Published in final edited form as:

Pain. 2016 July ; 157(7): 1575–1576. doi:10.1097/j.pain.0000000000000546.

Reply to: What reliability can and cannot tell us about pain report and pain neuroimaging

Janelle E. Letzen, M.S.^a, Jeff Boissoneault, Ph.D.^a, Landrew S. Sevel, M.S.^a, and Michael E. Robinson, Ph.D.^a

^aDepartment of Clinical and Health Psychology, University of Florida, Gainesville, FL

As noted in Woo and Wager's commentary [27], we agree that pain neuroimaging has been invaluable in understanding the neurobiological processes underlying pain perception. Given the heterogeneity across individuals' pain experience, we further agree with Woo and Wager that neuroimaging helps answer mechanistic questions about these individual differences and, importantly, how they pertain to clinical pain conditions. Indeed, we side with the authors that the true value of neuroimaging-based biomarker studies is not in replacing pain self-report, but rather in addressing dynamic interactions among behavioral and neurobiological factors contributing to the experience of pain.

Yet, most studies proposing neuroimaging-based biomarkers have focused on their ability to classify individuals into clinical or non-clinical groups, or ability to predict pain intensity ratings, rather than provide novel explanatory information [1,3,6,10,11,19,22]. Woo and Wager [27] did not report on the fact that there are a number of well-validated self-report measures assessing multidimensional aspects of pain that go well beyond simple measures of pain intensity [2,4,5,7,8,13,15,17]. It should be noted that the impetus for neuroimaging-based biomarker studies generated from the notion that self-report is a flawed measure, and a barrier to effective diagnosis and treatment [1,12,22,24]. Specifically, pain self-report has been termed "unreliable" [12], even though predictable variability in pain ratings has been widely demonstrated [9,16,18,20,21,23,25,26]. The reliability or reproducibility of neural pain markers, on the other hand, remains undetermined. We agree that reliability is only one psychometric property to be considered in neuroimaging biomarker development. However, because this property places a cap on validity, and consequently applicability, we strongly reiterate that it is potentially the most important property to establish [14].

Woo and Wager [27] correctly explain that systematic factors influencing between-person variability are not considered error and can positively impact reliability. A key point here is to be sure to recognize the difference between variability and reliability. Variability that is neither systematic nor predictable is error, and thus leads to lower reliability. Increased systematic, predictable variability can indeed lead to increased reliability. To this end, the authors imply, through examples, that the reliability of pain ratings and neuroimaging data might be higher due to contextual factors (e.g., effect of experimenter clarity in providing

Corresponding Author: Michael E. Robinson, PhD, University of Florida, 101 South Newell Drive, Rm 3151, P.O. Box 100165, Gainesville, FL 32610-9165, Tel: 325-273-5220, Fax: 352-273-6156, merobin@ufl.edu.

There is no conflict of interest among authors.

instructions and undesirable data quality factors, respectively), whereas consistencies across individuals in activation of a certain brain region would decrease reliability. We would like to point out that these circumstances are examples of poor experimental design, or poor standardization of procedures, just as likely to influence neuroimaging as any other measure. In these cases, the reliability of such measures is not artificially biased, but attributable to a set of circumstances (e.g., low measurement error, adequate range, sufficient variability). Additionally, excessive homogeneity (i.e., lack of range or variability) will lead to lower reliability of any measure, including both self-report and neuroimaging-based measures.

They further note that psychosocial factors, such as cultural acceptance of pain expression, can systematically affect how participants rate pain, resulting in reliability differences between heterogeneous vs. homogenous samples. If neuroimaging markers are truly representative of the symptom in question, then they should be similarly influenced by the same factors that influence self-report. Neural responses are not immune to behavioral factors that affect pain (which by definition, is a subjective experience), and there is a large body of literature demonstrating such a relationship, as noted above. We caution against relying on a dualistic perspective of pain perception (i.e., pain experienced in the brain vs. by the self). Reliability influenced by error, on the other hand, would account for incongruent reliability comparing self-report and neuroimaging data.

Regarding reliability in machine-learning models, we agree that these models should be optimized to include the most relevant and reliable features. However, to our knowledge, the impact of feature selection on model reliability has not been assessed in pain neuroimaging studies. Furthermore, the relevance of the features selected in optimized models is based upon their relationship to the criterion variable or label of interest (i.e., pain self-report). Given the numerous additional constructs that do influence pain, as cited by Woo and Wager [27], perhaps it would be more valid to construct individual component signatures of each of these subprocesses, with features selected based upon their relevance to that component process (e.g., pain-related communication tendency).

Our study examined brain regions that had an empirical, peer-reviewed, published basis for selection, and not an exploratory, whole brain analysis. While we do agree that a focus on large-scale, distributed networks is advantageous in refining our understanding of the neurobiological aspects of pain, we do not believe that this approach should come at the expense of *a priori* hypothesis testing, even in the context of marker discovery. Doing so comes along with the risk of capitalizing on spurious relationships (from the 5 trillion potential paired connections, as pointed out by Woo and Wager) and noise in the data, which again highlights the utility of reliability and validity assessment. The value of brain markers may indeed be in measuring component neurophysiological processes in pain, but this does not make them exempt from standard assessments of psychometric utility.

Finally, we have always contended that reliability is not a metric of what pain ratings actually measure. Reliability is the minimal basis upon which other measurement properties (e.g., validity) are based. Without good reliability, validity is compromised and limited. Woo and Wager [27] neglect to discuss that there are many well-validated self-report measures of the components of pain report, and experimental manipulations that capture what these

ratings actually measure. The authors' statements such as "Reliability: More is not always better" and "The value of brain markers, whether reliable or not, is in measuring neurophysiological processes that are closely and consistently aligned with particular 'ingredients' of pain..." are misleading. If not reliable, neuroimaging-based markers, no matter how computationally sophisticated, cannot be "closely and consistently" aligned with anything.

Acknowledgments

This research was funded by the National Center for Complementary and Integrative Health through grants to MER (R01AT001424-05A2) and JEL (F31AT007898-03), as well as the National Institute of Nursing Research to MER (1R01NR015314-01A1).

References

1. Apkarian AV, Hashmi JA, Baliki MN. Pain and the brain: specificity and plasticity of the brain in clinical chronic pain. *Pain*. 2011; 152:S49–S64. [PubMed: 21146929]
2. Bennell K, Bartam S, Crossley K, Green S. Outcome measures in patellofemoral pain syndrome: test retest reliability and inter-relationships. *Phys. Ther. Sport*. 2000; 1:32–41.
3. Callan D, Mills L, Nott C, England R, England S. A Tool for Classifying Individuals with Chronic Back Pain: Using Multivariate Pattern Analysis with Functional Magnetic Resonance Imaging Data. *PLoS One*. 2014; 9:e98007. Available: <http://dx.doi.org/10.1371/journal.pone.0098007>. [PubMed: 24905072]
4. Ferrell BA, Stein WM, Beck JC. The Geriatric Pain Measure: validity, reliability and factor analysis. *J. Am. Geriatr. Soc*. 2000; 48:1669–1673. [PubMed: 11129760]
5. Grafton KV, Foster NE, Wright CC. Test-Retest Reliability of the Short-Form McGill Pain Questionnaire: Assessment of Intraclass Correlation Coefficients and Limits of Agreement in Patients With Osteoarthritis. *Clin. J. Pain*. 2005; 21 Available: http://journals.lww.com/clinicalpain/Fulltext/2005/01000/Test_Retest_Reliability_of_the_Short_Form_McGill.9.aspx.
6. Green, aL; Wang, S.; Stein, JF; Pereira, EaC; Kringelbach, ML.; Liu, X.; Brittain, JS.; Aziz, TZ. Neural signatures in patients with neuropathic pain. *Neurology*. 2009; 72:569–571. [PubMed: 19204269]
7. Gronblad M. Interrelation and Test—Retest Reliability of the Pain Disability Index (PDI) and the Oswestry Disability. *Clin J Pain*. 1993; 9:189–195. [PubMed: 8219519]
8. Jensen MP. The validity and reliability of pain measures in adults with cancer. *J. Pain*. 2003; 4:2–21. [PubMed: 14622723]
9. Koyama T, McHaffie JG, Laurienti PJ, Coghill RC. The subjective experience of pain: where expectations become reality. *Proc. Natl. Acad. Sci. U. S. A*. 2005; 102:12950–12955. [PubMed: 16150703]
10. Lin C-S. Brain signature of chronic orofacial pain: a systematic review and meta-analysis on neuroimaging research of trigeminal neuropathic pain and temporomandibular joint disorders. *PLoS One*. 2014; 9:e94300. [PubMed: 24759798]
11. Loggia ML, Kim J, Gollub RL, Vangel MG, Kirsch I, Kong J, Wasan AD, Napadow V. Default mode network connectivity encodes clinical pain: an arterial spin labeling study. *PAIN®*. 2013; 154:24–33. [PubMed: 23111164]
12. Lu Y, Klein GT, Wang MY. Can pain be measured objectively? *Neurosurgery*. 2013; 73:24–25.
13. Margolis RB, Chibnall JT, Tait RC. Test-retest reliability of the pain drawing instrument. *Pain*. 1988; 33:49–51. [PubMed: 3380550]
14. Mayeux R. Biomarkers: potential uses and limitations. *NeuroRx*. 2004; 1:182–188. [PubMed: 15717018]
15. Osman A, Breitenstein JL, Barrios FX, Gutierrez PM, Kopper BA. The Fear of Pain Questionnaire-III: further reliability and validity with nonclinical samples. *J. Behav. Med*. 2002; 25:155–173. [PubMed: 11977436]

16. Schmid J, Bingel U, Ritter C, Benson S, Schedlowski M, Gramsch C, Forsting M, Elsenbruch S. Neural underpinnings of placebo hyperalgesia in visceral pain: A fMRI study in healthy volunteers. *Neuroimage*. 2015; 120:114–122. [PubMed: 26123378]
17. Swinkels-Meewisse E, Swinkels R, Verbeek ALM, Vlaeyen JWS, Oostendorp RAB. Psychometric properties of the Tampa Scale for kinesiophobia and the fear-avoidance beliefs questionnaire in acute low back pain. *Man. Ther.* 2003; 8:29–36. [PubMed: 12586559]
18. Tang NKY, Salkovskis PM, Hodges A, Wright KJ, Hanna M, Hester J. Effects of mood on pain responses and pain tolerance: an experimental study in chronic back pain patients. *Pain*. 2008; 138:392–401. [PubMed: 18325674]
19. Ung H, Brown JE, Johnson KA, Younger J, Hush J, Mackey S. Multivariate Classification of Structural MRI Data Detects Chronic Low Back Pain. *Cereb. Cortex*. 2012
20. Vase L, Robinson ME, Verne GN, Price DD. The contributions of suggestion, desire, and expectation to placebo effects in irritable bowel syndrome patients: An empirical investigation. *Pain*. 2003; 105:17–25. [PubMed: 14499416]
21. Villemure C, Bushnell MC. Mood influences supraspinal pain processing separately from attention. *J. Neurosci*. 2009; 29:705–715. [PubMed: 19158297]
22. Wager TD, Atlas LY, Lindquist Ma, Roy M, Woo C-W, Kross E. An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med*. 2013; 368:1388–1397. [PubMed: 23574118]
23. Wagner G, Koschke M, Leuf T, Schlösser R, Bär K-J. Reduced heat pain thresholds after sad-mood induction are associated with changes in thalamic activity. *Neuropsychologia*. 2009; 47:980–987. [PubMed: 19027763]
24. Wartolowska K. How neuroimaging can help us to visualise and quantify pain? *Eur. J. Pain Suppl*. 2011; 5:323–327.
25. Wiech K, Tracey I. The influence of negative emotions on pain: behavioral effects and neural mechanisms. *Neuroimage*. 2009; 47:987–994. [PubMed: 19481610]
26. Wiech K, Vandekerckhove J, Zaman J, Tuerlinckx F, Vlaeyen JWS, Tracey I. Influence of prior information on pain involves biased perceptual decision-making. *Curr. Biol*. 2014; 24:R679–R681. [PubMed: 25093555]
27. Woo C-W, Wager TD. What reliability can and cannot tell us about pain report and pain neuroimaging. *Pain*. 2016