2020.09.15%20FINAL%20737%20MAX%20Report%20for%20Public%20Release.pdf. [Accessed 19 October 2020]
8. U.S. Food and Drug Administration. *Artificial intelligence and machine learning in software as a medical device*. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device. [Accessed 26 August 2020]
9. Cosgriff CV, Stone DJ, Weissman G, et al. The clinical artificial intelligence department: a prerequisite for success. *BMJ Health Care Inform* 2020; **27**, e100183
10. Liu Y, Chen PC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. *JAMA* 2019; **322**: 1806–16
11. Schamberg G, Badgeley M, Brown EN. Controlling level of unconsciousness by titrating propofol with deep reinforcement learning. In: Michalowski M, Moskovitch R, editors. *Artificial intelligence in medicine*. *AIME*vol. 12299. Cham: Springer; 2020. Lecture Notes in Computer Science

doi: 10.1016/j.bja.2020.10.033
Advance Access Publication Date: 27 November 2020
Published by Elsevier Ltd on behalf of British Journal of Anaesthesia.

Variability in experimental pain studies: nuisance or opportunity?

Victoria J. Madden[1,2,3,*], Peter R. Kamerman[4,5], Mark J. Catley[3], Valeria Bellan[3,6], Leslie N. Russek[3,7], Danny Camfferman[3] and G. Lorimer Moseley[3]

[1]Department of Anaesthesia and Perioperative Medicine, Neuroscience Institute, University of Cape Town, Cape Town, Western Cape, South Africa, [2]Department of Psychiatry and Mental Health, Neuroscience Institute, University of Cape Town, Cape Town, Western Cape, South Africa, [3]IIMPACT in Health, University of South Australia, Adelaide, South Australia, Australia, [4]School of Physiology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, Gauteng, South Africa, [5]School of Pharmacy and Biomedical Sciences, Faculty of Health Sciences, Curtin University, Perth, Western Australia, Australia, [6]Cognitive and Systems Neuroscience Research Hub (CSN-RH), University of South Australia, Adelaide, South Australia, Australia and [7]Department of Physical Therapy, Clarkson University, Potsdam, NY, USA

*Corresponding author. E-mail: torymadden@gmail.com

Keywords: analgesia; pain; statistics; study design; variability

Editor—Pain is a variable experience, even in studies that use controlled nociceptive stimuli in healthy humans.[1] This variability is unsurprising considering that nociception (the neural process of encoding noxious stimuli) and pain (a conscious experience) do not share an isomorphic relationship.[2] Pain is influenced by a broad range of biological, cognitive, contextual, and mood-related factors that may vary from moment to moment.[3] Even a well-controlled experimental laboratory environment presents limited scope to control all these factors, and it is rare for pain to be elicited with high consistency.

The intraindividual variability in pain reporting may reflect important personal features that are relevant to our understanding of pain and the impact of analgesic strategies. With this in mind, several clinical studies have used high frequency pain assessment in longitudinal designs to examine the possibility that intraindividual variability may be relevant to clinical outcomes. They report that intraindividual variability in pain ratings may be related to depression,[4,5] self-efficacy,[4] emotional and physical functioning,[6] and may predict benefits from sham medication or active treatments.[7–9]

Nevertheless, experimental pain research in humans has largely neglected to acknowledge explicitly the importance of intraindividual variability in pain reporting. To our knowledge, only a handful of studies have attended to intraindividual variability.[1,10–14] Instead, the common practice is to analyse averages gained from repeated measurements and thus smooth out variability. Raw data are seldom presented. Thus, intraindividual variability is considered a nuisance, rather than a feature worthy of attention.

Obscuring variability has practical disadvantages. Many experimental studies rely on calibrating stimulus intensities to each individual participant and then assuming that subsequent stimuli are experienced at an intensity that reflects the data from that initial calibration. Intraindividual variability in trial-by-trial pain reports undermines this assumption, and points to contemporary experimental designs that allow for drift in stimulus-response relationships and detect shifts over time in, for example, the effect of one stimulus-response on a subsequent stimulus-response. Accounting for variability in research design or statistical analyses will allow greater confidence in interpreting the effect of analgesic interventions on stimulus-response relationships.

Obscuring variability has external validity disadvantages. The common approach to experimental pain studies is sequential averaging, where the averaged intraindividual

measurements are then averaged across individuals to realise a group average. It reduces variance but makes two flawed assumptions: (1) that the appropriate measure of central location (e.g. arithmetic mean, median) was used to capture the centre of location for the intraindividual variation and (2) that only interindividual variation is important. The cost of sequential averaging is that both intraindividual variation and interindividual variation are lost (Fig. 1). This removes potentially fertile data as to the effect of any analgesic strategy on variability itself – a treatment that decreases variability of a stimulus-response without reducing the average response within individuals will not be differentiated from a treatment that has no effect on either. It may be that decreased variability, or removal of peaks in pain, is a clinically worthwhile outcome. Moreover, sequential averaging has less relevance to clinical practice. Patients with pain typically seek relief of *their own* pain, not of the pain of a group and there may well be no individuals who have this 'average response'.

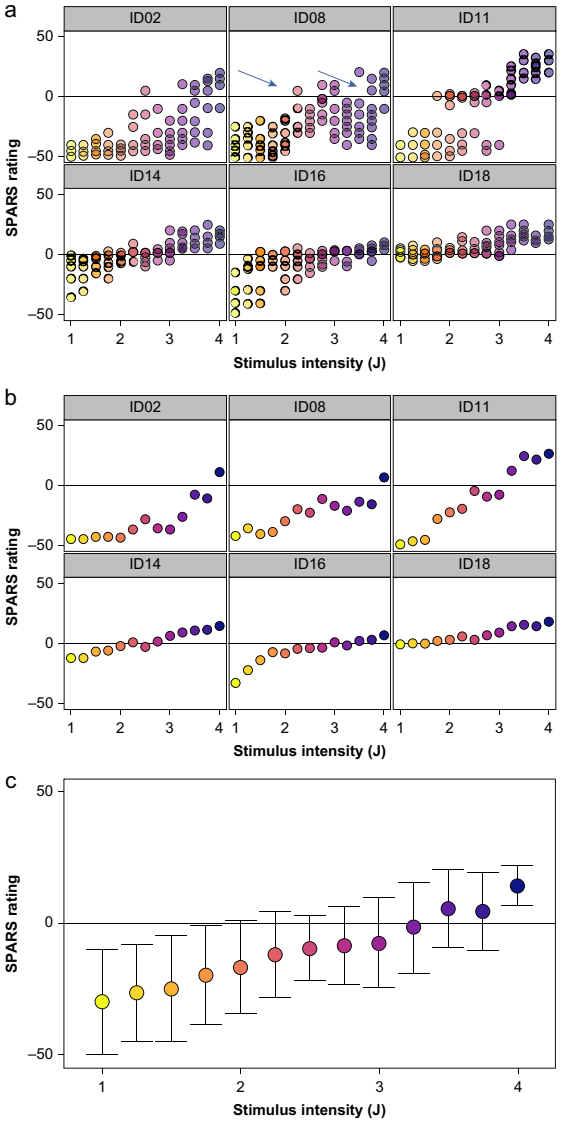

Fig 1. Selected data (from Madden and colleagues[15]) used to illustrate the consequences of averaging. (a) Rating given by six participants at each stimulus intensity. (b) Average rating from the same six participants at each stimulus intensity. (c) Average rating across all six participants at each stimulus intensity. A sequential loss of information is observed across the three panels as variability is obscured by averaging. Moreover, that the two stimuli marked with arrows (a) are associated with identical ratings illustrates the risk that, if only one reading were taken per stimulus intensity, an erroneous result is highly possible.

Obscuring variability has important implications for power calculation simulations that are used when designing complex studies. This has ethical and resource implications because concealing the extent of variance in reporting will lead to underpowered studies. Such studies are known to be associated with spurious findings as to the direction and magnitude of an intervention's effect.[16,17]

We suggest three remedies for experimental studies:

(1) Report variability data at the individual level, whether in the main body of a manuscript or in supplementary data. Ideally, raw data should be made available, and the data displayed using appropriate graphical methods.
(2) Use statistical techniques that allow for individual-level variability in pain reporting. Modern analytical techniques allow for hierarchical nesting of repeated measures within individuals and of individuals within groups. Such an approach would seem warranted when analysing experimental pain reports. Considering that experimental studies on pain lay the groundwork for clinical studies that inform treatments, acknowledging individual fluctuations in pain report in the analysis of experimental data stands to improve translation between experimental and clinical research studies.
(3) Ask participants to report the percept on every stimulation trial, so as to verify the stimulus–response relationship on each trial and verify or modify calibrations, or evaluate drift and fluctuation in stimulus–response over repeated trials.

Experimental examination of intraindividual variability in pain responses should improve clinical phenotyping. There have been calls to consider matching treatment to patients defined by their responses to psychophysical phenotyping procedures.[18–21] In the context of the growing clinical application of these phenotyping procedures to inform treatment decisions, clarifying the true variability in individual responses is of some urgency. For example, it is unclear how the three-replicate method used for some modalities in the DFNS Quantitative Sensory Testing paradigm[22] was derived, and whether it is sufficient. Further, the substantial overlap of ratings that can exist across a wide range of stimulus intensities (as visible in Fig. 1a) indicates that, in some cases, three replicates may be too few for confidence.

The relevance to clinically meaningful outcomes of intraindividual variability when reporting pain could be clarified with appropriate designs. For example, it would be interesting to understand whether intraindividual variability differs systematically between people who recover from a painful episode and those who do not. That is, does variability have utility as a risk factor for chronicity? We recently speculated that the variability with which an individual rates different, intensity-matched stimulation trials could reflect the flexibility of his/her perceptual processing[15] and could represent a

beneficial feature of a sensory processing system. Indeed, preliminary data infer the possibility that pain thresholds are more variable in people who do not have persistent pain than in people who do.[23] The idea that variability could be beneficial also loosely reflects the recent recognition that cognitive flexibility may be beneficial with regard to persistent pain.[24] The variability with which an individual reports on painful experimental stimuli could represent a useful biomarker of risk or vulnerability, or actual or predicted response to treatment. We speculate that clinical variables relevant to perioperative care (e.g. anxiety) may also influence variability in clinical pain report, in which case addressing those influential variables could yield beneficial outcomes. Experimental designs which study variability in pain reports and how that variability relates to other, clinically relevant individual features, may be the method of choice for these and other lines of inquiry.

Fully acknowledging and reporting intraindividual variability in reports of experimentally induced pain, and handling such variability appropriately, stand to move the field closer to understanding pain. Individual variability in stimulus-response may reflect important phenotypic features to inform clinical subgrouping and predict treatment outcomes. If the role of experimental work in humans is to clarify the contributions of various variables to pain outcomes, acknowledging intraindividual variability will be key to optimising the translation of experimental research to the clinical context.

Declarations of interest

VJM receives speaker's fees for lectures on pain and rehabilitation. PK is on retainer for Partners in Research, and receives speaker's fees for lectures and professional development courses on pain. GLM has received support from Pfizer, Kaiser Permanente, Workers' Compensation Boards in Australia, Europe, and North America, the International Olympic Committee, the Port Adelaide Football Club, and the Arsenal Football Club. He receives royalties for books on pain and speaker's fees for talks and professional development courses on pain and rehabilitation. The authors declare no other competing interests related to this work.

References

1. Rosier EM, Iadarola MJ, Coghill RC. Reproducibility of pain measurement and pain perception. *Pain* 2002; **98**: 205–16
2. Wall PD, McMahon SB. The relationship of perceived pain to afferent nerve impulses. *Trends Neurosci* 1986; **9**: 254–5
3. Moseley G, Butler D. *Explain pain supercharged*. Adelaide: Noigroup Publications; 2017
4. Schneider S, Junghaenel DU, Keefe FJ, et al. Individual differences in the day-to-day variability of pain, fatigue, and well-being in patients with rheumatic disease: associations with psychological variables. *Pain* 2012; **153**: 813–22
5. Zakoscielna KM, Parmelee PA. Pain variability and its predictors in older adults: depression, cognition, functional status, health, and pain. *J Aging Health* 2013; **25**: 1329–39
6. Schneider S, Junghaenel DU, Ono M, et al. Temporal dynamics of pain: an application of regime-switching models to ecological momentary assessments in patients with rheumatic diseases. *Pain* 2018; **159**: 1346–58
7. Harris RE, Williams DA, McLean SA, et al. Characterization and consequences of pain variability in individuals with fibromyalgia. *Arthritis Rheum* 2005; **52**: 3670–4
8. Farrar JT, Troxel AB, Haynes K, et al. Effect of variability in the 7-day baseline pain diary on the assay sensitivity of neuropathic pain randomized clinical trials: an ACTTION study. *Pain* 2014; **155**: 1622–31
9. Martini CH, Yassen A, Krebs-Brown A, et al. A novel approach to identify responder subgroups and predictors of response to low- and high-dose capsaicin patches in postherpetic neuralgia. *Eur J Pain* 2013; **17**: 1491–501
10. van Wijk AJ, Lobbezoo F, Hoogstraten J. Reliability and validity of a continuous pain registration procedure. *Eur J Pain* 2013; **17**: 394–401
11. Boly M, Balteau E, Schnakers C, et al. Baseline brain activity fluctuations predict somatosensory perception in humans. *Proc Natl Acad Sci U S A* 2007; **104**: 12187–92
12. Quiton RL, Greenspan JD. Across- and within-session variability of ratings of painful contact heat stimuli. *Pain* 2008; **137**: 245–56
13. Schulz E, Tiemann L, Schuster T, et al. Neurophysiological coding of traits and states in the perception of pain. *Cereb Cortex* 2011; **21**: 2408–14
14. Svensson P, Bjerring P, Arendt-Nielsen L, et al. Comparison of four laser types for experimental pain stimulation on oral mucosa and hairy skin. *Lasers Surg Med* 1991; **11**: 313–24
15. Madden VJ, Kamerman PR, Catley MJ, et al. Rethinking pain threshold as a zone of uncertainty. *bioRxiv* June 10 2019. https://doi.org/10.1101/521302
16. Gelman A, Carlin J. Beyond power calculations: assessing Type S (sign) and Type M (magnitude) errors. *Perspect Psychol Sci* 2014; **9**: 641–51
17. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013; **14**: 365–76
18. Edwards RR, Dworkin RH, Turk DC, et al. Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. *Pain* 2016; **157**: 1851–71
19. Baron R, Förster M, Binder A. Subgrouping of patients with neuropathic pain according to pain-related sensory abnormalities: a first step to a stratified treatment approach. *Lancet Neurol* 2012; **11**: 999–1005
20. Giesecke T, Williams DA, Harris RE, et al. Subgrouping of fibromyalgia patients on the basis of pressure-pain thresholds and psychological factors. *Arthritis Rheum* 2003; **48**: 2916–22
21. Vollert J, Maier C, Attal N, et al. Stratifying patients with peripheral neuropathic pain based on sensory profiles: algorithm and sample size recommendations. *Pain* 2017; **158**: 1446–55
22. Rolke R, Baron R, Maier C, et al. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. *Pain* 2006; **123**: 231–43

23. Lund I, Lundeberg T, Kowalski J, et al. Evaluation of variations in sensory and pain threshold assessments by electrocutaneous stimulation. *Physiother Theory Pract* 2005; **21**: 81–92

24. McCracken LM, Morley S. The psychological flexibility model: a basis for integration and progress in psychological approaches to chronic pain management. *J Pain* 2014; **15**: 221–34

doi: 10.1016/j.bja.2020.11.005
Advance Access Publication Date: 17 December 2020

Nociception level-guided fentanyl titration: potential impact of multimodal anaesthesia and false positives. Comment on *Br J Anaesth* 2020; 125: 1070–8

Sean Coeckelenbergh[1,2,*], Hugo E. M. Vereecke[3,4] and Philippe Richebé[5,6]

[1]Department of Anaesthesiology and Perioperative Medicine, Ghent University Hospital, Ghent University, Ghent, Belgium, [2]Department of Anaesthesiology, Erasme University Hospital, Université Libre de Bruxelles, Brussels, Belgium, [3]Department of Anaesthesiology, University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands, [4]Department of Anaesthesiology, AZ Sint-Jan Brugge-Oostende AV, Brugge, Belgium, [5]Department of Anaesthesiology and Pain Medicine, University of Montreal, Montreal, QC, Canada and [6]Maisonneuve-Rosemont Hospital, CIUSSS de l'Est de l'Ile de Montreal, University of Montreal, Montreal, QC, Canada

*Corresponding author. E-mail: sean.coeckelenbergh@ulb.be

Keywords: analgesia; dexamethasone; goal-directed therapy; monitoring; multimodal analgesia; nociception; opioid; pain

Editor—Meijer and colleagues[1] recently reported in the *British Journal of Anaesthesia* a two-centre RCT investigating the capacity of the nociception level (NOL) index to guide fentanyl administration during laparoscopic/robotic abdominal surgery. They reported that this strategy significantly decreased postoperative pain scores during the first 90 min of recovery in the PACU. In accordance to the work of Funcke and colleagues,[2] the stress response, which Meijer and colleagues measured through serial analyses of serum cortisol and adrenocorticotropic hormone levels, decreased significantly in the NOL index group. This study shows potential advantages of guiding intraoperative fentanyl dosing based on the NOL index. However, we think certain factors may have had an impact on the studied outcomes and that they should be clarified.

(i) It is unclear if dexamethasone, NSAIDs, or any other components of multimodal analgesia were administered. The authors explicitly state that patients were scheduled for surgery 'without epidural anaesthesia, local blocks, or infiltration', removing useful adjuvants to control postoperative pain. With the exception of paracetamol, their anti-nociception protocol seems to be exclusively opioid based (i.e. fentanyl-boluses-maintained NOL index or haemodynamic targets, remifentanil infusion if fentanyl was insufficient, and a transition dose for all patients of either piritramide or morphine at the end of surgery). Despite such a strong opioid strategy, rather high pain scores were observed in the control group. Dexamethasone, an NSAID, and trocar site infiltration are in many centres standard care, and could have possibly improved the immediate postoperative baseline conditions in both groups and led to lower initial pain scores, less nausea, and a modified stress response. Dexamethasone, for example, has become an essential perioperative drug, as its prophylactic administration is linked to decreased postoperative pain, nausea, and vomiting.[3] The majority of patients in this study were women, and all patients were expected to require postoperative opioids. Yet, the incidence of nausea were 28% and 36% of the patients in the NOL-guided and standard care groups, respectively, which suggests that dexamethasone was not administered. Furthermore, dexamethasone could have had an impact on the outcome of stress hormone release, one of their statistically significant findings.

(ii) Another point of discussion is the possibility of false-positive NOL index values (i.e. NOL index values >25 despite adequate anti-nociception). Three patients required remifentanil in addition to fentanyl because either blood pressure (standard care group) or NOL index (NOL-guided group) remained above target. Perhaps targets were not reached in the NOL-guided group because factors other than nociception caused the index to increase. High arterial CO_2, which can occur during laparoscopic surgery, leads to increased sympathetic tone and may cause arrhythmias.[4] In addition, plethysmographic variation has been used to predict fluid responsiveness,[5] and such variations could perhaps influence another parameter of the NOL index: photoplethysmogram amplitude. Non-nociceptive-related changes in the NOL index are possible, for example, after a bolus of phenylephrine,[6] and clinicians should be aware of potential cofounders.

Meijer and colleagues[1] showed several benefits of personalising intraoperative anti-nociception. Most notably, they were able to target outliers who required higher or lower