SUPRASPINAL PROCESSING OF SENSORY ASPECTS OF PAIN

Statistical Analysis Plan

Version 1.0

February 10, 2023

Principal Investigator:
Robert C. Coghill, Ph.D., Professor of Pediatrics Director, Pediatric Pain Research Center Division of Behavioral Medicine and Clinical Psychology Cincinnati Children's Hospital Medical Center

Supported by:
The National Institute of Neurological Disorders and Stroke (R01NS039426)

Statistical Analysis Plan

Behavioral data: Mplus[1] will be used to conduct all behavioral analyses to: 1) handle missing data via either maximum likelihood estimation or multiple imputation, with both allowing auxiliary correlate variable inclusion[2,3]; an attrition analysis will be performed using SAS PROC MI to identify possible non-random attrition dependent variable values (MNAR). If dropout is systematic, Selection and Pattern Mixture Models, and their newer mixture versions, [4,5] will be used to appropriately address MNAR attrition[2], 2) utilize several default parameter estimation algorithms (e.g., MLR, MLF, WLSMV) robust to Type-1 errors arising from non-normal response data, and 3) to allow additional options (e.g., start values, Cholesky decomposition) in the unlikely event of parameter estimation non-convergence.

Neuroimaging Data: The FSL (FMRIB's Software Library, Oxford, UK) software package will be used for the vast majority of image processing operations and statistical analyses. FSL analyses will be augmented by AFNI and/or CONN[6] and python scripts.

Spatial processing and transformation: T1-weighted images will be brain extracted and then non-linearly warped into standard anatomic space (MNI152). EPI images will be motion-corrected, unwarped, and registered to the high resolution T1 structural image and then nonlinearly warped to standard space. To facilitate intersubject comparisons and to reduce the number of statistically independent comparisons, BOLD images will be smoothed with a 5 mm FWHM filter.

Processing of pCASL images into CBF images: A single fully quantified CBF volume (ml/100g/min.) will be calculated from each 4D series of PCASL images[7] following motion correction, tag-control subtraction, and assessment of T1 signal. CBF images will be transformed to standard space as described above.

Statistical analysis of imaging data: CBF data from each individual will be motion corrected, and ratio normalized to minimize the impact of fluctuations in global CBF. A first level fixed effects analysis will be executed within FEAT to identify within subject effects. A second level random effects analysis will be executed within FEAT to identify between group effects. Clusters of activation will be identified using a threshold of Z>3.1 and their statistical significance will be estimated according to Gaussian random field theory.[8]

BOLD / Connectivity analyses: The aCompCor approach will first be used to reduce variability due to physiological and scanner noise[9] using a processing pipeline integrating modules from FSL and AFNI. Denoised data will then be imported into FEAT. Time courses of activity will be extracted from seed regions. These seed regions will be objectively defined on the basis of each participant's anatomy. Next, first level, fixed-effects analyses will be run for each BOLD series to identify voxels that have time courses that are significantly correlated with that of the seed. Second level analyses will examine effects across imaging series, but within subjects. Finally third level random effects analyses will identify differences in functional connectivity according to groups. Clusters of activation will be identified using a threshold of Z>3.1 and their statistical significance will be estimated according to Gaussian random field theory.[8] Age and sex will be added as covariates.

Conjunction analyses: Conjunction analyses will be performed on both CBF and BOLD data in order to determine if activation (or connectivity) overlaps between chronic pain groups.[10] This analysis tests the null hypothesis of no overlap, and as such, is an optimal method to test for similar patterns (but not magnitudes) of activation or connectivity. Statistical significance of overlapping clusters will be determined according to Gaussian Random Field theory.[8]

Structural analyses: T1-weighted structural data will be analyzed with FSL-VBM as we have done previously.[11] A regression analysis will be performed using a general linear model to examine the relationship between treatment type and grey matter differences across the whole brain. Age and sex will be added as covariates. Permutation-based nonparametric testing

(10,000 permutations) will be used to evaluate this relationship in a voxel-wise fashion. Threshold-free cluster enhancement will be utilized to define significant clusters. A familywise error corrected P value of P < .05 will be applied to correct for multiple comparisons and to identify clusters exhibiting a significant relationship between grey matter density and treatment type.

Multi-task deep Ensemble learning model: We will design the novel multi-task deep *Ensemble* learning model to be a two-level ensemble model (Fig. 6), combining the predictive power of both state-of-the-art deep learning and traditional machine learning. We will 1) first build a diverse model library. The diversity plays a key role, and it is a necessary and sufficient condition in building a powerful stacking ensemble model.[12-14] Each input data type (*i.e.*, features extracted from fMRI, QST, and psychological assessments) will be used to create a series of unique machine learning models. We will build a model library that will consist of a diverse set of multiple traditional models, including SVM,[15] Artificial Neural Networks (ANN),[16] random forest (RF),[17] LR,[18] Ridge[19] and least absolute shrinkage and selection operator (LASSO).[20] Multiple models will be trained with different hyperparameter settings and training datasets. 2) We will then integrate the multiple machine learning classifiers from the model library using our multi-channel deep neural network (DNN) as a fusion model.[21,22] The number of channels is designed based on the number of models in model library. Each input channel will contain several neural network blocks. The multiple input channels will be eventually fused into one output channel through a fusion block. Each block will consist of a fully connected layer, a batch normalization layer, and a dropout regularization layer. Followed by the fusion block, a softmax output layer will be used to predict chronic pain conditions (i.e., migraine, complex regional pain syndrome, musculoskeletal pain, functional abdominal pain, as well as heathy controls) and pain trajectory. We will perform nested k-fold cross-validation (i.e., training, validation, and testing dataset split method) to evaluate our model with multiple metrics, including multi-class accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). We will perform data augmentation on training data to prevent model overfitting using our prior method.[23] Hyperparameters of the model will be optimized based on validation data before testing on unseen test datasets. For feature ranking, we will apply a connection weights method[24] to identify the most discriminative features for each chronic pain condition or trajectory of recovery. The deep *Ensemble* learning model will be implemented using Python, scikit-learn, and Tensorflow package. Prior chronic pain studies[25] demonstrated that a robust deep learning model can be obtained using ~200 samples. Thus, we expect that the sample size (500 subjects) in this work, combined with data augmentation strategy, is sufficient for our deep learning model. During the data collection period, we will develop/optimize analysis pipelines with existing patient and control data.

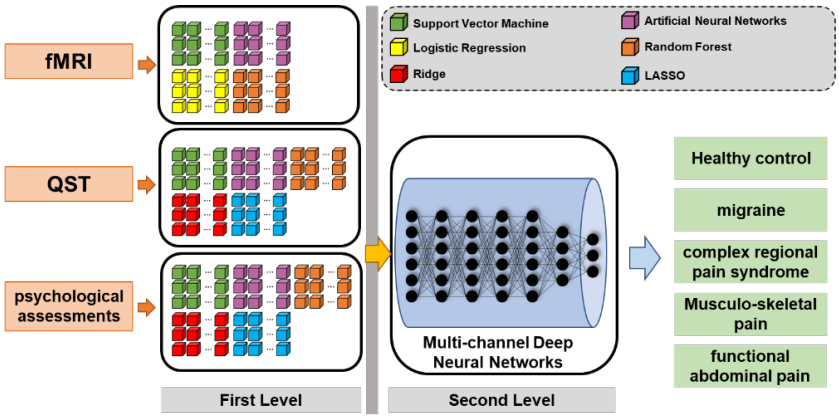

Figure 1. Schematic view of the multi-task deep Ensemble learning model.

<u>*Power analysis and missing data:*</u> Power calculations were performed for ***neuroimaging data*** to ensure that we have an adequate sample size 1) to detect brain activity changes in hypothesis-directed analyses and 2) to identify relevant brain mechanisms through deep Ensemble learning techniques. Power calculations for neuroimaging data are challenging since such calculations depend crucially on effect size as well as properties of the imaging data and statistical approach

used to deal with the multiple comparisons of >20,000 voxels. We used the NeuroPower tool[26] to calculate statistical power and sample sizes based on our preliminary BOLD in youth with migraine. Based on ***between group*** comparisons of **BOLD** functional connectivity data between migraine patients and controls, 36 participants (18 participants/group) would be required for 80% power in between group comparisons. However, contrasts between patient groups would likely need greater numbers due to potentially more subtle differences. These comparisons were calculated using z-transformed statistical images of the whole brain, a cluster-forming threshold of $z>3.1$ and $p<0.05$, isotropic smoothness of 5mm, and voxel sizes of 2x2x2mm, and a Gaussian Random Field theory-based approach for multiple comparisons. For these complex data, statistical power is defined as an 80% probability of correctly detecting an active peak for all peaks above the cluster-forming threshold. Power calculations for the deep Ensemble learning techniques are nearly impossible to develop given the nature of the analyses, however, analogous machine learning approaches with pain data required 109 participants to develop a marker for a single group.[27] Accordingly, we estimate that 100 participants/group would provide adequate power for both hypothesis-directed analyses and machine learning analyses.

Plan for Robust and Unbiased Results: Our group has a history of producing highly reproducible imaging and psychophysical studies. Our original psychophysical finding of offset analgesia[28] has been replicated in more than 22 papers by laboratories across the world. Our original finding of anterior insular activation during pain[29] has been replicated by hundreds of brain imaging studies.[30-33] Our imaging studies have been highly reproducible, in part, because we always use whole-brain searches rather than region of interest (ROI) analyses, consistent corrections for multiple comparisons using conservative cluster-forming thresholds, and random effects statistical models to increase generalizability and to diminish outlier effects. This highly-powered data set will be analyzed with a conservative, statistically rigorous approach designed to maximize reproducibility.[34] Towards this end, all analyses will be performed across the entire brain and both positive and negative relationships will be assessed and reported.[35,36] These analyses will be controlled for multiple comparisons by cluster-based methods, such that family-wise error rates will be held to a $p<0.05$. Region of interest approaches will be avoided in order to minimize errors due to confirmation bias and "double-dipping."[35,36] Analyses will be conducted by individuals blinded to group assignment in order to further minimize biases.

1. Muthen LK. Mplus user's guide. . 7th Edition ed. Los Angeles, CA: Muthen & Muthen; 2015.
2. Enders CK. Applied missing data analysis. New York: Guilford Press; 2010.
3. Graham JW. Missing data: Analysis and design. New York: Springer; 2012.
4. Gottfredson NC, Bauer DJ, Baldwin SA. Modeling Change in the Presence of Non-Randomly Missing Data: Evaluating A Shared Parameter Mixture Model. Struct Equ Modeling. 2014;21(2):196-209. Epub 2014/07/12. doi: 10.1080/10705511.2014.882666. PubMed PMID: 25013354; PMCID: PMC4084916.
5. Muthen B, Asparouhov T, Hunter AM, Leuchter AF. Growth modeling with nonignorable dropout: alternative analyses of the STAR*D antidepressant trial. Psychol Methods. 2011;16(1):17-33. Epub 2011/03/09. doi: 10.1037/a0022634. PubMed PMID: 21381817; PMCID: PMC3060937.
6. Whitfield-Gabrieli S, Nieto-Castanon A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connect. 2012;2(3):125-41. doi: 10.1089/brain.2012.0073. PubMed PMID: 22642651.
7. Wang J, Alsop DC, Li L, Listerud J, Gonzalez-At JB, Schnall MD, Detre JA. Comparison of quantitative perfusion imaging using arterial spin labeling at 1.5 and 4.0 Tesla. Magn Reson Med. 2002;48(2):242-54. doi: 10.1002/mrm.10211. PubMed PMID: 12210932.

8. Worsley KJ, Evans AC, Marrett S, Neelin P. A three-dimensional statistical analysis for CBF activation studies in human brain. JCerebBlood Flow Metab. 1992;12:900-18.
9. Behzadi Y, Restom K, Liau J, Liu TT. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage. 2007;37(1):90-101. doi: 10.1016/j.neuroimage.2007.04.042. PubMed PMID: 17560126; PMCID: 2214855.
10. Nichols T, Brett M, Andersson J, Wager T, Poline JB. Valid conjunction inference with the minimum statistic. Neuroimage. 2005;25(3):653-60. Epub 2005/04/06. doi: S1053-8119(04)00750-5 [pii]
10.1016/j.neuroimage.2004.12.005. PubMed PMID: 15808966.
11. Emerson NM, Zeidan F, Lobanov OV, Hadsel MS, Martucci KT, Quevedo AS, Starr CJ, Nahman-Averbuch H, Weissman-Fogel I, Granovsky Y, Yarnitsky D, Coghill RC. Pain sensitivity is inversely related to regional grey matter density in the brain. Pain. 2014;155(3):566-73. Epub 2013/12/18. doi: 10.1016/j.pain.2013.12.004. PubMed PMID: 24333778; PMCID: PMC3944887.
12. Dietterich TG. Ensemble methods in machine learning. International workshop on multiple classifier systems: Springer; 2000. p. 1-15.
13. Dietterich TG. Ensemble learning. The handbook of brain theory and neural networks. 2002;2:110-25.
14. Güneş F, Wolfinger R, Tan P-Y. Stacked ensemble models for improved prediction accuracy. Proceedings of Statistical Annual Symposium2017. p. 1-19.
15. Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995;20(3):273-97.
16. Bishop CM. Neural networks for pattern recognition: Oxford university press; 1995.
17. Ho TK, editor. Random decision forests. Proceedings of 3rd international conference on document analysis and recognition; 1995: IEEE.
18. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons; 2013.
19. Hoerl AE, Kennard RWJT. Ridge regression: Biased estimation for nonorthogonal problems1970;12(1):55-67.
20. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996:267-88.
21. He L, Li H, Wang J, Chen M, Gozdas E, Dillman JR, Parikh NA. A multi-task, multi-stage deep transfer learning model for early prediction of neurodevelopment in very preterm infants. Sci Rep. 2020;10(1):15072. Epub 2020/09/17. doi: 10.1038/s41598-020-71914-x. PubMed PMID: 32934282; PMCID: PMC7492237.
22. Chen M, Li H, Wang J, Dillman JR, Parikh NA, He L. A Multichannel Deep Neural Network Model Analyzing Multiscale Functional Brain Connectome Data for Attention Deficit Hyperactivity Disorder Detection. Radiology Artificial intelligence. 2019;2(1):e190012. Epub 2020/02/23. doi: 10.1148/ryai.2019190012. PubMed PMID: 32076663; PMCID: PMC6996597.
23. Chen M, Li H, Wang J, Yuan W, Altaye M, Parikh NA, He L. Early Prediction of Cognitive Deficit in Very Preterm Infants Using Brain Structural Connectome With Transfer Learning Enhanced Deep Convolutional Neural Networks. Frontiers in neuroscience. 2020;14:858. Epub 2020/10/13. doi: 10.3389/fnins.2020.00858. PubMed PMID: 33041749; PMCID: PMC7530168.
24. Olden JD, Jackson DA. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. Ecological modelling. 2002;154(1-2):135-50.
25. Santana AN, Cifre I, de Santana CN, Montoya P. Using Deep Learning and Resting-State fMRI to Classify Chronic Pain Conditions. Frontiers in neuroscience. 2019;13:1313. Epub 2020/01/11. doi: 10.3389/fnins.2019.01313. PubMed PMID: 31920483; PMCID: PMC6929667.

26. Durnez J, Degryse J, B M, R S, V S, RA P, TE. N. Power and sample size calculations for fMRI studies based on the prevalence of active peaks. bioRxiv. 2016.
27. Lee JJ, Kim HJ, Ceko M, Park BY, Lee SA, Park H, Roy M, Kim SG, Wager TD, Woo CW. A neuroimaging biomarker for sustained experimental and clinical pain. Nat Med. 2021;27(1):174-82. Epub 2021/01/06. doi: 10.1038/s41591-020-1142-7. PubMed PMID: 33398159; PMCID: PMC8447264.
28. Grill JD, Coghill RC. Transient analgesia evoked by noxious stimulus offset. J Neurophysiol. 2002;87(4):2205-8. Epub 2002/04/04. doi: 10.1152/jn.00730.2001. PubMed PMID: 11929939.
29. Coghill RC, Talbot JD, Evans AC, Meyer E, Gjedde A, Bushnell MC, Duncan GH. Distributed processing of pain and vibration by the human brain. J Neurosci. 1994;14(7):4095-108. Epub 1994/07/01. PubMed PMID: 8027764; PMCID: PMC6577049.
30. Coghill RC, Gilron I, Iadarola MJ. Hemispheric lateralization of somatosensory processing. J Neurophysiol. 2001;85(6):2602-12. Epub 2001/06/02. doi: 10.1152/jn.2001.85.6.2602. PubMed PMID: 11387404.
31. Oshiro Y, Quevedo AS, McHaffie JG, Kraft RA, Coghill RC. Brain mechanisms supporting spatial discrimination of pain. J Neurosci. 2007;27(13):3388-94. Epub 2007/03/30. doi: 10.1523/JNEUROSCI.5128-06.2007. PubMed PMID: 17392455; PMCID: PMC6672117.
32. Lobanov OV, Quevedo AS, Hadsel MS, Kraft RA, Coghill RC. Frontoparietal mechanisms supporting attention to location and intensity of painful stimuli. Pain. 2013;154(9):1758-68. Epub 2013/05/29. doi: 10.1016/j.pain.2013.05.030. PubMed PMID: 23711484; PMCID: PMC3755961.
33. Lobanov OV, Zeidan F, McHaffie JG, Kraft RA, Coghill RC. From cue to meaning: brain mechanisms supporting the construction of expectations of pain. Pain. 2014;155(1):129-36. Epub 2013/09/24. doi: 10.1016/j.pain.2013.09.014. PubMed PMID: 24055334; PMCID: PMC3947355.
34. Poldrack RA, Baker CI, Durnez J, Gorgolewski KJ, Matthews PM, Munafo MR, Nichols TE, Poline JB, Vul E, Yarkoni T. Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat Rev Neurosci. 2017;18(2):115-26. Epub 2017/01/06. doi: 10.1038/nrn.2016.167. PubMed PMID: 28053326; PMCID: PMC6910649.
35. Bennett CM, Wolford GL, Miller MB. The principled control of false positives in neuroimaging. Soc Cogn Affect Neurosci. 2009;4(4):417-22. Epub 2010/01/01. doi: 10.1093/scan/nsp053. PubMed PMID: 20042432; PMCID: PMC2799957.
36. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci. 2009;12(5):535-40. Epub 2009/04/28. doi: 10.1038/nn.2303. PubMed PMID: 19396166; PMCID: PMC2841687.