A protocol for modelling chronic pain using the UK Biobank Brain Imaging Data

Stefan Pszczolkowski, Duncan Hodkinson, Marianne Drabek, Ali-Reza Mohammadi-Nezhad, Stamatios Sotiropoulos, Dorothee P. Auer, Christopher R. Tench.

Objectives

Our ultimate objective is to utilise brain magnetic resonance imaging (MRI) in a representative population sample to estimate novel chronic pain-related brain features that may inform on mechanisms of pain progression and contribution from associated comorbidity. This is to describe the neural representation of the complex experience and consequences of persistent pain.

Design and methods

The approach we are taking is to start from a two-group classification task with inclusion/exclusion criteria and covariates designed in such a way that predictive models optimised for individual classification can serve a ***discovery strategy for brain signatures*** linked to the experience of living with persistent musculoskeletal pain, minimizing the risk of co-selecting neural representations linked to pain risk factors of no interest for this study (e.g. lifestyle, metabolic syndrome, medication, socio-economic) while preserving the effects of relevant comorbidities (e.g., depression, anxiety, education, adverse life events). Two data driven models will be built before an a-priori knowledge constrained model is considered.

We consider brain *signatures* to reflect composite neural representations of complex mental states identified through association strength with a given subject-reported characteristic of such state and may thus be subdivided further. Hence, we ultimately aim to *decompose the chronic pain* signature into subcomponents along available reports/phenotypes along the main domains of pain experience (sensory, affective, cognitive, behavioural). Of note, the nature of the pain conditions of interest, the UK Biobank scan protocol and available phenotypic data preclude the investigation of the sensory domain or to link brain features to specific cognitive appraisal of pain. In fact, the brain signature will

not be a signature of pain intensity as we do not know and cannot assume that participants are in pain during the fMRI acquisition. A flow chart depicting the modelling process is shown in Figure 1.

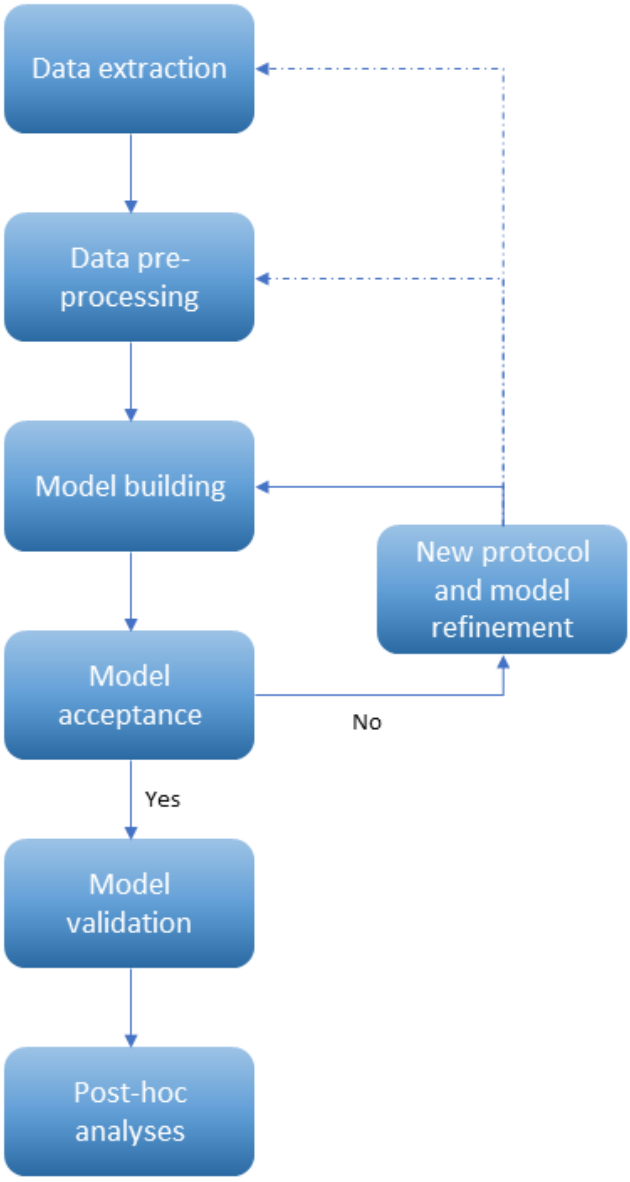

Figure 1: Modelling flow-chart.

Participant selection and group assignment

A subset of the UK Biobank cohort [1], [2] will be used as primary dataset for model training and initial testing. Two groups of subjects will be considered: chronic osteoarthritis (OA) pain and controls. The inclusion and exclusion criteria are summarised in Table 1.

Table 1: Inclusion and exclusion criteria for chronic OA pain subjects and controls.

Chronic OA pain		**Controls**	
Inclusion	**Exclusion**	**Inclusion**	**Exclusion**
• Hip or knee pain for +3 months. • OA diagnosis. • T1-weighted structural MRI passing QC. • Resting state fMRI passing QC.	• General pain for +3 months. • Neck/shoulder pain for +3 months. • Back pain for +3 months. • Stomach/abdominal pain for +3 months. • Headaches for +3 months. • Facial pains for +3 months.	• T1-weighted structural MRI passing QC. • Resting state fMRI passing QC.	• General pain for +3 months. • Neck/shoulder pain for +3 months. • Hip pain for +3 months. • Back pain for +3 months. • Stomach/abdominal pain for +3 months. • Knee pain for +3 months. • Headaches for +3 months. • Facial pains for +3 months. • Medication for pain relief. • OA diagnosis.

From an initial total dataset of approximately 42,000 subjects, we will identify all UK Biobank participants that comply with the inclusion criteria of chronic OA pain. The number of participants complying with the inclusion criteria for controls will potentially be in the few tens of thousands, which makes the pre-processing of all of them impractical. For this reason, we will select, randomly, a number of controls similar to that of chronic OA pain participants, with the added benefit of balancing the classes for the classification problem. The controls will be matched as close as possible to the chronic OA pain subjects by age and gender.

The projected final number of imaging datasets for the UK Biobank is 100,000. Therefore, if the constructed models are deemed successful, we will retrain these models using any additional chronic OA pain participants, pain phenotyping, and matching number of controls once they are available.

Magnetic resonance imaging

The UK biobank magnetic resonance imaging protocol includes several modalities. These are T1-weighted imaging, T2-FLAIR imaging, susceptibility-weighted imaging, diffusion-weighted imaging, BOLD resting-state fMRI, and task fMRI. The acquisitions were performed on three different centres equipped with identical scanners (3T Siemens Skyra, software VD13) and a standard Siemens 32-channel receive head coil. For the purpose of the proposed predictive model, only the T1-weighted and BOLD resting-state fMRI sequences will be used. The high resolution T1-weighted images were acquired in sagittal plane using a 3D MPRAGE sequence (TR=2000ms, TI=880ms, slice thickness=1mm, matrix=208x256x256, in-plane acceleration factor=2, voxel resolution=1mm^3). The BOLD resting-state fMRI data consisted of 490 volumes acquired over 6 minutes 10 seconds whilst participants were asked to keep their eyes open looking at a fixation cross and relax (TE/TR=39/735ms, slice thickness=2.4mm, flip angle=52°, matrix=88x88x64, multiband factor=8, in-plane acceleration factor=1, voxel resolution=2.4x2.4x2.4mm).

The imaging data that will be used for validation was acquired at the University of Nottingham on a GE Healthcare Discovery 750 scanner using a 32-channel head coil and includes T1-weighted and BOLD resting-state fMRI imaging. The high resolution T1-weighted images were acquired in the sagittal plane using a fast spoiled gradient echo (FSPGR) sequence (TE/TR=3.164/8.132ms, TI=450ms, slice thickness=1mm, matrix=256x256x256, flip angle=12°, voxel resolution=1mm^3). The BOLD resting-state fMRI data consisted of 205 volumes acquired over 6 minutes 50 seconds whilst participants were asked to keep their eyes open looking at a fixation cross and relax (TE/TR=30/2000ms, interleaved acquisition, slice thickness=3mm, slice gap=0.5mm, flip angle=77°, matrix=64x64x37, voxel resolution=3x3x3.5mm).

Imaging data pre-processing

To extract meaningful information from imaging data, several pre-processing steps must be carried out. We will use the Human Connectome Project (HCP) minimal processing pipelines [3] on the T1-weighted structural and resting state fMRI data. The main reason to use this processing pipeline is that it includes surface-based registration on the resting state fMRI data, leading to better alignment of sulci and gyri.

Feature extraction

To train and test a predictive model, features of interest must be extracted from the pre-processed MRI data. We will combine the Glasser and Cole-Anticevic parcellations [4], [5] to produce a whole brain (cortical + subcortical) parcellation. From this parcellation we will derive a resting state fMRI functional connectivity matrix using the HCP Connectome Workbench software [6]. Model features will correspond to the lower or upper triangle of the functional connectivity matrix plus clinico-demographic covariates. These covariates are age, gender, body mass index (BMI) and metabolic syndromes (diabetes, hypertension, dyslipidaemia, and cardiovascular). Additional covariates of no interest include average head motion, brain size, intercranial volume, variables (x, y, z, table) related to bed position in scanner, imaging centre and confounds modelling slow date-related drift.

Model construction

Logistic regression will be used to discriminate between chronic OA pain patients and controls, i.e., the regression output corresponds to the probability of belonging to the OA pain group. For this purpose, two slightly different data-driven models are to be constructed.

Elastic-net Generalised Linear Models (GLM) [7] will be used because its main purpose is to handle correlated covariates (the features), which are problematic in other schemes. These are expected given that the atlas has a high degree of granularity that we cannot a-priori guarantee is necessary but should avoid inadequate detail for the model. Elastic net then performs grouping region-to-region connections if they do not seem to contribute independently to the modelling problem. Elastic net has two parameters, λ and α, that promote sparsity in the solution (remove covariates that do not seem to contribute to the model) and propensity to group correlated covariates.

In the first instance, the whole brain parcellation will be used to generate a functional connectivity matrix and extract the functional features. Then, a 5-fold cross validation grid search with log-loss as measure will be run to determine the optimal values for parameters λ and α.

A limitation of the grouping effect of elastic-net in the context of resting state fMRI data is that it does not consider spatial information. This means that grouping is not necessarily performed on connections involving anatomically neighbouring regions, which is considered a common-sense anatomical constraint. Additionally, the grouping effect is true for features correlated both positively and negatively. To overcome these limitations, an alternative model will also be constructed based

on hierarchical merging of anatomically neighbouring regions on the whole brain parcellation. Then, the same elastic-net procedure will be run, but only on the λ parameter, since a value of 1 will be assigned to the mixing parameter α. This is equivalent to performing regression with LASSO regularisation [8]. The main difference between LASSO and elastic net is that LASSO does not form grouped covariates, which in this case is instead performed in an anatomically constrained way by the hierarchical merging. The cross-validation will be used to identify the best estimate of λ and the threshold for merging of neighbouring regions.

A second a-priori evidence-based parcellation will also be constructed based on a priori selection of brain regions with reported pain-relevance. Once the new repository-based features are selected, they will be used to create an elastic net model and a LASSO model in the same way detailed above. This evidence-based parcellation will be determined by searching for the terms “chronic pain ” or “pain” in the Neurosynth platform (neurosynth.org). The resulting spatial pattern of functional associations with pain and chronic pain will determine what parcels in the combined Glasser and Cole-Anticevic parcellations will be considered.

Model acceptance

The resulting model will be compared with other models already published at that time in internationally leading journals. If the predictive accuracy of the proposed model, estimated using cross validation, is at least as good as any of such publications, the model will be deemed ready for validation. This is considered an all-or-nothing step in that the validation data cannot be reused without explicitly becoming part of the model creation process. On the other hand, if the proposed model does not reach a comparable predictive accuracy, it will undergo a refinement process (e.g., addition of imaging modalities such as diffusion-weighted imaging) and a new protocol will be produced to inform a new model building process and potentially new data extraction and pre-processing steps. This procedure will be iteratively repeated until the model is considered ready for validation using independent data.

Model validation

The final model will be validated using the UK Biobank testing dataset. To perform this validation, the data will be randomly split into two age and gender matched partitions. These partitions will be disjoint, where one will serve for model construction (80% of the data) and the other for validation (20% of the data). Several classification measures such as accuracy, sensitivity, specificity, positive

predictive value and negative predictive value will be measured and reported. As a further validation procedure, we will test the model on a separate dataset of other types of pain in UKB, and in chronic OA pain patients and controls recruited at the University of Nottingham. As with the UK Biobank dataset, we will use the HCP processing pipeline and Connectome Workbench software to extract the relevant features from these data and will report the same previously mentioned classification measures.

Post-hoc analyses

As predefined post-hoc analyses, we will look at moderation or mediation from clinical and health markers thought to convey central pain sensitivity or resilience. These include educational level, cognition, anxiety and negative mood. Selection of the most relevant features from the rich information available in the UKB will be undertaken by experts considering the explanatory power of the final model (to define the number of possible moderator/mediator factors) and documented prior to analysis.

References

[1] C. Sudlow *et al.*, 'UK Biobank: An open access resource for Identifying the causes of a wide range of complex diseases of middle and old Age', *PLOS Med.*, vol. 12, no. 3, p. e1001779, Mar. 2015, doi: 10/bfn8.

[2] K. L. Miller *et al.*, 'Multimodal population brain imaging in the UK Biobank prospective epidemiological study', *Nat. Neurosci.*, vol. 19, no. 11, pp. 1523–1536, Nov. 2016, doi: 10/f89khc.

[3] M. F. Glasser *et al.*, 'The minimal preprocessing pipelines for the Human Connectome Project', *NeuroImage*, vol. 80, pp. 105–124, Oct. 2013, doi: 10/f46nj4.

[4] J. L. Ji, M. Spronk, K. Kulkarni, G. Repovš, A. Anticevic, and M. W. Cole, 'Mapping the human brain's cortical-subcortical functional network organization', *NeuroImage*, vol. 185, pp. 35–57, Jan. 2019, doi: 10/ggd8dm.

[5] M. F. Glasser *et al.*, 'A multi-modal parcellation of human cerebral cortex', *Nature*, vol. 536, no. 7615, pp. 171–178, Aug. 2016, doi: 10/f8z3gb.

[6] D. S. Marcus *et al.*, 'Informatics and data mining tools and strategies for the Human Connectome Project', *Front. Neuroinformatics*, vol. 5, 2011, doi: 10/fp2d3q.

[7] H. Zou and T. Hastie, 'Regularization and variable selection via the elastic net', *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, Apr. 2005, doi: 10/b8cwwr.

[8] R. Tibshirani, 'Regression shrinkage and selection via the lasso', *J. R. Stat. Soc. Ser. B Methodol.*, vol. 58, no. 1, pp. 267–288, Jan. 1996, doi: 10/gfn45m.