**Test-retest reliability of a multivariate brain pattern in task-fMRI:** an analysis of the Neurological Pain Signature (NPS)

**Abstract** (144 words)

There is an increasing movement towards establishing neuroimaging-based biomarkers with pattern recognition techniques related to mental processes. As the field gets serious about using the brain to predict behavior and health outcomes, reliability and other measurement properties will become increasingly important. We tested one widely validated marker - the Neurologic Pain Signature (NPS), a multivariate brain measure trained to track pain induced by nociceptive input, across more than 17,000 fMRI images related to single-stimulus epochs collected in over 320 participants from 9 studies. NPS showed a decent performance in test-retest reliability across different studies (ICC = 0.83 on average, ranging from 0.68 to 0.91). Our analyses also revealed several factors that might influence the test-retest reliability, such as trial numbers, the effect size of the stimuli and the measurement calculation strategy; while other factors had barely influence, such as the time interval between sessions.

**Introduction** (655 words)

In functional magnetic resonance imaging (fMRI) studies, there is a growing move toward developing neuroimaging-based biomarkers for clinical purposes. Large samples combined with machine learning techniques have made the goal finally seem within reach (Woo et al., 2017). As the field gets serious about using the brain to predict health outcomes, efforts to establish reliability of candidate biomarkers will become more and more important (Elliott et al., 2020). Reliability is the ability of a measure to give consistent results under similar conditions. It places an upper limit on the validity of a measurement and constraints on the utility of a measurement for assessing individual differences. Previous studies have evaluated the test-retest reliability of task fMRI measures based on the traditional mass-univariate analyses, and demonstrated to be not ideal in general. For example, Elliott et al (2020) showed converging evidence of the poor test-retest reliability of the common task-fMRI measures based on the average of univariate activities in the regions of interest (ROIs). The test-retest reliability of the emerging multivariate brain patterns developed by the machine learning algorithms, however, hasn't been examined systematically.

Several significant differences between univariate measures and multivariate brain patterns might influence their performance in the test-retest reliability. Firstly, univariate measures typically have not specified the location and topography with sufficient precision, thus limiting direct replications. For example, in meta-analyses, there is no consensus regarding how close findings should be considered as replications. By contrast, multivariate pattern signatures can specify a precise set of voxels and the topography of the relative expected activity levels across voxels, providing a basis for exact replication. Secondly, region of interest analyses, i.e., average of univariate responses within individual brain regions, are usually large, encompassing neurons with different functions, which dilutes signal and reduces their functional specificity. The multivariate pattern signatures, however, could capture fine-grained functional organization and can more accurately predict perceptions and behaviours. Lastly, traditional univariate analyses mainly focus on which local regions highly related to stimuli or clinical outcomes. However, many features of neurologic and psychiatric disorders (e.g., pain, negative emotions, cognitive and social processes) are more likely encoded in distributed neural systems involving networks of many regions (Wager et al., 2013; Chang et al., 2015). In such situations, multivariate pattern signatures that could integrate contributions from different brain areas will likely be required for accurate prediction. In sum, all these differences indicate the test-retest reliability of the traditional univariate measures might be not applied to the multivariate pattern signatures, which is worthy to be examined systematically.

In the current study, we examined the test-retest reliability of one widely validated multivariate pattern signature - the Neurologic Pain Signature (NPS) - whose weights in each voxel are optimized to be maximally predictive of pain based on fMRI signal (Wager et al., 2013). The NPS could accurately predict pain experience in response to noxious thermal, mechanical, and electrical stimuli, but does not respond to non-noxious warm stimuli, threat cues, social rejection-related stimuli, observed pain, or aversive images. The signature does not measure a disorder, but rather a basic mental process regarding negative sensory and affective processes, which can serve as intermediate features that are altered in various disorders. For example, Lopez-Sola et al. (2016) found that enhanced NPS responses, combined with another brain signature related to non-painful sensory processing, discriminated fibromyalgia from pain-free controls with 93% accuracy. Before testing the reliability of NPS, we first validate the NPS response to pain stimuli by examining the effect size of NPS in response to noxious thermal stimuli in a large dataset - the single-trial dataset - with 9 published studies, more than 17,000 fMRI images related to single-stimulus epochs collected in over 320 participants. We then tested the reliability of the internal consistency of NPS in the single-trial dataset, with the first and second half trials collected within one session. Lastly, we examined the test-retest reliability of NPS with data collected in multiple sessions of another two unpublished studies.