Neuroimaging-based biomarker discovery and validation

Choong-Wan Woo¹,²*, Tor D. Wager¹,²

¹Department of Psychology and Neuroscience, University of Colorado, Boulder
²Institute of Cognitive Science, University of Colorado, Boulder

* Please address correspondence to:
Choong-Wan Woo
Department of Psychology and Neuroscience
University of Colorado Boulder
345 UCB, Boulder, CO 80305
Email: Choongwan.Woo@Colorado.Edu
Telephone: (720) 443-3640
Though developing biological markers for chronic pain has been a major goal of the field for decades, such biomarkers have not yet made their way into clinical practice. However, given the potential uses of biomarkers in multiple aspects of prevention and treatment—such as pain and risk factor assessment, diagnosis, prognosis, treatment selection, drug discovery, and more—efforts to discover new pain biomarkers have been expanding [5; 6; 8; 30].

Recent advances in human neuroimaging, including functional and structural Magnetic Resonance Imaging (fMRI/sMRI) combined with machine learning techniques, are bringing us closer to the goal of developing objective, brain-based markers of the neural functions and neuropathology that underlie chronic pain [2; 7; 25; 33]. These brain measures are particularly promising as biomarkers for chronic pain. Though pain is reliably induced by peripheral nociceptive input, many forms of chronic pain may arise from neuropathology in the supra-spinal circuits that govern the construction of pain experience and long-term motivation [1; 14; 26; 32].

Particularly, structural neuroimaging measures could provide more stable markers of neuropathology of chronic pain, including stable features underlying pain risk and resilience [2; 3; 11; 19; 28; 29]. Gray-matter changes have also been associated with a number of conditions that are often co-morbid with chronic pain, including depression [4; 22; 24], stress [10; 12; 20], post-traumatic stress disorder [17; 21; 27], and early-life adversity [13; 18; 23; 31]. Therefore, structural measures may provide important clues about supra-spinal contributions to both pain and related risk factors (Figure).

In this issue, Labus et al. [16] developed a new neuroimaging biomarker for irritable bowel syndrome (IBS) using structural MRI data, based on a relatively large sample of 80 IBS patients and 80 healthy controls. They used sparse Partial Least Squares-Discriminant Analysis (sPLS-DA), a method that allowed them to both develop a classification model based on brain structure and identify the regions that make the most important contributions to the classification. They subsequently tested the predictive model on a “holdout” sample of 26 IBS patients and 26 healthy controls. The model discriminated
patients from controls with 70% accuracy (compared to a chance accuracy of 50%), providing a moderate but reliable morphological brain signature for IBS.

Rather than being the end of the story, this study serves as a starting point for biomarker discovery and validation. Like other brain ‘signatures’ [30], the signature they identified can become a ‘research product’ that can be tested on multiple samples from different laboratories, and validated or challenged in various ways. The more the marker for IBS status or IBS risk holds up to the scrutiny of being characterized across samples and populations, the more useful it will become.

Importantly, there is a set of desirable characteristics that a useful neuroimaging biomarker should demonstrate throughout the biomarker development process. We briefly describe several such characteristics (summarized in Table), and then relate them to the findings of Labus et al. [16].

**Criterion 1. Diagnosticity**

Good biomarkers should produce high diagnostic performance in classification or prediction. Diagnostic performance can be evaluated by sensitivity and specificity. Sensitivity concerns whether a model can correctly detect signal when signal exists. Effect size is a closely related concept; larger effect sizes are related to higher sensitivity. Specificity concerns whether the model produces negative results when there is no signal. Specificity can be evaluated relative to a range of specific, alternative conditions that may be confusable with the condition of interest.

**Criterion 2. Interpretability**

Brain-based biomarkers should be meaningful and interpretable in terms of neuroscience, including prior neuroimaging studies and converging evidence from multiple sources (e.g., animal models, lesion studies, etc.). One potential pitfall in developing neuroimaging biomarkers is that classification or prediction models can capitalize on confounding variables that are not neuroscientifically meaningful or interesting at all (e.g., in-scanner head movement [9]). Therefore, neuroimaging biomarkers should be evaluated and interpreted in the light of existing neuroscientific.
Criterion 3. Deployability

Once the classification or outcome-prediction model has been developed as a neuroimaging biomarker, the model and the testing procedure should be precisely defined so that it can be prospectively applied to new data. Any flexibility in the testing procedures could introduce potential over-optimistic biases into test results, rendering them useless and potentially misleading. For example, “amygdala activity” cannot be a good neuroimaging biomarker without a precise definition of which ‘voxels’ in the amygdala should be activated and the relative expected intensity of activity across each voxel. A well-defined model and standardized testing procedure are crucial aspects of turning neuroimaging results into a ‘research product,’ a biomarker that can be shared and tested across laboratories.

Criterion 4. Generalizability

Clinically useful neuroimaging biomarkers aim to provide predictions about new individuals. Therefore, they should be validated through prospective testing to prove that their performance is generalizable across different laboratories, different scanners or scanning procedures, different populations, and variants of testing conditions (e.g., other types of chronic pain). Generalizability tests inherently require multi-study and multisite efforts. With precisely defined model and standardized testing procedure (Criterion 3), we can easily test the generalizability of biomarkers and define the boundary conditions under which they are valid and useful.

Evaluating the neuroimaging biomarker for IBS by Labus et al.

We hope that more studies will use criteria such as those described above to evaluate existing and new biomarkers. Here, we apply our criteria to Labus et al.’s new neuroimaging biomarker for IBS, and in so doing point towards some opportunities for future development.

Criterion 1. Labus et al.’s sMRI signature for IBS showed 68% sensitivity, 71% specificity, and 70% classification accuracy in holdout test data. While this accuracy level is similar to other brain structure-based tests (e.g., 73% accuracy in [2]), it is not high enough
to be used as a biomarker for IBS, as Labus et al. acknowledged. However, the signature could be still useful as a marker for a potential risk factor for IBS, in combination with other measures, or as a probe for resilience given a brain propensity for IBS. There are avenues for potential improvement, including refinement of the algorithm, generation and selection of important brain features, data quality control, multi-modal assessment, and refined phenotyping (i.e., using multiple functional or symptomatic outcomes rather than diagnostic categories). Labus et al. tested healthy controls, but later studies could also evaluate specificity relative to other types of chronic pain or other mental health conditions that may share similar brain features (e.g., depression).

**Criterion 2.** Through stability analysis and variable importance in projection scores, Labus et al. tried to obtain an interpretable classification model and discussed the brain findings based on prior literature. However, we still need more evidence to fully understand the roles of these brain structures in IBS or chronic pain broadly, and to know which features are related to pain versus other co-morbid risk factors. Converging evidence from different approaches (e.g., fMRI, animal models) and across patient groups will be helpful.

**Criterion 3.** Labus et al. developed their model on 160 participants, and then they applied the *a priori* model on new holdout test data. They also reported precise model weights. These are strong features of the study. The research community could facilitate deployment across laboratories and patient groups using new innovations in technology. For example, Labus et al. could provide an online platform where researchers can upload structural images that they want to test the signature on, and signature scores can be sent to the researchers. In this way, we can maximize ease of deployment and minimize the number of choices researchers can make in applying Labus et al.’s results to their data.

**Criterion 4.** Like the vast majority of studies, Labus et al. used data only from one laboratory and one scanner. However, importantly, they used data obtained from six different acquisition sequences, which could help generalize their findings across different sequences. They included only female participants in this study, so the results cannot be generalized to males and/or different types of visceral pain disorders yet.
Therefore, the next step could include testing their *a priori* signature on new data from different laboratories and different scanners, and also on male participants and patients with other chronic pain conditions (including other types of chronic *visceral* pain and other types of chronic pain, such as chronic low back pain).

**Conclusion**

Labus et al. [16] took an exciting step toward a neuroimaging biomarker for IBS, and more broadly, chronic visceral pain. Taking Labus et al. as a starting point, collaborative, multi-site efforts will help facilitate the biomarker development process, particularly focusing on the criteria above. We believe that Pain and Interoception Imaging Network (PAIN; painrepository.org [15]) repository will provide great resources for the biomarker discovery and validation process.

**Conflict of interest statement**

The authors have no conflicts of interest or financial relationships relevant to this commentary to disclose.

**Acknowledgments**

This work was funded by NIDA R01DA035484-01 (T.D.W., P.I.).
Figure Legend

**Figure.** Key common brain regions that show structural changes across different conditions related to chronic pain, including depression, stress, post-traumatic stress disorder (PTSD), and early-life adversity.

Table Legend

**Table.** Desirable characteristics of neuroimaging biomarkers

References


**Table.** Desirable characteristics of neuroimaging biomarkers

<table>
<thead>
<tr>
<th>Development Stages</th>
<th>Criteria</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery</td>
<td>1 Diagnosticity</td>
<td>Sensitivity: Positive results when there is signal, effect size</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specificity: Negative results when there is no signal</td>
</tr>
<tr>
<td></td>
<td>2 Interpretability</td>
<td>Neuroscientifically interpretable model</td>
</tr>
<tr>
<td>Validation</td>
<td>3 Deployability</td>
<td>Precisely defined model and standardized testing procedure (well-described,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clear and easy to deploy across research groups/clinics)</td>
</tr>
<tr>
<td></td>
<td>4 Generalizability</td>
<td>Generalizable results across different laboratories, scanners, populations,</td>
</tr>
</tbody>
</table>