Measuring the Quality of Resting State fMRI Data

Zarrar Shehzad[1,2], Qingyang Li[1], Yassine Benhajali[4,5], Daniel Lurie[1], David O'Connor[1,2], Chao Gan Yan[2,6], Zhen Yang[1,2], Michael P Milham[1,2], Pierre Bellec[4], and R Cameron Craddock[1,2]

1 Center for the Developing Brain, Child Mind Institute, New York, NY; **2** Nathan Kline Institute for Psychiatric Research, Orangeburg , NY; **4** Dépt d'anthropologie, Université de Montréal, Montréal, QC; **5** Centre de recherche de l'institut de gériatrie de Montréal, Montréal, QC; **6** Dept of Child & Adolescent Psychiatry, New York University, New York, NY; **7** Dépt d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, QC

cameron.craddock@childmind.org / http://preprocessed-connectomes-project.github.io/quality-assessment-protocol

Introduction

- Many measures have been proposed to assess MRI data quality[1]
- However, no clear guidance is given on which measures to choose or the range of values that constitute 'good' or 'bad' data
- Here, we provide a toolbox and normative distributions for spatial and temporal measures of data quality on two large resting-state datasets
- And we assess the validity and reproducibility of our measures

Methods

- For details on the quality assessment (QA) metrics, see website and code[1,2]
- *ABIDE dataset*: 1,112 participants across 20 sites
- *CoRR dataset*: 1,439 participants across 31 sites

Spatial Measures (anatomical or mean functional):

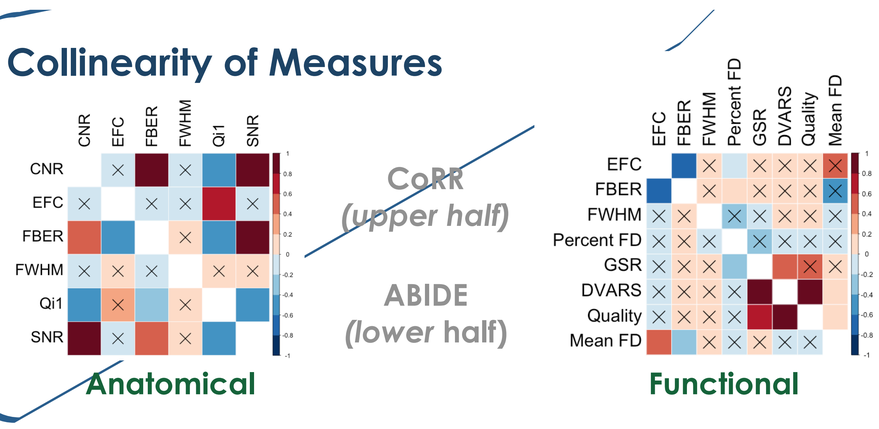
- Contrast to Noise Ratio (CNR): Mean of the gray matter (GM) values minus the mean of the white matter (WM) values, divided by the standard deviation of the air values - *higher is better*
- Entropy Focus Criterion (EFC): Shannon entropy of voxel intensities indicating ghosting & blurring induced by head motion - *lower is better*
- Foreground to Background Energy Ratio (FBER): Mean energy of values within the head relative to outside the head - *higher is better*
- Smoothness of Voxels (FWHM): Full-width half maximum (FWHM) of the spatial distribution of intensity values in units of voxels - *lower is better*
- Percent of Artifact Voxels (Qi1): Prop. of voxels with intensity corrupted by artifacts normalized by # of voxels in background - *lower is better*
- Signal-to-Noise Ratio (SNR): Mean of values within GM divided by the standard deviation of values within air - *higher is better*
- Ghost to Signal Ratio (GSR): Mean signal in the 'ghost' image (signal present outside the brain due to acquisition in phase encoding direction) relative to mean signal within the brain - *lower is better*

Temporal Measures (functional time-series):

- Standardized DVARS: Spatial standard deviation of the temporal derivative, normalized by the temporal standard deviation and temporal autocorrelation - *lower is better*
- Median Distance Index: Mean distance (1 – spearman's rho) between each time-point's volume and the median volume using AFNI's 3dTqual command - *lower is better*
- Mean Fractional Displacement (FD) - Jenkinson: Measure of subject head motion, which compares motion between current and previous volumes. Sum the absolute value of displacement changes in x, y, z directions and rotational changes about those 3 axes. Rotational changes given distance values based on changes across the surface of a 80mm radius sphere - *lower is better*
- Percent of volumes with FD greater than 0.2mm: *lower is better*

Results

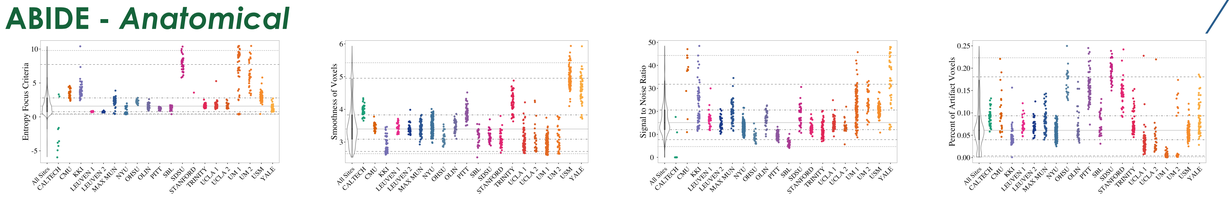
Collinearity of Measures



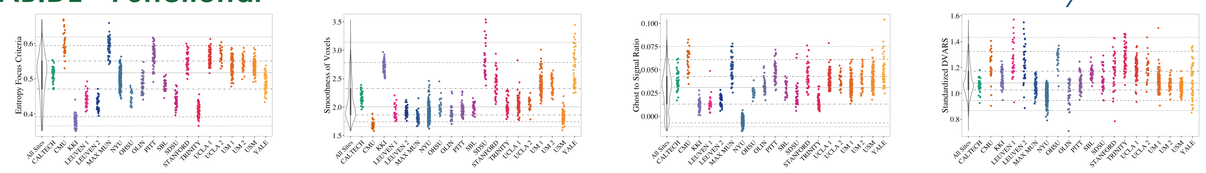
x = non-significant

Normative Distributions of Select Measures

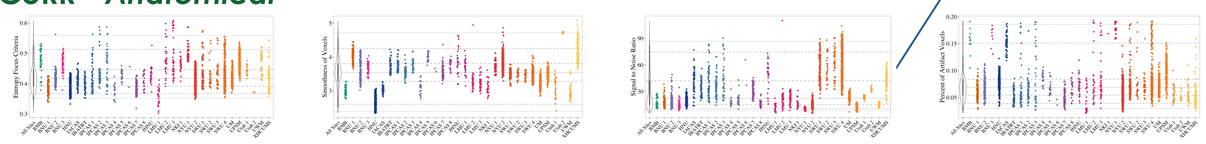
ABIDE - *Anatomical*



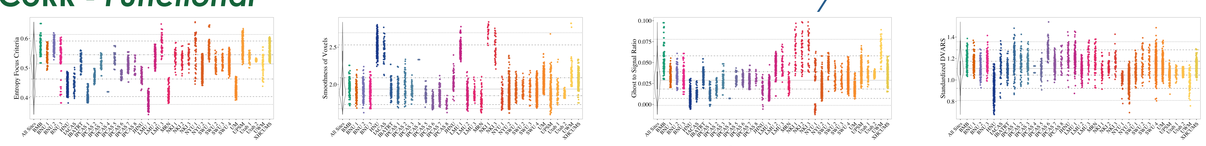
ABIDE - *Functional*



CoRR - *Anatomical*



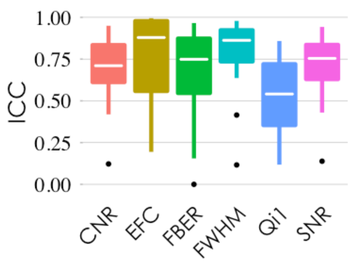
CoRR - *Functional*



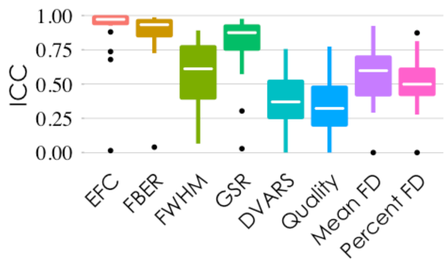
EFC FWHM SNR Qi1

Reliability of Each Measure

CoRR - *Anatomical*



CoRR - *Functional*

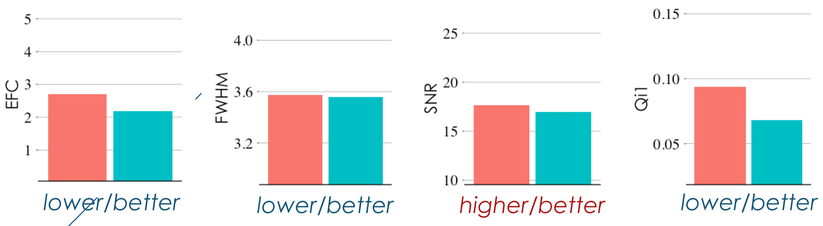


Majority of sites had high ICC (>0.5) across most measures

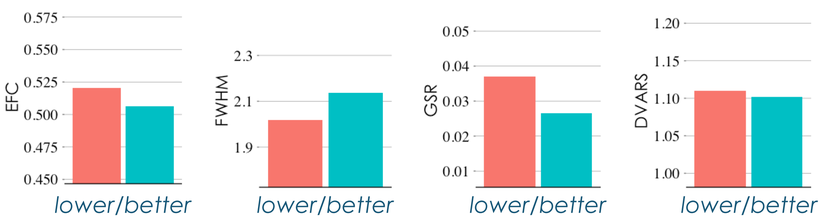
Lower ICCs were found for measures related to motion

Validity of Each Measure

ABIDE - *Anatomical*



ABIDE - *Functional*



Manual QA Fail Pass
Consensus among 4 raters

Qi1 and SNR were significant predictors of the manual anatomical QA ratings

EFC, FWHM, Percent FD, and GSR were significant predictors of the manual functional QA ratings

Discussion

- Assembled a diverse set of QA metrics in the *qap* python toolbox[1]
- Built normative distributions for each metric using the ABIDE and CoRR datasets; These distributions are shared online[1,2]
- Measures with high reliability tended to reflect scanner-specific noise; Measures with low reliability tended to reflect subject-specific noise
- Spatial QA metrics were most predictive of labels obtained from visual inspection

References

1. http://preprocessed-connectomes-project.github.io/quality-assessment-protocol
2. http://github.com/czarrar/qap_poster