Journal of Neuroscience Methods 253 (2015) 38–46

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Clinical neuroscience

# Test–retest reliability of evoked heat stimulation BOLD fMRI

Jaymin Upadhyay[a,b,1], Jordan Lemme[b,1], Julie Anderson[a,b], David Bleakman[a,c], Thomas Large[a,d], Jeffrey L. Evelhoch[a,e], Richard Hargreaves[a,f], David Borsook[a,b], Lino Becerra[a,b,*]

[a] Imaging Consortium for Drug Development, Boston, MA 02115, USA
[b] Pain/Analgesia Imaging Neuroscience (P.A.I.N.) Group, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA
[c] Neuroscience Research, Eli Lilly and Company, Indianapolis, IN 46285, USA
[d] Preclinial Research and Translational Medicine, Sunovion Inc., Marlborough, MA 01752, USA
[e] Imaging, Merck Research Laboratories, West Point, PA 19486-0004</postal>, USA
[f] Discovery Sciences, Biogen-Idec, Cambridge, MA 02142, USA

## HIGHLIGHTS

- We observed a graded response of increasing magnitude for both VAS and fMRI data.
- Test–retest reliability was highest for VAS and fMRI data during the 7/10 pain.
- 7/10 Pain stimuli yielded the greatest number of ROIs that were reproducible.
- ICC of behavioral measures and ROIs.
- Data Acquired in a setting corresponding to a drug trial.

## ARTICLE INFO

*Article history:*
Received 16 March 2015
Received in revised form 1 June 2015
Accepted 3 June 2015
Available online 10 June 2015

*Keywords:*
fMRI
Test–retest
Evoked heat stimulation
Pain

## ABSTRACT

To date, the blood oxygenated-level dependent (BOLD) functional magnetic resonance imaging (fMRI) technique has enabled an objective and deeper understanding of pain processing mechanisms embedded within the human central nervous system (CNS). In order to further comprehend the benefits and limitations of BOLD fMRI in the context of pain as well as the corresponding subjective pain ratings, we evaluated the univariate response, test–retest reliability and confidence intervals (CIs) at the 95% level of both data types collected during evoked stimulation of 40 °C (non-noxious), 44 °C (mildly noxious) and a subject-specific temperature eliciting a 7/10 pain rating. The test–retest reliability between two scanning sessions was determined by calculating group-level interclass correlation coefficients (ICCs) and at the single-subject level. Across the three stimuli, we initially observed a graded response of increasing magnitude for both VAS (visual analog score) pain ratings and fMRI data. Test–retest reliability was observed to be highest for VAS pain ratings obtained during the 7/10 pain stimulation (ICC = 0.938), while ICC values of pain fMRI data for a distribution of CNS structures ranged from 0.5 to 0.859 ($p < 0.05$). Importantly, the upper and lower confidence interval CI bounds reported herein could be utilized in subsequent trials involving healthy volunteers to hypothesize the magnitude of effect required to overcome inherent variability of either VAS pain ratings or BOLD responses evoked during innocuous or noxious thermal stimulation.



## 1. Introduction

With blood oxygenated-level dependent (BOLD) functional magnetic resonance imaging (fMRI) methodology, the behavior of pain pathway structures within the central nervous system (CNS) has been characterized in states of acute or chronic pain (Jensen et al., 2013; Lebel et al., 2008; Lopez-Sola et al., 2014), during therapeutic intervention (Gustin et al., 2010; Upadhyay et al., 2011;

* Corresponding author at: Harvard Medical School, Pain/Analgesia Imaging Neuroscience (P.A.I.N.) Group, Boston Children's Hospital, 9 Hope Avenue, Waltham, MA 02453, USA. Tel.: +1 781 216 1182.
*E-mail address:* lino.becerra@childrens.harvard.edu (L. Becerra).
[1] These authors contributed equally to the manuscript.

http://dx.doi.org/10.1016/j.jneumeth.2015.06.001
0165-0270/

Wager et al., 2013) and during the placebo response (Bingel et al., 2011; Jensen et al., 2014). From work stemming from the pain fMRI field, it is better known which CNS structures can potentially be targeted, directly or indirectly, in order to obtain a possible analgesic effect (e.g., insula in fibromyalgia patients (Harris et al., 2013)). However, despite the large body of pain fMRI work, there remain fundamental unknowns regarding the utility of this methodology, as a means to characterize pathology in distinct pain patient populations or to evaluate novel therapies, be they pharmacological or behavioral in nature. One area believed to require further attention is a more in depth assessment of test–retest reliability of pain fMRI experimental paradigms (Letzen et al., 2014; Quiton et al., 2014). With the test–retest reliability of pain fMRI as well as the inherent variability of this methodology in hand, the utility of pain fMRI, particularly during longitudinal, proof-of-concept clinical trials can be further realized.

Across pain fMRI studies, the stimulation paradigm has indeed varied from study to study. These variations derive from study site preferences as well as which pain stimuli may be most suitable to implement in order to probe the etiology and pathology of a particular patient population. However, thermal heat pain is one type of evoked stimulation widely used in not only fMRI studies, but also behavioral work aimed solely at quantitative sensory testing (QST) in healthy subjects or pain patients. To date, it is well known that thermal heat pain induces robust BOLD activation and deactivation in multiple ascending and descending pain pathway structures, which process sensory and affective components of pain (Upadhyay et al., 2010). Multiple studies have dissected the properties of the BOLD signal elicited by noxious thermal stimuli and in turn, have characterized what features (e.g., biphasic response) of the BOLD signal are specific to evoked pain (Loggia et al., 2012; Moulton et al., 2012; Upadhyay et al., 2010). Moreover, cross-sectional investigations utilizing thermal heat pain have clearly demonstrated that a substantial dynamic range exists in the BOLD signal such that a treatment effect, acute or chronic, can be measured.

Given the *a priori* knowledge of thermal pain responses as measured by BOLD fMRI methodology, this healthy subject study focused on a quantitative assessment of test–retest reliability of pain fMRI during the presence of heat stimuli. Based on the variability of what temperatures are considered painful from one individual to another, a subject-specific heat stimuli yielding a 7 out of 10 rating on the visual analogue scale (VAS) was utilized. The objective of our study was to further evaluate the univariate response, test–retest reliability of pain fMRI (as assessed by intraclass correlation coefficients (ICCs)) and confidence intervals (CIs) at the 95% level across CNS structures previously implicated in pain processing. These three measures were calculated and compared for 40 °C, 44 °C and subject-specific temperatures corresponding to 7/10 pain rating, where the latter was considered the 'pain condition'. Furthermore, the current study was carried out with the specific needs and requirements of a pharmacological fMRI study in mind (Schwarz et al., 2011a,b). Therefore, steps enabling better standardization of study procedures were executed as well as onsite 'drug' administration of a placebo pill at each scan session.

## 2. Materials and methods

### 2.1. Study participants

This investigation was approved by the McLean Hospital Institutional Review Board and subsequently carried at McLean Hospital, Harvard Medical School. The study conformed with the Helsinki agreement for experimentation in humans (http://www.wma.net/en/30publications/10policies/b3/). Twelve healthy, right-handed male participants were included in this study (mean age ± standard deviation: 31 ± 8.8 years old). Each subject made three study site visits (prescreen and 2 MRI scanning sessions termed Scans 1 and 2). The prescreen visit consisted of the following procedures: (1) introduction to 40 °C and 44 °C stimuli (dorsum or left foot) in order to become familiar with the Medoc Thermal Sensory Analyzer (TSA) probe and stimuli itself, (2) quantitative sensory testing to determine subject-specific temperature (maximum temperature = 50 °C) corresponding to a VAS of 7 out of 10 pain rating, (3) training on pain reporting equipment utilized during fMRI, (4) review of complete medical history and (5) physician assessment, which included physiological measurements. Enrolled participants had no presence of physical or mental illness, no presence chronic pain disorders, consumed no medications and reported no history of alcoholism or drug abuse. Subsequent to prescreening procedures, each individual underwent 2 scanning sessions that were 2–3 weeks apart. All volunteers gave informed consent prior to study participation. Subjects in the current study were enrolled as part of a larger, placebo-controlled drug (buprenorphine) study. Volunteers were told that they would receive placebo or drug in a double-blinded, cross-over study; however, placebo was only administered to the current set of healthy subjects. After completion of the study, each subject was told that only placebo was administered.

### 2.2. Pre-scanning procedures

For each subject and each scanning session, a detailed checklist was kept to ensure all predefined standard operating procedures (SOP) were adhered to. Within this checklist, deviations from the SOP were logged (e.g., ancillary equipment malfunction) as well as timestamps of distinct components of the study (e.g., baseline vitals or start of anatomical MRI). Drug test were first completed before each scan session to verify the subjects' eligibility. Following drug testing, subjects received (1) a brief re-training of VAS pain rating equipment and review of the study procedures, (2) physician assessment, (3) placement of an intravenous line (subjects' left arm) for subsequent within MRI scanner blood draw (not actually performed) and (4) oral drug administration (sublingual vitamin B-12 pill) prior to subject placement within the scanner. Once subjects were positioned within the MRI monitoring of physiology (heart rate, end-tidal $CO_2$, $PO_2$, respiratory rate) was initiated.

### 2.3. Heat stimulation

During the fMRI scan, a mixed heat stimulation paradigm was used to assess the reproducibility of 40 °C (non-noxious), 44 °C (mildly noxious) and the subject-specific 7/10 noxious heat pain stimulation. A 32 $cm^2$ TSA thermode was attached to the subjects left foot via Velcro strap with a baseline temperature of 32 °C. The stimulation paradigm consisted of 25 s off and 15 s on cycle where each temperature (i.e., 40 °C, 44 °C and 7/10) was randomly presented five times. During heat stimulation and fMRI data collection, VAS pain ratings were continuously and simultaneously collected from the subjects.

### 2.4. MRI data acquisition

MRI data were collected on a 3 Tesla Siemens Trio scanner with an 8-channel phased array head coil (Erlangen, Germany). fMRI data were collected using a gradient echo-echo planar pulse sequence (GE-EPI) with 3.5 × 3.5 × 3.5 $mm^3$ resolution. GE-EPI Parameters: Time of Repetition (TR) = 2500 ms, Time of Echo (TE) = 30 ms, Field of View (FOV) = 224 × 224, Flip Angle (FA) = 90°, No. of Slices = 41 axial slices, No. of Volumes = 283. Magnetization-Prepared Rapid Acquisition Gradient-Echo (MPRAGE) anatomical images were collected. MPRAGE Parameters: TR = 2100 ms,

**Table 1**
Regions of interest for which evoked BOLD responses and test–retest reliability were quantified during 40 °C, 44 °C and 7/10 thermal pain stimulation paradigms. Quantification was performed in left and right components of each ROI unless specified otherwise. * Quantification of metrics from the primary somatosensory cortex was restricted to the right hemisphere (contralateral to stimulation site) and foot representation.

| Regions of interest | |
|---|---|
| Amygdala | Inferior frontal triangular |
| Anterior cingulate | Middle frontal |
| Anterior insula | Middle frontal orbitalis |
| Caudate | Middle-anterior cingulate |
| Cerebellum crus I | Middle-posterior cingulate |
| Cerebellum crus II | Nucleus accumbens |
| Cerebellum III | Posterior cingulate |
| Cerebellum IV & V | Posterior insula |
| Cerebellum IX | Primary somatosensory cortex (right)* |
| Cerebellum VI | Putamen |
| Cerebellum VIIb | Sensory thalamus (bilateral) |
| Cerebellum VIII | Superior frontal |
| Cerebellum X | Superior frontal orbitalis |
| Hypothalamus | Superior medial frontal |
| Inferior frontal opercularis | Supplemental motor area |
| Inferior frontal orbitalis | Thalamus |

TE = 2.74 ms, Time of Inversion (TI) = 1100 ms, FA = 12°, 128 sagittal slices.

### 2.5. Quality assurance (QA) of MRI data

QA of all fMRI data was performed using a MatLab (MathWorks, Sherbon, MA, USA)-based functional Biomedical Informatics Research Network (fBIRN) analysis algorithm. Each fMRI dataset was also motion corrected using FSL's mcflirt algorithm (FMRIB Software Library (FSL 5) (www.fmrib.ax.ac.uk/fsl)), and evaluated to determine if the maximum deviation due to motion was greater than 2.5 mm.

### 2.6. fMRI data processing

Subsequent single-subject fMRI data analyses were performed using FSL 5 (www.fmrib.ax.ac.uk/fsl). Preprocessing steps for fMRI data have been described elsewhere (Upadhyay et al., 2010, 2012). Moreover, a dual explanatory variable (EV) general linear model (GLM) analysis of single-subject fMRI data was utilized to quantify early (stimulus-locked) and late phase BOLD responses identified and characterized in earlier evoked pain fMRI work. Within group average analyses were performed using a mixed-effects analyses (Woolrich et al., 2004). Each group average statistical map was set to a threshold of $z > 2.3$ and cluster-size corrected (Smith and Nichols, 2009).

### 2.7. Intra-class correlation coefficients (ICC) and confidence intervals (CI)

To evaluate the test–retest reliability of VAS pain rating and BOLD fMRI data between Scans 1 and 2 and data stemming from 40 °C, 44 °C and 7/10 heat stimuli, the ICC values were calculated using SPSS v21.0 (SPSS Inc., Chicago, IL, USA). For BOLD fMRI data, multiple regions of interest (ROI, Table 1) were selected using data provided in earlier pain fMRI work (Denk et al., 2014) as a guide and anatomically defined using the WFU PickAtlas (WFU Pickatlas, v2.4). The PickAtlas ROI for mid-cingulum was modified based on earlier findings within this structure (Vogt, 2005). The sensory thalamus ROI was defined using FSL's DTI-based segmentation of the thalamus. The ROI corresponding to the foot representation within the primary somatosensory cortex has been defined previously. From each atlas defined ROI, the mean BOLD response, defined in terms of parameter estimates, was extracted from each subject and each scanning session. ICCs of absolute agreement were then calculated using a two-way mixed model that provided a measure of consistency through a ratio of between subject variance to total variance (Caceres et al., 2009). This statistical practice has been also been used in recent work aiming to test the reliability of fMRI data (Letzen et al., 2014). In addition, CIs at a 95% confidence level were calculated for all VAS pain ratings and all ROI evaluated.

## 3. Results

### 3.1. Subjective pain ratings

During fMRI experimentation, participants rated their subjective pain experience felt during each heat stimuli on a VAS pain scale (0 = no pain and 10 = maximal pain). The group-average pain ratings validated that 40 °C, 44 °C and 7/10 heat stimuli would respectively be non-noxious, mildly noxious and noxious (Table 2). A significant difference between the VAS ratings of pain intensity for the three temperatures was found ($p < 0.05$; ANOVA). As expected, the 40 °C stimulus and 7/10 heat stimuli were predominately reported as being non-noxious and noxious, respectively. Moreover, CIs at the 95% confidence level demonstrated similar magnitudes of lower and upper CI bounds across the three thermal stimulation types.

Test–retest reliability for VAS pain ratings for each temperature was obtained by calculating ICC values between scans 1 and 2 (Table 2). For the 44 °C and 7/10 heat pain stimuli, ICC between scans 1 and 2 were high, while significant ICC was not observed for the 40 °C. The latter was believed to result from little or no pain being perceived and reported by healthy subjects. The ICC stemming from the 7/10 subjective pain rating was the highest ICC value obtained in this study.

Single-subject, VAS responses corresponding to 40 °C, 44 °C and 7/10 heat stimuli are shown in Fig. 1. While no significant correlation in behavioral pain responses between scans 1 and 2 were observed, significant correlation and reproducibility in pain ratings were measured for both 44 °C and 7/10 heat stimuli.

### 3.2. QA of MRI data

All fMRI data surpassed the fBIRN QA procedure. However, upon inspection of motion correction results, data from 2 of the

**Table 2**
Summary of subjective reports obtained during 40 °C, 44 °C and 7/10 thermal pain stimulation paradigms. VAS pain ratings are given in terms of mean ± SE. CIs are reported at the 95% confidence level from a repeated measures test. * Indicates no test–retest reliability between scans 1 and 2. SE—standard error; ICC—intra-class correlation coefficient; CI—confidence interval.

| Stimulation | Scan number | VAS Mean | ICC | *F* value | *p*-Value | CI lower | CI upper |
|---|---|---|---|---|---|---|---|
| 40 °C | 1 | 0.34 ± 0.4 | −0.308* | 0.762 | 0.654 | −1.254 | 0.460 |
| | 2 | 0.74 ± 1.0 | | | | | |
| 44 °C | 1 | 2.21 ± 1.4 | 0.860 | 7.787 | 0.003 | −1.240 | 0.275 |
| | 2 | 2.70 ± 1.7 | | | | | |
| 7/10 | 1 | 7.18 ± 2.3 | 0.938 | 15.694 | 0.001 | −1.089 | 0.477 |
| | 2 | 7.48 ± 2.1 | | | | | |

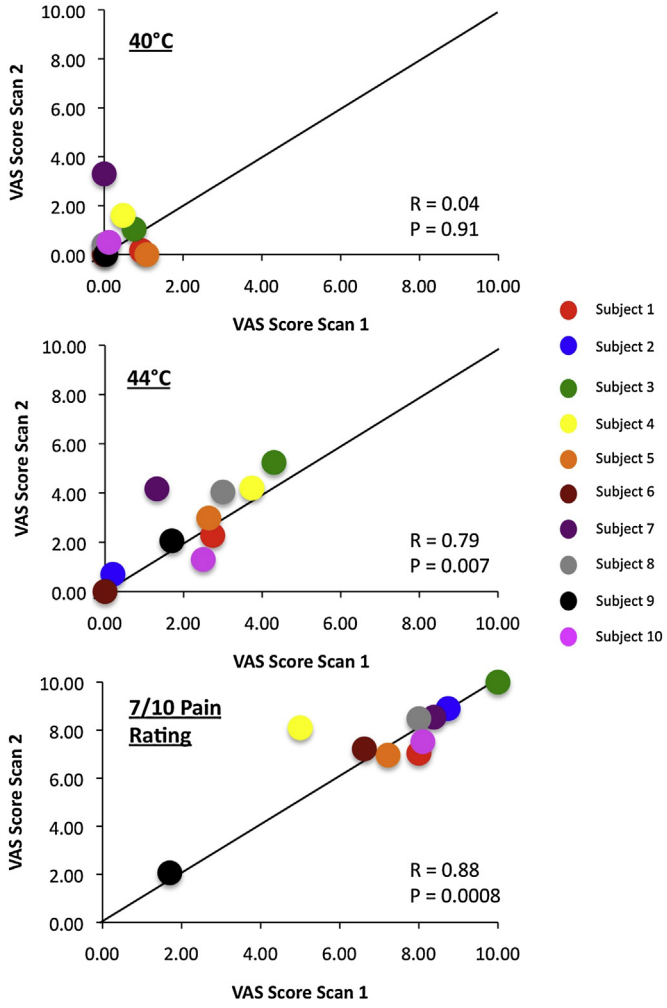


**Fig. 1.** Single-subject evoked VAS pain ratings. Single-subject pain ratings for 40 °C, 44 °C and 7/10 pain stimuli for scans 1 and 2 are given in scatter plot format. Greater reproducibility at the single-subject level can be ascertained from data points projected closer to the centerline possessing a slope of 1.

12 subjects was observed to have head motion greater than 2.5 mm. VAS pain rating and BOLD fMRI results provided and discussed correspond to an $N = 10$ dataset.

### 3.3. Early and late phase evoked BOLD responses

Similar to VAS pain rating data, statistical BOLD activation maps corresponding to early (Fig. 2A) and late (Fig. 2C) phase BOLD responses showed an overall pattern of increased activation when progressing from 40 °C, 44 °C and 7/10 heat stimuli. This progression was clearly observed in cortical (insula) and deeper subcortical (striatum) structures within the supraspinal pain processing pathways. To further elucidate the behavior of pain processing CNS structures during early and late phase evoked BOLD responses, parameter estimates from the anterior insula (bilateral, atlas defined) were extracted and averaged across subjects and scanning sessions. Interestingly, while a graded group average response was detected in the early phase within the anterior insula (Fig. 2B), the anterior insula's late phase response (Fig. 2D) there was no measurable response for the 40 °C and 44 °C stimuli and a significant BOLD response was detected for the 7/10 heat pain stimuli.

Single-subject, BOLD responses within cortical and subcortical ROIs as well as between scans 1 and 2 were evaluated for the 7/10 heat pain stimuli (Fig. 3). Overall, across CNS structures and subjects good reliability can be visually observed with regards to the magnitude of the early (Fig. 3A) and late (Fig. 3B) phase BOLD responses. A comparison across the structures evaluated demonstrates higher consistency of BOLD responses between the two scan sessions for structures such as the anterior insula (early phase) and posterior cingulate (late phase) and lower consistency for the primary somatosensory cortex (late phase). To detect significant differences in early and late phase BOLD responses between scans 1 and 2, paired comparisons were performed for each of the three heat stimulation paradigms. The results of the paired, *t*-tests are summarized in Supplemental Tables 1–3.

### 3.4. ICC values and CI for evoked BOLD responses

ICCs were quantified across all ROIs noted in Table 1 and for the three heat stimulation conditions, 40 °C, 44 °C and 7/10 heat stimuli (Table 3). Very few CNS structures surpassed the $p < 0.05$ threshold (corresponding to ICC of ~ 0.5) for the 40 °C (Table 3A and B) or 44 °C (Table 3C and D) experimental conditions. Moreover, the ICC values for 40 °C and 44 °C stimuli ranged from 0.520 to 0.682. In sharp contrast, the 7/10 heat pain stimuli (Table 3E and F) yielded a substantially greater number of CNS structures surpassing the $p < 0.05$ threshold (ICC range: 0.515–0.859) for both early and late phase BOLD responses.

When evaluating CIs at the 95% confidence level in relation to the corresponding ICC value, a relationship between the magnitudes of upper or lower CI bounds with ICC values was not observed. Based on CIs, low variance was observed for early phase BOLD responses in structures such as the anterior and posterior insula, putamen and thalamus (including sensory thalamus), while higher variability was calculated for late phase BOLD responses such as the posterior cingulate and primary somatosensory cortex. However, the magnitudes of CIs (lower and upper bounds) for early and late phase BOLD responses were overall of similar extent.

## 4. Discussion

### 4.1. Summary of findings

The current study was design to reflect a standard randomized clinical trial with a cross-over design of drug and placebo. The study assessed the graded, group-level response and importantly, the test–retest reliability of VAS pain ratings and CNS BOLD responses measured during administration of three, evoked thermal stimulations; 40 °C, 44 °C and a subject-specific temperature eliciting a subjective pain level of 7/10. Based on ICC values calculated between two scanning sessions and across study endpoints, reproducibility was observed to be highest for VAS pain ratings obtained during the 7/10 pain stimulation (ICC = 0.938). Importantly, the level of reproducibility for behavioral and fMRI measures could change in similar studies executed in pain patient populations rather than in healthy subjects.

With respect to test–retest reliability of pain fMRI data, the 7/10 pain stimuli also yielded the greatest number of CNS structures possessing ICC levels of ~0.5 or more ($p < 0.05$; ICC range: 0.5–0.859), thus demonstrating sufficient reproducibility exists for pain fMRI methodology. The latter observation further supports the use of fMRI measures as reliable endpoints that can be utilized to evaluate an analgesic's mechanistic effects on CNS function. For all VAS pain ratings and BOLD fMRI data points surpassing the threshold of $p < 0.05$/ICC of ~0.5, CIs were also calculated. In future trials where an analgesic might be evaluated in healthy male subjects; the upper and lower CI bounds could be utilized to determine the magnitude of therapeutic effect necessary to overcome inherent variability of

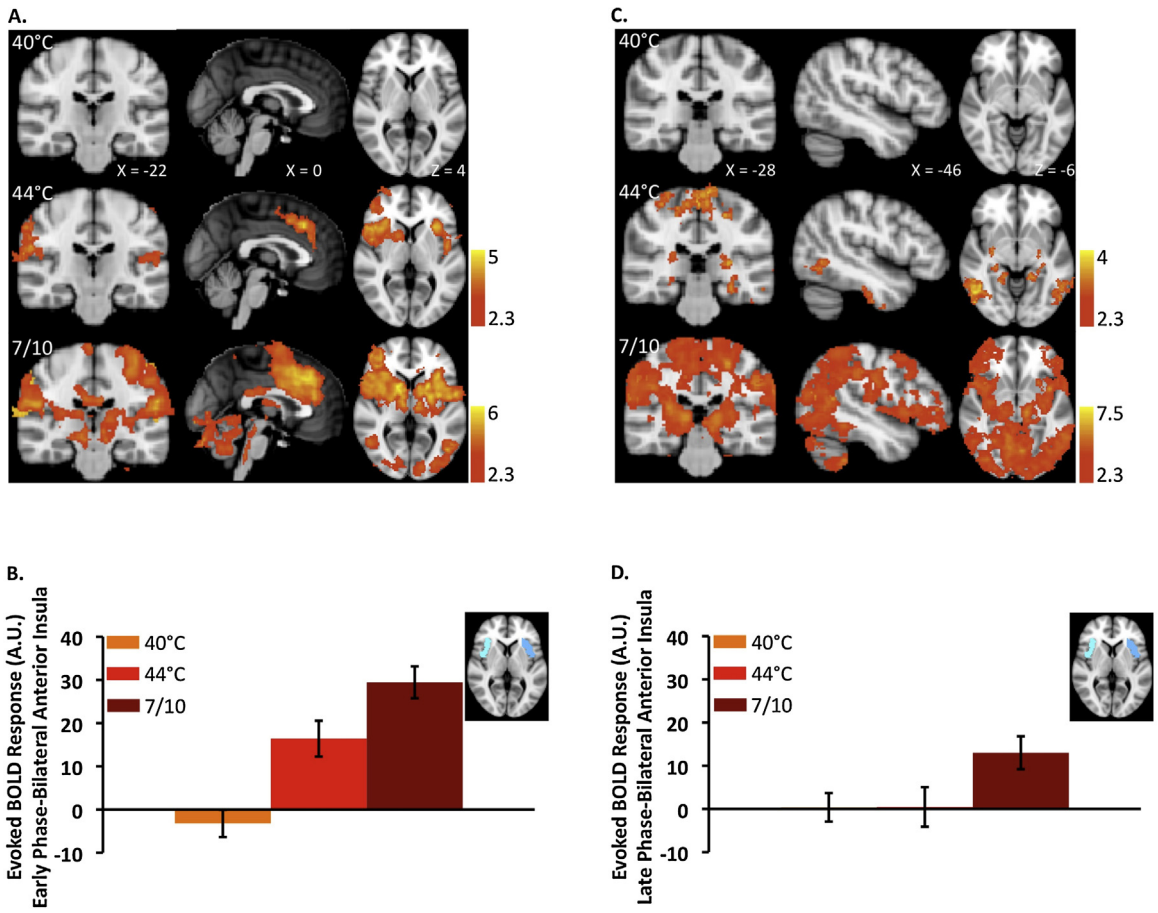


**Fig. 2.** Group-average evoked BOLD responses to 40 °C, 44 °C and 7/10 pain stimulation. Within group average response were obtained by combining fMRI data across all subjects ($N = 10$) as well as across scan sessions 1 and 2. Statistical maps obtained from mixed-effects analyses ($z > 2.3$ and cluster size corrected) are shown for each stimulation type and for early (A) and late phase (B) BOLD responses. Summary statistics extracted from the anterior insula show a graded response during the early phase (C). The ANOVA test for the anterior insula, early phase BOLD responses revealed significant between condition differences ($F_{2, 57} = 19.88$, $p = 2.88E-07$). Late phase BOLD responses in the anterior insula (D) showed little or no response during the 40 °C and 44 °C stimulation, while the 7/10 pain stimulation was comparably more robust. The ANOVA test revealed significant between condition differences ($F_{2, 57} = 3.42$, $p = 0.04$).

either VAS pain ratings or BOLD responses evoked during innocuous or noxious thermal stimulation.

### 4.2. Comparison of pain fMRI test–retest results

Similar to previous investigations measuring test–retest reliability of pain fMRI measures (Letzen et al., 2014; Quiton et al., 2014; Taylor and Davis, 2009), ICC values of BOLD responses evoked during noxious thermal stimulation were calculated across CNS structures. A common theme amongst studies was the fact that high ICC values within BOLD fMRI data collected during noxious conditions were observed in sub-regions within the cingulate, insular and frontal cortices–regions known to process affective components of pain. For example, for the anterior middle cingulate cortex, the current study, Letzen et al. (2014), Quiton et al. (2014) all observed ICCs between ~0.7 and ~0.8. Further congruence in terms of ICCs for BOLD fMRI data may have been present between this and past studies; however, in earlier work, the CNS ROI evaluated did not include regions such as the amygdala, striatum or cerebellum. One notable difference between the work by Quinton et al. and this study pertains to the reproducibility of BOLD responses within the sensory thalamus and primary somatosensory cortex during noxious stimulation. While the earlier study found non-significant or low test–retest reliability within the thalamus or primary somatosensory cortex, we observed that early and late phase BOLD responses possessed high ICC values of similar magnitude (0.7–0.75) in comparison to affective pain processing structures such as the cingulate, insular and frontal cortices. These differences may arise from how the thalamus and primary somatosensory were defined or the statistical modeling procedures implemented to analyze BOLD fMRI data (see below discussion on GLM analysis). For instance, within the current study, ICC values were measured for both the sensory thalamus, as defined by a DTI-based segmentation of the thalamus, as well as the thalamus as a whole. Comparatively, Quiton et al., solely evaluated reliability metrics for the thalamus as a whole as defined by Tailairach and Tournoux (1988). Moreover, in accord with Letzen et al., we observed higher intersession reliability for subjective pain ratings compared to evoked BOLD responses in general.

The evaluation of VAS pain ratings and BOLD responses for a range of thermal stimuli of increasing pain intensity enabled a graded CNS BOLD response to be detected, but importantly, demonstrated that insignificant or low test–retest reliability was present for innocuous or mildly noxious thermal stimuli (i.e., 40 °C and 44 °C) compared to the 7/10 pain stimulus. It cannot be simply concluded that compared to low, non-noxious temperatures, higher temperatures will conventionally yield higher reproducibility of BOLD fMRI data. It is likely that while some individuals can withstand, for example, a 50 °C stimuli, others may find this temperature too intense, possess head motion and subsequently introduce variability into the data. Thus, as performed in this and earlier work, to obtain significant test–retest reliability of VAS pain ratings and

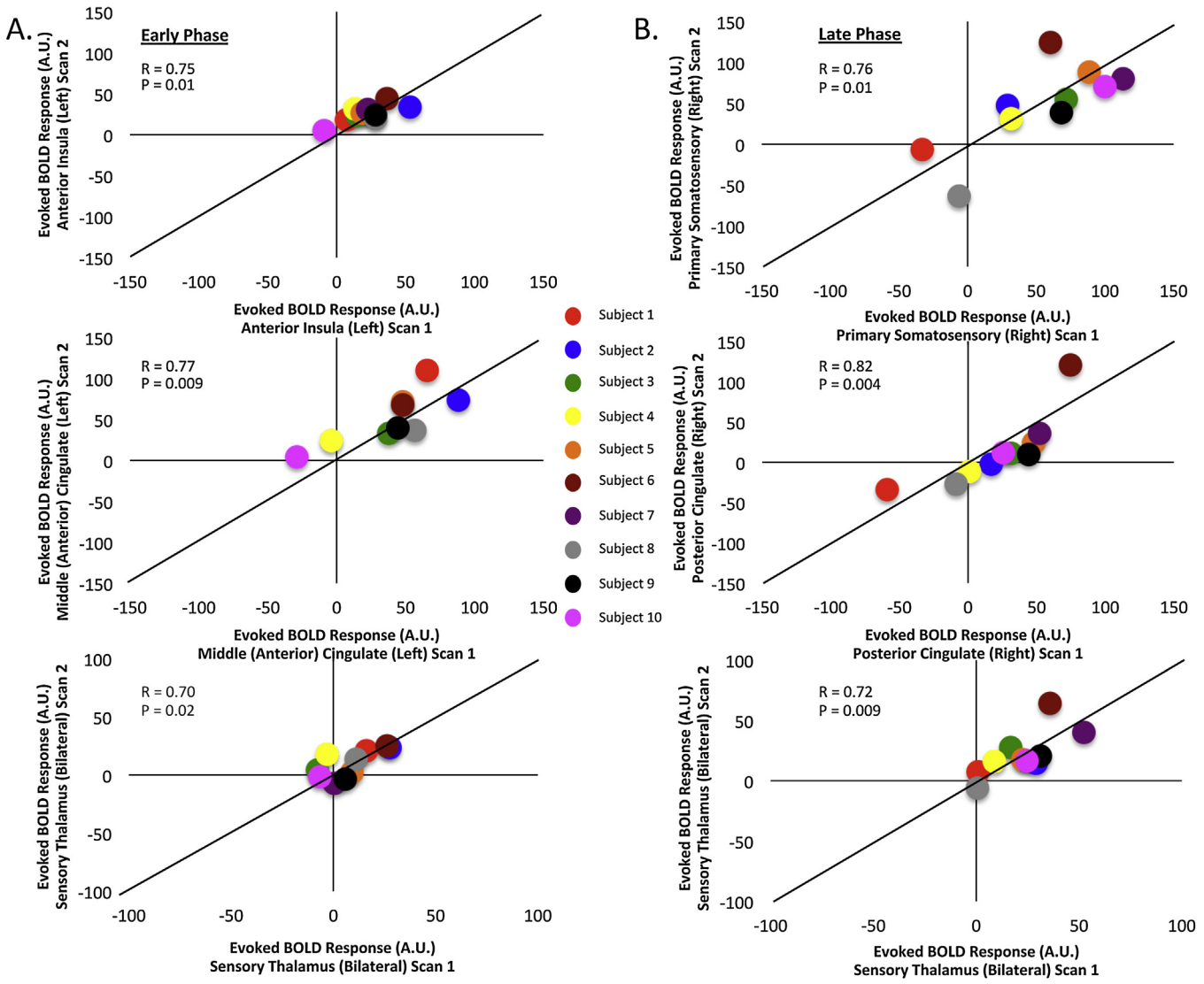


**Fig. 3.** Single-subject evoked BOLD responses to 7/10 pain stimulation. Single-subject early (A) and late (B) phase BOLD responses (defined by parameter estimates) for scans 1 and 2 are given in scatter plot format. Greater reproducibility at the single-subject level can be ascertained from data points projected closer to the centerline possessing a slope of 1. Parameter estimates were extracted from atlas defined cortical and subcortical ROIs. Corresponding ICC values for each ROI have been given in Table 3E and F.

BOLD fMRI data, it is necessary to have a balance between a noxious stimuli as defined by the subject or patient, which is intense enough to remain noxious between 2 or more scanning session, yet not noxious to the point, where study volunteers may find the stimulation paradigm unbearable.

It is noted that comparing reproducibility results across pain fMRI investigations is not necessarily a straightforward process given methodological differences each study incorporated in terms of evoked pain stimulation paradigms, data acquisition and data analyses. For example, the evoked thermal pain stimuli utilized by Quinton et al., were 48 °C and 50 °C heat stimuli, while Letzen et al. used a 'tolerated' temperature between 43 and 51 °C. This is in contrast to the 7/10 subject-specific temperature utilized herein. In the reports by Letzen et al. as well as Quinton et al., reliability of early and late phase BOLD responses throughout the CNS were not assessed. Given the differences of the explanatory variables used in GLM analyses of BOLD fMRI data, between-study differences in ICC values for a single CNS structure could arise strictly as a result of the analytical approach taken. Nonetheless, when considering the reproducibility results reported across studies and in pain processing CNS structures (i.e., cingulate, insular and frontal cortices) what can be said with confidence is that sufficient intersession reliability exist for pain fMRI, such that the method can be implemented to evaluate CNS pain processing in longitudinal studies involving healthy subjects. The utility of pain fMRI is further warranted when considering the work of Wager et al. (2013), where an fMRI-based signature of pain demonstrated high sensitivity and specificity as well as treatment response. Interestingly, the fMRI signature of pain identified in the latter study consisted of CNS structures such as the thalamus, insula and cingulate; structures that showed high reproducibility in the past and current fMRI investigation.

### 4.3. Study limitations

#### 4.3.1. Number of scan sessions

Intersession reliability of behavioral and fMRI data may vary based on the number of scan sessions used to assess methodological test–retest reliability. Previous studies before have used a comparison of three scan sessions to determine reproducibility (Letzen et al., 2014), (Quiton et al., 2014). In contrast, the design implemented in the current work was setup with a placebo-controlled,

**Table 3**
ICC values and CIs across ROIs possessing a $p$ value of 0.05 and lower. ICCs and CIs were calculated for early and late phase evoked BOLD responses for 40 °C ((A) early phase, (B) late phase.), 44 °C ((C) early phase, (D) late phase.) and 7/10 ((E) early phase, (F) late phase.) thermal pain stimulation. CIs are reported at the 95% confidence level from a repeated measures statistical analysis.

| (A) 40 °C | Early phase | | | | |
|---|---|---|---|---|---|
| Regions of Interest | ICC | *F*-value | *p*-Value | CI lower | CI upper |
| Thalamus (right) | 0.643 | 4.119 | 0.023 | −9.479 | 10.210 |

| (B) 40 °C | Late phase | | | | |
|---|---|---|---|---|---|
| Regions of interest | ICC | *F*-value | *p*-Value | CI lower | CI upper |
| Superior frontal (left) | 0.575 | 5.606 | 0.009 | 1.590 | 14.252 |
| Cerebellum VI (right) | 0.615 | 4.298 | 0.020 | −3.239 | 10.018 |
| Cerebellum crus I (right) | 0.532 | 3.291 | 0.045 | −5.427 | 14.514 |
| Supplementary motor area (left) | 0.580 | 3.715 | 0.032 | −7.576 | 17.669 |
| Sensory thalamus (bilateral) | 0.627 | 4.159 | 0.023 | −3.802 | 6.721 |
| Thalamus (right) | 0.614 | 4.441 | 0.018 | −3.071 | 12.037 |

| (C) 44 °C | Early phase | | | | |
|---|---|---|---|---|---|
| Regions of interest | ICC | *F*-value | *p*-Value | CI lower | CI upper |
| Inferior frontal triangular (right) | 0.552 | 3.252 | 0.047 | −9.506 | 13.161 |
| Inferior frontal opercularis (right) | 0.682 | 4.885 | 0.014 | −11.438 | 14.152 |
| Anterior insula (right) | 0.620 | 4.092 | 0.024 | −19.850 | 10.500 |

| (D) 44 °C | Late phase | | | | |
|---|---|---|---|---|---|
| Regions of interest | ICC | *F*-value | *p*-Value | CI lower | CI upper |
| Cerebellum VIII (right) | 0.536 | 3.213 | 0.049 | −5.309 | 10.628 |
| Cerebellum VIII (left) | 0.546 | 3.296 | 0.045 | −6.366 | 12.600 |
| Cerebellum X (left) | 0.563 | 3.398 | 0.041 | −10.032 | 16.464 |

| (E) 7/10 VAS pain rating | Early phase | | | | |
|---|---|---|---|---|---|
| Region of interest | ICC | *F*-value | *p*-Value | CI lower | CI upper |
| Medial superior frontal (left) | 0.556 | 3.282 | 0.046 | −19.423 | 14.487 |
| Medial superior frontal (right) | 0.515 | 2.909 | 0.064 | −13.836 | 14.654 |
| Superior frontal (left) | 0.642 | 4.240 | 0.021 | −10.093 | 11.270 |
| Superior frontal (right) | 0.637 | 4.338 | 0.020 | −10.696 | 5.535 |
| Superior frontal orbitalis (left) | 0.773 | 7.386 | 0.003 | −17.986 | 10.261 |
| Inferior frontal orbitalis (left) | 0.787 | 7.796 | 0.003 | −7.546 | 11.020 |
| Inferior frontal orbitalis (right) | 0.625 | 4.362 | 0.019 | −11.203 | 4.176 |
| Anterior insula (left) | 0.657 | 5.014 | 0.012 | −12.674 | 3.834 |
| Anterior middle cingulate (left) | 0.748 | 7.490 | 0.003 | −25.797 | 6.222 |
| Anterior middle cingulate (right) | 0.646 | 5.697 | 0.008 | −28.793 | 1.971 |
| Supplemental motor area (left) | 0.585 | 3.656 | 0.033 | −27.117 | 15.050 |
| Supplemental motor area (right) | 0.566 | 3.735 | 0.031 | −29.396 | 8.858 |
| Inferior frontal opercularis (right) | 0.706 | 7.668 | 0.003 | −21.464 | 0.252 |
| Putamen (left) | 0.568 | 3.402 | 0.041 | −11.492 | 8.118 |
| Sensory thalamus (bilateral) | 0.715 | 5.661 | 0.008 | −8.450 | 5.225 |
| Posterior middle cingulate (left) | 0.648 | 4.449 | 0.018 | −15.646 | 9.053 |
| Posterior middle cingulate (right) | 0.695 | 5.644 | 0.008 | −16.453 | 5.797 |
| Posterior insula (left) | 0.715 | 5.805 | 0.008 | −9.371 | 4.698 |
| Posterior insula (right) | 0.551 | 3.313 | 0.045 | −13.932 | 7.769 |
| Hypothalamus (left) | 0.489 | 3.273 | 0.046 | −45.560 | 6.474 |
| Thalamus (left) | 0.637 | 4.378 | 0.019 | −12.957 | 6.254 |
| Thalamus (right) | 0.726 | 7.430 | 0.003 | −13.155 | 1.588 |
| Cerebellum IV & V (right) | 0.823 | 9.532 | 0.001 | −9.282 | 6.568 |
| Cerebellum IV & V (left) | 0.675 | 4.958 | 0.013 | −7.909 | 15.432 |
| Cerebellum VI (left) | 0.619 | 3.924 | 0.027 | −17.981 | 17.207 |
| Cerebellum VI (right) | 0.578 | 3.472 | 0.039 | −17.800 | 14.839 |
| Cerebellum VIIb (left) | 0.750 | 6.556 | 0.005 | −14.648 | 9.181 |
| Cerebellum VIIb (right) | 0.644 | 4.294 | 0.020 | −15.593 | 11.592 |
| Cerebellum VIII (left) | 0.701 | 5.221 | 0.011 | −13.742 | 12.159 |
| Cerebellum VIII (right) | 0.708 | 5.366 | 0.010 | −10.109 | 11.356 |

| (F) 7/10 VAS pain rating | Late phase | | | | |
|---|---|---|---|---|---|
| Region of interest | ICC | *F*-value | *p*-Value | CI lower | CI upper |
| Amygdala (left) | 0.724 | 5.796 | 0.008 | −18.169 | 24.764 |
| Sensory thalamus (bilateral) | 0.731 | 5.906 | 0.007 | −9.158 | 10.161 |
| Posterior middle cingulate (left) | 0.559 | 3.419 | 0.041 | −12.348 | 24.559 |
| Posterior insula (left) | 0.665 | 4.618 | 0.016 | −9.478 | 12.579 |
| Thalamus (left) | 0.633 | 4.163 | 0.023 | −12.246 | 18.051 |
| Thalamus (right) | 0.649 | 4.413 | 0.019 | −16.370 | 10.784 |
| Primary somatosensory cortex (right) | 0.761 | 6.908 | 0.004 | −19.523 | 31.063 |

Table 3 (*Continued*)

| (F) 7/10 VAS pain rating | Late phase | | | | |
|---|---|---|---|---|---|
| Region of interest | ICC | *F*-value | *p*-Value | CI lower | CI upper |
| Posterior cingulate (left) | 0.623 | 5.019 | 0.012 | −4.245 | 34.890 |
| Posterior cingulate (right) | 0.812 | 9.813 | 0.001 | −9.051 | 26.390 |
| Cerebellum III (left) | 0.587 | 3.568 | 0.036 | −12.501 | 14.971 |
| Cerebellum IV & V (left) | 0.744 | 6.302 | 0.006 | −13.191 | 9.722 |
| Cerebellum IV & V (right) | 0.759 | 6.712 | 0.005 | −10.097 | 8.386 |
| Cerebellum VI (left) | 0.816 | 9.440 | 0.001 | −18.293 | 9.608 |
| Cerebellum VI (right) | 0.859 | 11.973 | 0.001 | −10.926 | 10.592 |
| Cerebellum VIIb (left) | 0.703 | 5.407 | 0.010 | −14.726 | 9.057 |
| Cerebellum VIII (left) | 0.835 | 10.207 | 0.001 | −6.938 | 9.104 |
| Cerebellum VIII (right) | 0.678 | 4.818 | 0.014 | −9.773 | 12.435 |
| Cerebellum IX (left) | 0.587 | 5.563 | 0.009 | 1.894 | 23.381 |
| Cerebellum IX (right) | 0.678 | 4.818 | 0.014 | −3.145 | 30.200 |
| Cerebellum X (right) | 0.564 | 3.605 | 0.035 | −11.967 | 32.231 |
| Cerebellum Crus I (left) | 0.645 | 4.323 | 0.020 | −24.706 | 17.437 |
| Cerebellum Crus I (right) | 0.757 | 6.637 | 0.005 | −17.112 | 19.650 |

cross over study paradigm in mind, given its utility in trials evaluating pharmacological therapies. Nonetheless, the inclusion of additional scanning sessions or a longer intersession time interval may enable a better evaluation of reliability and stability of behavioral and fMRI measures.

#### 4.3.2. Subjects

Compared to earlier trials investigating test–retest reliability of evoked pain fMRI, the cohort size of the present study was considered small ($N = 10$). The univariate analysis results for BOLD fMRI data did indeed elicit the expected robust response within the supraspinal CNS structures mediated pain, particularly for the 7/10 pain stimulation. However, with a larger study population, a more accurate account of variance for both pain ratings and BOLD fMRI data may be achieved.

The primary goal of this study was to evaluate the reproducibility of evoked pain methodology itself, and therefore, a 'homogeneous' population of healthy male subjects was enrolled. Given the known differences in pain processing between males and females, the projection of the current results to a female population or a broader population encompassing both males and females should be done with caution (Moulton et al., 2006).

#### 4.3.3. Habituation

As with most fMRI studies, including those investigating evoked pain responses, stimuli are often repeatedly presented in order to increase the SNR of the BOLD signal. This repetition could induce a habituation effect, and in turn introduce variance into the data. Here, each stimulus (40 °C, 44 °C and 7/10 heat stimuli) was repeated 5 times, but in a randomized manner. The randomization is believed to have minimized the risk of habituation to any one particular stimulus as evidenced by the graded responses measured across the three stimuli as well as the group-average pain rating of ∼7 obtained for the 7/10 pain stimulus. However, a complete negation of habituation would be difficult.

### 4.4. Future directions

A key step to further understanding the benefits and limitations of pain fMRI measures, is to utilize the technique in a longitudinal manner in a pain patient population. In patients, fluctuations in disease specific pain, from magnitude, location and frequency perspectives, as well as other related disease factors (e.g., fatigue, depression or frequency of medications consumed) can be present across time. To what extent these patient-specific fluctuations impact the reliability and reproducibility of measures obtained in an fMRI study are in large part unknown. Moreover, by implementing pain fMRI longitudinally in patients, if and how fMRI measures track with core symptoms, which are of key clinical interest during therapeutic evaluation, may also be quantitatively understood. It may be the case that the natural waxing and waning of endogenous pain levels or other symptoms may introduce variability and thus lessen intersession test–retest reliability of evoked pain endpoints. However, it may be of value to elucidate the underlying pathophysiology of pain or symptom fluctuation in patients, as the former could be easier to target within the translational medicine setting.

In patients, much is unknown regarding how much potentiation or attenuation of the BOLD response measured in a CNS structure or network is necessary to overcome the inherent variance as well as to induce a meaningful clinical effect. By longitudinally measuring the pharmacodynamics of a standard of care for a specific pain patient population, the magnitude of effect and its clinical relevance can be better realized. Given the diverse functional, structural and neurochemical changes that can be present between distinct pain conditions, it is likely that a single answer does not exists in terms how and what CNS brain region(s) should be therapeutically targeted.

## Acknowledgments

We thank former members of the Imaging Consortium for Drug Development at McLean Hospital and financial support to D. Borsook and L. Becerra from the Herlands Fund for Pain.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jneumeth.2015.06.001

## References

Bingel U, Wanigasekera V, Wiech K, Ni Mhuircheartaigh R, Lee MC, Ploner M, et al. The effect of treatment expectation on drug efficacy: imaging the analgesic benefit of the opioid remifentanil. Sci Transl Med 2011;3:1–9 (70ra14).

Caceres A, Hall DL, Zelaya FO, Williams SC, Mehta MA. Measuring fMRI reliability with the intra-class correlation coefficient. NeuroImage 2009;45:758–68.

Denk F, McMahon SB, Tracey I. Pain vulnerability: a neurobiological perspective. Nat Neurosci 2014;17:192–200.

Gustin SM, Schwarz A, Birbaumer N, Sines N, Schmidt AC, Veit R, et al. NMDA-receptor antagonist and morphine decrease CRPS-pain and cerebral pain representation. Pain 2010;151:69–76.

Harris RE, Napadow V, Huggins JP, Pauer L, Kim J, Hampson J, et al. Pregabalin rectifies aberrant brain chemistry, connectivity, and functional response in chronic pain patients. Anesthesiology 2013;119:1453–64.

Jensen KB, Petzke F, Carville S, Choy E, Fransson P, Gracely RH, et al. Segregating the cerebral mechanisms of antidepressants and placebo in fibromyalgia. J Pain 2014:1328–37 (official journal of the American Pain Society).

Jensen KB, Srinivasan P, Spaeth R, Tan Y, Kosek E, Petzke F, et al. Overlapping structural and functional brain changes in patients with long-term exposure to fibromyalgia pain. Arthritis Rheum 2013;65:3293–303.

Lebel A, Becerra L, Wallin D, Moulton EA, Morris S, Pendse G, et al. fMRI reveals distinct CNS processing during symptomatic and recovered complex regional pain syndrome in children. Brain 2008;131:1854–79 (a journal of neurology).

Letzen JE, Sevel LS, Gay CW, O'Shea AM, Craggs JG, Price DD, et al. Test–retest reliability of pain-related brain activity in healthy controls undergoing experimental thermal pain. J Pain 2014;15:1008–14 (Official journal of the American Pain Society).

Loggia ML, Edwards RR, Kim J, Vangel MG, Wasan AD, Gollub RL, et al. Disentangling linear and nonlinear brain responses to evoked deep tissue pain. Pain 2012;153:2140–51.

Lopez-Sola M, Pujol J, Wager TD, Garcia-Fontanals A, Blanco-Hinojo L, Garcia-Blanco S, et al. Altered functional magnetic resonance imaging responses to nonpainful sensory stimulation in fibromyalgia patients. Arthritis Rheumatol 2014;66:3200–9.

Moulton EA, Keaser ML, Gullapalli RP, Maitra R, Greenspan JD. Sex differences in the cerebral BOLD signal response to painful heat stimuli. Am J Physiol Regul Integr Comp Physiol 2006;291:R257–67.

Moulton EA, Pendse G, Becerra LR, Borsook D. BOLD responses in somatosensory cortices better reflect heat sensation than pain. J Neurosci 2012;32:6024–31 (the official journal of the Society for Neuroscience).

Quiton RL, Keaser ML, Zhuo J, Gullapalli RP, Greenspan JD. Intersession reliability of fMRI activation for heat pain and motor tasks. NeuroImage Clin 2014;5:309–21.

Schwarz AJ, Becerra L, Upadhyay J, Anderson J, Baumgartner R, Coimbra A, et al. A procedural framework for good imaging practice in pharmacological fMRI studies applied to drug development #1: processes and requirements. Drug Discov Today 2011a;16:583–93.

Schwarz AJ, Becerra L, Upadhyay J, Anderson J, Baumgartner R, Coimbra A, et al. A procedural framework for good imaging practice in pharmacological fMRI studies applied to drug development #2: protocol optimization and best practices. Drug Discov Today 2011b;16:671–82.

Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage 2009;44:83–98.

Tailairach J, Tournoux P. Co-planar stereotaxic atlas of the Human BrainThieme Medical Publishers. New York, NY: Human BrainThieme Medical Publishers; 1988.

Taylor KS, Davis KD. Stability of tactile- and pain-related fMRI brain activations: an examination of threshold-dependent and threshold-independent methods. Hum Brain Mapp 2009;30:1947–62.

Upadhyay J, Anderson J, Baumgartner R, Coimbra A, Schwarz AJ, Pendse G, et al. Modulation of CNS pain circuitry by intravenous and sublingual doses of buprenorphine. NeuroImage 2012;59:3762–73.

Upadhyay J, Anderson J, Schwarz AJ, Coimbra A, Baumgartner R, Pendse G, et al. Imaging drugs with and without clinical analgesic efficacy. Neuropsychopharmacology 2011;36:2659–73 (official publication of the American College of Neuropsychopharmacology).

Upadhyay J, Pendse G, Anderson J, Schwarz AJ, Baumgartner R, Coimbra A, et al. Improved characterization of BOLD responses for evoked sensory stimuli. NeuroImage 2010;49:2275–86.

Vogt BA. Pain and emotion interactions in subregions of the cingulate gyrus. Nat Rev Neurosci 2005;6:533–44.

Wager TD, Atlas LY, Lindquist MA, Roy M, Woo CW, Kross E. An fMRI-based neurologic signature of physical pain. New Eng J Med 2013;368:1388–97.

Woolrich MW, Behrens TE, Beckmann CF, Jenkinson M, Smith SM. Multilevel linear modelling for FMRI group analysis using Bayesian inference. NeuroImage 2004;21:1732–47.