Establishing Consensus in Categorizing Resting State fMRI Brain Networks: An Inter-Rater Reliability Study

Maithili Kulkarni, Sriya Darsi, William R. Reuther, MS, Emilio G. Cediel, MD, Varina L. Boerwinkle, MD

UNC rs-fMRI Clinical Service And Neuroscience Laboratory | UNC School of Medicine, Department of Neurology

Abstract

Successful clinical use of resting state functional magnetic resonance imaging (rs-fMRI) analysis hinges upon accurate categorization of intrinsic brain networks, yet variability in interpretation among clinicians poses a challenge. We present an inter-rater reliability study aimed at establishing a gold standard for categorizing rs-fMRI-derived brain networks. Leveraging an independent component analysis (ICA) pipeline, raters categorized networks using an internal report generation tool, reflecting interpretations across nine anatomical regions and four specific categories including “normal”, “atypical”, “atypical resting state”, and “noise”. Preliminary findings reveal variability among raters, underscoring the need for a consensus framework to ensure consistency in clinical interpretation. Our analysis revealed moderate to substantial concordance among raters regarding the classification of “atypical” and “noise,” whereas agreement was minimal to absent for the “other” category. In most instances, a sole rater assigned the “other” classification for a participant's data, while the remaining raters consistently concurred on decisions related to “atypical” or “noise,”. Our study highlights the importance of interdisciplinary collaboration in refining categorization protocols to enhance clinical reliability and reproducibility in neurological research.

Background

Resting state fMRI is a valuable tool for studying intrinsic brain activity without task-based paradigms[1]. However, variability in interpreting rs-fMRI data challenges its reliability, especially in clinical settings[2]. Standardizing protocols is insufficient, highlighting the need for methods to improve inter-rater reliability. This study focuses on patients with epilepsy, aiming to identify commonalities and discrepancies in network classification among raters using ICA. Establishing a gold standard for network categorization is crucial for clinical interpretation and advancing understanding of brain function in health and disease.

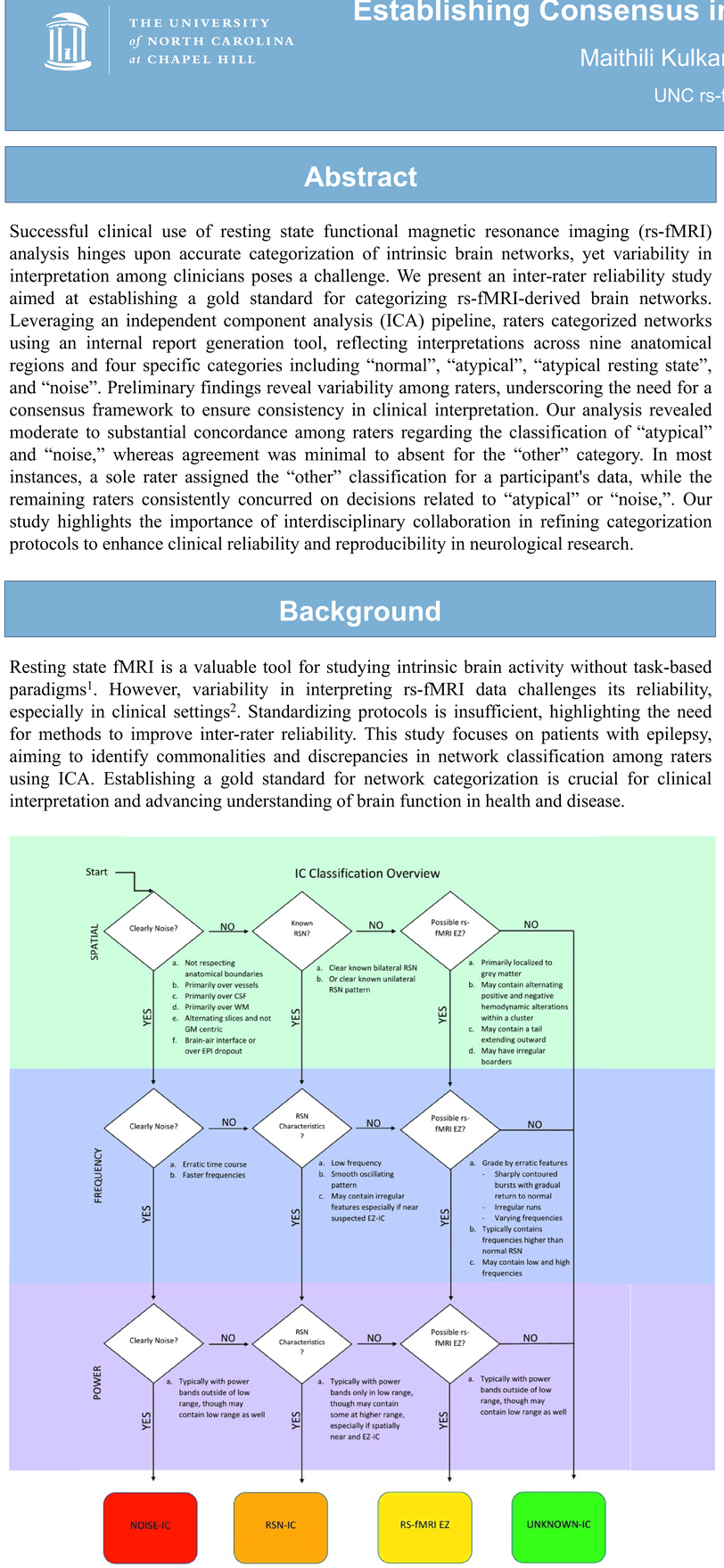

Fig. 1 – ICA classification decision making pipeline. *Image courtesy of Boerwinkle et al. 2017*[3]

Methods

From a total sample of two hundred and fifty-five, a pilot proportion of sixteen patients diagnosed with epilepsy have participated in this study so far. Participants were selected based on the availability of rs-fMRI data acquired as part of their clinical evaluation at the University of North Carolina Main Hospital.

Demographic and Clinical Data	Results (n = 255)
Age, range (years)	0 - 68
Sex, n (%)	
Male	119 (46.67)
Female	136 (53.33
Clinical Diagnosis, n (%)	
Epilepsy	52 (20.39)
Acute Brain Injury	58 (22.75)
Other	18 (7.06)
Combination	127 (49.80)

Fig. 2 - Demographic and clinical characteristics study participants.

Three experienced raters were recruited for the initial phase of the study, with plans to expand to a total of ten raters in subsequent analyses. Each rater independently categorized rs-fMRI-derived brain networks using an established independent component analysis (ICA) pipeline[4]. Thirteen predefined network categories were considered, encompassing separate and unique domains: "Atypical," "Atypical resting state," "Association," "Deep grey," "Frontal," "Language," "Modulatory," "Motor," "Parietal," "Temporal," "Vision," "Noise," and "Other."

Fig. 3 - ICA analysis rater comparison protocol

Objectives

Assess Inter-Rater Reliability: Evaluate the inter-rater reliability in the categorization of rs-fMRI-derived brain networks among researchers and clinicians, quantify the degree of agreement, and identify areas of discrepancy in network classification.

Establish Consensus Framework: Establish a consensus framework for categorizing rs-fMRI-derived brain networks through iterative analysis and discussion among raters.

Validate Categorization Protocols: Validate the effectiveness and generalizability of the categorization protocols by examining agreement among raters across multiple anatomical regions and predefined network categories.

Inform Clinical Practice: Translate our findings into clinical practice by providing insights and recommendations for improving the reliability and reproducibility of rs-fMRI analysis.

Discussion

Through the evaluation of inter-rater agreement across multiple anatomical regions and predefined network categories, we identified areas of both moderate to substantial concordance and minimal to absent agreement. Specifically, we observed higher agreement for anatomical categories such as “Motor" and “Frontal," while discrepancies were more pronounced for the "Other" category. These findings emphasize the importance of refining categorization protocols and establishing common criteria for network classification.

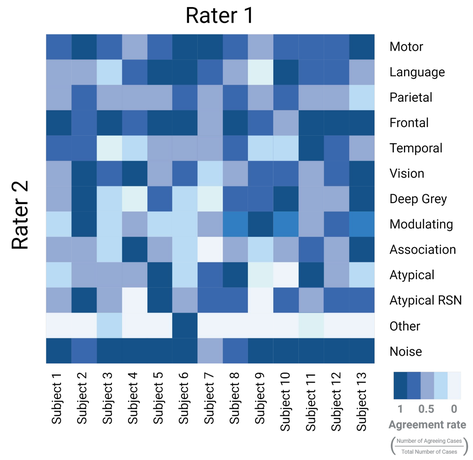

Fig. 4 – Sample rater agreement comparison between 13 subjects and 13 network classifications (Left).

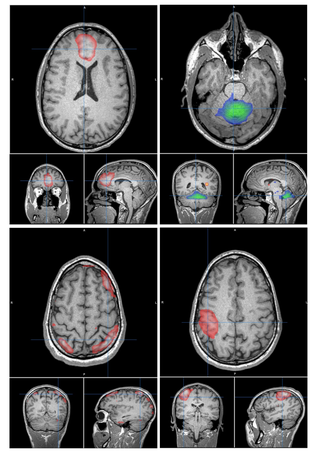

Fig. 5 – Example of 4 networks with strong agreement: Frontal (Top left), Deep grey (Top right), Noise (Bottom left), Motor (Bottom right)

While this study represents a significant step towards understanding inter-rater reliability in rs-fMRI analysis, it is important to acknowledge that our work is ongoing. In future phases of this project, we plan to expand our cohort to include a larger number of raters and participants, as well as bolster our data analysis by way of Cohen’s Kappa coefficient measurement. Additionally, we intend to explore the impact of clinical variables and imaging parameters on network interpretation, providing valuable insights for improving the clinical utility of rs-fMRI in epilepsy and other neurological disorders.

This study highlights the complexity of network interpretation in rs-fMRI and underscores the importance of standardized protocols and guidelines to ensure consistency and reproducibility in clinical practice. By establishing a consensus framework for network classification, we can advance our understanding of brain function and dysfunction in health and disease, paving the way for improved patient care and management in neurological research.

References

1. Boerwinkle VL, Mirea L, Gaillard WD, Sussman BL, Larocque D, Bonnell A, Ronecker JS, Troester MM, Kerrigan JF, Foldes ST, Appavu B, Jarrar R, Williams K, Wilfong AA, Adelson PD. Resting-state functional MRI connectivity impact on epilepsy surgery plan and surgical candidacy: prospective clinical work. J Neurosurg Pediatr. 2020 Mar 20:1-8. doi: 10.3171/2020.1.PEDS19695. Epub ahead of print. PMID: 32197251.
2. Cole David, Smith Stephen, Beckmann Christian. Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. Frontiers in Systems Neuroscience. Vol. 4, 2010. DOI; 10.3389/fnsys.2010.00008. ISSN=1662-5137
3. Boerwinkle VL, Mohanty D, Foldes ST, Guffey D, Minard CG, Vedantam A, Raskin JS, Lam S, Bond M, Mirea L, Adelson PD, Wilfong AA, Curry DJ. Correlating Resting-State Functional Magnetic Resonance Imaging Connectivity by Independent Component Analysis-Based Epileptogenic Zones with Intracranial Electroencephalogram Localized Seizure Onset Zones and Surgical Outcomes in Prospective Pediatric Intractable Epilepsy Study. Brain Connect. 2017 Sep;7(7):424-442. doi: 10.1089/brain.2016.0479. PMID: 28782373; PMCID: PMC5647510.
4. M.W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, S.M. Smith. Bayesian analysis of neuroimaging data in FSL. NeuroImage, 45:S173-86, 2009