HHS Public Access

Author manuscript

Proc IEEE Int Symp Biomed Imaging. Author manuscript; available in PMC 2019 November 18.

Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2019 April ; 2019: 1581–1584. doi:10.1109/ISBI.2019.8759526.

Sensitivity of derived clinical biomarkers to rs-fMRI preprocessing software versions

Kevin P. Nguyen[a], **Cherise Chin Fatt**[b], **Cooper Mellema**[a], **Madhukar H. Trivedi**[b], **Albert Montillo**[a]

[a]Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center, Dallas, Texas, USA

[b]Center for Depression Research and Clinical Care, Department of Psychiatry, UT Southwestern Medical Center, Dallas, Texas, USA

Abstract

When common software packages (CONN and SPM) are used to process fMRI, results such as functional connectivity measures can substantially differ depending on the versions of the packages used and the tools used to convert image formats such as DICOM to NIFTI. The significance of these differences are illustrated within the context of a realistic research application: finding moderators of antidepressant response from a large psychiatric study of 288 major depressive disorder (MDD) patients. Significant differences in functional connectivity measurements and discrepancies in derived moderators were found between nearly all software configurations. These results should encourage researchers to be vigilant of software versions during fMRI preprocessing, to maintain consistency throughout each project, and to carefully report versions to facilitate reproducibility.

Keywords

fMRI; SPM; preprocessing; connectivity; biomarkers

1. INTRODUCTION

Functional magnetic resonance imaging (fMRI) has become a crucial tool for the quantitative study of neurological and psychiatric diseases. However, obtaining interpretable, replicable, and noise-free data from raw images requires extensive preprocessing. It has been shown that selection of parameters for preprocessing steps such as detrending and bandpass filtering can significantly affect results derived from fMRI data, such as functional connectomes and discriminability between healthy and diseased subjects [1, 2].

This study examines the effect of more subtle and often overlooked preprocessing decisions: the version of MRI preprocessing software packages and the method for converting raw DICOM to NIFTI format, which one might assume is already standardized. The sensitivity of derived clinical biomarkers to such processing alterations will be examined and put into practical context in a psychiatric data analysis task.

The first preprocessing tool is CONN, a MATLAB-based toolbox that performs registration, artifact detection, and denoising of fMRI images and then extracts measures of functional

connectivity (FC) between brain regions [3]. At the time of writing, it is the sixth most downloaded software and the top downloaded fMRI preprocessing tool on the NeuroImaging Tools & Resources Collaboratory (NITRC) website, which currently lists about 50% more downloads for version 2017f than for the more recent version 2018a. The second tool to be investigated is the widely used Statistical Parametric Mapping (SPM) toolbox [4]. SPM can be used as a standalone tool but is also employed by the CONN backend for various initial preprocessing steps, such as coregistration and realignment. SPM12, released October 2014, and SPM8, released April 2009, are the two commonly used versions at time of writing. The latest upgrade from SPM8 to SPM12 includes several algorithmic changes, e.g. updated coregistration and spatial normalization methods, which could very well affect clinical conclusions drawn from preprocessed data.

MRI scanners typically output data in DICOM format, which is not fully standardized among scanner manufacturers. Thus, conversion of DICOM files to the more universal NIFTI format is required before preprocessing in software such as CONN and SPM. However, the discrepancies in the DICOM format among manufacturers are handled differently by various DICOM conversion tools and can ultimately affect preprocessing steps such as registration [5]. Consequently, two methods for performing this conversion will be compared here.

To demonstrate how these preprocessing decisions can affect the neurophysiological interpretation and clinical application of fMRI results, this paper will compare preprocessing configurations in the context of data from a large, multi-site clinical study of antidepressant treatment. The data preprocessed with each configuration will be used to identify moderators of antidepressant response in major depressive disorder (MDD) subjects, and the sensitivity of the derived clinical measures to the choice of software versions and DICOM conversion method will be evaluated.

2. METHODS

Data acquisition

Resting state fMRI (rs-fMRI) data were obtained as part of the Establishing Moderators and Biosignatures of Antidepressant Response in Clinical Care (EM-BARC) study, a double-blind trial in which MDD subjects received either the antidepressant sertraline or placebo; the subset of 288 subjects who underwent rs-fMRI scanning was selected for this work [6]. Imaging was performed at 4 sites on 3-Tesla scanners, with one site using a GE scanner, one using a Siemens scanner, and two using Philips scanners. T1-weighted structural images were obtained, followed by resting-state blood oxygen level dependent images via gradient-echo echo-planar imaging: TR 2000 ms, TE 28 ms, 64 × 64 × 39 voxels, 180 frames, and 3.2 × 3.2 × 3.1 mm slice thickness. Subjects were evaluated with the Hamilton Rating Scale for Depression throughout the 8 week treatment period to measure MDD severity [7].

Preprocessing

The various preprocessing configurations compared are as follows (see Table 1):

DICOM conversion: rs-fMRI and T1-weighted DICOM files were first converted to NIfTI format with either the built-in conversion tool in SPM12 (used by default when loading DICOM files into CONN) or the dcm2niix tool [4, 5].

CONN Preprocessing: The converted NIfTI files were then preprocessed with either CONN 2017f or 2018a in MATLAB 2017a. The default preprocessing pipeline in CONN was used, which includes the following steps run in the SPM12 (build 7219) backend:

For T1-weighted images: centering → segmentation → normalization to MNI space

For BOLD images: unwarping and realignment → slicetiming correction, normalization to MNI space → smoothing with 8 mm Gaussian kernel

CONN 2017f was also tested with SPM8 (build 6313) as the backend. Afterwards, mean BOLD signal timeseries were extracted for 100 regions-of-interest (ROIs) from the Schaefer cortical parcellation [8] plus 21 subcortical ROIs including the amygdala, hippocampus, thalamus, and striatum. The timeseries were then denoised using the CompCor method with 5 white matter and 5 CSF components and 6-parameter head motion regression [9]. Frames with significant motion artifact were scrubbed with the ART method and the resultant timeseries were bandpass filtered at 0.008 to 0.09 Hz [10]. Finally, a functional connectivity (FC) matrix was constructed for each subject by computing the Fisher transformed Pearson correlation coefficient between the timeseries of each pair of ROIs.

Moderator analysis

The FC matrices were used to search for moderators of antidepressant response via linear mixed effects modeling, similar to an approach previously used in the EMBARC study [Chin Fatt et al. 2018, in review]. For each of the 7260 ROI-ROI connectivity values in the upper triangle of the FC matrix, a linear mixed effects model was fitted to regress the HAMD score. Independent variables included the ROI-ROI connectivity value, time (week in the study), treatment group, study site, and the interaction terms between these variables. The modeling was performed in the *statsmodels* Python package and cross-validated using the *nlme* package in R. After fitting each model, a 1-way ANOVA was performed to test the functional connectivity measure × time × treatment group interaction term. Finally, FDR correction was performed on all ANOVA p-values, and ROI-ROI connections with corrected p-value < 0.05 were selected as moderators of antidepressant response.

3. RESULTS

FC matrix comparison

Pairwise comparisons of the 6 preprocessing pipeline configurations were performed. For each pair of configurations, FC matrices were subtracted element-wise and the mean scalar difference, $\bar{d}$, for each subject was computed. Figure 1 illustrates the distribution of $\bar{d}$ across subjects for each pairwise comparison. A paired two-tailed t-test was used to test the hypothesis that the grand mean $\mu_{\bar{d}} \neq 0$. After FDR correction, several configuration comparisons were significantly different at α =0.05:

- DICOM conversion method: While there were nonzero differences between dcm2niix and the built-in SPM converter, the differences were not statistically significant.
- SPM version: There were significant differences in FC matrix values when comparing SPM8 configurations with SPM12 configurations.
- CONN version: There were significant differences in FC matrix values when comparing CONN 2017f configurations with CONN 2018a configurations. Specifically, FC matrix values were lower with CONN 2018a than with 2017f ($\bar{d} < 0$).

Comparison of spatial normalization in SPM8 vs. SPM12

FC matrix values involving ROIs in the occipital lobe demonstrated the greatest discrepancy between SPM8 and SPM12. The penultimate and final BOLD image preprocessing steps (slice timing correction and spatial normalization to MNI space, respectively) by the SPM backend in CONN are compared in Figure 2 for configurations CONN17 + SPM8 + D2N and CONN17 + SPM12 + D2N. While the outputs of the two SPM versions are comparable for the slice timing correction step, they diverge after spatial normalization; there are qualitative differences in the occipital and prefrontal regions and in the perimeter near the cortical surface.

Moderator analyses

As shown in Table 2, the number of FC moderators of antidepressant response obtained from linear mixed models depended upon the preprocessing configuration. The number of total moderators found varied between 128 and 236 and the proportion of moderators in common between any pair of configurations was as low as 12% (CONN17 + SPM8 + SPM and CONN17 + SPM12 + SPM). The connectograms shown in Figure 3 illustrate the between-network FC moderators found with each preprocessing configuration (within-network moderators were omitted for clarity). Notably, many more somatomotor network-related moderators were found with CONN17 + SPM8 + D2N than with any of the other configurations.

4. DISCUSSION AND CONCLUSION

Significant differences in generated FC matrices were found when using SPM8 vs. SPM12 and CONN 2017f vs. 2018a. The effect of the spatial normalization algorithm update from SPM8 to SPM12 was apparent upon inspection of the BOLD preprocessing outputs (Figure 2) and was likely a major contributor to the difference in FC matrices. Though the change from CONN 2017f to 2018a may be less drastic, the FC matrices still demonstrated a discrepancy; specifically, values tended to be smaller in magnitude in 2018a. Tests showed that the newer version performed stronger denoising, leading to smaller amplitudes in the ROI timeseries, which would alter the connectivity values compared to 2017f.

While the effect of DICOM conversion method on FC matrix values was non-zero but non-significant, manual inspection revealed a difference in intensities of converted Philips

images. This was attributed to a linear scaling factor embedded in Philips DICOM metadata, which is used by the SPM converter but not used by default by dcm2niix. While this scaling would not be expected to affect correlation measures that depend on relative intensities, conversion to integer values in the NIFTI outputs may cause precision loss that could account for the non-zero difference in FC matrices.

Future work will involve 1) evaluating the neurobiological plausibility of the different moderator sets with psychiatry collaborators and 2) employing multivariate modeling (e.g. machine learning) to compare the performance of each moderator set in predicting treatment response, which will provide evidence for the optimal preprocessing configuration. In the meantime, the results presented here should encourage researchers to be vigilant of differences in software versions when preprocessing fMRI data and to maintain consistency throughout each project. Finally, to ensure reproducibility of fMRI-based results, authors should detail in their manuscripts both the versions and configurations of processing tools used.

5. REFERENCES

[1]. Borchardt Viola, Lord Anton Richard, Meng Li, van der Meer Johan, Heinze Hans Jochen, et al., “Preprocessing strategy influences graph-based exploration of altered functional networks in major depression,” Human Brain Mapping, 2016.

[2]. Shirer William R., Jiang Heidi, Price Collin M., Ng Bernard, and Greicius Michael D., “Optimization of rsfMRI Pre-processing for Enhanced Signal-Noise Separation, Test-Retest Reliability, and Group Discrimination,” NeuroImage, 2015.

[3]. Whitfield-Gabrieli Susan and Nieto-Castanon Alfonso, “Conn : A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks,” Brain Connectivity, vol. 2, no. 3, pp. 125–141, 2012. [PubMed: 22642651]

[4]. Friston K. J. (Karl J.), John Ashburner, Stefan Kiebel, Thomas Nichols, and Penny William D., Statistical parametric mapping : the analysis of funtional brain images, Elsevier/Academic Press, 2007.

[5]. Li Xiangrui, Morgan Paul S., Ashburner John, Smith Jolinda, and Rorden Christopher, “The first step for neuroimaging data analysis: DICOM to NIfTI conversion,” J. Neuroscience Methods, vol. 264, pp. 47–56, 2016.

[6]. Trivedi Madhukar H., McGrath Patrick J., Maurizio Fava, Parsey Ramin V., Kurian Benji T., et al., “Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design,” J. Psychiatric Research, vol. 78, pp. 11–23, 2016.

[7]. Hamilton M, “A rating scale for depression.,” J. Neurology, Neurosurgery, and Psychiatry, vol. 23, no. 1, pp. 56–62, 1960.

[8]. Schaefer Alexander, Kong Ru, Gordon Evan M., Laumann Timothy O., Zuo Xi-Nian, et al., “Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI,” Cerebral Cortex, 2017.

[9]. Behzadi Yashar, Restom Khaled, Liau Joy, and Liu Thomas T, “A component based noise correction method (CompCor) for BOLD and perfusion based fMRI.,” NeuroImage, vol. 37, no. 1, pp. 90–101, 2007. [PubMed: 17560126]

[10]. Mazaika Paul K, Susan Whitfield, and Cooper Jeffrey C, “Detection and Repair of Transient Artifacts in fMRI Data,” NeuroImage, vol. 26, pp. S36, 2005.

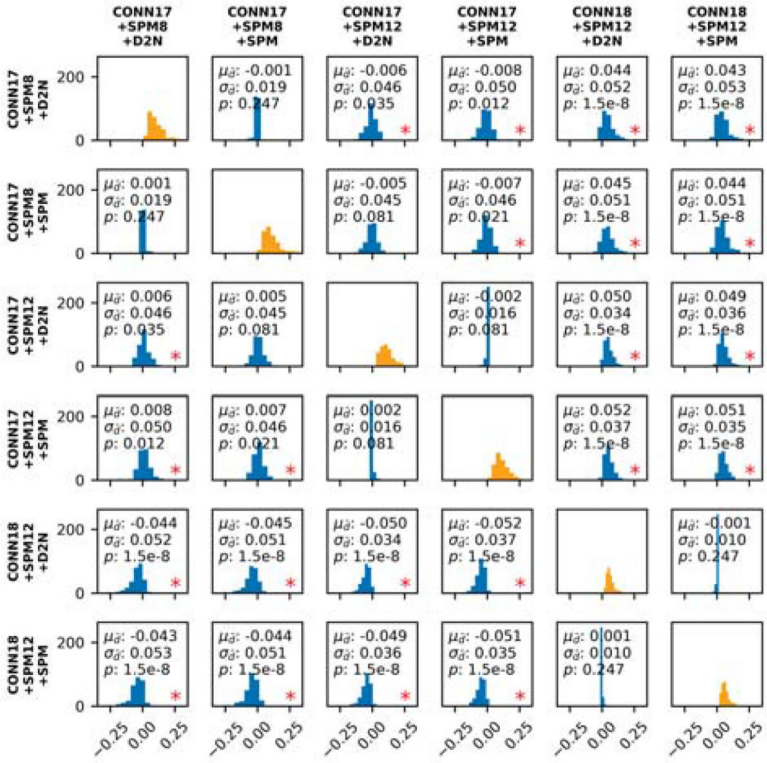

Fig. 1.

Histograms on the diagonal (in orange) display the distribution of mean FC matrix values of each of the 288 subjects generated with each tested preprocessing configuration. Non-diagonal histograms (in blue) display the distribution of the mean difference in FC matrix values, $\bar{d}$, generated using each pair of configurations (row config. vs. column config.). Text annotations list the grand mean and standard deviation of of $\bar{d}$ over all subjects, as well as the FDR-corrected p for the two-tailed t-test of $\mu_{\bar{d}} \neq 0$. Statistically significant differences are annotated with an asterisk ($p < 0.05$).

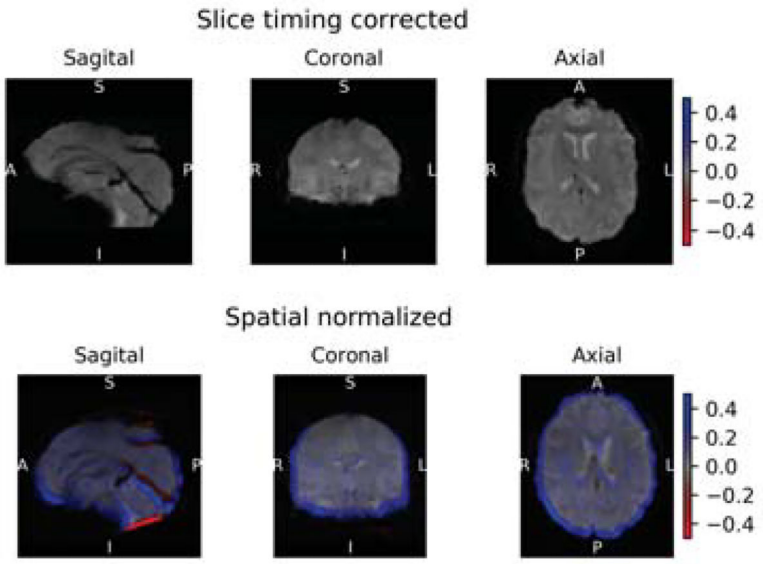

Fig. 2.

Orthogonal slice views of the difference between CONN17 + SPM8 + D2N and CONN17 + SPM12 + D2N (SPM12) preprocessed intermediate images after the penultimate (slice-timing correction, *top*) and final (spatial normalizaiton, *bottom*) steps of SPM preprocessing, for the subject with the greatest difference between FC matrices from these configurations. Blue areas have higher intensity in the SPM8 image while red areas have higher intensity in the SPM12 image.

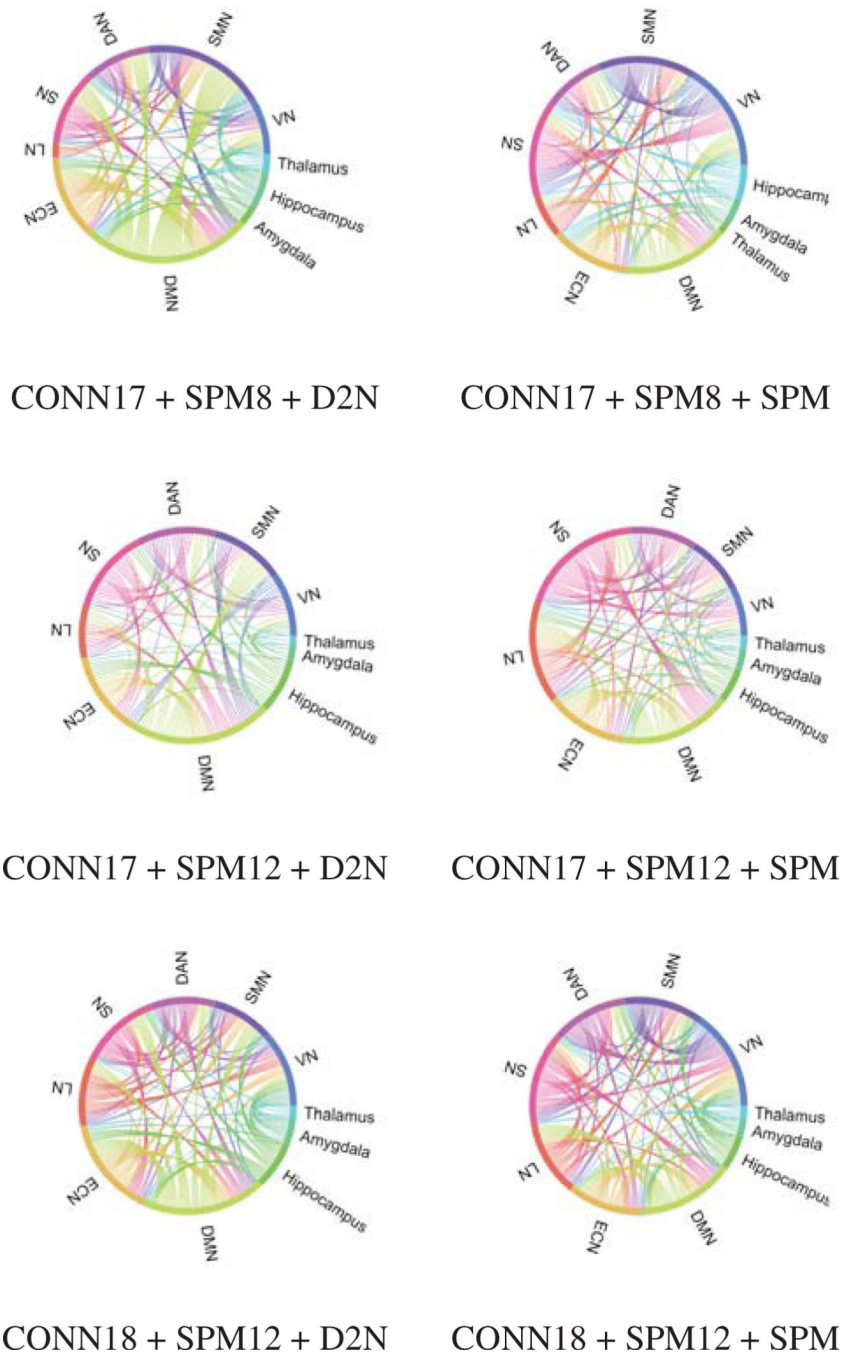

Fig. 3.

Between-network antidepressant moderators found from FC matrices generated by each configuration. VN: visual network; SMN: somatomotor network; DAN: dorsal attention network; SN: salience network; LN: limbic network; ECN: executive control network; DMN: default mode network.

Table 1.

Preprocessing configurations tested

Configuration name	CONN version	SPM version	DICOM conversion
CONN17+SPM8+D2N	17f	8	dcm2niix
CONN17+SPM8+SPM	17f	8	SPM converter
CONN17+SPM12+D2N	17f	12	dcm2niix
CONN17+SPM12+SPM	17f	12	SPM converter
CONN18+SPM12+D2N	18a	12	dcm2niix
CONN18+SPM12+SPM	18a	12	SPM converter

Table 2.

Moderators found via linear mixed effects modeling, using the FC matrices generated by each configuration, and number of moderators found in common between configuration.

		Moderators in common					
Config. name	**Total Mods.**	**CONN17+SPM8+D2N**	**CONN17+SPM8+SPM**	**CONN17+SPM12+D2N**	**CONN17+SPM12+SPM**	**CONN18+SPM12+D2N**	**CONN18+SPM12+SPM**
CONN17+SPM8+D2N	215	-	55	37	25	53	34
CONN17+SPM8+SPM	158		-	20	28	27	24
CONN17+SPM12+D2N	128			-	58	93	66
CONN17+SPM12+SPM	140				-	64	98
CONN18+SPM12+D2N	236					-	116
CONN18+SPM12+SPM	228						-