

Denoising scanner effects from multimodal MRI data using linked independent component analysis

Huanjie Li^{a,b,c}, Stephen M. Smith^d, Staci Gruber^{b,c}, Scott E. Lukas^{b,c}, Marisa M. Silveri^{b,c}, Kevin P. Hill^e, William D.S. Killgore^f, Lisa D. Nickerson^{b,c,*}

^a School of Biomedical Engineering, Dalian University of Technology, Dalian, China

^b McLean Imaging Center, McLean Hospital, Belmont, MA, United States

^c Department of Psychiatry, Harvard Medical School, Boston, MA, United States

^d FMRIB (Oxford University Centre for Functional MRI of the Brain), Department Clinical Neurology, University of Oxford, UK

^e Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

^f Department of Psychiatry, University of Arizona, Tucson, AZ, United States

ARTICLE INFO

Keywords:

Linked independent component analysis

Data fusion

Multimodal

Multivariate regression

ABSTRACT

Pooling magnetic resonance imaging (MRI) data across research studies, or utilizing shared data from imaging repositories, presents exceptional opportunities to advance and enhance reproducibility of neuroscience research. However, scanner confounds hinder pooling data collected on different scanners or across software and hardware upgrades on the same scanner, even when all acquisition protocols are harmonized. These confounds reduce power and can lead to spurious findings. Unfortunately, methods to address this problem are scant. In this study, we propose a novel denoising approach that implements a data-driven linked independent component analysis (LICA) to identify scanner-related effects for removal from multimodal MRI to denoise scanner effects. We utilized multi-study data to test our proposed method that were collected on a single 3T scanner, pre- and post-software and major hardware upgrades and using different acquisition parameters. Our proposed denoising method shows a greater reduction of scanner-related variance compared with standard GLM confound regression or ICA-based single-modality denoising. Although we did not test it here, for combining data across different scanners, LICA should prove even better at identifying scanner effects as between-scanner variability is generally much larger than within-scanner variability. Our method has great promise for denoising scanner effects in multi-study and in large-scale multi-site studies that may be confounded by scanner differences.

1. Introduction

Neuroimaging studies aimed at understanding healthy brain structure and function and diseases of the brain have proliferated with the ubiquity of magnetic resonance imaging (MRI) as a tool for neuroscience research. While the vast majority of neuroimaging studies have been conducted within a single research study, it is now being recognized that pooling multi-study data, either through multiple single-site studies, large-scale multi-site trials or through data sharing initiatives, has great potential for obviating some of the shortcomings of single studies. For example, increasing sample size by pooling multi-study MRI data can greatly enhance the statistical power to detect subtle effects or to study rare conditions and can enable the analysis of subgroups within a cohort (Button et al., 2013; Varoquaux, 2018). Pooling data also offers increased

reliability and confidence about the size of an effect by averaging out unforeseen biases of individual studies and taking advantage of the wider variety of patient types and disease etiologies that are represented in multi-site studies (Van Horn et al., 2009). For these reasons, data sharing for pooling MRI data is gaining importance in the field of neuroimaging (Eickhoff et al., 2016).

While combining multiple datasets across individual studies or sites to create large datasets is a powerful approach for increasing sample size, combining MRI datasets presents formidable challenges. Differences in scanners, hardware and software upgrades, and differences in data acquisition parameters, may introduce systematic scanner variability that can confound true effects of interest and complicate the interpretation of the results (Jovicich et al., 2009). This is not only a problem for combining datasets from different sites or studies, but can also be

* Corresponding author. McLean Imaging Center, McLean Hospital, Department of Psychiatry, Harvard Medical School, Belmont, MA, 02478, United States
E-mail address: lisa_nickerson@hms.harvard.edu (L.D. Nickerson).

<https://doi.org/10.1016/j.neuroimage.2019.116388>

Received 21 September 2019; Received in revised form 14 November 2019; Accepted 20 November 2019

Available online 23 November 2019

1053-8119/© 2019 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

problematic for long-term longitudinal studies using a single scanner. Scanner hardware or software upgrades during the study that are specifically intended to improve the performance of the scanner can result in differences in signal to noise over a prolonged study period, an effect that is compounded in multi-site longitudinal studies (Takao et al., 2011; Venkatraman et al., 2015). Scanner effects have been shown to impact all MRI modalities, including high spatial resolution structural imaging outcomes (Focke et al., 2011; Isca et al., 2015; Keihaninejad et al., 2010; Littmann et al., 2006; Takao et al., 2011), diffusion tensor imaging (DTI) measurements (i.e., fractional anisotropy (FA) and mean diffusivity (MD)) (Bartsch et al., 2014; Huisman et al., 2006; Kochunov et al., 2015; Pagani et al., 2010; Vollmar et al., 2010; Pohl et al., 2016; Venkatraman et al., 2015) and fMRI activation during tasks and functional connectivity during rest (Casey et al., 1998; Zivadinov and Cox, 2008; Costafreda et al., 2007; Friedman et al., 2008; Wegner et al., 2008; Zou et al., 2005).

Approaches to minimize the effects of scanner variability when combining MRI data have been proposed for each MRI technique: structural (Chen et al., 2014; Fennema-Notestine et al., 2007; Keihaninejad et al., 2010; Pardoe et al., 2016; Takao et al., 2011); diffusion (Pohl et al., 2016; Venkatraman et al., 2015; Fortin et al., 2017); and fMRI (Feis et al., 2015). Two statistical methods for *post hoc* correction have emerged as the most common approaches to combining MRI datasets across studies/sites. The first, and perhaps most common approach, is the inclusion of site/study covariates in the higher-level general linear model (GLM) implemented to assess group differences (Glover et al., 2012; Chen et al., 2014; Fennema-Notestine et al., 2007; Venkatraman et al., 2015). A second, less common approach for denoising is modality-specific independent component analysis (ICA). ICA is a data-driven technique that attempts to decompose a multivariate signal into independent non-Gaussian signals (Bell and Sejnowski, 1995) and identify signals of interest versus processes that are related to motion and other artifacts. Chen et al. (2014) applied ICA to multi-study structural MRI data and investigated the associations of the resulting spatial patterns with various scanning parameters to assess the influence of site on the data. Components that were related to scanner effects were then eliminated from the original data to “denoise” the data. They also compared the ICA-based denoising approach with the standard approach of including site as a covariate in the GLM and found that both the GLM approach and the ICA-based denoising method mitigated scanner effects. However, the ICA-based approach was better able to handle collinear effects.

The over-arching strategy of these two methods is to first specify an accurate “model” of the scanner effects and then to remove or statistically control for those effects, typically through regression. In the case of the GLM approach, the “model” is a simple regressor of values corresponding to site/scanner/study in the group-level GLM to statistically control for potential differences *post hoc*. In the case of ICA-based denoising, the “model” is the spatial pattern of a noise-related independent components for an individual modality and, similarly, regression is applied to remove it from the data.

In the present study, we propose a novel multimodal MRI denoising method based on data fusion of multiple MRI modalities to identify multimodal spatial patterns corresponding to scanner effects in a purely data-driven fashion that can be used to denoise the scanner effects from each individual modality. First, linked ICA (LICA) (Groves et al., 2011, 2012) is applied to multimodal MRI data, for example, morphometric maps, diffusion outcomes, and brain activation or connectivity maps all together, to decompose the data into a set of independent components. LICA components are multi-modal spatial maps and corresponding subject loadings that reflect the sources of variability (signal and noise) in the data. Once the components corresponding to scanner effects have been identified, they are removed from the original measurements using a regression-based method to provide a “clean” set of measurements for each modality that can be used for further statistical analyses.

LICA offers several unique advantages compared with other

methods for identifying scanner effects for denoising: 1) LICA more efficiently models the common variance among the multiple measurements (modalities) in each subject. This improves its ability to identify between-subject sources of variability in the data and to separate out interesting between-subject effects from uninteresting effects (such as those related to scanner variability). 2) LICA can identify patterns that are multi-modal or that are sparse in modalities. That is, some scanner effects may impact several modalities, for example, all DTI outcomes, while others may affect only a single modality (fMRI). 3) LICA explicitly takes into account the spatial correlation of each modality to ensure that information is correctly balanced across modalities. LICA applies a correction for this spatial smoothness that reflects the effective degrees of freedom instead of the actual number of voxels and uses adaptive modality-weighting to address the different scalings of the signal in each modality. The implication is that modalities can have different numbers of voxels and different degrees of smoothness, which may not be the case with other data fusion approaches. 4) LICA specifically identifies patterns that reflect *variability* in the data. Thus, the *more* scanner variability, the easier it should be to detect using LICA. Groves et al. (2012) showed that LICA could effectively identify scanner effects due to software upgrades, which may represent a relatively small source of scanner variability as compared with hardware upgrades, acquisition differences, or variability due to different scanners (e.g., from multi-site data). Thus, we expect an approach based on LICA to identify our “model” of the scanner-related noise to easily detect multi-site scanner variability, while also retaining sensitivity to within-scanner variability (e.g., due to software and hardware upgrades).

We conducted an analysis to demonstrate this new approach using multimodal MRI data collected across six different studies, all using the same scanner, but with a major software and hardware upgrade during the data collection periods and using different acquisition parameters across studies. The datasets we selected were chosen as part of a larger study to investigate structural and functional MRI outcomes in marijuana users using multi-study data to increase the sample size (neurobiological findings will be presented elsewhere). The method presented here is the approach we developed to address the scanner upgrades that occurred during those studies. Given that all of our data were collected on the same scanner, our dataset is a more conservative test of LICA for identifying scanner effects than using multi-site data. Notably, scanner-related components have been reported previously using LICA of single-scanner data with software upgrades only (Groves et al., 2012), thus LICA is sensitive to these subtler scanner effects.

We applied LICA to the data to derive multimodal components and a set of subject loadings for each component. The subject loadings for all participants were used to identify scanner/study-related noise components by testing the associations of the loadings for each component with scanner/study variables. Once the noise components were identified, either the spatial patterns or the loadings for these noise components were used for denoising. We tested three regression methods for denoising scanner effects via spatial patterns or loadings (Griffanti et al., 2014). These denoising methods were applied to the original data (fed into the LICA) to derive denoised data that can be used for further statistical analyses. The performance of our proposed method and two other common methods for addressing scanner variability, a higher-level GLM with a site/study covariate included in the group-level model and modality-specific ICA-based denoising, were then tested and compared using datasets that were artificially constructed from our existing data for all the healthy participants, but with group (dataset) differences in scanner/sequence parameters. While we demonstrate a proof of concept of our method using DTI, structural, and fMRI modalities, the LICA is versatile and can be applied to any combination of these modalities (e.g., DTI alone, DTI + structural, structural + fMRI, etc.).

2. Methods

2.1. Study data

Data from 185 subjects collected from six different studies were used for the present work. Initial quality assurance of the structural images was done to identify head motion, artifacts, and other problems with the data. Data from 29 participants were removed from the multi-modal data, including participants who did not have a structural MRI (1 subject), head motion (>1 voxel size, 4 subjects), or visible intensity effects, such as cerebellum pathology (7 subjects), missing cortex (from positioning, 2 subjects), extreme nonuniform image intensity, radiofrequency noise (2 subjects), severe ghosting and ringing (13 subjects). Thus, MRI data from 156 subjects were used for the present study. Data included high-resolution structural images, diffusion data, and fMRI data, although not all data were collected in each participant, as each study collected data from different modalities. This resulted in “missing data” for some participants for some modalities when considering all modalities together. Data from 73 chronic heavy marijuana (MJ) users (near daily use and positive THC screen on test day) (age = 24.6 ± 6.8 , 62 male and 11 female) and 83 non-MJ using healthy controls (HC) (age = 24.2 ± 5.2 , 50 male and 33 female) were included in our analyses.

All data were collected using the same Siemens 3T Trio, but with 3 different scanner software versions (SSWV), and over the course of a major hardware upgrade. Versions VA23A and VA25A were used prior to a major hardware and software upgrade of the Trio (TIM upgrade), while VB17A was used post-TIM. Acquisition sequences also differed across the studies (Table 1), thus, these data contain some similar sources of scanner variability as multi-site/multi-study data, including scanner hardware differences, scanner software versions, and data acquisition differences.

2.2. MRI acquisition parameters

High-resolution structural MRI data for all six studies were collected using a T1-weighted MPRAGE with 128 slices and flip angle (FA) = 12° . Studies 1 and 5 utilized echo time (TE)/repetition time (TR) = 2.74 ms/2100 ms, $1.5 \times 1.0 \times 1.3$ mm³ voxel size. Studies 2, 3 and 4 utilized TE/TR = 2.15 ms/2000 ms, $1.5 \times 1.0 \times 1.3$ mm³ voxel size. Study 6 utilized TE/TR = 2.25 ms/2100 ms, $1.0 \times 1.0 \times 1.0$ mm³ voxel size.

Diffusion tensor imaging data for studies 2 and 3 were collected using an echo-planar imaging (EPI) sequence with scan acquisition parameters: TE/TR = 89 ms/9300 ms, 60 slices, $2.0 \times 2.0 \times 2.0$ mm³, 48 directions, b-value = 700 s/mm² (three b = 0 images without diffusion weighting were acquired).

Task fMRI data for studies 2, 3, 4 and 6 were collected with gradient-echo EPI. Acquisition imaging sequences for studies 2, 3 and 4 were: TE/

TR/FA = 30 ms/3000 ms/90°, 40 coronal slices, $3.1 \times 3.1 \times 5.0$ (no gap) mm³ voxel size, interleaved acquisition. For study 6, TE/TR/FA = 30 ms/2000 ms/90°, 34 coronal slices, $3.5 \times 3.5 \times 3.5$ (no gap) mm³ voxel size, with interleaved acquisition. fMRI data were collected while participants performed a multi-source interference task (MSIT) of inhibitory processing during the fMRI scan. The MSIT task is described in more detail in Gruber et al. (2012). Scanner and imaging related details are summarized in Table 1.

2.3. Data processing

All data processing was done using FMRIB Software Library, FSL (Smith et al., 2004 <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>), FreeSurfer (<https://surfer.nmr.mgh.harvard.edu>) and Matlab (Mathworks, Inc., Natick, MA). Optimized modality-specific preprocessing pipelines, including quality assurance to identify data with excessive motion or other artifacts, were used to produce standard-space outcome images for each subject for a given modality, including: 1) modulated gray matter (GM) images generated by FSL-VBM (voxel-based morphometry; Douaud et al., 2007), 2) vertex-wise cortical thickness (CT) and pial surface area (PSA) maps estimated using FreeSurfer by means of automated surface reconstruction scheme (Dale et al., 1999; Fischl and Dale, 2000; Fischl et al., 1999a, 1999b, 2001, 2002, 2004), 3) FA, MD and tensor mode (MO) images calculated using FSL FDT (FMRIB's Diffusion Toolbox; Smith et al., 2006), and 4) MSIT brain activation maps estimated by FSL FEAT (FMRI Expert Analysis Tool; <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FEAT>) analysis of the MSIT task fMRI data. For GM and fMRI data, the images were registered to $2 \times 2 \times 2$ mm³ MNI (Montreal Neurologic Institute) standard space; FA, MD and MO images were registered to $1 \times 1 \times 1$ mm³ MNI standard space. For each outcome that was derived from the structural MRI, DTI and MSIT fMRI data, a “subject series” was created by concatenating the resulting spatial maps across all participants into a single 4D (volume x subjects) data file. These outcomes are hereafter also referred to as modalities even though they are derived quantities.

2.4. Structural MRI

GM images: For all high-resolution structural MR images that passed through quality assurance (e.g., with no significant motion or artifacts observed in the structural images), non-brain tissue was removed, followed by GM-segmentation and registration to the MNI 152 standard space using non-linear registration via FNIRT (FMRIB's Nonlinear Image Registration Tool; Anderson et al., 2007a,b). The resulting images were then averaged together and flipped along the x-axis to create a left-right symmetric, study-specific gray matter template. All native gray matter images were then non-linearly registered to this study-specific template and “modulated” to correct for local expansion (or contraction) due to the non-linear component of the spatial transformation. Modulated GM images were then smoothed using an isotropic Gaussian kernel with a sigma of 4 mm.

CT and PSA images: Using the same high-resolution structural MR images that were used for GM, CT measurements were obtained by reconstructing representations of the gray/white boundary and the pial surface and then calculating the distance between the surfaces at each vertex across the cortical mantle (Dale and Sereno, 1993; Dale et al., 1999). PSA was estimated by registering each subject's reconstructed surfaces to a common template, and the relative amount of expansion or compression at each vertex was used as a proxy for regional arealization. PSA were resampled and mapped to a common coordinate system using a non-rigid high-dimensional spherical averaging method to align cortical folding patterns (Fischl et al., 1999a, 1999b, 2008). Both CT and PSA images were smoothed with a Gaussian kernel (full width of half maximum = 10 mm).

Table 1

The acquisition parameters for each modality for the six different studies.

Study	T1-Weighted Image TR/ TE Voxel Size	DTI TR/TE Voxel Size B-value, Directions	MSIT fMRI TR/TE Voxel Size
1	2100/2.74 ms $1.5 \times 1.0 \times 1.3$ mm ³	— —	— —
2	2000/2.15 ms $1.5 \times 1.0 \times 1.3$ mm ³	9300/89 ms $2.0 \times 2.0 \times 2.0$ mm ³ 700 s/mm ² , 48	3000/30 ms $3.1 \times 3.1 \times 5.0$ mm ³
3	2000/2.15 ms $1.5 \times 1.0 \times 1.3$ mm ³	9300/89 ms $2.0 \times 2.0 \times 2.0$ mm ³ 700 s/mm ² , 48	3000/30 ms $3.1 \times 3.1 \times 5.0$ mm ³
4	2000/2.15 ms $1.5 \times 1.0 \times 1.3$ mm ³	— —	3000/30 ms $3.1 \times 3.1 \times 5.0$ mm ³
5	2100/2.74 ms $1.5 \times 1.0 \times 1.3$ mm ³	— —	— —
6	2100/2.25 ms $1.0 \times 1.0 \times 1.0$ mm ³	— —	2000/30 ms $3.5 \times 3.5 \times 3.5$ mm ³

2.5. DTI data

FSL's FDT and TBSS (Tract-Based Spatial Statistics; Smith et al., 2006) tools were used to generate FA, MD, and MO maps for white matter for each subject. FSL eddy was used to correct DTI data for motion and eddy currents (Andersson et al., 2016). After removal of non-brain tissue, fitting of diffusion tensors on corrected data was applied using *dtifit* to compute FA, MD, and MO maps (Smith, 2002). Each subject's FA map was transformed into common space by warping to the FMRIB58_FA template using FMRIB's nonlinear image registration tool (FNIRT, Andersson et al., 2007a,b). The mean FA volume over all individuals was computed and then thinned to create a mean FA skeleton. The mean skeleton was thresholded at FA exceeding 0.2 to minimize partial volume effects at the boundaries between tissue classes. Individual FA maps were multiplied by the mean skeleton to compute skeletonized FA maps for each subject. MD and MO maps for each subject were processed similarly to result in a standard space skeletonized FA, MD, and MO map for each subject.

2.6. MSIT fMRI data

The fMRI data was processed to estimate subject-specific task spatial maps (TSM) corresponding to brain activation during the interference condition relative to the control condition using a first-level GLM implemented in FSL FEAT v6.00. Each fMRI dataset underwent quality assurance to identify excess subject motion and other artifacts followed by standard pre-processing, which included: motion correction, slice-timing correction, non-brain tissue removal, spatial smoothing using a Gaussian kernel of FWHM 7.0 mm, grand-mean intensity normalization of the entire dataset and high-pass temporal filtering with a cut-off of sigma = 42.0 s. Registration of fMRI data to high-resolution structural images was carried out using FLIRT (FMRIB's Linear Image Registration Tool; Jenkinson et al., 2002). Registration of high-resolution structural images to standard space was first done using FLIRT, then further refined using FNIRT nonlinear registration. Time-series statistical analysis was carried out via GLM with regressors corresponding to the inhibitory control condition convolved with a gamma hemodynamic response function using FILM with local autocorrelation correction (Woolrich et al., 2001). The Z-stat image from the first-level analysis (the contrast of parameter estimate map for interference – control) was thresholded at $Z = 2.3$ and -2.3 , resulting in the subject-specific TSM. These images were registered to standard space using the FLIRT/FNIRT transformation matrices in a single step.

2.7. Linked ICA

Subject series were constructed for each imaging outcome, or modality, by concatenating images for that modality across subjects, using the same subject order for each modality. To account for the fact that participants from different studies had different measurements, a volume of zeros was used to represent a subject's "missing" data for a given modality. This allows the features that a participant does have to contribute to the estimation of the spatial patterns and the subject loadings, which are shared across modalities. As such, retaining more data in other modalities improves the estimation, even though other modalities will have less data. Thus, seven 4D datasets, one for each of the seven modalities (FA, MD, MO, GM, CT, PSA, and TSM), were constructed with 156 subject volumes for each modality. The subject-series for all 7 modalities were analyzed simultaneously using LICA implemented in Matlab (free software available at <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLICA>). In applying LICA, participants with high noise variance (1/precision) will influence a significant fraction of the LICA components (Groves et al., 2012). These participants should be removed from the LICA to achieve robust results. This is similar in practice to conducting group independent component analysis (GICA) of fMRI data. Namely, GICA maps can reveal problems in the subject data that lead to corruption

of GICA maps – these data must be remediated or removed to obtain improved GICA maps. Initial LICA results showed that 23 subjects had high noise variance that were clearly identified as their noise variance was orders of magnitude larger than other participants. All data from these participants were removed, resulting in data from 133 subjects being included in the final LICA to identify scanner-related components (62 chronic heavy MJ users and 71 non-MJ using HC). Table 2 shows details of the 133 subject final dataset that was used, including MRI scan-types that were collected for each of the six studies, the number of subjects from each study, and the SSWV information.

LICA identifies spatial components, with each component being an aggregate of spatial patterns from each modality, along with a set of subject loadings, one for each component. Loadings for a given component are shared between all of the modalities represented in that component, and indicate the degree to which that multimodal component is present in any individual subject. The subject loadings for each component were assessed for relationships with SSWV, acquisition parameter differences (referred to as a "study" effect), and participant variables (demographic, drug use, and task performance measures) using a single multiple regression with loadings for each component as the dependent variable and participant variables as the independent variables in each regression. This is the same approach used by Chen et al. (2014) to identify ICA components associated with multi-site scanner effects in GM and utilized to examine group differences between individuals with schizophrenia and healthy controls in the loadings on ICA components (Xu et al., 2009) to identify gray matter differences.

While many components were identified that related to participant variables, only those components whose subject loadings related *only* to SSWV/study were identified as noise components (evaluated with multiple comparison corrected p-value < 0.05). Components with loadings that related to both scanner variables and any participant variable were not selected for removal. Using a stringent threshold and ensuring that loadings were not related to participant factors provides some protection against removing interesting neurobiological variability, while being somewhat conservative in the removal of scanner effects. These criteria parallel the more general use of ICA for denoising individual modalities.

Table 2

For each study, the number of HC and MJ subjects with each modality is shown (upper and lower row for each modality for each study, respectively). Also shown is the total number of HC/MJ and all subjects for each modality, with SSWV in parenthesis (VA23A/VA25A/VB17A). A '—' indicates that data for that modality and software version were not collected for that group (HC/MJ) in the study. HC: healthy control; MJ: marijuana; SSWV: scanner software version. VA23A and VA25A are pre-TIM software versions; VB17A is the post-TIM software version. Missing data indicates how many participants do not have data for that modality. Publications from the original studies provide additional details about each study and related study findings.

Study	Structural	DTI	MSIT fMRI	Publication
	HC (VA23A/VA25A/VB17A)			
	MJ (VA23A/VA25A/VB17A)			
1	7 (4/3/—)	—	—	
	8 (5/3/—)	—	—	
2	16 (—/9/7)	15 (—/8/7)	16 (—/9/7)	Gruber et al. (2012, 2014, 2015)
	21 (—/10/11)	20 (—/9/11)	15 (—/8/7)	
3	—	—	—	Gruber et al. (2012, 2014, 2015)
	16 (—/—/16)	16 (—/—/16)	14 (—/—/14)	Hill et al. (2017)
4	—	—	—	
	9 (—/—/9)	—	9 (—/—/9)	
5	13 (—/13/—)	—	—	Mashhoon et al. (2015)
	8 (—/8/—)	—	—	
6	35 (—/—/35)	—	30 (—/—/30)	Cui et al. (2015)
	—	—	—	
Total HC	71 (4/25/42)	15 (—/8/7)	46 (—/9/37)	
Total MJ	62 (5/21/36)	36 (—/9/27)	38 (—/8/30)	
Total All	133 (9/46/78)	51 (—/17/34)	84 (—/17/67)	
Missing Data	0	82	49	

For example, for fMRI, ICA-based denoising involves classifying each component as noise or signal using various heuristics, such as correlation of the component timecourse with a motion timecourse. We chose to select components that have loadings that correlate with scanner variables, but not with participant variables. Of note, the participant variables we were interested in were those related to marijuana use (for our future work). We were not concerned with other participant characteristics that could be collinear with scanner/study variables as it would actually be desirable in many cases to remove such effects (indeed this is the case for our future work with respect to sex and age), as long as the effects of interest are retained. Ultimately, more sophisticated automated approaches similar to ICA-AROMA (Pruim et al., 2015) or FSL FIX (Salimi-Khorshidi et al., 2014) for fMRI data denoising may be possible with further development of our method and availability of large multi-site data.

Similar to how we apply GICA to group fMRI data (Smith et al., 2009; Nickerson, 2018; Abou-Elseoud et al., 2010; Groves et al., 2012), we ran the LICA decomposition at several different dimensionalities (L) of 12, 14, 15, 17, 20, 25 and 30 components to “tune” the LICA to identify the dimensionality that resulted in components that were stable across most dimensionalities (e.g., patterns that appeared across dimensionalities with high spatial correlation in each modality) and that had loadings most clearly associated with SSWV/study (and not with participant variables). Resulting multimodal spatial patterns from LICA were converted to pseudo-Z-statistics by accounting for the scaling of the variables and the SNR for individual modalities. LICA spatial maps were thresholded at $Z = 2.3$ for visualization, based on assuming an explicit spherical noise model during the LICA decomposition (Groves et al., 2012).

Details on how to implement LICA and on how to do denoising as described in the next section, including code, are supplied in Supplement File titled “LICA Denoising Manual”.

2.8. Data denoising

Once the LICA components associated with scanner/study effects were identified, they were used to denoise the original data using multivariate regression. Similar to Griffanti et al. (2014), who tested both hard and soft regression for denoising single-subject fMRI data, we tested three approaches we developed for denoising the multimodal MRI data using the LICA results. All regressions presented in Equations (1)–(5) are easily implemented using FSL command line tools (fsl_glm and fsl_regfilt).

2.9. LICA-R1 denoising: hard regression using LICA subject loadings

A single multivariate regression of the subject loadings for only the scanner components against the original data was implemented to remove scanner effects from each modality as follows:

$$Y_{orig} = B_{LICA_noise} \hat{S}_{est} + E_{LICA_R1_denoised} \quad (1)$$

Y_{orig} is the 4D subject-series for a given modality reorganized into a 2D data matrix (M subjects x N voxels), B_{LICA_noise} are the matrix of subject loadings for the scanner components identified from the LICA (M subjects x P components). \hat{S}_{est} are the estimated spatial maps of regression coefficients (P components x M subjects). $E_{LICA_R1_denoised}$ are the residuals from the regression, which represent clean data using the LICA-R1 denoising method. Each modality was denoised separately.

2.10. LICA-R2 denoising: hard regression via dual regression using LICA spatial maps

A multivariate dual regression procedure (Nickerson et al., 2017) using the LICA component spatial maps was implemented to remove scanner components as follows:

$$Y'_{orig} = S'_{LICA_all} \hat{B}'_{est_all} + E \quad (2)$$

$$Y_{orig} = B_{est_noise} \hat{S}'_{est_noise} + E_{LICA_R2_denoised} \quad (3)$$

Y_{orig} is the original 4D data, reorganized into a 2D data matrix, similar to above, with the prime indicating transpose. In Eq. (2), S_{LICA_all} are the spatial maps of all the components from the LICA. \hat{B}_{est_all} are the estimated regression coefficients for all components. B_{est_noise} are the estimated regression coefficients for noise (scanner) components, which are similar but not equal to the subject loadings for the noise components that come straight from LICA. In the second regression (Eq. (3)), B_{est_noise} (e.g., from \hat{B}_{est_all}) is regressed against Y_{orig} . \hat{S}_{est_noise} are the estimated spatial maps of regression coefficients. $E_{LICA_R2_denoised}$ are the residuals from the regression, which represent clean data using the LICA-R2 denoising method. Each modality was denoised separately.

For each scanner component identified by LICA, the correlations between the subject loadings, B_{LICA_noise} from Equation (1), and B_{est_noise} from Equation (2) were calculated to investigate differences in performance between LICA-R1 and LICA-R2 denoising methods.

2.11. LICA-SR denoising: soft regression using LICA subject loadings

First, multiple regression of the full set of LICA subject loadings against the original subject-series was done to estimate the contribution of all components (i.e., signal and noise). The variance of the noise was removed by subtracting the contribution of the noise components from the original data (Griffanti et al., 2014):

$$\hat{\beta}_{all} = \text{pinv}(B_{LICA_all}) Y_{orig} \quad (4)$$

$$Y_{LICA_SR_denoised} = Y_{orig} - B_{LICA_noise} \hat{\beta}_{noise} \quad (5)$$

B_{LICA_all} are the subject loadings of all the components from the LICA. $\hat{\beta}_{all}$ are the estimated contributions of both signal and noise components. B_{LICA_noise} are the subject loadings of the noise components and $\hat{\beta}_{noise}$ are the regression coefficients. $Y_{LICA_SR_denoised}$ represents the clean data using LICA-SR. Each modality was denoised separately.

2.12. GLM and ICA-Based methods

We compared the performance of LICA-R1/R2/SR for denoising our multi-study data with two other statistical methods for addressing scanner effects: a higher-level GLM with a site/study covariate included in the group-level model, and modality-specific ICA-based denoising.

For the GLM analyses, the scanner and study (e.g., to account for different acquisition parameters across studies) variables were coded as categorical variables, similar to coding used in ANOVA factor effects models. For example, the scanner variable has three levels (i.e., VA23A, VA25A and VB17A), thus we use two coding variables to represent the scanner variable. VA23A was chosen as the reference level, thus for each coding variable, rows of the design corresponding to VA23A will have -1 . The first coding variable is -1 for VA23A, 1 for VA25A and 0 for VB17A; the second coding variable is -1 for VA23A, 0 for VA25A and 1 for VB17A.

ICA-based denoising of individual modalities was done using FSL MELODIC (Multivariate Exploratory Linear Optimized Decomposition into Independent Components; Beckmann and Smith, 2004) with automatic dimensionally estimation to estimate spatial maps and “time-series”, which are in this case subject loadings. The dimensionality of each single-modality ICA was “tuned” to achieve separation between signals (e.g., components that resemble standard structural covariance networks (Xu et al., 2009; Guo et al., 2015)). The dimensionality that resulted in scanner components most clearly identified based on correlations of subject loadings for each component with SSWV and study variables while separating out signals (components related to participant

variables) was selected as the optimal dimensionality. This approach of correlating subject loadings on ICA patterns to identify what each pattern may be capturing (signal or noise) is the same approach applied by Chen et al. (2014) to identify multi-site effects in structural MRI ICA components, and used by others to study group differences (Xu et al., 2009) in loadings on ICA components. Once the optimal dimensionality was selected, all noise components were identified and `fsl_regfilt` was used to remove the noise components from the modality series.

2.13. Assessing performance of denoising strategies using multi-study MRI data

The HC data were split into two “groups” that were defined based on pre- and post-TIM upgrade or on study variables. Group differences were then assessed pre- and post-denoising using the various strategies via two-group *t*-tests. Since the groups are artificially constructed based on SSWV/study variables from age matched HC subjects, any group differences may be attributed to scanner effects. Effective denoising should result in removal of these effects, with a subsequent reduction of group differences observed in the no-denoising scenario. Group differences will ideally be reduced to zero if all other participant characteristics are matched. GM and CT maps from HCs were separated in two groups based on SSWV (pre-vs. post-TIM). For pre-TIM, age = 24.0 ± 5.2 years, 9 females and 17 males; for post-TIM, age = 25.2 ± 5.5 years, 20 females and 22 males). Most of the fMRI data were collected post-TIM, therefore TSM maps from healthy controls were divided in two groups based on different acquisition parameters: Group 1 comprised data from studies 2, 3, and 4, which used the same acquisition parameters (age = 23.3 ± 3.6 years, 10 females and 6 males), and Group 2 comprised data from study 6 (age = 25.1 ± 4.9 years, 16 females and 14 males), which used different acquisition parameters. For DTI outcomes, there were only 15 HC subjects, thus we were not able to construct groups to assess denoising of DTI measures. While we endeavored to use all available data for these tests, we have also included results with a smaller set of data with groups that were additionally matched for sex and sample size in the Supplement Material (results were similar to those presented for the larger sample sizes). While the HC had similar demographics and limited ranges for many variables (IQ, alcohol drinking) across the set of studies included, this simple analysis did not consider these other characteristics (e.g., that were not matched upon specifically) that could contribute to between-group variance.

Group differences in each modality were assessed for statistical significance before and after denoising using two-group *t*-tests using non-parametric permutation testing with threshold-free cluster enhancement (TFCE, Smith and Nichols, 2009) in FSL's Randomise (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Randomise>), with 5000 permutations, to achieve a significance level of $p = 0.05$, corrected for family-wise error. Regions that showed statistically significant group differences were selected as regions of interest (ROI) to illustrate associations between group differences, SSWV, and study variables in these areas before and after denoising. The mean Z-value and the standard deviation of voxels within the ROI were calculated (from the outputs of the two-group *t*-test) to illustrate the effectiveness of each method for removing SSWV effects (e.g., to show that apparent differences in the maps are not simply due to thresholding effects). In this case, Z-values for non-denoised data reflect the differences due to SSWV/STUDY, and the Z-values for the denoising methods should move closer to 0 if the denoising method is working well.

While the HC data are well-suited for demonstrating the ability of LICA denoising to remediate scanner effects, it is not very well-suited for assessing the ability of LICA denoising to retain effects of interest (e.g., specificity of LICA denoising). Although marijuana effects on brain structure and function are not the focus of the present work, the data we used for testing LICA denoising was collected in HC participants and chronic heavy marijuana users as part of our funded study aimed at this goal. As such, we can use the data to demonstrate a very basic assessment of the specificity of LICA denoising. We show in the Supplement File,

“LICA Denoising Performance”, a data fusion component that was associated with marijuana variables both pre-and post LICA-R1 denoising (hence demonstrating specificity to retain effects of interest), and that we also independently observed in a LICA of multi-modal MRI data to investigate alcohol abuse disorder (in which some participants also had a diagnosis of cannabis disorder).

3. Results

From the final LICA, three noise components had loadings that were strongly associated with scanner effects (SSWV and/or the study variable), but were not associated with participant variables. While LICA dimensionality $L = 15$ identified these three noise components with loadings that were most clearly associated with SSWV/study, the spatial patterns were also observed at different dimensionalities (e.g., the LICA spatial patterns were stable).

Fig. 1 shows a component with loadings that were strongly associated with both SSWV ($p = 1.91\text{e-}08$) and study ($p = 1.88\text{e-}04$). This component showed global effects in FA and MD and region-specific effects in GM, TSM, CT, and PSA. Each modality had different contributions to this component (FA (29%), MD (61%) MO (3%), GM (1%), TSM (0%), CT (5%) and PSA (1%)), thus, the scanner effect primarily impacted FA and MD. Similarly, Groves et al. (2012) reported two LICA components heavily weighted on diffusion modalities that were associated with scanner operating system version. Fig. 2 shows a second multimodal component with loadings that had a strong linear association ($p = 2.15\text{e-}04$) with SSWV. This component showed region-specific effects in FA, MD, GM, CT and PSA, with weights for each modality given by FA (3%), MD (11%), GM (31%), TSM (2%), CT (12%) and PSA (36%), indicating a more “multimodal” scanner effect. Fig. 3 shows a third component heavily weighted to GM only, with loadings that were strongly related to study, e.g., different acquisition parameters ($p = 1.17\text{e-}06$).

Fig. 4 shows the results of the two-group *t*-test to assess differences in GM between the two HC groups, before and after denoising. Non-denoised results were obtained using the original GM data without any noise variable regression or data denoising. The first row in the figure corresponds to the group differences in GM associated with pre- and post-TIM upgrade. Group differences were observed in widespread brain regions, including insula, sub-cortical brain regions, and occipital cortex. Many of these regions are also seen in the GM spatial maps from the noise components identified using LICA, especially the map in Fig. 3 (the values in the LICA maps are reversed, however, they are in the same direction when considering the sign of the loadings). We do not know for certain what scanner effect (software/hardware or acquisition parameters) is driving the group differences in the first row of Fig. 4. However, considering that study six has the highest loadings on the pattern in Fig. 3, and that study six has a smaller voxel size for the structural MRI scan than the other studies, we can speculate that this component is related to acquisition differences. Thus, the group differences in the first row of Fig. 4 could be related to these effects. The second row of Fig. 4 shows that statistically controlling for SSWV and study using GLM regression is modestly effective at removing the effects, with some remaining differences (and the appearance of new regions with group differences).

The third row shows the results using modality-specific ICA-denoising. For GM, ICA identified 34 components (with automatic dimensionally estimation). Twenty components were found to have subject loadings that were significantly related to SSWV and study variables. Thus, these 20 components were removed from the original data using regression (`fsl_glm`). Rows 4–6 show the results using the LICA-SR, LICA-R2 and LICA-R1 denoising methods to remove the three scanner components (from LICA, shown in Figs. 1–3). Modality-specific ICA-denoising, LICA-SR and LICA-R2 denoising methods were also modestly effective at denoising, removing most of the effects that were related to scanner variables with some remaining differences still apparent. The

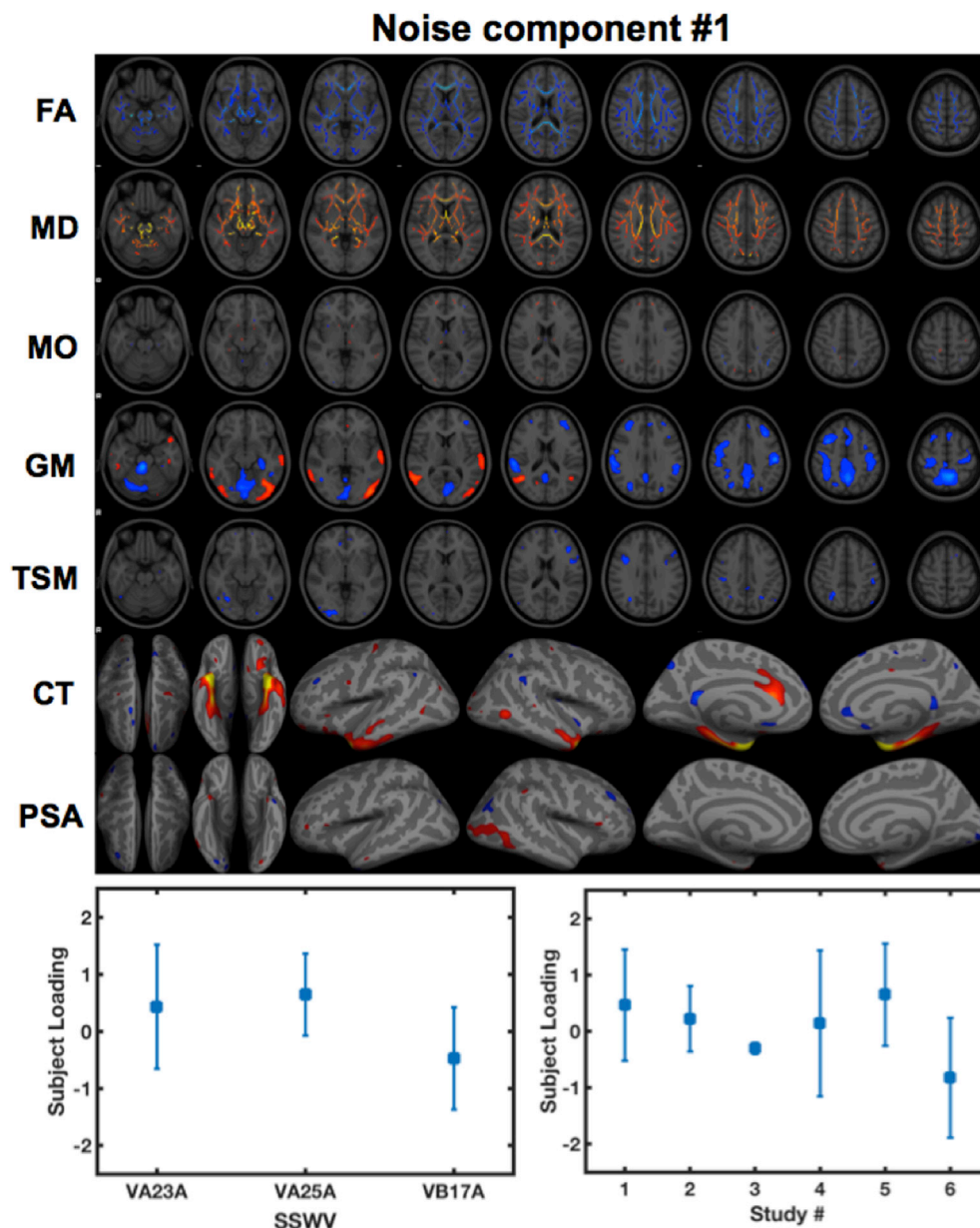


Fig. 1. A multimodal noise component identified from LICA. Subject loadings for this component were strongly associated with both SSWV ($p = 1.91 \times 10^{-8}$) and STUDY ($p = 1.88 \times 10^{-4}$) variables. The spatial pattern shows global effects in FA and MD and region-specific effects in GM, fMRI, CT and PSA. The spatial maps were thresholded at $Z = 2.3$.

hard aggressive denoising methods LICA-R2/R1 show better denoising performance than the soft regression-based denoising method LICA-SR. Strikingly, the LICA-R1 denoising method reduced all scanner effects, while not introducing false positives. Thus, LICA-R1 denoising demonstrated superior performance for denoising scanner/study effects from GM data.

We randomly chose two regions with significant group differences for the non-denoised data (with groups constructed based on scanner differences) to illustrate the denoising performance of each method. With no denoising, group differences in these regions (Z-values) are statistically significant. The LICA-R1/R2 based methods show a greater reduction in Z-values, with LICA-R1 perhaps having the best by subjective assessment of the whole map. For example, the mean Z-value plots suggest that the LICA-based methods are similar in performance, although LICA-R1 is the only method resulting in no significant differences in the spatial map.

Fig. 5 shows the group differences in TSM before and after data

denoising. The procedure for each denoising method was the same as for the GM analyses. For modality-specific ICA, 48 components were estimated using automatic dimensionally estimation. Ten components with loadings that were highly correlated with scanner variables were removed using `fsl.glm`. The denoising performance of each method for fMRI outcomes was consistent with the performance of each method for GM. With no denoising, there were some regions with statistically significant group differences. GLM regression appears to remove scanner-related group differences from many regions, although other regions appear with group differences. ICA, LICA-SR and LICA-R2 denoising methods remove most scanner effects and reduced group differences in most regions. LICA-R1 appears to perform slightly better than other methods with all scanner-related group differences no longer significant after denoising.

Fig. 6 shows the group differences in CT between the two HC groups before and after data denoising. The procedure for each denoising method was the same as for GM/TSM analyses. The denoising results are

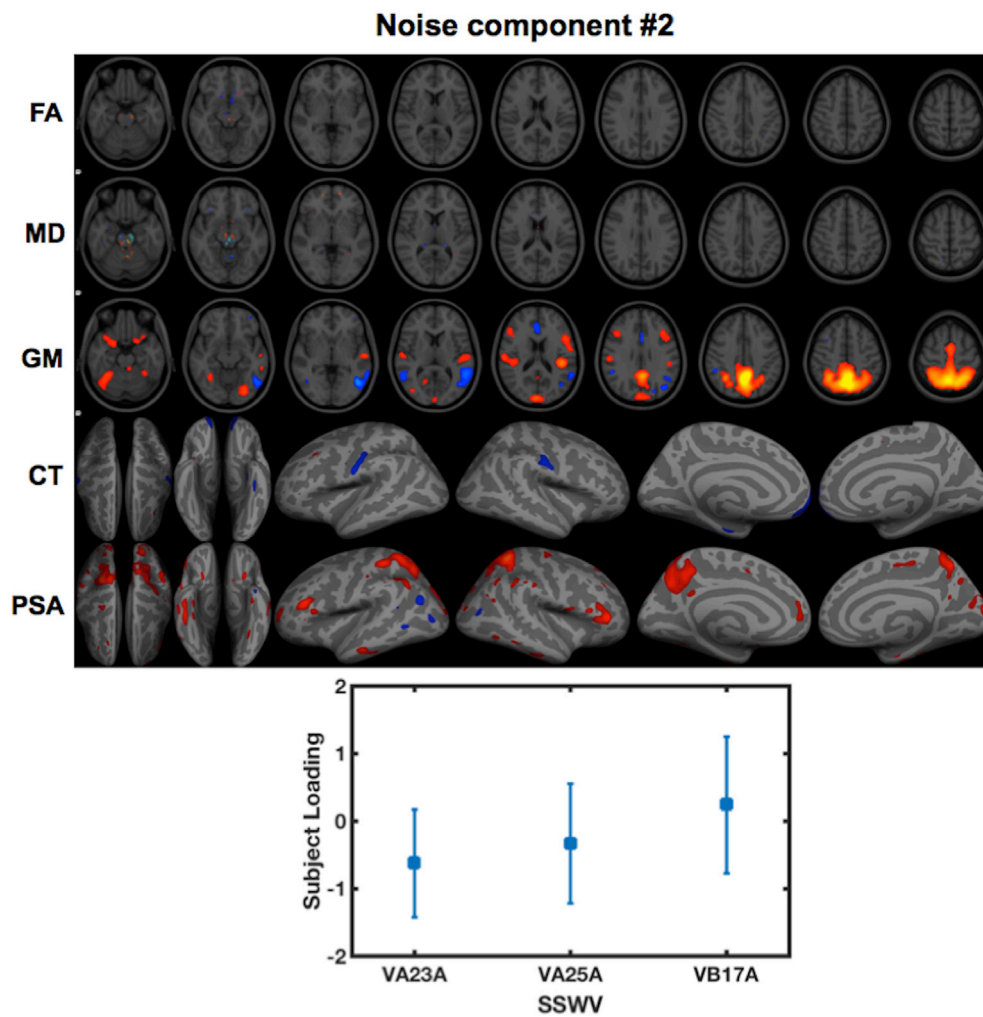


Fig. 2. A second multimodal noise component identified from LICA. Loadings for this component were only associated with SSWV ($p = 2.15 \times 10^{-4}$). The spatial pattern shows region-specific effects in FA, MD, GM, CT and PSA. The spatial maps were thresholded at $Z = 2.3$.

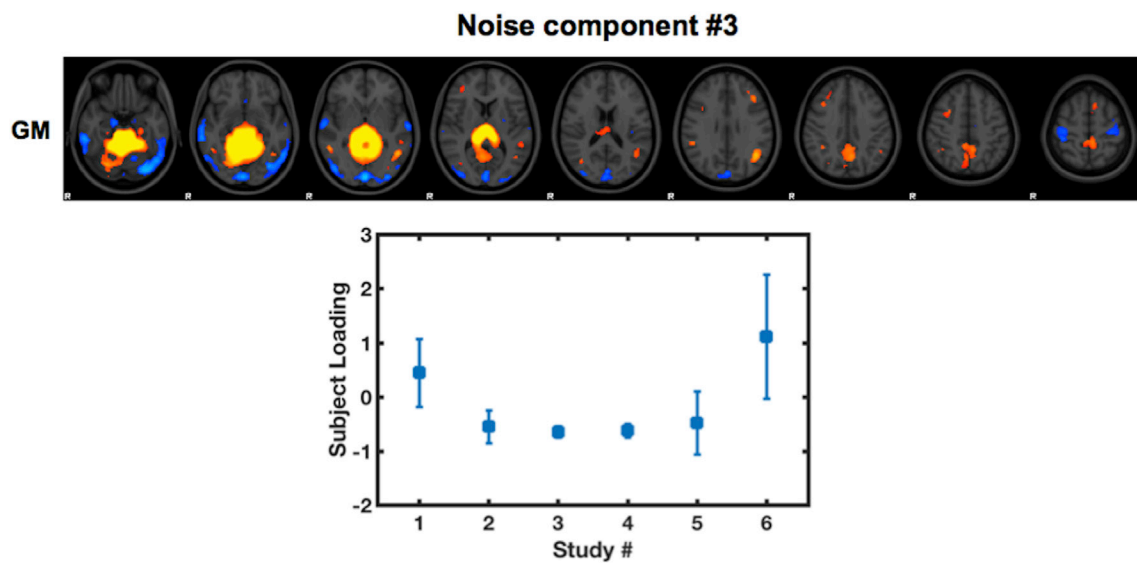


Fig. 3. A third noise component identified from LICA. Loadings for this component were only associated with STUDY ($p = 1.17 \times 10^{-6}$). The spatial pattern was weighted very high for GM, with no or small weights for the other modalities (which are not shown). The spatial map was thresholded at $Z = 2.3$.

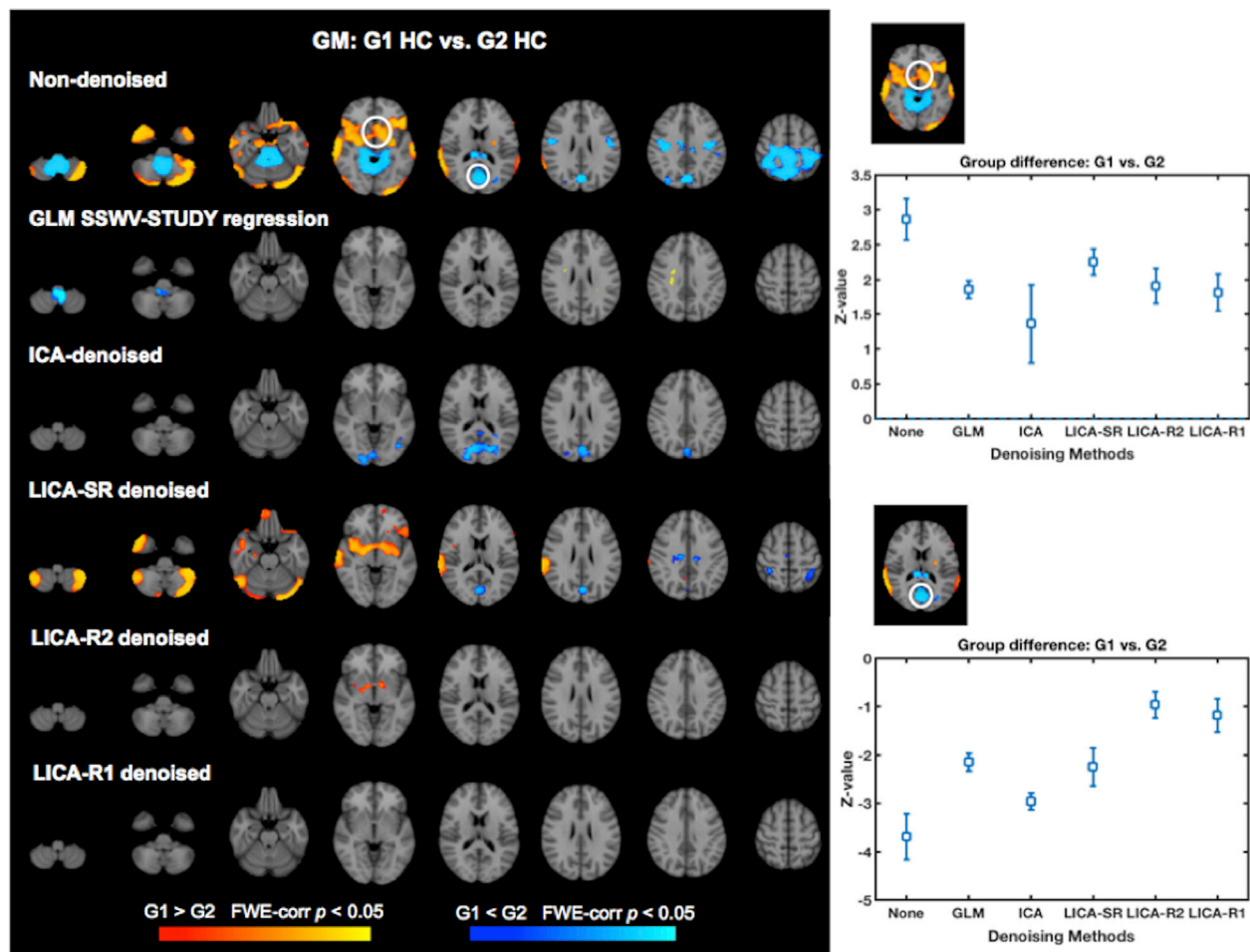


Fig. 4. Group-level analysis of GM maps before and after data denoising. Data were constructed for two HC groups, each with different SSWV. G1 contains 26 subjects with data acquired pre-TIM upgrade; G2 contains 42 subjects with data acquired post-TIM upgrade. The first row shows the results of the group comparison without any data de-noising or regression (i.e., using original GM data). The second row shows statistically significant group differences with SSWV-STUDY regression. The third row shows group differences after GM maps have undergone modality-specific ICA denoising. The bottom three rows show the group differences for GM data that have been denoised with LICA-SR, -R2, and -R1 methods. All spatial maps are thresholded at $p < 0.05$, corrected. The plots on the right show the mean Z-value and standard deviation from voxels with significant group differences shown in the white circles (first row). Group differences (Z-values) are greatest with no denoising and are reduced in both regions (e.g., toward $Z = 0$) using each of the denoising methods, although LICA with hard denoising using loadings shows a clean spatial map, suggesting this method has the best performance.

also quite consistent with the GM and fMRI denoising results. For ICA-denoising, only one component was detected by automatic dimensionally estimation, thus we set the dimensionality to 15 and 30 and found that the 30 component analysis more clearly defined components with loadings related to scanner variables. Therefore, ICA denoising of CT was based on the ICA with 30 components. Eleven components for right hemisphere CT and 9 components for left hemisphere CT were identified as scanner effects and were thus removed from the original data. With no denoising, there were many regions with significant group differences. GLM regression, ICA, and LICA-SR denoising methods remove some scanner-related group differences. LICA-R2 shows better denoising performance with only a few regions with significant group differences remaining after LICA-R2 denoising, while LICA-R1 denoising results in the cleanest maps.

Fig. 7 shows the correlation between subject loadings and subject-specific regression weights obtained from the first stage regression of the LICA noise spatial maps against each modality's subject series conducted for LICA-R2 (\hat{B}_{est_all} from Equation (2)) for the three noise components from the LICA. Subject-specific regression weights from the first stage regression of the noise LICA component spatial maps against each the data for each individual modality (e.g., with a different set of regression weights for component and for each modality) are not

equivalent to the component subject loadings, but there is a high correlation between them. For the first two noise components, which are multimodal, the correlation levels were very high for each modality. For the third noise component, the correlation is especially high for GM (with r-value nearly equal to 1), with less correlation between subject loadings and regression weights for the other modalities, which is due to the strong weighting toward GM for this pattern with very little contribution from other modalities.

To provide some insight into LICA denoising specificity, we identified a multimodal component from the LICA of the non-denoised data with loadings related *only* to MJ effects (Fig. S2a in the Supplement) that also had some spatial overlap with one of the noise components (Fig. 3). This component was evident in a second LICA using data that had been denoised with LICA-R1 (Fig. S2b), demonstrating that LICA denoising of overlapping spatial effects does not remove effects of interest. Furthermore, this same component was replicated using an independent dataset (Fig. S2c).

4. Discussion

A data denoising approach was proposed to remove scanner variability from multimodal MRI data to facilitate combining MRI data

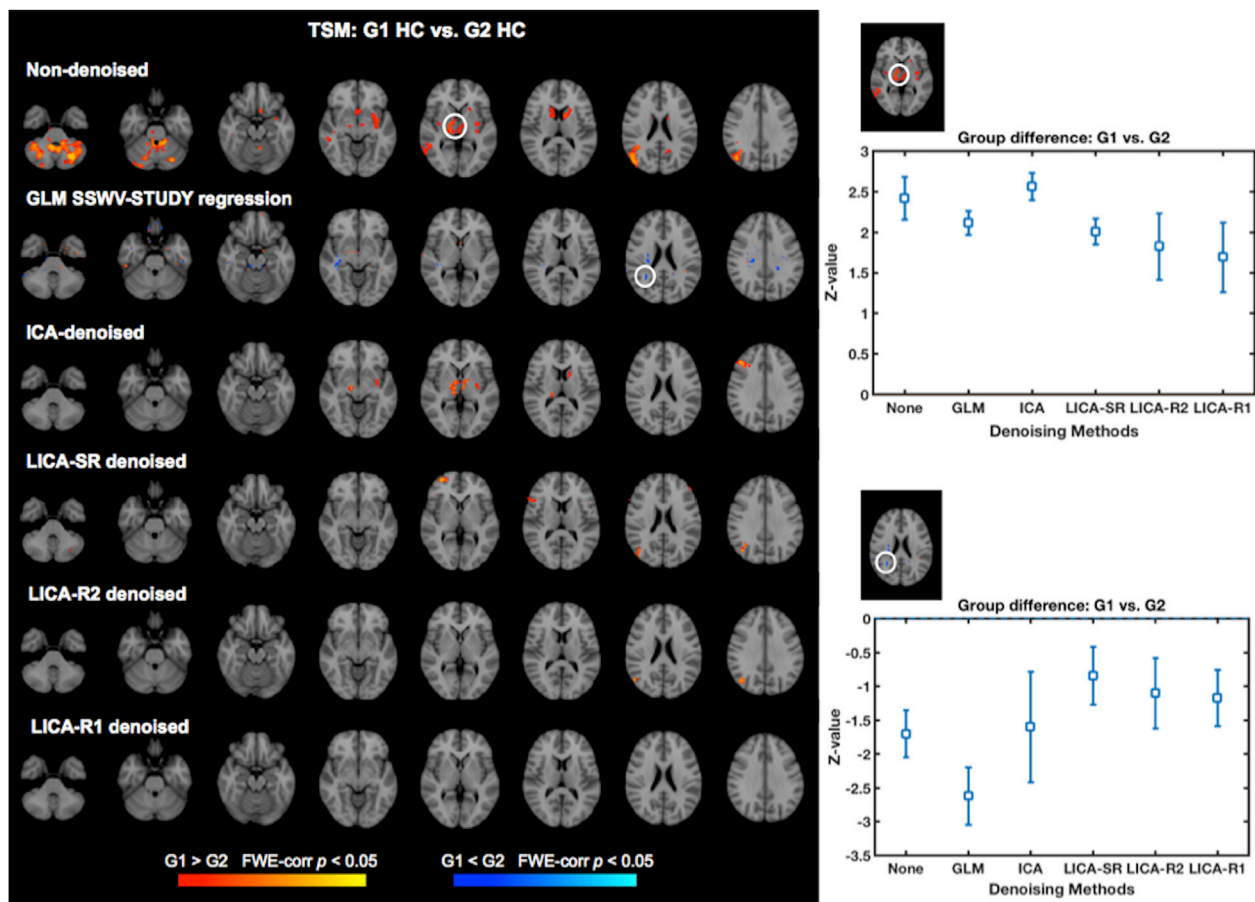


Fig. 5. Group-level analysis of fMRI task spatial maps (TSM) before and after data denoising. The two HC groups were constructed based on SSWV. G1 contains 16 subjects with data acquired pre-TIM upgrade; G2 contains 30 subjects with data acquired post-TIM upgrade. Resulting group difference maps without any data denoising (first row) and with data denoising (rows 2–6) and plots (right) of the mean Z-values and standard deviations extracted from voxels with significant group differences (shown in the white circles) show that LICA-R1 denoising of TSM achieves the best noise removal.

collected across scanner upgrades, or from different scanners. We demonstrated this approach for the former (e.g., pre- and post-TIM upgrade of both gradient and head coils, operating system, and other hardware), which represents a more conservative test of our method than using multi-site data. E.g., LICA is sensitive to variability in the data and between-scanner variability is generally greater than within-scanner variability, thus LICA should be more effective at identifying scanner-related effects for multi-site data than data collected on a single scanner.

We tested our method for structural, DTI, and fMRI modalities. However, LICA can be used with any combination of these modalities (or outcomes from fewer or single modalities) and may also be capable of denoising structural and functional connectome modalities, although LICA will need to be tested and possibly optimized for these other modalities. For example, LICA may perform better using spatial connectivity maps such as those from dual regression (Nickerson et al., 2017) rather than association matrices, but this remains to be investigated. Because either LICA spatial patterns or subject loadings can be used to denoise the data, we assessed three different strategies for removing scanner-related components identified by LICA. We also compared the performance of LICA-based denoising with two other statistical methods frequently employed to address scanner effects, GLM regression (that includes scanner covariates in the group-level model), and single-modality ICA denoising prior to group level statistical analysis. Overall, LICA-R1, hard regression using LICA subject loadings, performed well across all modalities and brain regions, by reducing scanner-related variance, while not introducing false positives.

LICA-based denoising generally performed better than the GLM

method for several possible reasons. The first is that scanner-related effects are better captured by LICA and removed using multivariate regression. The standard GLM approach includes scanner confound regressors in the higher-level voxel-wise GLM that represent simple models for scanner effects. These regressors have the same value for all participants for the same scanner or scanner state, and are therefore unable to capture day-to-day variations in scanner performance. Furthermore, ignoring the spatial covariance when identifying scanner-related effects is an unrealistic assumption that likely results in reduced sensitivity to detect effects when compared with multivariate methods that utilize spatial covariance information. The GLM approach also treats each voxel as being independent of all other voxels when controlling for such effects via implementation of a massively univariate GLM for statistical modeling. In contrast to multivariate regression employed for LICA-based denoising that is able to disentangle overlapping effects that may contribute to a voxel's signal intensity (e.g., a scanner artifact may also contribute to a voxel that has meaningful signal), using the GLM approach controls for scanner effects in a voxel-wise fashion and is therefore not able to disentangle any overlapping effects.

LICA denoising also offers advantages for multimodal MRI data as compared with modality-specific ICA-based denoising. Modality-specific ICA-based denoising identifies noise components based only on a single modality and thus does not capitalize on shared covariance across modalities to identify scanner effects. In addition, while LICA and single-modality ICA both require estimating the dimensionality (or number of components) and manually identifying scanner-related components, this only has to be done once for LICA, whereas for single-modality ICA, it has

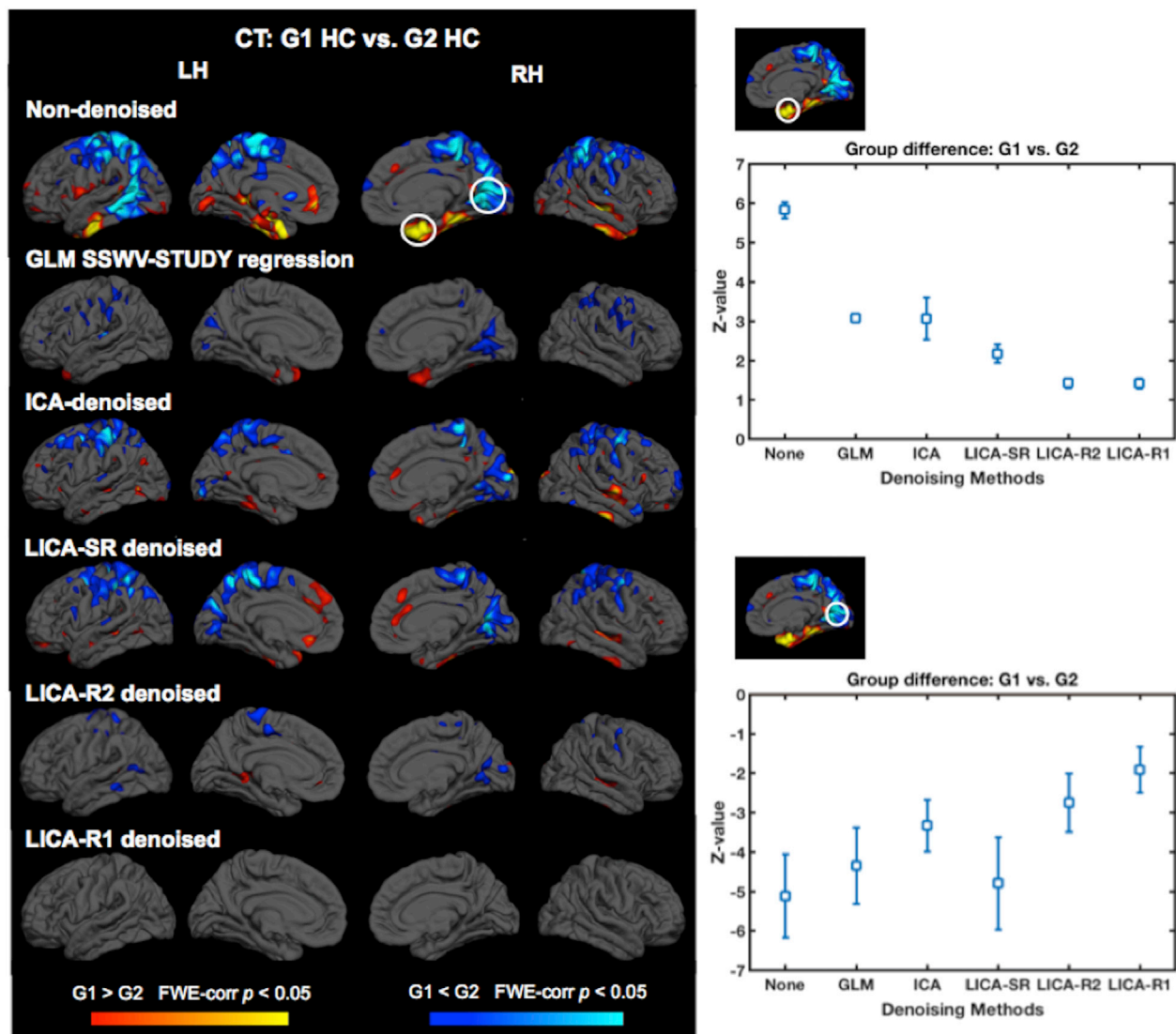


Fig. 6. Group-level analysis of CT data before and after denoising. Two groups of HC data were constructed based on SSWV. G1 contains 29 subjects with data acquired pre-TIM upgrade; G2 contains 42 subjects with data acquired post-TIM upgrade. Resulting group difference maps without any data denoising (first row) and with data denoising (rows 2–6) and plots (right) of the mean Z-values and standard deviations extracted from voxels with significant group differences (shown in the white circles) show results that are similar to the GM and fMRI results in that every denoising procedure removes some of the effects of SSWV, while LICA-R1 denoising achieves the cleanest group difference maps.

to be done for each modality separately. Ideally both LICA and ICA dimensionalities should be “tuned” to give the best separation of signal and noise into separate components (that is, run at different dimensionalities to identify the best separation). This process is tricky in either case to ensure that effects of interest are not also removed, yet relatively easier to do for LICA, as compared with doing this identification process multiple times for many individual modalities. It should also be noted that for fMRI data, it is common to denoise each subject’s fMRI timeseries data prior to statistical analyses (Griffanti et al., 2014; Pruim et al., 2015; Salimi-Khorshidi et al., 2014), which only removes within-subject sources of noise, such as motion and artifacts. While scanner-related noise during that fMRI run may be identified and removed by ICA denoising, the ICA would not be able to capture between subject variability associated with different scanners or with software/hardware upgrades. Thus, if using single-modality ICA to denoise scanner effects from fMRI data, one should apply this method to the subject-series (activation maps for all subjects together), not the single-subject fMRI timeseries itself. However, our findings show that single-modality ICA denoising is not as effective as multimodal denoising for removing scanner effects.

We found that aggressive denoising using the subject loadings (LICA-R1) of only the scanner-related components was able to remove scanner effects somewhat better than aggressive denoising using the LICA spatial maps (LICA-R2) of scanner components and non-aggressive denoising using the subject loadings for all LICA components (LICA-SR). LICA-R2 may be less effective than LICA-R1 because the subject-specific regression weights from the first stage regression (\hat{B}_{est_all} from Eq. (2)) in LICA-R2 are modality-specific, whereas the LICA loadings used in LICA-R1 are shared across modalities. The correlation between the LICA loadings and the subject-specific regression weights were extremely high when the noise component being removed was heavily weighted to a single modality that was the same as the modality being denoised, but was reduced when using these same components to denoise other modalities that did not contribute as much to the LICA noise pattern. Thus, while LICA-R2 may be effective, we recommend using LICA-R1 as the loadings are more accurate for multi-modal effects.

We identified LICA scanner-related components as those with loadings that were associated with scanner variables and not with participant variables, thus hard regression (LICA-R1) is interpreted to remove solely

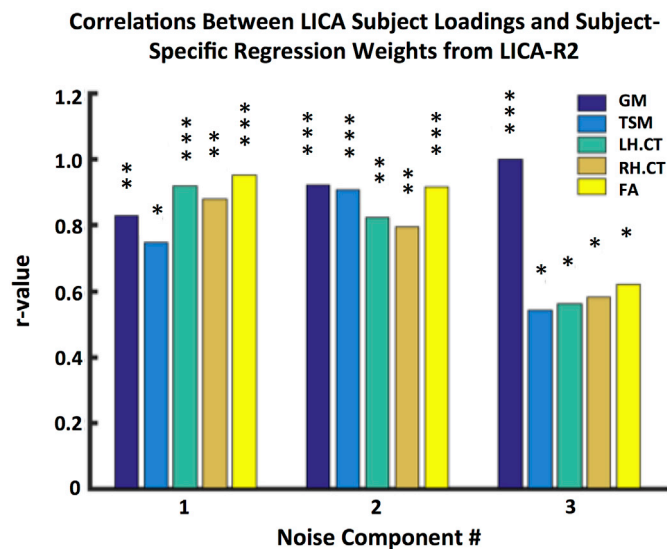


Fig. 7. Correlations between LICA subject loadings for each noise component and the subject-specific regression weights for each modality obtained from the first stage regression of the LICA noise spatial maps against each modality's subject series conducted for LICA-R2. Subject-specific regression weights are not equivalent to the LICA subject loadings, but they are highly correlated for modalities represented in any given component. For example, for multimodal noise components #1 and #2, correlations were very high for all modalities. For noise component #3, which is primarily weighted on GM, the correlation is very high for GM, but reduced for other modalities. * $p < 10^{-5}$, ** $p < 10^{-30}$, *** $p < 10^{-50}$.

scanner effects from the two-group constructed data we used to assess the performance of each method. Furthermore, with these components, we would expect LICA-SR, which removes only the unique variance related to the scanner components to give similar results. However, we found that this approach is not as effective in removing scanner noise, possibly due to the reduction in degrees of freedom (e.g., by including more regressors in the first step regression).

We demonstrated our proposed method using multi-study data collected on the same scanner, but across major hardware and software upgrades and utilizing different acquisition parameters. We showed that, similar to Groves et al. (2012), we could identify components associated with these scanner effects. Importantly, LICA should be more effective at identifying cross-site scanner variability since cross-site variations likely exceed the within-scanner variations in the data used for the present study. Thus, the proposed LICA-based denoising method has great potential for pooling multi-site MRI data.

There are some limitations of this study. First, while we used the p-value associated with the regression of scanner and other variables against the component loadings to identify noise components, there may be other more effective approaches for identifying noise components. For example, with enough training data, more sophisticated automated machine-learning based approaches similar to ICA-AROMA (Pruim et al., 2015) or FSL FIX (Salimi-Khorshidi et al., 2014) for fMRI data denoising may be possible. Further work is necessary in this regard. In addition, although we tested the performance of different denoising methods by separating (only) healthy control data into two groups based on scanner variables, the groups were matched for age, but not other variables, i.e., sex, that may have had an effect on the differences between groups in these comparisons (observed using non-denoised data, first rows of Figs. 4–6). The group differences observed in GM, fMRI, and CT maps between the two HC groups were not significant after LICA-based denoising of components that were related only to scanner/study effects so we are confident that we accurately constructed simulated groups from the HC data that reflect primarily pre- and post-TIM effects, and do not likely reflect the impact of demographic variables. However, since use of all HC data resulted in an imbalance between the groups in sample

size and sex for some modalities, we also show in the Supplementary File, “LICA Denoising Performance”, that using a smaller subset of the HC data to construct pre- and post-TIM groups of equal sample size that are matched for both age and sex produced similar denoising results. Thus, group differences observed with non-denoised data are likely not due to sex or age. It is of note however, that we did not consider other participant demographic variables in the analyses (IQ, etc) that may have contributed variance, which will be done in future studies. While we demonstrated our method for multi-modal MRI data comprised of structural MRI, DTI and functional MRI, this method is applicable to any combination of modalities that are specific to a particular study (Groves et al., 2011), with Groves et al. (2012) also identifying scanner-related components using a different combination of modalities than in our study. Last, we demonstrated our technique using a dataset that was heterogeneous in the modalities collected for each study, leading to “missing” data for some participants (e.g., for a modality that was not collected for that participant). While this does not impact the goal of the present study – to show how our method is applied – the linked ICA is expected to perform better with lower levels of “missing” data.

A final limitation of our study is that while we present evidence that LICA-based R1 and R2 methods perform better for removing noise effects, to demonstrate the specificity of the approach, simulations are needed (with known ground truth) to assess how well effects of interest are retained upon denoising. However, a preliminary assessment of specificity presented in the Supplementary File “LICA Denoising Performance” (Figs. S2a–c) shows that LICA denoising does not remove effects of interest and that LICA findings are reproducible. As an aside, aging components reported in Groves et al. (2012) were also replicated using our data and using an independent dataset, which provides further evidence that LICA results are reproducible. While future work is needed to better establish the specificity of our method, our findings provide evidence that LICA-based denoising is a sensitive tool for removing scanner variability from multi-modal MRI data.

The Supplement file “Manual for LICA-based Denoising”, provides instructions and code for implementing LICA-based denoising methods.

5. Conclusions

Combining MRI data collected using different MRI scanners, across scanner software and hardware upgrades, or using different acquisition parameters, is a complicated endeavor because scanner-related variability can confound statistical analyses of the combined data. Linked ICA identifies multi-modal patterns associated with sources of variability in multi-modal MRI data, thus it is well-suited for identifying scanner effects, which can be large sources of variability. We present a proof-of-concept study that demonstrates linked ICA-based denoising for removing these scanner effects from multimodal MRI data collected on the same scanner, but spanning major software and hardware upgrades. Our proposed approach overcomes limitations of existing methods we tested that are frequently applied for combining MRI data. We demonstrate this approach in the case of denoising data combined across a major scanner upgrade. Although we did not test it here, for combining data across different scanners, LICA should prove even better at identifying scanner effects, as between-scanner variability is generally much larger than within-scanner variability. Thus our proposed approach may be of great value for combining data in numerous existing large multi-site neuroimaging studies, including the Human Connectome studies (Van Essen et al., 2013; <https://www.humanconnectome.org>), the Adolescent Brain Cognitive Development (ABCD; Volkow et al., 2018) and UK Biobank (Miller et al., 2016), and for combining data from repositories.

Author contributions

LDN conceived and designed the study, oversaw data analysis and manuscript preparation. HL was responsible for data analysis and manuscript preparation. SM provided statistical expertise and edited the

manuscript. SG, SEL, MMS, WDSK, and KPH contributed data and edited the manuscript.

Acknowledgements

This work was supported by the National Institutes of Health grant DA037265 (LDN). Data collection was supported by National Institutes of Health grants DA016695 (SG), DA021241 (SG), DA024007 (SEL), DA029115 (KPH) and AA014651 (MMS), and by DARPA-12-12-11-YFA11-FP-029 (WDSK).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2019.116388>.

References

- Abou-Elseoud, A., Starck, T., Remes, J., Nikkinen, J., Tervonen, O., Kiviniemi, V., 2010. The effect of model order selection in group PICA. *Hum. Brain Mapp.* 31, 1207–1216.
- Andersson, J.L., Graham, M.S., Zsoldos, E., Sotiropoulos, S.N., 2016. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage* 141, 556–572.
- Andersson, J.L.R., Jenkinson, M., Smith, S., 2007a. Non-linear Optimization. FMRIB technical report TR07JA1 from. www.fmrib.ox.ac.uk/analysis/techrep.
- Andersson, J.L.R., Jenkinson, M., Smith, S., 2007b. Non-linear Registration, Aka Spatial Normalization. FMRIB technical report TR07JA2 from. www.fmrib.ox.ac.uk/analysis/techrep.
- Bartsch, H., Thompson, W.K., Jernigan, T.L., Dale, A.M., 2014. A web-portal for interactive data exploration, visualization, and hypothesis testing. *Front. Neuroinf.* 8, 25.
- Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* 23, 137–152.
- Bell, A.J., Sejnowski, T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159.
- Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., et al., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376.
- Casey, B.J., Cohen, J.D., O'Craven, K., Davidson, R.J., Irwin, W., Nelson, C.A., et al., 1998. Reproducibility of fMRI results across four institutions using a spatial working memory task. *Neuroimage* 8, 249–261.
- Chen, J., Liu, J., Calhoun, V.D., Arias-Vasquez, A., Zwiers, M.P., Gupta, C.N., et al., 2014. Exploration of scanning effects in multi-site structural MRI studies. *J. Neurosci. Methods* 230, 37–50.
- Costafreda, S.G., Brammer, M.J., Vêncio, R.Z.N., Mourão, M.L., Portela, L.A.P., de Castro, C.C., et al., 2007. Multisite fMRI reproducibility of a motor task using identical MR systems. *J. Magn. Reson. Imaging* 26, 1122–1126.
- Cui, J., Tkachenko, O., Gogel, H., Kipman, M., Preer, L.A., Weber, M., et al., 2015. Microstructure of frontoparietal connections predicts individual resistance to sleep deprivation. *Neuroimage* 106, 123–133.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194.
- Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. *J. Cogn. Neurosci.* 5, 162–176.
- Douaud, G., Smith, S., Jenkinson, M., Behrens, T., Johansen-Berg, H., Vickers, J., et al., 2007. Anatomically related grey and white matter abnormalities in adolescent-onset schizophrenia. *Brain* 130, 2375–2386.
- Eickhoff, S., Nichols, T.E., Van Horn, J.D., Turner, J.A., 2016. Sharing the wealth: neuroimaging data repositories. *Neuroimage* 124, 1065–1068.
- Feis, R.A., Smith, S.M., Filippini, N., Douaud, G., Dopper, E.G., Heise, V., et al., 2015. ICA-based artifact removal diminishes scan site differences in multi-center resting-state fMRI. *Front. Neurosci.* 9, 395.
- Fennema-Notestine, C., Gamst, A.C., Quinn, B.T., Pacheco, J., Jernigan, T.L., Thal, L., et al., 2007. Feasibility of multi-site clinical structural neuroimaging studies of aging using legacy data. *Neuroinformatics* 5, 235–245.
- Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11050–11055.
- Fischl, B., Sereno, M.I., Dale, A.M., 1999a. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207.
- Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8, 272–284.
- Fischl, B., Liu, A., Dale, A.M., 2001. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans. Med. Imaging* 20, 70–80.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., et al., 2004. Automatically parcellating the human cerebral cortex. *Cerebr. Cortex* 14, 11–22.
- Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T., et al., 2008. Cortical folding patterns and predicting cytoarchitecture. *Cerebr. Cortex* 18, 1973–1980.
- Friedman, L., Stern, H., Brown, G.G., Mathalon, D.H., Turner, J., Glover, G.H., et al., 2008. Test-retest and between-site reliability in a multicenter fMRI study. *Hum. Brain Mapp.* 29, 958–972.
- Focke, N.K., Helms, G., Kaspar, S., Diederich, C., Toth, V., Dechent, P., et al., 2011. Multi-site voxel-based morphometry-not quite there yet. *Neuroimage* 56, 1164–1170.
- Fortin, J.P., Parker, D., Tunç, B., Watanabe, T., Elliott, M.A., Ruparel, K., et al., 2017. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* 161, 149–170.
- Glover, G.H., Mueller, B.A., Turner, J.A., van Erp, T.G.M., Liu, T.T., Greve, D.N., et al., 2012. Function biomedical informatics research network recommendations for prospective multicenter functional MRI studies. *J. Magn. Reson. Imaging* 36, 39–54.
- Griffanti, L., Salimi-Khorshidi, G., Beckmann, C.F., Auerbach, E.J., Douaud, G., Sexton, C.E., et al., 2014. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage* 95, 232–247.
- Groves, A.R., Beckmann, C.F., Smith, S.M., Woolrich, M.W., 2011. Linked independent component analysis for multimodal data fusion. *Neuroimage* 54, 2198–2217.
- Groves, A.R., Smith, S.M., Fjell, A.M., Tamnes, C.K., Walhovd, K.B., Douaud, G., et al., 2012. Benefits of multi-modal fusion analysis on a large-scale dataset: life-span patterns of inter-subject variability in cortical morphometry and white matter microstructure. *Neuroimage* 63, 365–380.
- Gruber, S.A., Sagar, K.A., Dahlgren, M.K., Gonenc, A., Conn, N.A., Winer, J.P., et al., 2015. Citalinone treatment improves measures of impulsivity and task performance in chronic marijuana smokers: a pilot BOLD fMRI study. *Int. J. Neurol. Neurother.* 2, 1–8.
- Gruber, S.A., Dahlgren, M.K., Sagar, K.A., Gonenc, A., Lukas, S.E., 2014. Worth the wait: effects of age of onset of marijuana use on white matter and impulsivity. *Psychopharmacology* 231, 1455–1465.
- Gruber, S.A., Dahlgren, M.K., Sagar, K.A., Gonenc, A., Killgore, W.D.S., 2012. Age of onset of marijuana use impacts inhibitory processing. *Neurosci. Lett.* 511, 89–94.
- Guo, X., Wang, Y., Guo, T., Chen, K., Zhang, J., Li, K., Jin, Z., Yao, L., 2015. Structural covariance networks across healthy young adults and their consistency. *J. Magn. Reson. Imaging* 42, 261–268.
- Hill, K.P., Palastro, M.D., Gruber, S.A., Fitzmaurice, G.M., Greenfield, S.F., Lukas, S.E., et al., 2017. Nabilone pharmacotherapy for cannabis dependence: a randomized, controlled pilot study. *Am. J. Addict.* 26, 795–801.
- Huisman, T.A., Loenneker, T., Barta, G., Bellemann, M.E., Hennig, J., Fischer, J.E., et al., 2006. Quantitative diffusion tensor MR imaging of the brain: field strength related variance of apparent diffusion coefficient (ADC) and fractional anisotropy (FA) scalars. *Eur. Radiol.* 16, 1651–1658.
- Iscan, Z., Jin, T.B., Kendrick, A., Szeglin, B., Lu, H., Trivedi, M., et al., 2015. Test-retest reliability of freesurfer measurements within and between sites: effects of visual approval process. *Hum. Brain Mapp.* 36, 3472–3485.
- Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841.
- Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., et al., 2009. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *Neuroimage* 46, 177–192.
- Keihaninejad, S., Heckemann, R.A., Fagiolo, G., Symms, M.R., Hajnal, J.V., Hammers, A., 2010. A robust method to estimate the intracranial volume across MRI field strengths (1.5T and 3T). *Neuroimage* 50, 1427–1437.
- Kochunov, P., Jahanshad, N., Marcus, D., Winkler, A., Sprooten, E., Nichols, T.E., et al., 2015. Heritability of fractional anisotropy in human white matter: a comparison of Human Connectome Project and ENIGMA-DTI data. *Neuroimage* 111, 300–311.
- Littmann, A., Guehring, J., Buechel, C., Stiehl, H.S., 2006. Acquisition-related morphological variability in structural MRI. *Acad. Radiol.* 13, 1055–1061.
- Mashhoon, Y., Sava, S., Sneider, J.T., Nickerson, L.D., Silveri, M.M., 2015. Cortical thickness and volume differences associated with marijuana abuse in emerging adults. *Drug Alcohol Depend.* 155, 275–283.
- Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J.Q., et al., 2016. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* 19, 1523–1536.
- Nickerson, L.D., 2018. Replication of resting state-task network correspondence and novel findings on brain network activation during task fMRI in the human connectome project study. *Sci. Rep.* 8, 17543.
- Nickerson, L.D., Smith, S.M., Öngür, D., Beckmann, C.F., 2017. Using dual regression to investigate network shape and amplitude in functional connectivity analyses. *Front. Neurosci.* 11, 115.
- Pagani, E., Hirsch, J.G., Pouwels, P.J., Horsfield, M.A., Perego, E., et al., 2010. Intercenter differences in diffusion tensor MRI acquisition. *J. Magn. Reson. Imaging* 31, 1458–1468.
- Pardoe, H.R., Cutter, G.R., Alter, R., Hiess, R.K., Semmelroch, M., Parker, D., et al., 2016. Pooling morphometric estimates: a statistical equivalence approach. *J. Neuroimaging* 26, 109–115.
- Pohl, K.M., Sullivan, E.V., Rohlfing, T., Chu, W., Kwon, D., Nichols, B.N., et al., 2016. Harmonizing DTI measurements across scanners to examine the development of white matter microstructure in 803 adolescents of the NCANDA study. *Neuroimage* 130, 194–213.

- Pruim, R.H.R., Mennes, M., Rooij, D.V., Llera, A., Buitelaar, J.K., Beckmann, C.F., 2015. ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* 112, 267–277.
- Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* 90, 449–468.
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., et al., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23 (Suppl. 1), S208–S219.
- Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., et al., 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* 31, 1487–1505.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155.
- Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44, 83–98.
- Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., et al., 2009. Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci.* 106, 13040–13045.
- Takao, H., Hayashi, N., Ohtomo, K., 2011. Effect of scanner in longitudinal studies of brain volume changes. *J. Magn. Reson. Imaging* 34, 438–444.
- Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., for the WU-Minn HCP Consortium, 2013. The Wu-Minn human connectome project: an overview. *Neuroimage* 80, 62–79.
- Van Horn, J.D., Toga, A.W., 2009. Multisite neuroimaging trials. *Curr. Opin. Neurol.* 22, 370–378.
- Varoquaux, G., 2018. Cross-validation failure: small sample sizes lead to large error bars. *Neuroimage* 180, 68–77.
- Venkatraman, V.K., Gonzalez, C.E., Landman, B., Goh, J., Reiter, D.A., An, Y., et al., 2015. Region of interest correction factors improve reliability of diffusion imaging measures within and across scanners and field strengths. *Neuroimage* 119, 406–416.
- Volkow, N.D., Koob, G.F., Croyle, R.T., Bianchi, D.W., Gordon, J.A., Koroshetz, W.J., et al., 2018. The conception of the ABCD study: from substance abuse to a broad NIH collaboration. *Dev. Cogn. Neurosci.* 32, 4–7.
- Vollmar, C., O'Muircheartaigh, J., Barker, G.J., Symms, M.R., Thompson, P., Kumari, V., et al., 2010. Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0 T scanners. *Neuroimage* 51, 1384–1394.
- Wegner, C., Filippi, M., Korteweg, T., Beckmann, C., Ciccarelli, O., De Stefano, N., et al., 2008. Relating functional changes during hand movement to clinical parameters in patients with multiple sclerosis in a multi-centre fMRI study. *Eur. J. Neurol.* 15, 113–122.
- Woolrich, M.W., Ripley, B.D., Brady, J.M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modelling of FMRI data. *Neuroimage* 14, 1370–1386.
- Xu, L., Groth, K.M., Pearlson, G., Schretlen, D.J., Calhoun, V.D., 2009. Source-based morphometry: the use of independent component analysis to identify gray matter differences with application to schizophrenia. *Hum. Brain Mapp.* 30, 711–724.
- Zivadinov, R., Cox, J.L., 2008. Is functional MRI feasible for multi-center studies on multiple sclerosis? *Eur. J. Neurol.* 15, 109–110.
- Zou, K.H., Greve, D.N., Wang, M., Pieper, S.D., Warfield, S.K., White, N.S., 2005. Reproducibility of functional MR Imaging: preliminary results of prospective multi-institutional study performed by biomedical informatics research network. *Radiology* 237, 781–789.