the T3 interval, while the intensity at T2 is slightly higher. Collecting pain data continuously allows to track pain dynamically over time. In a typical OA effect, one can observe a significant drop in pain during the application of the T3 temperature (a so-called disproportional pain reduction). The OA effect can be observed either in the individualized (Fig 1, left) or fixed approach (Fig 1, right). The individualized temperature - even though calibrated for pain 50 - lead to a significant variance in pain ratings, however less than in the fixed approach. The consequence of the fixed approach is a relatively large number of volunteers needed to detect a difference in the effect under investigation. This is caused by the smaller effect size to be detected i.e., Cohen’s d. Thus, sample size underestimation is more likely when designing studies using a fixed approach but calculating the sample size based on data from the other approach.

Fig 2 presents simulations of p values derived from 1000 paired student t tests comparing pain in the T1 vs the T3 interval (Fig 1). Scatterplot indicates that the likelihood for rejecting the null hypothesis is about 2 times higher if using a fixed approach (left, simulations for small samples, n=10), the likelihood was much reduced with simulations of larger samples (right, n=20). In this example, the problem vanishes with sufficiently large sample e.g., with n of 40 power in either approach was >0.99. The assumptions for simulations were that i) OA is a true effect as experimental studies clearly replicated that observation (see for review), ii) the probability-density-function can be Gaussian described by two parameters: mean(s) and SD(s) for calibrated and fixed T1 and T3, respectively.

Figure 2. A scatterplot showing p-values distribution obtained from 1000 simulated student t tests. Each simulated test compared pain from T1 to T3 interval in offset analgesia (OA) paradigm obtained via calibrated (red) and fixed intensity (blue). In general, the magnitude of OA is huge (see pain that drops over time on Figure 1). Note that the likelihood for not rejecting a null hypothesis is 3.72 times higher if using a fixed approach (left, simulations for small samples, n=10), the likelihood was much reduced with simulations of larger samples (right, n=20). In this example, the problem vanishes with sufficiently large sample e.g., with n of 40 power in either approach was >0.99. The assumptions for simulations were that i) OA is a true effect as experimental studies clearly replicated that observation (see for review), ii) the probability-density-function can be Gaussian described by two parameters: mean(s) and SD(s) for calibrated and fixed T1 and T3, respectively.
Addressing the Dilemma

One possibility to tackle the dilemma is a scenario which implies resource challenges. With a large sample size, it would be possible to divide subjects in subgroups by stimulus intensity as well as by pain intensity. A scenario that does not always seem feasible. Alternatively, it seems logical to implement a variety of stimulus ranges applied in a fixed manner to obtain both types of data: calibrated and fixed. In fact, many studies followed such a design.\(^1,3,7,8\) The reverse is also possible. For example, in a study by Weissman-Fogel et al.\(^7\), participants received tonic noxious heat stimulation calibrated to induce pain of 2, 4 and 6 out of 10. In the end, authors were able to dissect the effect of individualization from the fixed intensity stimulus. Namely, data from subgroups of individual participants that received their comparable stimulus intensities (temperature) were pooled and the absolute effect of temperatures on pain perception was analyzed. As such, a reversed scenario is possible, too. Applying a variety of intensities and subgrouping subjects into different perceptual categories.

Concluding Remarks

In summary, both approaches have their own pros and cons (Appendix 2), which should be considered when deciding on the study design, the aim, and the primary outcome. Whether there is a different variance depending on the pain modality should be systematically investigated in further studies. It is advisable to consider multiple factors (Table 1) when deciding if to individualize or fix the intensity. Apart from the solution based on an application of graded intensities, authors should clearly weight pros and cons in the context of their research question(s). For instance, if the main outcome of interest is objective, the fixed approach can be the method of choice. If the sample size is small, due to e.g., invasiveness of the procedure then calibration might be more suitable. Finally, we do agree that novel pain models and stimulation techniques must be developed that reduce the variability in the pain responses while keeping the intensity constant and since no gold standard exist in this methodological step, it is advisable to conduct reproducibility studies with the unused approach (calibration in the secondary report, if fixed was used in the original study). Furthermore, the awareness of the sample size problem and the generalization of results from a given study must be highlighted and future experimental studies are needed to determine the superiority of one method over another.

Acknowledgments

None.

Supplementary data

Supplementary data related to this article can be found at [https://doi.org/10.1016/j.jpain.2022.07.007](https://doi.org/10.1016/j.jpain.2022.07.007).

References


20. Fust J, Lalouni M, Vadenmark-Lundqvist V, Wärnberg E, Jensen KB: Offset analgesia and onset hyperalgesia with different stimulus ranges. PAIN Reports 6:e914, 2021


34. Levy D, Abdian L, Dekel-Steinkeller M, Defrin R: Experimental evidence for weaker endogenous inhibition of trigeminal pain than extra-trigeminal pain in healthy individuals. Cephalalgia 38:1307-1315, 2018


47. Nickel MM, May ES, Tiemann L, Postorino M, Ta-Dinh S, Ploner M: Autonomic responses to tonic pain are more
closely related to stimulus intensity than to pain intensity. Pain 158:2129-2136, 2017


73. Tansley SN, MacIntyre LC, Diamond L, Sotocinal SG, George N, Meluban L, Austin JS, Coderre TJ, Martin LJ, Mogil JS: Conditioned pain modulation in rodents can feature hyperalgesia or hypoalgesia depending on test stimulus intensity. Pain 160:784-792, 2019


