bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Title

Concordance between the fast and efficient mixed-effects algorithm (FEMA) and conventional mixed-effects modeling for whole-brain connectivity in longitudinal chronic pain

Authors

Natalie McLain[1], Chelsea Kaplan[1], Steven Harte[1], Andrew Schrepf[1]

mcnatali@umich.edu, chelsmar@med.umich.edu, seharte@med.umich.edu, aschrepf@med.umich.edu

1. Chronic Pain and Fatigue Research Center, Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, MI 48105, USA.

Corresponding author

Natalie McLain, Chronic Pain and Fatigue Research Center, Department of Anesthesiology, University of Michigan Medical School, 24 Frank Lloyd Wright Drive, Ann Arbor, MI 48105, USA. Email: mcnatali@umich.edu

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Abstract

Researchers increasingly have access to neuroimaging data with repeated measures within subjects. Linear mixed-effects modeling offers a valuable way to analyze such data, but its application remains limited in the neuroimaging literature due to high computational cost and a lack of mainstream analysis packages. Here we benchmark a recently published algorithm, the Fast and Efficient Mixed-Effects Algorithm (FEMA), against a conventional implementation (MATLAB's fitlme, restricted maximum likelihood) across 94,830 functional connectivity edges. In contrast to the Adolescent Brain Cognitive Development (ABCD) Study Dataset used in the development and testing of FEMA, the dataset used in the current analysis has a smaller sample size (n = 378) and comprises individuals with chronic pelvic pain scanned at up to four visits over three years. Despite these differences, the two implementations produced highly concordant results: fixed-effect estimates and test statistics correlated at $r \geq 0.999$ (Lin's concordance correlation coefficient ≥ 0.999), and the two approaches reached the same statistical conclusion for 99.98% of edges. FEMA completed each analysis roughly 20 times faster. Agreement was weaker for the estimated variance components, where FEMA attributed systematically less variance to the participant random effect, yielding a slightly lower intraclass correlation in approximately 82% of edges. We further show that most residual disagreement reflected FEMA's default variance-grid resolution rather than the estimator itself, and could be eliminated at negligible computational cost. These findings support the use of FEMA for connectome-wide analysis of repeated-measures neuroimaging data.

Keywords

linear mixed-effects models, functional connectivity, longitudinal neuroimaging, benchmarking, resting-state fMRI, chronic pain

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

1 Introduction

Identifying robust relationships between brain function and chronic pain remains a central challenge in neuroimaging. While cross-sectional analyses have provided valuable insight about group differences between individuals with chronic pain and healthy controls, these approaches often obscure substantial within-subject variability. In addition to being biologically meaningful, this variability impacts the reliability of functional magnetic resonance imaging (fMRI) measures (Flournoy et al., 2024; Nakuci et al., 2023; Noble et al., 2019, 2021; Tozzi et al., 2020). Indeed, functional connectivity specifically has shown overall poor reliability, with the largest meta-analysis of test-retest reliability in functional connectivity studies showing a field-wide average of approximately 0.3 (Noble et al., 2019). These concerns are especially relevant in the context of longitudinal cohort studies that image tens of thousands of individuals (Fiúza-Fernandes et al., 2025; Henn et al., 2023; Kim et al., 2021; Tanasescu et al., 2016).

Increasingly, neuroimaging resources such as the UK Biobank and the Adolescent Brain Cognitive Development (ABCD) study include repeated measurements and nested data structures alongside rich clinical phenotyping (Hsu et al., 2025; Kaplan et al., 2022; McQueenie et al., 2021). This presents opportunities to examine symptom-brain relationships across and within individuals over time. However, conventional analytic approaches often treat repeated observations as independent or rely on summary measures that discard within-subject information, limiting sensitivity to clinically relevant effects (Madhyastha et al., 2018).

One viable method for addressing this issue is linear mixed-effects modeling (LMEs) paired with repeated measures designs. By explicitly modeling subject-level dependence, LMEs can account for repeated measures, hierarchical data structures, and unbalanced designs (Chen et al., 2013). However, the substantial advantages offered by LMEs are rarely employed in neuroimaging analyses because of intensive computational demands and a lack of mainstream toolboxes (Chen et al., 2013; Madhyastha et al., 2018; Parekh et al., 2024; Russman Block et al., 2023). This is

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

especially true in applications such as whole-brain analyses which can involve tens of thousands of models.

Recently, the Fast and Efficient Mixed-Effects Algorithm (FEMA) was introduced to address these limitations (Parekh et al., 2024). FEMA enables mass-univariate LME analyses by dramatically reducing computation time through a combination of method-of-moments variance estimation and generalized least squares, while preserving the ability to model complex longitudinal and hierarchical data structures (Parekh et al., 2024). In its original validation, FEMA demonstrated strong agreement with conventional likelihood-based mixed-effects solvers, alongside orders-of-magnitude improvements in computational efficiency. FEMA has since been recommended for longitudinal analyses within the ABCD study (Hawes et al., 2025), the dataset it was originally validated and developed in.

Although FEMA shows strong promise, it has seen limited application outside of the dataset used in its original development. Specifically, it has yet to be demonstrated whether FEMA produces results that are concordant with conventional mixed-effects approaches when applied to repeated-measures functional connectivity data in smaller data sets of chronic pain patients, where effect sizes are often modest and within-subject variability is substantial.

Here, we apply FEMA to a longitudinal dataset of individuals with chronic pelvic pain and compare its performance against a conventional linear mixed-effects implementation using MATLAB's fitlme function. We focus on whole-brain, node-wise functional connectivity and examine two complementary dimensions of the pain experience: pain reported at the time of scanning and recalled average pain over the preceding week. Our primary goal is not to introduce new neurobiological claims, but rather to evaluate the extent to which FEMA reproduces established mixed-effects results (McLain et al., 2026) with a smaller number of subjects in a dataset and modeling context that were not part of its original validation. Specifically, we assess concordance between methods in terms of identified pain-related connectivity patterns,

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

fixed-effect estimates, and statistical inference, while also quantifying differences in computational efficiency.

By benchmarking FEMA in a longitudinal chronic pain dataset and in the context of connectome-wide functional connectivity, this work aims to inform the broader adoption of mixed-effects modeling in neuroimaging.

2 Materials and Methods

This work is based, in-part, on an earlier publication (McLain et al., 2026). The current approach allowed us to compare FEMA against a known set of results in a dataset/set of models not previously tested or used in FEMA's development. Methods for the fitlme approach were described in the earlier publication (McLain et al., 2026). Here we briefly describe the previous methods as well as methods that were added to or differ from the original approach.

2.1 Participants

Data were drawn from the Multidisciplinary Approach to the Study of Pelvic Pain (MAPP) Symptom Patterns Study (Clemens et al., 2020) where participants with urologic chronic pelvic pain scanned repeatedly (2-4 visits) over 36 months. Of the 492 chronic pain patients, the final number of subject included in this analysis was 378. Filtering criteria can be found in section 2.6 Data Filtering. At each visit, resting-state fMRI was acquired during two naturalistic bladder states: a "fuller bladder" scan following controlled water intake and an "empty bladder" scan after voiding. Participants reported pain at the time of each scan and recalled average pain over the preceding week.

2.1.1 Ethics Statement

All study protocols were approved by the Institutional Review Boards at the six study sites. All study protocols were followed according to the Declaration of Helsinki. All participants provided informed consent for participating in the study.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

2.1.2 Inclusion and Exclusion Criteria

Inclusion and exclusion criteria have been previously published (Clemens et al., 2020). Briefly, UCPPS inclusion criteria were 1) UCPPS symptoms present for a majority of the time during most recent 3 months; 2) age ≥18 years; and 3) response ≥1 on the bladder/prostate or pelvic pain/pressure/discomfort scale during past 2 weeks. Exclusion criteria include symptomatic urethral stricture, neurological disease or disorder affecting the bladder, bladder fistula, a history of cystitis caused by tuberculosis, radiation therapy or chemotherapy, prior augmentation cystoplasty or cystectomy, active autoimmune or infectious disorder, history of pelvic cancer, current major psychiatric disorder, severe cardiac, pulmonary, renal, or hepatic disease, unilateral orchialgia (without pelvic symptoms), and prior prostate procedures (transurethral microwave thermotherapy (TUMT), transurethral needle ablation (TUNA), balloon dilation, prostate cryo-surgery, or laser procedure).

2.2 MRI Protocol

The SPS MRI acquisition protocol has been described previously (Clemens et al., 2020; Mawla et al., 2020). Briefly, participants emptied their bladders and then consumed 350cc of water. After approximately 40 minutes, a resting state fMRI (rs-fMRI) scan labelled “fuller bladder” (rs-FB) was performed (10 min acquisition). Following rs-FB, participants exited the scanner and emptied their bladders. Returning to the scanner, an “empty bladder” resting state (rs-EB) scan was performed (10 min acquisition), followed by a T1-weighted structural scan.

2.2.1 MRI Acquisition

The Neuroimaging Core of the G. Oppenheimer Center for Neurobiology of Stress and Resilience (CNSR) at UCLA operated as the neuroimaging data coordinating hub for the MAPP Research Network. Scanning took place at six data collection sites: Northwestern University (NU), Chicago, Illinois; University of California/University of Southern California, Los Angeles (UCLA/USC); University of Iowa (UI), Iowa City; University of Michigan (UM), Ann Arbor; University of Washington (UW), Seattle; and Washington University (WashU), St. Louis, Missouri. The full scanning parameters and

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

quality control procedures for multi-site rs-fMRI and T1 imaging have been described previously (Mawla et al., 2020).

Here we briefly describe the nominal rs-fMRI and T1 parameters: rs-fMRI scans were acquired with a single shot echo planar imaging (EPI) pulse sequences with conventional rectangular Cartesian sampling. Basic pulse sequence parameters were as follows TR = 2000 ms, TE = 30 ms, Flip angle = 77°, FoV = 220 mm × 220 mm, Resolution = 64 × 64, Phase encode direction = A > P, Slice thickness = 4.0 mm, Slice gap = 0.5 mm, Slice acquisition = ascending (not interleaved), Slices per volume = 34-40 to cover entire brain, Phased array acceleration factor = 2, Bandwidth = maximum to accommodate resolution specifications, Orientation = axial-oblique parallel to the line between the anterior and posterior commissures, Number of volumes = 300 (10 min acquisition).

The MP-RAGE pulse sequence was used for high-resolution T1-weighted 3D volume imaging. The equivalent pulse sequence on a GE scanner was the 3D FSPGR IR. Basic parameters were as follows: TR = 2300 ms, TE = 2.98 ms, TI = 900 ms, Flip angle = GE: 11°, Siemens: 9°, Philips: 9°, FoV = 256 mm × 256 mm, Resolution = 256 × 256, Slices per volume = 240 (or maximum available while maintaining all other parameters), Slice thickness = 1 mm, Inversion = Slice Selective, parallel imaging acceleration factor = 2, Phase encode direction = left-right and superior-inferior, Orientation = axial-oblique parallel to the line between the anterior and posterior commissures.

2.2.2 rs-fMRI and T1 Preprocessing

rs-fMRI and T1 preprocessing were performed at the Chronic Pain and Fatigue Research Center, University of Michigan, Ann Arbor, MI. Results included in this manuscript come from preprocessing performed using fMRIPrep 20.2.0 (Esteban et al., 2019); RRID:SCR 016216, which is based on Nipype 1.5.1 (Esteban et al., 2022; Gorgolewski et al., 2011); RRID:SCR 002502.

2.2.3 T1 Preprocessing

The T1-weighted (T1w) image underwent intensity correction using N4BiasFieldCorrection (ANTs 2.2.0) (Tustison et al., 2010) and skull-stripping using antsBrainExtraction.sh (ANTs 2.2.0). Surface reconstruction was performed with recon-

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

all (FreeSurfer 6.0.1) (Dale et al., 1999), with additional refinement of the brainmask (Klein et al., 2017). Nonlinear registration and spatial normalization to the 2009c ICBM152 template was performed with antsRegistration (ANTs 2.2.0) (Avants et al., 2008). Tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM), and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9).

2.2.4 Functional Data Preprocessing

For each of the 2 BOLD runs per subject (rs-EB and rs-FB scans at each visit), the following preprocessing was performed. Reference images were co-registered with 9 degrees-of-freedom to the T1w reference using boundary-based registration (bbregister, FreeSurfer) (Greve & Fischl, 2009). Head-motion realignment was performed using mcflirt (FSL 5.0.9) (Jenkinson et al., 2002). Images were warped to MNI152NLin2009cAsym standard space and resampled to 2 × 2 × 2 mm voxel dimension to allow for cross-subject comparison. Framewise Displacement (FD) was calculated using Nipype (Power et al., 2014). Six physiological regressors were extracted for principal component-based noise correction based on anatomical CSF and WM masks computed in native space (aCompCor) (Behzadi et al., 2007). Following the fMRIprep minimally preprocessed pipeline, the preproc.nii images were skull-stripped using the subject-specific brain mask generated by fMRIPrep in MNI152NLin2009cAsym space, which was slightly dilated using fslmaths to ensure full cortical coverage. Based on recent recommendations (Lindquist et al., 2019), six head-motion parameters from mcflirt, six aCompCor regressors, and high-pass temporal filtering (0.01 Hz) were simultaneously regressed out of the BOLD time series using AFNI's 3dTproject function. Finally, 3DBlurToFWHM was used to estimate smoothness of each image followed by iterative smoothing until the images reached a target smoothness of 6 mm FWHM.

2.3 Motion Assessment

Two motion criteria were applied to exclude high-motion scans: maximum framewise displacement (FD) greater than 3 mm and mean FD greater than 0.5 mm. Out of 3,108 rs-FB and rs-EB runs, 10 subjects and 681 scans were excluded based on motion filtering.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

2.4 Connectivity Matrix Calculation

Connectivity matrices were calculated using the Schaefer+Brainnetome Subcortex parcellation scheme, which consists of 436 nodes: 400 cortical nodes derived from the Schaefer atlas (Schaefer et al., 2018) and 36 subcortical nodes derived from the Brainnetome atlas (Fan et al., 2016). Each node is designated to one of 8 networks: Visual (Vis), Somatosensory/Motor (SomMot), Limbic, Dorsal Attention (DorsAttn), Salience/Ventral Attention (SalVentAttn), Executive Control (Cont), and Default Mode (Default) from Schaefer and Subcortex for nodes from the Brainnetome atlas. Nodes have associated MNI-space centroids for visualization.

Post-processed, four-dimensional, rs-FB and rs-EB images were entered into nilearn. Functions under nilearn.connectome (e.g., ‘ConnectivityMeasure’) was used to compute Fisher z-transformed bivariate correlation (Pearson's r) matrices of the 436x436 density matrix. In the resulting density matrix, each index represents the measure of functional connectivity, f, calculated between a pair of nodes. Each matrix contained f values for a total of 94,830 unique edges.

2.5 Pain Assessment

Two measures of pain were used in this analysis. Recalled pain was the average self-reported genitourinary pain over the week prior to the scanning visit as assessed with responses on the Genitourinary Pain Index (GUPI), question number four: “Which number best describes your AVERAGE pain or discomfort on the days you had it, over the last week?” Possible scores range from 0-10 (Clemens et al., 2009). Pain at time of scan for both empty and full scans was calculated as an average of pain intensity ratings (0-10) measured immediately before and immediately after each scan: “Several times during the session we will ask you to rate pain and urgency to urinate. Please rate pain from 0 (no pain) to 10 (worst pain imaginable) and urgency from 0 (no urge to urinate) to 10 (worst urge imaginable)."

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

2.6 Data Filtering

Analyses were restricted to participants meeting motion criteria (worst framewise displacement <3 mm; mean <0.5 mm, see motion assessment section above). 10 participants (681 scans) were dropped as a result of this filtering step. This left a total of 482 participants (2427 scans).

In a repeated measure, mixed effects model, the variance associated with the slope of a participant with only one data point is large and, as a result, their data will have little leverage on the group-level slope estimate. As a precautionary measure and to avoid inflating the degrees of freedom, we included only a subset of the original sample: individuals who completed at least two of the four study visits. After excluding the 104 remaining individuals who completed only one visit, 378 participants remained (2234 observations).

A complete-case filtering procedure was applied globally to remove scans with missing imaging values or missing model covariates, ensuring consistent sample sizes across all edges. 6 scans were dropped as a result of this complete-case filtering, but the number of subjects stayed the same. Site effects were harmonized using ComBat with biological covariates preserved. The final dataset comprised 378 participants (2228 scans). This final dataset was used in both the fitlme and FEMA implementations (described below), so both approaches were using identical data.

2.7 Statistical Approach

Two linear mixed-effects models were fit separately for pain at the time of scan and recalled pain, adjusting for age, sex, scan type, and head motion, with a random intercept for participant.

Model 1: f ~ sex + age + MeanFD + scantype + pain at time of scan + (1|pid)

Model 2: f ~ sex + age + MeanFD + scantype + recalled pain+ (1|pid)

We ran Models 1 and 2 using both FEMA and a more basic set of scripts that relied on the fitlme function in MATLAB.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

The fitlme approach relies on the base linear mixed effects modeling function in Matlab which uses restricted maximum likelihood (REML) for variance estimation. Fitlme estimates the full random-effects covariance matrix via a Cholesky parameterization. The maximum number of optimization iterations was set to the default of 1000. Additionally, all model outputs from the fitlme approach were assessed to confirm convergence. Of the 94,830 models per pain measure, none failed to converge; 183 (pain at time of scan) and 164 (recalled pain) returned a “local minimum possible” exit status, indicating that the gradient tolerance was satisfied but optimality could not be certified. None of the “local minimum possible” edges survived FDR correction (described below). Degrees of freedom for fixed effects were approximated using the Satterthwaite method, which has been shown to produce consistent Type I error rates and be more conservative than other common approaches for approximating p-values in linear mixed-effects models (Luke, 2017).

FEMA implementation was based on the same models, Models 1 and 2. A random intercept for participant ID was specified, with the family and individual identifiers set to the participant level. Model fitting was performed using the FEMA_fit function from the CMIG Tools FEMA implementation which uses generalized least squares (GLS) to estimate mixed-effects parameters, and a method-of-moments (MoM) approach to estimate the variance components for random effects. Unlike iterative likelihood-based optimizers used for ML/REML in standard mixed-effects solvers, FEMA’s default MoM + GLS approach does not rely on iterative likelihood maximization (ML/REML) (Parekh et al., 2024). Further details on the FEMA function and its modeling approach are comprehensively outlined in the original publication (Parekh et al., 2024) and associated materials on the authors’ github repository (https://github.com/cmig-research-group/cmig_tools).

For both implementations, fitlme and FEMA, p-values for the fixed effects of pain at time of scan and recalled pain were false discovery rate (FDR) corrected across all edges (94,830) using the Benjamini-Hochberg procedure (via the mafdr function in MATLAB), with a significance threshold of $q < 0.01$.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

2.8 Assessing Concordance

Agreement between the two implementations was assessed at two levels: across the complete set of 94,830 edgewise models, independent of statistical significance, and again examining only the set of significant edges from each approach.

For the complete set of edges, we compared the standardized fixed effect of pain, its standard error, the corresponding test statistic, the signed −log10 p value, and the estimated variance components. Test statistics were computed from each implementation's own standard error (Satterthwaite-adjusted t for fitlme, Wald z for FEMA). Agreement was quantified using both the Pearson correlation, a measure of precision, and Lin's concordance correlation coefficient (CCC) (L. I.-K. Lin, 1989), which factors in both accuracy and precision coefficients (L. Lin et al., 2002; L. I.-K. Lin, 1989). Differences between implementations (FEMA minus fitlme) were further summarized as the median, interquartile range, and 95% limits of agreement (mean difference ± 1.96 SD), and displayed using Bland-Altman plots, in which the difference between methods was plotted against their mean (Bland & Altman, 1986; Giavarina, 2015; Mansournia et al., 2021).

To examine only the significant edges, we quantified set overlap with the Dice and Jaccard coefficients. Significant effects were then visualized using custom MATLAB scripts to create glass-brain representations, with edge color indicating effect direction and magnitude, and node size reflecting the number of significant connections per region.

FEMA discretizes the estimated random-effect variance into a grid and computes a single generalized least squares solution for all outcomes falling within the same grid cell, rather than solving separately for each outcome. The default resolution is set to 20 bins, but can be adjusted by the user. To determine whether this approximation contributed to any differences observed between implementations, the FEMA analyses were repeated at grid resolutions of 100, 1000, and 0 bins (one solution per edge, no binning). All other settings were held constant, and each was compared against the original/ fitlme reference.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

2.9 Use of AI

The original analysis pipeline was designed and written by the authors. Claude (Anthropic, model `claude-opus-5`) was used as an assistant throughout to review the code and identify bugs, parameterize an older version of the code to be more flexible for future applications, to add derived summary statistics during preparation of the associated manuscript, and to clean and package the repository for publication. All AI generated/edited code was compared against a known set of results. Specific applications of AI for each code file can be found in the repository README.

3 Results

MAPP-II (N=378)		Male (N=138)		Female (N=240)		2020 US Census
			(SD)		(SD)	
Age	(years)	48.5	15.4	42.6	15.6	
Symptom Duration	(years)	10.8	10.4	12.8	12.5	
Race						
		(N)	(%)	(N)	(%)	(%)
	American Indian	2	1.4%	1	0.4%	1.3%
	Asian	0	0.0%	3	1.3%	6.1%
	Black	6	4.3%	12	5%	13.6%
	Native Hawaiian	0	0.0%	1	0.4%	0.3%
	White	124	89.9%	213	88.8%	75.8%
	Multi Race	3	2.2%	9	3.8%	2.9%
	Other	3	2.2%	1	0.4%	
	Unknown	0	0%	0	0%	
Ethnicity						
	Hispanic	10	7.2%	16	6.8%	18.9%
	Non-Hispanic	128	92.8%	223	92.7%	59.3%

Table 1. The demographic information for the 378 subjects included in the analysis. Mean age (±SD) across the entire study population was 44.7±15.7 years and mean symptom duration across the entire study population (±SD) was 12.1±11.8 years.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

3.1 Demographic Information

378 subjects with a minimum of two visits were entered into this analysis (240 female / 138 male). At baseline, mean age (mean±SD) across this study population was 44.7±15.7 years and mean symptom duration was 12.1±11.8 years. Mean age was 42.6±15.6 for females and 48.5±15.4 for males. Mean duration of symptoms was 10.8±10.4 years for males and 12.8±12.5 years for females. Further demographic data is summarized in Table 1.

	Pain at time of scan		Recalled pain	
	Pearson r	**CCC**	**Pearson r**	**CCC**
Standardized β (fixed effect of pain)	0.9996	0.9996	0.9993	0.9993
Test statistic (t / z)	0.9995	0.9995	0.9993	0.9993
Signed $-\log_{10}$ p	0.9990	0.9990	0.9989	0.9989
Variance explained by fixed effects (marginal R^2)	0.9999	0.9999	0.9999	0.9999
Residual variance	0.9926	0.9878	0.9927	0.9877
Standard error of β	0.9867	0.9852	0.9865	0.9851
Variance explained by full model (conditional R^2)	0.9862	0.9753	0.9865	0.9751
Intraclass correlation coefficient (ICC)	0.9826	0.9689	0.9830	0.9686
Random-effect (participant) variance	0.9745	0.9525	0.9750	0.9520

Table 2. Correlation between implementations. Pearson r indexes precision; Lin's concordance correlation coefficient (CCC) is included for its sensitivity to systematic differences in location or scale. Test statistics are Satterthwaite-adjusted t for fitlme and Wald z for FEMA, each computed from that implementation's own standard error.

3.2 Concordance Across All Edges

We compared the two implementations across all 94,830 edgewise models, independent of significance (Table 2, Table 3, Figure 1). Standardized B weights correlated at r = 0.9996 for pain at time of scan and r = 0.9993 for recalled pain. The corresponding test statistics, Satterthwaite-adjusted t for fitlme and Wald z for FEMA, correlated at r = 0.9995 and r = 0.9993 (Figure 1A, 1D), and signed −log10 p values at r

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

= 0.9990 and r = 0.9989. For each of these quantities CCC equaled the Pearson correlation to four decimal places (B, recalled pain: r = 0.9993, CCC = 0.9993), indicating that the small discrepancies between implementations reflect random scatter rather than systematic bias. The median difference in B (FEMA-fitlme) was +0.00004 in both models, with 95% limits of agreement of [−0.0016, +0.0019] and [−0.0017, +0.0020] respectively (Table 3, Figure 1B, 1E). The largest absolute difference in B anywhere in the brain was 0.0095 (pain at time of scan) and 0.0112 (recalled pain), against typical effect magnitudes of 0.11 to 0.13. Marginal R^2 was highly similar between the implementations (r = 0.9999, CCC = 0.9999 in both models).

	Pain at time of scan		**Recalled pain**	
	Median (Q1, Q3)	**95% LoA**	**Median (Q1, Q3)**	**95% LoA**
Standardized β	+0.00004 (−0.00022, +0.00045)	−0.0016, +0.0019	+0.00004 (−0.00022, +0.00047)	−0.0017, +0.0020
Standard error of β	−0.000025 (−0.000095, +0.000063)	−0.00029, +0.00029	−0.000017 (−0.000088, +0.000078)	−0.00029, +0.00031
ICC	−0.0125 (−0.0220, −0.0032)	−0.0406, +0.0150	−0.0127 (−0.0223, −0.0037)	−0.0407, +0.0144
Random-effect variance	−0.0146 (−0.0257, −0.0042)	−0.0477, +0.0169	−0.0150 (−0.0260, −0.0047)	−0.0477, +0.0163
Residual variance	+0.0076 (+0.0012, +0.0143)	−0.0115, +0.0273	+0.0079 (+0.0015, +0.0145)	−0.0110, +0.0274

Table 3. Difference between implementations (FEMA minus fitlme). Distribution of the edgewise difference between implementations. LoA, limits of agreement (mean difference ± 1.96 SD).

Agreement in the estimated variance components was weaker and directionally biased. Intraclass correlation coefficients correlated at r = 0.983 in both models, but CCC was lower (0.969 and 0.969). Random-effect variance agreement was also lower (r = 0.975 and 0.975; CCC = 0.953 and 0.952). The gap between the two coefficients indicates a consistent offset. FEMA attributed less variance to the participant random effect than REML (median difference −0.0146 and −0.0150) and more to residual error (median difference +0.0076 and +0.0079), yielding a lower ICC in approximately 82% of edges in both models (median difference −0.0125 and −0.0127). Bland-Altman comparison (Figure 1C, 1F) showed this offset was uniform across the range of estimated ICC values.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

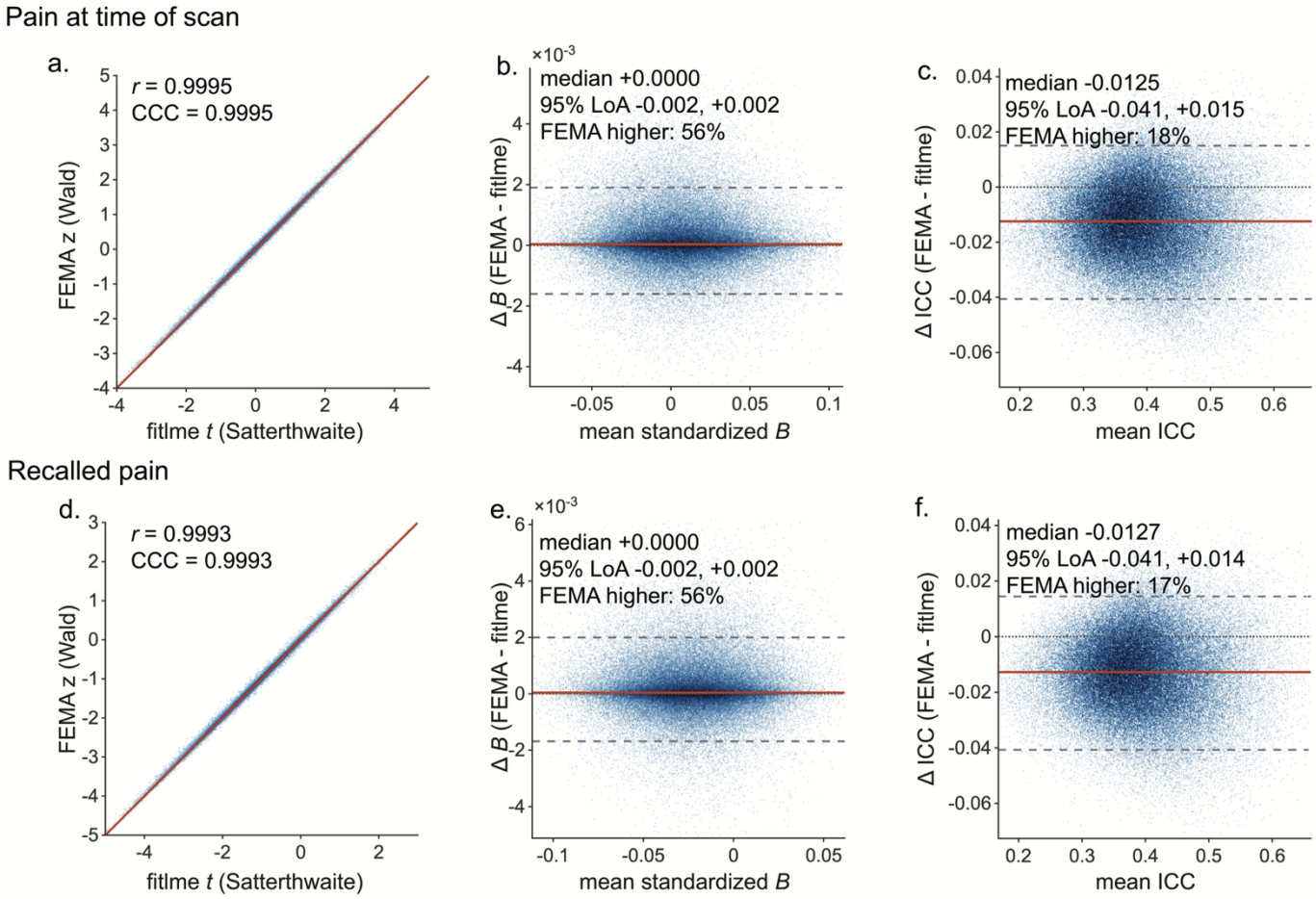


Figure 1. Agreement between the fitlme and FEMA implementations across all 94,830 edgewise models. Top row, pain at time of scan; bottom row, recalled pain (a, d). Test statistics from the two implementations, plotted against the identity line (red) (b, e). Bland-Altman comparison of the standardized fixed effect of pain: the difference between implementations (FEMA minus fitlme) is plotted against their mean (c, f). Bland-Altman comparison of the intraclass correlation coefficient. In panels b, c, e and f the solid orange line marks the median difference, dashed grey lines the 95% limits of agreement (mean ± 1.96 SD), and the dotted line zero.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

3.3 Concordance Across Significant Edges

Both FEMA and the fitlme approach identified connectivity within a sparse set of sensory-discriminative regions as being associated with pain at time of scan, while past-week recalled pain was associated with distributed patterns across regions implicated in memory, salience, and affective processing (Figure 2).

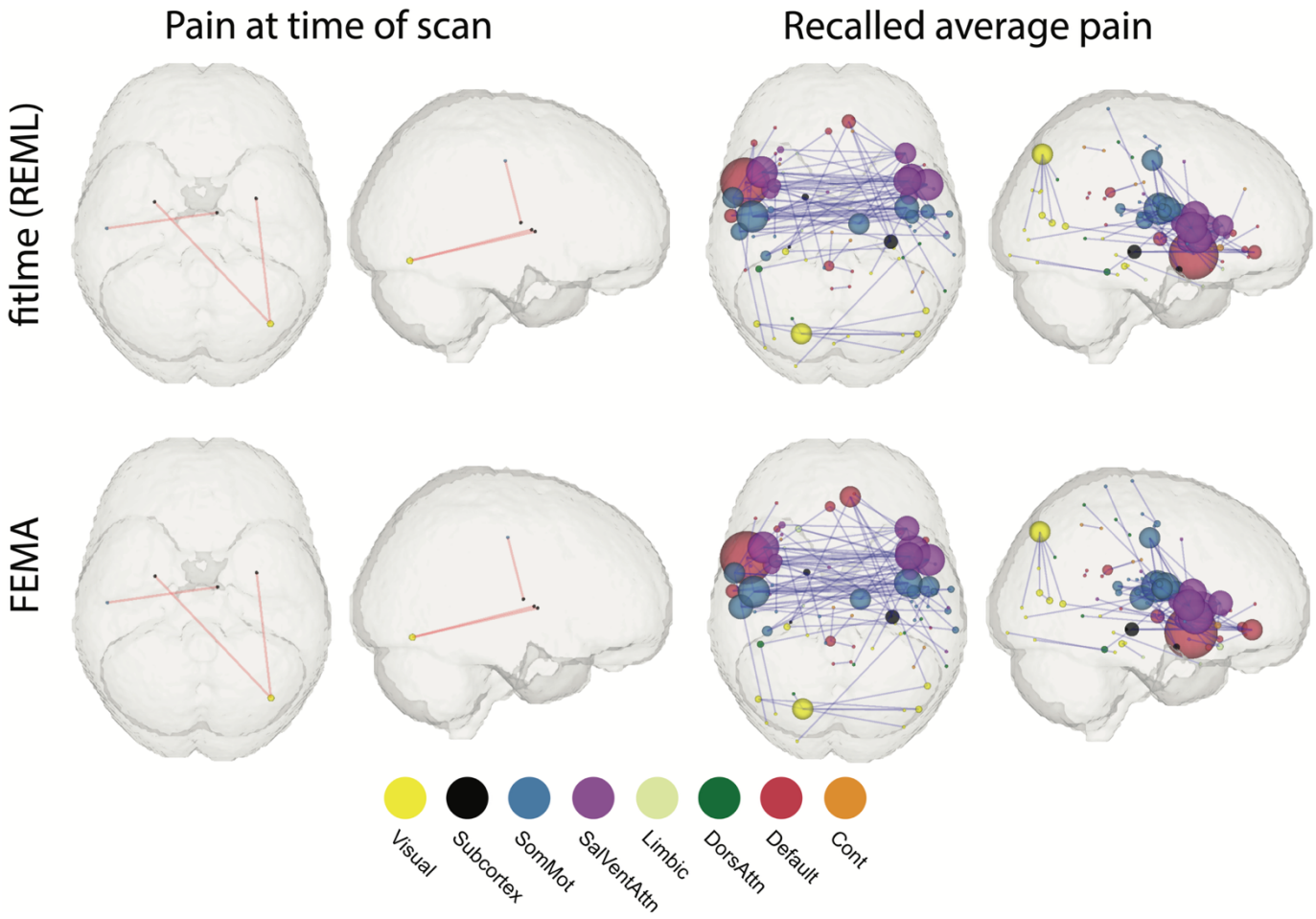


Figure 2. Comparison of significant edges identified in the pain at time of scan and recalled pain analyses for the two approaches (fitlme vs FEMA). Vis=Visual, SomMot=somatosensory/motor, SalVentAttn=Salience Ventral Attention, Cont=Executive Control. Spheres at each node are scaled by the number of significant edges they belong to. Spheres are color coded based on the network they belong to in the Schaefer/Brainnetome atlas. Line color indicates positive (red) or negative (blue) beta weight.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

For the pain at time of scan analysis, fitlme and FEMA both produced three, overlapping significant edges. The three edges implicated in the pain at time of scan analysis were the right occipital fusiform gyrus to bilateral basal ganglia and left postcentral gyrus to right thalamus, involving a total of five unique nodes. For the recalled pain analysis, fitlme produced 104 significant edges, and FEMA produced 113. 99 of these edges were overlapping between the two methods. There were 5 edges unique to the fitlme approach and 14 unique to FEMA. Set-level agreement was complete for pain at time of scan (Dice = 1.00) and high for recalled pain (Dice = 0.912; Jaccard = 0.839); treating the fitlme implementation as the reference, FEMA recovered 95.2% of significant edges, and the two implementations reached the same conclusion for 99.98% of the 94,830 edges tested (Supplementary Table 3). Summary statistics for overlapping significant edges can be found in Table 4. Model information for non-overlapping edges from the recalled pain analysis can be found in Supplementary Table 1.

	Pain at time of scan median (range), n = 3		**Recalled pain median (Q1, Q3), n = 99**	
	fitlme	**FEMA**	**fitlme**	**FEMA**
Standardized β, fixed effect of pain	0.1321 (0.1256 - 0.1356)	0.1326 (0.1257 - 0.1355)	−0.1117 (−0.1162, −0.1080)	−0.1112 (−0.1162, −0.1080)
Standard error of β	0.0252 (0.0242 - 0.0252)	0.0250 (0.0238 - 0.0253)	0.0236 (0.0233, 0.0242)	0.0237 (0.0235, 0.0242)
Intraclass correlation coefficient (ICC)	0.3884 (0.3802 - 0.4452)	0.4055 (0.3959 - 0.4517)	0.4611 (0.4148, 0.4881)	0.4358 (0.4021, 0.4701)
Variance explained by fixed effects (marginal R^2)	0.0229 (0.0222 - 0.0671)	0.0228 (0.0222 - 0.0669)	0.0381 (0.0273, 0.0560)	0.0383 (0.0275, 0.0572)
Variance explained by full model (conditional R^2)	0.4020 (0.3944 - 0.4824)	0.4187 (0.4097 - 0.4884)	0.4909 (0.4403, 0.5130)	0.4685 (0.4246, 0.4971)

Table 4. Model statistics for concordantly significant edges. Model statistics summarized across edges reaching significance under both implementations ($q < 0.01$): three edges for pain at time of scan and 99 for recalled pain. Because only three edges were concordantly

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

significant in pain at time of scan analysis, values for that model are reported as median (range). Recalled pain values are median (Q1, Q3).

We additionally examined the unique edges from the recalled pain analysis in the set of model outputs from the opposite model approach: that is, for edges that were only significant in the fitlme analysis, we examined the model statistics for that same edge in the FEMA analysis. We did the same for the edges uniquely significant in the FEMA analysis. The significant connections unique to FEMA were all edges in which models successfully converged in the fitlme analysis. Upon inspection of the model statistics for these edges in fitlme, all 14 edges were approaching significance, with q values between 0.0100 and 0.0118. The same was true of the 5 edges unique to the fitlme analysis. Model statistics for these non-overlapping edges have been included as Supplementary Table 1.

Overlapping significant edges were highly similar for the fixed effects of pain. FEMA and fitlme agreed on direction of effects for all overlapping edges in both the pain at time of scan and recalled pain analyses. Because only three edges reached significance in the pain at time of scan analysis, correlations computed across this set are uninformative; whole-brain concordance for this model is reported above. For recalled pain, the p values, q values, and normalized B weights across the 99 concordantly significant edges were also highly correlated ($r = 0.933$, 0.956, and 0.996, respectively), although these values are attenuated by range restriction relative to the whole-brain estimates.

In addition to arriving at a similar set of findings, FEMA offered a reduction in processing time: FEMA completed each whole-brain analysis in 21-35 seconds, compared with 8-10 minutes (473-602 s) for a fitlme implementation parallelized across 14 workers. With no parallel processing, the fitlme approach took approximately 65 minutes per whole-brain analysis.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

3.4 Sensitivity to FEMA's Variance-Grid Resolution

Estimated variance components were numerically identical at every grid resolution tested but standard errors changed. Agreement with the fitlme implementation in the estimated standard error improved from r = 0.987 at the default resolution (20 bins) to r = 0.996 at 100 bins, with negligible further change at 1000 bins and no binning.

For recalled pain, FEMA identified 113 significant edges at the default resolution and 100 at all finer resolutions, against the fitlme reference of 104. The number of discordant edges fell from 19 (5 unique to fitlme, 14 unique to FEMA) to 8 (6 and 2, respectively), and the Dice coefficient rose from 0.912 to 0.961. Twelve of the 14 edges uniquely identified by FEMA at the default resolution were therefore attributable to grid discretization rather than to a difference in model behavior. For pain at time of scan, the same three edges were identified at every resolution.

Computation time increased from 21-35 seconds at the default resolution to 40-58 seconds at 100 bins and 89-101 seconds with binning disabled, all of which remained substantially faster than the parallelized fitlme implementation. Full results are given in Supplementary Table 2.

4 Discussion

Here we apply FEMA (Fast and Efficient Mixed-Effects Algorithm for large-sample whole-brain imaging data) to a non-ABCD dataset to analyze whole-brain functional connectivity in individuals with chronic pain. Across two complementary dimensions of the pain experience (pain reported at the time of scanning and recalled average pain over the preceding week) FEMA produced results that were highly concordant with those obtained using a likelihood-based mixed-effects solver. In addition to producing highly similar results, FEMA offered a substantial reduction in computation time. Whole-brain analyses that required several minutes using a parallelized fitlme implementation were completed in seconds using FEMA. These findings extend the original validation of FEMA (Parekh et al., 2024) by demonstrating its reliability in a non-ABCD dataset, in a clinical population characterized by substantial within-subject variability, and in the

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

context of connectome-wide functional connectivity analyses. To date, FEMA has been applied to neuroimaging data in one other, non-ABCD dataset: the AHEAD study (Bayat et al., 2025; Curtis et al., 2025) based out of Florida International University in Miami, FL. Both papers implemented ABCD compliant preprocessing and processing approaches, and FEMA was used to analyze diffusion weighted imaging (DWI) magnetic resonance imaging (Curtis et al., 2025) and BOLD response to Kiddie-Continuous Performance Test (Bayat et al., 2025).

The neurobiological findings were not introduced as novel patterns here, but rather serve as a consistency check across modeling approaches in a dataset where these effects have been previously reported using standard mixed-effects methods (McLain et al., 2026). Briefly, both approaches identified sparse and overlapping patterns of functional connectivity associated with pain at the time of scan, primarily involving regions implicated in sensory-discriminative processing. In contrast, recalled pain over the preceding week was associated with a more distributed pattern of connectivity spanning regions commonly linked to affective, salience, and memory-related processes. This dissociation aligns with prior conceptualizations of pain as a multidimensional experience, with momentary pain more closely tied to sensory processing (Borsook et al., 2013; De Ridder et al., 2022; Liu & Kelliher, 2022) and recalled or clinical pain reflecting higher-order integrative systems (Barroso et al., 2021; Jefferson et al., 2021; Neugebauer, 2020; Stegemann et al., 2023; Waisman & Katz, 2024).

Model statistics for the fixed effects of pain (normalized *B* weight, p, and q values) were almost perfectly correlated between the two approaches. The significant findings in the pain at time of scan analysis were identical in terms of identified edges (Figure 2) and highly correlated across various model statistics (Table 4). There were some minor inconsistencies in the larger set of recalled pain findings. Out of 113 significant edges from FEMA and 104 edges from fitlme, 99 were overlapping (Table 4). This left 5 edges unique to the fitlme analysis and 14 edges unique to the FEMA analysis (Supplementary Table 1). Inspection of these non-overlapping edges revealed that they were uniformly near the significance threshold in both implementations, suggesting that

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

these differences likely reflect marginal variation in standard error estimation rather than meaningful divergence in model behavior. Several methodological differences between FEMA and conventional mixed-effects solvers may contribute to these small discrepancies.

FEMA and fitlme rely on fundamentally different estimation strategies, with fitlme using restricted maximum likelihood (REML) estimation and FEMA relying on a method-of-moments approach to estimate variance components, paired with generalized least squares for fixed-effect estimation. This is one of the core differences that allows FEMA to complete a similar set of analyses in magnitudes of order shorter time (Parekh et al., 2024). These differences in approach have the potential to produce different standard errors and variance estimates. Moment-based estimators are consistent but generally less efficient than REML in finite samples and can become biased when variance estimates are constrained to be non-negative (Swallow & Monahan, 1984). Fixed-effect point estimates, however, are generally robust to such differences, even when the estimated variance components diverge modestly (Kackar & Harville, 1984; Kenward & Roger, 1997; Parekh et al., 2024). Consistent with this expectation, we observed near-identical fixed-effect estimates across methods, alongside a small but systematic difference in the variance components: FEMA attributed less variance to the participant random effect and correspondingly more to residual error, yielding a lower ICC in approximately 82% of edges in both models.

Another potential source for the nonoverlapping findings in the recalled pain analysis are the differences in p value determination. The specific fitlme implementation described in this analysis relied on Satterthwaite degrees of freedom to estimate the final p values from fixed effects when derived from a mixed effects structure. The Satterthwaite p-values come from an approximate t/F distribution (Satterthwaite, 1946). By contrast, FEMA treats fixed-effect test statistics as Wald z-scores derived from the standard normal distribution. With moderate/large samples, Satterthwaite t and normal z are very similar (Luke, 2017), which is reflected in the p and q values identified in this analysis. Given the high correlation of p, q, and fixed effects for pain across the two

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

approaches, it seems more likely that the factor driving the difference is related to the calculation of standard error discussed in the previous paragraph.

A secondary analysis additionally indicated that the small number of edges on which the two implementations disagreed was largely attributable to FEMA's variance-grid approximation rather than to the estimator itself. Increasing the grid resolution from the default of 20 bins to 100 eliminated 12 of the 14 edges uniquely identified by FEMA, and reduced overall disagreement from 19 to 8 of 94,830 edges at a runtime cost of under a minute. Investigators whose conclusions depend on edges close to the correction threshold may therefore wish to increase this bin setting, although the default appears adequate for the type of connectome-wide analysis used in the present analysis. Notably, the systematic difference in estimated variance components was unaffected by grid resolution, indicating that it reflects a difference between moment-based and likelihood-based estimation rather than an artifact of the approximation.

Another distinction between the two approaches concerns model convergence. Conventional mixed-effects solvers require explicit convergence checks while the FEMA approach does not. Of the 94,830 models per pain measure, none failed to converge under the fitlme implementation, but 183 (pain at time of scan) and 164 (recalled pain) returned an exit status indicating that optimality could not be certified. However, none of these reached significance. Nonconvergence therefore cannot be a driver of the non-overlapping results.

Several limitations of the present work should be noted. While the present dataset includes repeated measures, the random-effects structure was limited to a random intercept for participant. In a previous version of the fitlme implementation, it was discovered that adding a random slope led to overfitting and poor convergence across the 94,830 models. More complex random-effects specifications may reveal additional differences between estimation approaches. Checks for overfitting should always be employed when applying any mixed effects approach, FEMA or otherwise.

In summary, these results demonstrate that FEMA produces results that closely match those obtained using conventional mixed-effects modeling while dramatically improving

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

computational efficiency. By benchmarking FEMA in a longitudinal chronic pain dataset and in a connectome-wide functional connectivity analysis, this work supports its use as a scalable and reliable alternative to traditional mixed-effects solvers. More broadly, FEMA lowers the practical threshold for incorporating mixed-effects modeling into whole-brain neuroimaging analyses, facilitating more rigorous treatment of within-subject variability in studies of complex clinical phenotypes.

Data and Code Availability

All MAPP neuroimaging data and associated clinical information are available through the NIDDK Central Repository (https://repository.niddk.nih.gov/). Code used to process, analyze, and generate figures from the data is available from the corresponding author on request, and publicly on publication. Code for the FEMA implementation was made available by its authors on GitHub at https://github.com/cmig-research-group/cmig_tools.

Author Contributions

Natalie McLain: Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization, Writing - original draft. Chelsea Kaplan: Methodology, Writing - review & editing. Steven Harte: Methodology, Writing - review & editing. Andrew Schrepf: Conceptualization, Methodology, Supervision, Funding acquisition, Writing - review & editing.

Funding

This work was supported by a cooperative agreement from the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases (grant numbers DK082370, DK082342, DK082315, DK082344, DK082325, DK082345, and

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

598 DK082316), as well as DK110669, DK121724, and DK123164. This work was
599 additionally supported by funding from the NIH HEAL Initiative Partnerships to Advance
600 INterdisciplinary (PAIN) Training in Clinical Pain Research (T90DE034663).

601

602 Declaration of Competing Interests

603 The authors declare no competing interests.

604

605 Acknowledgements

606 The authors thank all of the volunteers who participated in the Symptom Patterns Study.
607 The authors additionally thank the team who developed and made the FEMA algorithm
608 available for public use.

609

610 Supplementary Material

611 Supplementary Table 1: model statistics for non-overlapping significant edges in the
612 recalled pain analysis. Supplementary Table 2: full results of the variance-grid resolution
613 sensitivity analysis. Supplementary Table 3: edgewise concordance of statistical
614 conclusions across all 94,830 edges.

615

616

617

618

619

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

620

621

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

References

Avants, B., Epstein, C., Grossman, M., & Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Barroso, J., Branco, P., & Apkarian, A. V. (2021). Brain mechanisms of chronic pain: Critical role of translational approach. *Translational Research*, *238*, 76–89. https://doi.org/10.1016/j.trsl.2021.06.004

Bayat, M., Hernandez, M., Curzon, M., Garic, D., Graziano, P., & Dick, A. S. (2025). Reduced recruitment of inhibitory control regions in very young children with ADHD during a modified Kiddie Continuous Performance Task: A fMRI study. *Cortex*, *185*, 153–169. https://doi.org/10.1016/j.cortex.2024.11.025

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*(8476), 307–310.

Borsook, D., Edwards, R., Elman, I., Becerra, L., & Levine, J. (2013). Pain and analgesia: The value of salience circuits. *Progress in Neurobiology*, *104*, 93–105. https://doi.org/10.1016/j.pneurobio.2013.02.003

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

642 Chen, G., Saad, Z. S., Britton, J. C., Pine, D. S., & Cox, R. W. (2013). Linear mixed-effects
643 modeling approach to FMRI group analysis. *NeuroImage*, *73*, 176–190.
644 https://doi.org/10.1016/j.neuroimage.2013.01.047
645 Clemens, J. Q., Calhoun, E. A., Litwin, M. S., McNaughton-Collins, M., Kusek, J. W.,
646 Crowley, E. M., & Landis, J. R. (2009). Validation of a Modified National Institutes of
647 Health Chronic Prostatitis Symptom Index to Assess Genitourinary Pain in Both Men
648 and Women. *Urology*, *74*(5), 983-987.e3.
649 https://doi.org/10.1016/j.urology.2009.06.078
650 Clemens, J. Q., Kutch, J. J., Mayer, E. A., Naliboff, B. D., Rodriguez, L. V., Klumpp, D. J.,
651 Schaeffer, A. J., Kreder, K. J., Clauw, D. J., Harte, S. E., Schrepf, A. D., Williams, D.
652 A., Andriole, G. L., Lai, H. H., Buchwald, D., Lucia, M. S., Van Bokhoven, A., Mackey,
653 S., Moldwin, R. M., … Landis, J. R. (2020). The Multidisciplinary Approach to The
654 Study of Chronic Pelvic Pain (MAPP) Research Network*: Design and
655 implementation of the Symptom Patterns Study (SPS). *Neurourology and*
656 *Urodynamics*, *39*(6), 1803–1814. https://doi.org/10.1002/nau.24423
657 Curtis, M., Bayat, M., Garic, D., Alfano, A. R., Hernandez, M., Curzon, M., Bejarano, A.,
658 Tremblay, P., Pruden, S. M., Graziano, P., & Dick, A. S. (2025). Structural
659 Development of Speech Networks in Young Children: A Cross-Sectional Study.
660 *Neurobiology of Language*, *6*, nol_a_00168. https://doi.org/10.1162/nol_a_00168
661 Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis.
662 *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

De Ridder, D., Vanneste, S., Smith, M., & Adhia, D. (2022). Pain and the Triple Network Model. *Frontiers in Neurology*, *13*, 757241. https://doi.org/10.3389/fneur.2022.757241

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Esteban, O., Markiewicz, C. J., Burns, C., Goncalves, M., Jarecka, D., Ziegler, E., Berleant, S., Ellis, D. G., Pinsard, B., Madison, C., Waskom, M., Notter, M. P., Clark, D., Manhães-Savio, A., Clark, D., Jordan, K., Dayan, M., Halchenko, Y. O., Loney, F., … Ghosh, S. (2022). *nipy/nipype: 1.8.3* (Version 1.8.3) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.6834519

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., Fox, P. T., Eickhoff, S. B., Yu, C., & Jiang, T. (2016). The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cerebral Cortex*, *26*(8), 3508–3526. https://doi.org/10.1093/cercor/bhw157

Fiúza-Fernandes, J., Pereira-Mendes, J., Esteves, M., Radua, J., Picó-Pérez, M., & Leite-Almeida, H. (2025). Common neural correlates of chronic pain – A systematic review and meta-analysis of resting-state fMRI studies. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *138*, 111326. https://doi.org/10.1016/j.pnpbp.2025.111326

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

685 Flournoy, J. C., Bryce, N. V., Dennison, M. J., Rodman, A. M., McNeilly, E. A., Lurie, L. A.,
686 Bitran, D., Reid-Russell, A., Vidal Bustamante, C. M., Madhyastha, T., & McLaughlin,
687 K. A. (2024). A precision neuroscience approach to estimating reliability of neural
688 responses during emotion processing: Implications for task-fMRI. *NeuroImage*, *285*,
689 120503. https://doi.org/10.1016/j.neuroimage.2023.120503
690 Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica*, *25*(2), 141–
691 151. https://doi.org/10.11613/BM.2015.015
692 Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., &
693 Ghosh, S. S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging
694 Data Processing Framework in Python. *Frontiers in Neuroinformatics*, *5*.
695 https://doi.org/10.3389/fninf.2011.00013
696 Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using
697 boundary-based registration. *NeuroImage*, *48*(1), 63–72.
698 https://doi.org/10.1016/j.neuroimage.2009.06.060
699 Hawes, S. W., Littlefield, A. K., Lopez, D. A., Sher, K. J., Thompson, E. L., Gonzalez, R.,
700 Aguinaldo, L., Adams, A. R., Bayat, M., Byrd, A. L., Castro-de-Araujo, L. F., Dick, A.,
701 Heeringa, S. F., Kaiver, C. M., Lehman, S. M., Li, L., Linkersdörfer, J., Maullin-Sapey,
702 T. J., Neale, M. C., … Thompson, W. K. (2025). Longitudinal analysis of the ABCD®
703 study. *Developmental Cognitive Neuroscience*, *72*, 101518.
704 https://doi.org/10.1016/j.dcn.2025.101518
705 Henn, A. T., Larsen, B., Frahm, L., Xu, A., Adebimpe, A., Scott, J. C., Linguiti, S., Sharma, V.,
706 Basbaum, A. I., Corder, G., Dworkin, R. H., Edwards, R. R., Woolf, C. J., Habel, U.,

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Eickhoff, S. B., Eickhoff, C. R., Wagels, L., & Satterthwaite, T. D. (2023). Structural imaging studies of patients with chronic pain: An anatomical likelihood estimate meta-analysis. *Pain*, *164*(1), e10–e24. https://doi.org/10.1097/j.pain.0000000000002681

Hsu, T.-C. C., Whelan, P., Armitage, C. J., & McBeth, J. (2025). Identifying chronic pain subgroups in the UK biobank for persona development: A clustering analysis. *DIGITAL HEALTH*, *11*, 20552076251333497. https://doi.org/10.1177/20552076251333497

Jefferson, T., Kelly, C. J., & Martina, M. (2021). Differential Rearrangement of Excitatory Inputs to the Medial Prefrontal Cortex in Chronic Pain Models. *Frontiers in Neural Circuits*, *15*, 791043. https://doi.org/10.3389/fncir.2021.791043

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Kackar, R. N., & Harville, D. A. (1984). Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models. *Journal of the American Statistical Association*, *79*(388), 853–862. https://doi.org/10.1080/01621459.1984.10477102

Kaplan, C. M., Schrepf, A., Mawla, I., Ichesco, E., Boehnke, K. F., Beltz, A., Foxen-Craft, E., Puglia, M. P., Tsodikov, A., Williams, D. A., Hassett, A. L., Clauw, D. J., Harte, S. E., & Harris, R. E. (2022). Neurobiological antecedents of multisite pain in children. *Pain*, *163*(4), e596–e603. https://doi.org/10.1097/j.pain.0000000000002431

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

729 Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from
730 restricted maximum likelihood. *Biometrics*, *53*(3), 983–997.
731 Kim, D., Chae, Y., Park, H.-J., & Lee, I.-S. (2021). Effects of Chronic Pain Treatment on
732 Altered Functional and Metabolic Activities in the Brain: A Systematic Review and
733 Meta-Analysis of Functional Neuroimaging Studies. *Frontiers in Neuroscience*, *15*,
734 684926. https://doi.org/10.3389/fnins.2021.684926
735 Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter,
736 M., Chaibub Neto, E., & Keshavan, A. (2017). Mindboggling morphometry of human
737 brains. *PLOS Computational Biology*, *13*(2), e1005350.
738 https://doi.org/10.1371/journal.pcbi.1005350
739 Lin, L., Hedayat, A. S., Sinha, B., & Yang, M. (2002). Statistical Methods in Assessing
740 Agreement: Models, Issues, and Tools. *Journal of the American Statistical*
741 *Association*, *97*(457), 257–270. https://doi.org/10.1198/016214502753479392
742 Lin, L. I.-K. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility.
743 *Biometrics*, *45*(1), 255. https://doi.org/10.2307/2532051
744 Lindquist, M. A., Geuter, S., Wager, T. D., & Caffo, B. S. (2019). Modular preprocessing
745 pipelines can reintroduce artifacts into fMRI data. *Human Brain Mapping*, *40*(8),
746 2358–2376. https://doi.org/10.1002/hbm.24528
747 Liu, S., & Kelliher, L. (2022). Physiology of pain—A narrative review on the pain pathway and
748 its application in the pain management. *Digestive Medicine Research*, *5*, 56–56.
749 https://doi.org/10.21037/dmr-21-100

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Madhyastha, T., Peverill, M., Koh, N., McCabe, C., Flournoy, J., Mills, K., King, K., Pfeifer, J., & McLaughlin, K. A. (2018). Current methods and limitations for longitudinal fMRI analysis across development. *Developmental Cognitive Neuroscience*, *33*, 118–128. https://doi.org/10.1016/j.dcn.2017.11.006

Mansournia, M. A., Waters, R., Nazemipour, M., Bland, M., & Altman, D. G. (2021). Bland-Altman methods for comparing methods of measurement and response to criticisms. *Global Epidemiology*, *3*, 100045. https://doi.org/10.1016/j.gloepi.2020.100045

Mawla, I., Schrepf, A., Ichesco, E., Harte, S. E., Klumpp, D. J., Griffith, J. W., Strachan, E., Yang, C. C., Lai, H., Andriole, G., Magnotta, V. A., Kreder, K., Clauw, D. J., Harris, R. E., Clemens, J. Q., Landis, J. R., Mullins, C., Rodriguez, L. V., Mayer, E. A., & Kutch, J. J. (2020). Natural bladder filling alters resting brain function at multiple spatial scales: A proof-of-concept MAPP Network Neuroimaging Study. *Scientific Reports*, *10*(1), 19901. https://doi.org/10.1038/s41598-020-76857-x

McLain, N., Kaplan, C., Naliboff, B., Rodriguez, L. V., Moldwin, R. M., Williams, D. A., Farrar, J., Harris, R., Spitznagle, T., Clauw, D. J., Clemens, J. Q., Harte, S. E., Kutch, J., & Schrepf, A. (2026). Distinct connectivity patterns subserve two dimensions of the chronic pain experience. *Pain*. In press.

McQueenie, R., Jani, B. D., Siebert, S., McLoone, P., McCowan, C., Macdonald, S., Mair, F. S., & Nicholl, B. I. (2021). Prevalence of chronic pain in LTCs and multimorbidity: A

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

772 cross-sectional study using UK Biobank. *Journal of Multimorbidity and Comorbidity*,
773 *11*, 26335565211005870. https://doi.org/10.1177/26335565211005870
774 Nakuci, J., Wasylyshyn, N., Cieslak, M., Elliott, J. C., Bansal, K., Giesbrecht, B., Grafton, S.
775 T., Vettel, J. M., Garcia, J. O., & Muldoon, S. F. (2023). Within-subject reproducibility
776 varies in multi-modal, longitudinal brain networks. *Scientific Reports*, *13*(1), 6699.
777 https://doi.org/10.1038/s41598-023-33441-3
778 Neugebauer, V. (2020). Amygdala physiology in pain. In *Handbook of Behavioral*
779 *Neuroscience* (Vol. 26, pp. 101–113). Elsevier. https://doi.org/10.1016/B978-0-12-
780 815134-1.00004-0
781 Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of
782 functional connectivity: A systematic review and meta-analysis. *NeuroImage*, *203*,
783 116157. https://doi.org/10.1016/j.neuroimage.2019.116157
784 Noble, S., Scheinost, D., & Constable, R. T. (2021). A guide to the measurement and
785 interpretation of fMRI test-retest reliability. *Current Opinion in Behavioral Sciences*,
786 *40*, 27–32. https://doi.org/10.1016/j.cobeha.2020.12.012
787 Parekh, P., Fan, C. C., Frei, O., Palmer, C. E., Smith, D. M., Makowski, C., Iversen, J. R.,
788 Pecheva, D., Holland, D., Loughnan, R., Nedelec, P., Thompson, W. K., Hagler, D. J.,
789 Andreassen, O. A., Jernigan, T. L., Nichols, T. E., & Dale, A. M. (2024). FEMA: Fast
790 and efficient mixed-effects algorithm for large sample whole-brain imaging data.
791 *Human Brain Mapping*, *45*(2), e26579. https://doi.org/10.1002/hbm.26579

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

792 Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E.
793 (2014). Methods to detect, characterize, and remove motion artifact in resting state
794 fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048
795 Russman Block, S., Norman, L. J., Zhang, X., Mannella, K. A., Yang, H., Angstadt, M.,
796 Abelson, J. L., Himle, J. A., Taylor, S. F., & Fitzgerald, K. D. (2023). Resting-State
797 Connectivity and Response to Psychotherapy Treatment in Adolescents and Adults
798 With OCD: A Randomized Clinical Trial. *American Journal of Psychiatry*, *180*(1), 89–
799 99. https://doi.org/10.1176/appi.ajp.21111173
800 Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance
801 Components. *Biometrics Bulletin*, *2*(6), 110. https://doi.org/10.2307/3002019
802 Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S.
803 B., & Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex
804 from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114.
805 https://doi.org/10.1093/cercor/bhx179
806 Stegemann, A., Liu, S., Retana Romero, O. A., Oswald, M. J., Han, Y., Beretta, C. A., Gan, Z.,
807 Tan, L. L., Wisden, W., Gräff, J., & Kuner, R. (2023). Prefrontal engrams of long-term
808 fear memory perpetuate pain perception. *Nature Neuroscience*, *26*(5), 820–829.
809 https://doi.org/10.1038/s41593-023-01291-x
810 Swallow, W. H., & Monahan, J. F. (1984). Monte Carlo Comparison of ANOVA, MIVQUE,
811 REML, and ML Estimators of Variance Components. *Technometrics*, *26*(1), 47–57.
812 https://doi.org/10.1080/00401706.1984.10487921

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.20.753017; this version posted September 25, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Tanasescu, R., Cottam, W. J., Condon, L., Tench, C. R., & Auer, D. P. (2016). Functional reorganisation in chronic pain and neural correlates of pain sensitisation: A coordinate based meta-analysis of 266 cutaneous pain fMRI studies. *Neuroscience & Biobehavioral Reviews*, *68*, 120–133. https://doi.org/10.1016/j.neubiorev.2016.04.001

Tozzi, L., Fleming, S. L., Taylor, Z. D., Raterink, C. D., & Williams, L. M. (2020). Test-retest reliability of the human functional connectome over consecutive days: Identifying highly reliable portions and assessing the impact of methodological choices. *Network Neuroscience*, *4*(3), 925–945. https://doi.org/10.1162/netn_a_00148

Tustison, N. J., Avants, B. B., Cook, P. A., Yuanjie Zheng, Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Waisman, A., & Katz, J. (2024). The autobiographical memory system and chronic pain: A neurocognitive framework for the initiation and maintenance of chronic pain. *Neuroscience & Biobehavioral Reviews*, *162*, 105736. https://doi.org/10.1016/j.neubiorev.2024.105736