



ORIGINAL ARTICLE

Global multi-center and multi-modal magnetic resonance imaging study of obsessive-compulsive disorder: Harmonization and monitoring of protocols in healthy volunteers and phantoms

Petra J. W. Pouwels¹  | Chris Vriend² | Feng Liu³ | Niels T. de Joode² | Maria C. G. Otaduy⁴ | Bruno Pastorello⁴ | Frances C. Robertson⁵ | Ganesan Venkatasubramanian⁶  | Jonathan Ipser⁷ | Seonjoo Lee³ | Marcelo C. Batistuzzo^{8,9} | Marcelo Q. Hoexter⁸ | Christine Lochner¹⁰ | Euripedes C. Miguel⁸ | Janardhanan C. Narayanaswamy⁶ | Rashmi Rao⁶ | Y. C. Janardhan Reddy⁶ | Roseli G. Shavitt⁸ | Karthik Sheshachala⁶ | Dan J. Stein⁷ | Anton J. L. M. van Balkom¹¹ | Melanie Wall³ | Helen Blair Simpson³ | Odile A. van den Heuvel²

¹Department of Radiology and Nuclear Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Neuroscience, Amsterdam, The Netherlands

²Department of Psychiatry, Department of Anatomy and Neuroscience, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Neuroscience, Amsterdam, The Netherlands

³Columbia University Irving Medical Center, Columbia University, New York State Psychiatric Institute, New York, NY, USA

⁴Department of Radiology, LIM44, Institute, Hospital Das Clinicas-HCFMUSP, University of Sao Paulo Medical School, Sao Paulo, Brazil

⁵Cape Universities Body Imaging Centre, University of Cape Town, Cape Town, South Africa

⁶National Institute of Mental Health & Neurosciences (NIMHANS), Bangalore, India

⁷Department of Psychiatry, SAMRC Unit on Risk & Resilience in Mental Disorders, Neuroscience Institute, University of Cape Town, Cape Town, South Africa

⁸Obsessive-Compulsive Spectrum Disorders Program, Departamento de Psiquiatria da Faculdade de Medicina, LIM23, Hospital Das Clinicas HCFMUSP, Universidade de São Paulo, Sao Paulo, SP, Brazil

⁹Department of Methods and Techniques in Psychology, Pontifical Catholic University, Sao Paulo, SP, Brazil

¹⁰Department of Psychiatry, SAMRC Unit on Risk & Resilience in Mental Disorders, Stellenbosch University, Cape Town, South Africa

¹¹Department of Psychiatry, Amsterdam UMC, Vrije Universiteit, Amsterdam Public Health Research Institute, Specialised Mental Health Care, Amsterdam, The Netherlands

Correspondence

Petra J. W. Pouwels, Department of Radiology and Nuclear Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Neuroscience, de Boelelaan 1117, Amsterdam, The Netherlands.
Email: pjw.pouwels@amsterdamumc.nl

Abstract

Objectives: We describe the harmonized MRI acquisition and quality assessment of an ongoing global OCD study, with the aim to translate representative, well-powered neuroimaging findings in neuropsychiatric research to worldwide populations.

Petra J. W. Pouwels, Chris Vriend Shared first author.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. International Journal of Methods in Psychiatric Research published by John Wiley & Sons Ltd.

Funding information

National Institute of Mental Health, Grant/Award Number: R01 MH113250; The Wellcome Trust DBT India Alliance

Methods: We report on T1-weighted structural MRI, resting-state functional MRI, and multi-shell diffusion-weighted imaging of 140 healthy participants (28 per site), two traveling controls, and regular phantom scans.

Results: Human image quality measures (IQMs) and outcome measures showed smaller within-site variation than between-site variation. Outcome measures were less variable than IQMs, especially for the traveling controls. Phantom IQMs were stable regarding geometry, SNR, and mean diffusivity, while fMRI fluctuation was more variable between sites.

Conclusions: Variation in IQMs persists, even for an a priori harmonized data acquisition protocol, but after pre-processing they have less of an impact on the outcome measures. Continuous monitoring IQMs per site is valuable to detect potential artifacts and outliers. The inclusion of both cases and healthy participants at each site remains mandatory.

KEYWORDS

DWI, fMRI, image quality measure, multi-vendor, structural MRI

1 | INTRODUCTION

Neuroimaging has increased our understanding of the neurobiology of obsessive-compulsive disorder (OCD) (Stein et al., 2019), but studies generally used relatively small samples. International collaborative efforts, such as Enhancing Neuroimaging and Genetics through Meta-Analysis (ENIGMA), could increase statistical power by combining samples across sites (Thompson et al., 2020). ENIGMA-OCD meta- and mega-analyses have strengthened international collaboration and investigated whether structural alterations found in OCD reflect neurodevelopmental changes, vulnerability factors, effects of disease (chronicity), or medication (van den Heuvel et al., 2020). However, a methodological limitation of ENIGMA is the lack of prospective harmonization of acquisition protocols before data pooling. Moreover, due to between-site differences in clinical measures (e.g., course of illness, severity, comorbidity, treatment history, and symptom dimensions), pooling of these data is also limited.

To address this problem, five ENIGMA-OCD sites from five continents, launched the largest multimodal-imaging and neurocognitive study in medication-free OCD patients to date (i.e., the OCD Global Study, entitled 'Identifying Reproducible Brain Signatures of Obsessive-Compulsive Profiles' R01-MH113250), using harmonized methods for clinical phenotyping, neurocognitive testing and neuroimaging (Simpson et al., 2020). With this collaboration, we aim: (1) to identify reproducible brain signatures that distinguish OCD patients from their unaffected siblings and healthy control participants, and (2) to associate these brain signatures with OCD-related clinical and neurocognitive profiles. We developed detailed protocols to enhance cross-site reliability on clinical, neurocognitive, and MRI measures.

This manuscript describes the standardized MRI protocol we developed to acquire structural imaging, resting-state functional MRI (rsfMRI) and multi-shell diffusion-weighted imaging (DWI), within a clinically feasible timeframe (<1 h) and for clinical 3 T scanners from various vendors, and the methods used to monitor scan quality. We describe within and between-site variability in image quality and standard neuroimaging outcome measures (e.g., whole-brain morphometry and white matter diffusion) from physical phantoms, traveling human volunteers and 28 healthy participants from each site.

2 | METHODS

The OCD Global Study started data collection in 2018 across its five research sites in Brazil, India, the Netherlands, South Africa, and U.S.A (Simpson et al., 2020). Vendors and types of MRI scanners are listed in Table 1.

2.1 | Human participants

For details on in- and exclusion criteria of the OCD global study, see (Simpson et al., 2020). Written informed consent was obtained from participants according to the Declaration of Helsinki. Study protocols were approved by the five local Medical Ethical Committees.

We selected 28 healthy participants from each site by matching them on age and education using a weighted nearest-neighbor selection method (Szekér & Vathy-Fogarassy, 2020).

Two traveling healthy volunteers (not enrolled in the OCD Global study) visited each site and underwent the same MRI protocol

TABLE 1 MRI scanners and parameters per site

site	1	2	3	4	5
Scanner	Philips achieva 3.0 T	Philips ingenia 3.0 T CX	GE 3.0 T discovery MR750	Siemens MAGNETOM skyra 3.0 T	GE SIGNA 3.0 T premier
Head coil	32-Channel	32-Channel	32-Channel	32-Channel	48-Channel
T1w: 3D sagittal T1-weighted MPRAGE according to ADNI-3 protocol					
TR (ms) ^a	6.5	6.5	6.9	2300	2235
TI (ms)	900	900	900	900	900
TE (ms)	2.9	2.9	3	2	2.8
Flip angle (°)	9	9	9	9	9
Voxel size (mm)	1 × 1 × 1	1 × 1 × 1	1 × 1 × 1	1 × 1 × 1	1 × 1 × 1
Matrix	256 × 256	256 × 256	256 × 256	256 × 256	256 × 256
Resting-state fMRI: Gradient-echo echo-planar images while subjects are awake and keep their eyes closed; axial ascending slices parallel to line through the pituitary gland and 4 th ventricle; 10 min					
TR (ms)	2200	2200	2200	2200	2200
TE (ms)	28	28	28	28	28
Flip angle (°)	80	80	80	80	80
# Slices	44	44	42	42	42
# Volumes	275	275	275	272	272
Voxel size (mm)	3.3 × 3.3 × 3	3.3 × 3.3 × 3	3.3 × 3.3 × 3	3.3 × 3.3 × 3	3.3 × 3.3 × 3
Slice gap (mm)	0.3	0.3	0.3	0.3	0.3
Matrix	64 × 64	64 × 64	64 × 64	64 × 64	64 × 64
DWI: Multi-shell single spin-echo echo-planar images; parallel imaging factor 2; no multi-band or simultaneous multi-slice; axial interleaved slices parallel to line through the pituitary gland and 4th ventricle; 73 directions interleaved 25 b1000, 24 b2000, 24 b3000, 7 b0; sampling scheme according to Caruyer et al. (Caruyer et al., 2013)					
TR (ms)	7220	7220	6310	7300	7000
TE (ms)	100	100	81	100	74
Flip angle (°)	90	90	90	90	90
# Slices	56	56	56	56	56
Voxel size (mm)	2.5 × 2.5 × 2.5	2.5 × 2.5 × 2.5	2.5 × 2.5 × 2.5	2.5 × 2.5 × 2.5	2.5 × 2.5 × 2.5
Matrix	96 × 96	96 × 96	96 × 96	96 × 96	96 × 96

Abbreviations: DWI, diffusion-weighted imaging; MPRAGE, magnetization-prepared rapid acquisition gradient-echo; T1w, T1 weighted; TE, echo time; TI, inversion time; TR, repetition time.

^avalues for TR are highly variable due to different definitions of TR for this pulse sequence.

within a time frame of 6 months, except at site 5 where only one volunteer was scanned and with a delay of 15 months due to a system upgrade and COVID-19 restrictions.

2.2 | Phantoms

Each site used the ISMRM/NIST system phantom (henceforth: geometry phantom; CaliberMRI, formerly High Precision Devices, Boulder, CO, U.S.A.), to monitor structural geometry. Each site used its own ball-shaped (sites 1, 3, 5) or cylinder-shaped (sites 2, 4) agar phantom to monitor the quality of rsfMRI and DWI.

2.3 | MRI acquisition

2.3.1 | Human participants

See Table 1 for spatial and timing parameters. 3D T1-weighted structural images were acquired according to the ADNI-3 protocol (Weiner et al., 2017), including correction of 3D geometric distortion and intensity non-uniformity. Resting-state fMRI with eyes closed was acquired for 10 min. Multi-shell DWI was acquired with 73 diffusion-weighted directions (25 b1000, 24 b2000, and 24 b3000 s/mm²) and 7 interleaved non-diffusion-weighted volumes (b0 s/mm²). We used a reduced spatial resolution version of the DWI sequence of the human

connectome project (Sotiropoulos et al., 2013) that can be acquired within 8–10 min on a clinical scanner. For both rsfMRI and DWI, scans with opposite phase-encoding directions were acquired to correct for susceptibility-induced distortions. At each site, participants were guided through the MRI session according to a standardized protocol.

2.4 | Phantoms

To monitor stability, the agar phantom was scanned twice a month at each site with a 10-min protocol (supporting methods). This quality control (QC) protocol was based on protocols commonly used in multi-site fMRI studies (Casey et al., 2018; Friedman & Glover, 2006). To monitor geometry, the geometry phantom was scanned every 2 months as described in the ISMRM/NIST phantom manual and supporting methods. We included all phantom data obtained from start of the study until May 2022, with gaps due to COVID-19 restrictions.

2.5 | Image quality metrics (IQMs)

2.5.1 | Human participants

Images were converted from DICOM format to NIfTI, and minimally processed prior to calculation of the image quality metrics (IQMs) to monitor data quality.

We used tools from FMRIB Software Library (FSL, version 6.0.1) (Smith et al., 2004), MRI quality control tool (MRIQC) (Esteban et al., 2017), and fMRIPrep (v20.2.3) (Esteban et al., 2019). For an extensive description see supporting methods.

For rsfMRI we designed an additional IQM to describe the temporal variation of WM heterogeneity, because we observed artifacts in several fMRI scans that could not be quantified by measures provided by MRIQC (Figure 1).

DWI quality was evaluated using the EDDYQC tool (Andersson & Sotiropoulos, 2016; Bastiani et al., 2019). We used the median sum-of-squared-error (SSE) from DTIFIT per b-shell, as an additional DWI IQM. IQMs used in the current investigation are briefly

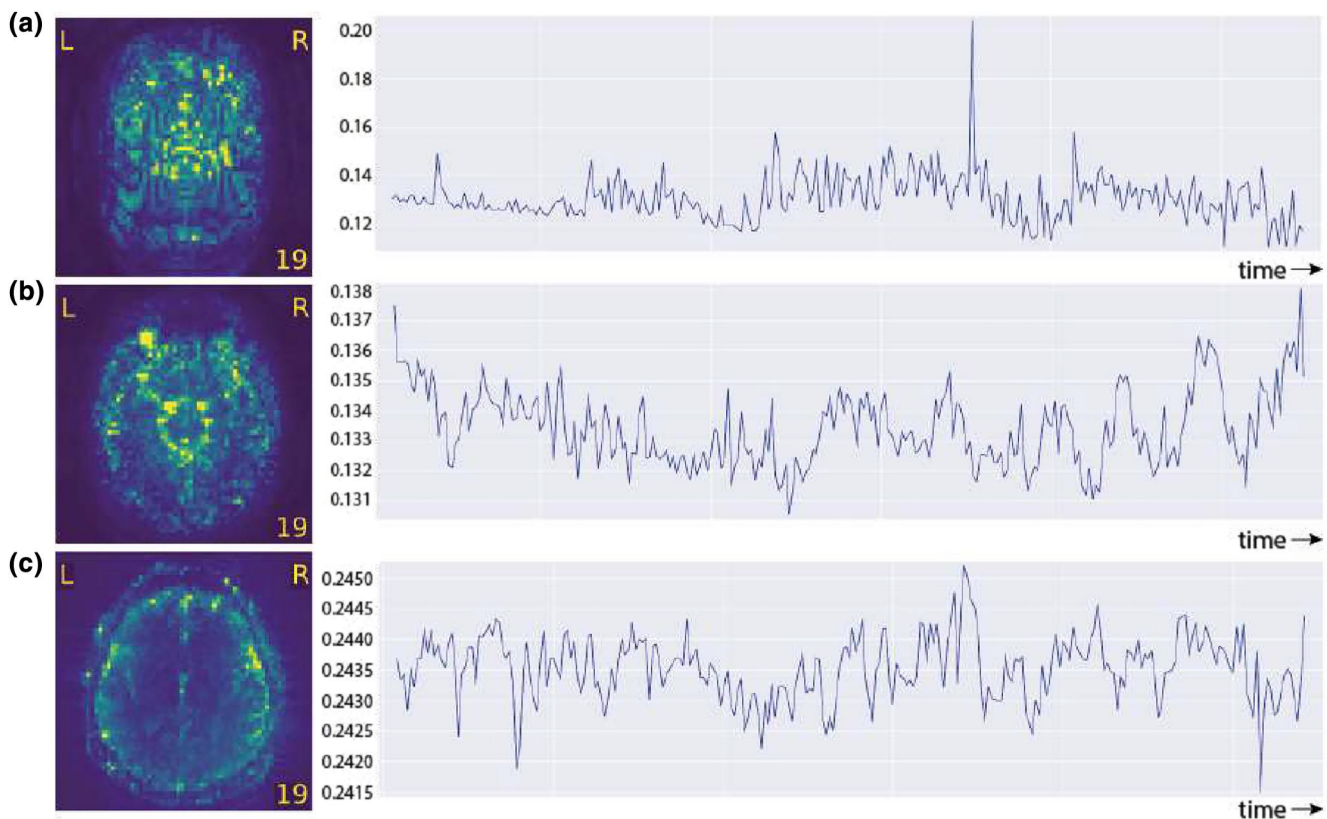


FIGURE 1 Temporal variation in WM heterogeneity. The differences in precision on the y-axis illustrate the differences between these three examples. (a) Axial slice 19 of a subject from site 1, illustrating severe artifacts on the temporal standard deviation (output from MRIQC). The corresponding time series show the temporal evolution of the SD within a WM mask. The temporal WM heterogeneity equals 7.46%. (b) a subject from site 1 with less severe artifacts, and temporal WM heterogeneity of 0.95%. (c) a subject from site 4 without artifacts, and temporal WM heterogeneity of 0.21%

described in Table 2. Calculation of motion-related IQMs and outlier detection in DWI data was performed on raw data, whereas further analysis was performed on motion-corrected data (Bastiani et al., 2019).

2.6 | Phantoms

For a detailed QC description, see supporting methods.

Because each site used its own agar phantom, the exact contents may differ between sites. Therefore, SNR values from agar phantoms are intended to evaluate intra-site stability. fMRI QC was determined based on an analysis for time-series stability (Friedman & Glover, 2006; Weisskoff, 1996). We determined static SNR (spatial mean within a center ROI), SFNR (signal-to-fluctuation-noise ratio within the ROI), ROI fluctuation, and trend. For DWI QA we checked

the consistency of mean diffusivity (MD) within a center ROI across the three b-values and the three gradient directions using the coefficient of variation (SD/mean). We determined within-site variation in MD at room temperature.

In images of the geometry phantoms, we determined the position of the 56 fiducial spheres and compared this with phantom specifications. Deviations in R-L, A-P, and S-I directions were expressed as percentage of the expected distance, and the largest deviations were used as QC parameters.

2.7 | Outcome measures

We extracted global morphometric measures from each T1-weighted image using FreeSurfer 7.1.1: volumes of whole brain (without ventricles), total cortical gray and white matter, and mean cortical

TABLE 2 Description of image quality metrics

IQM	Explanation	Modality
CJV	Coefficient of joint variation: a Measure of intensity variation between a WM and GM mask proposed by (Ganzetti et al., 2016). Higher values may be indicative of motion or intensity non-uniformity (INU) artifacts	T1
(t)SNR	Signal-to-noise ratio: The amount of real signal relative to the background noise. Because noise in the background is poorly defined with multi-channel receivers, and because some scanners completely suppress noise in the air background, SNR was calculated using the within tissue variance.	T1, fMRI
CNR	Contrast-to-noise ratio: An extension of the SNR that measures the separation of the tissue distributions of GM and WM (Magnotta et al., 2006). Diffusion angular CNR - the diffusion related variance versus the noise variance—separate per b-value.	DWI
EFC	Entropy-focus criterion: a Measure of the amount of entropy of the voxels in the image as a marker for ghosting and motion-related blurring (Atkinson et al., 1997). Lower values are better.	T1, fMRI
INU	Intensity non-uniformity: Spurious variability in voxel intensity due to imperfections in the acquisition process (Vovk et al., 2007). Location and spread of the bias field produced during the INU correction provides a measure of the data quality. Median values are reported. Values around 1.0 are better	T1
FWHM	Full-width half maximum: a Spatial distribution of the voxel intensity values in the image as a measure of blurriness. Lower is better.	T1, fMRI
DVARS	Spatial standard deviation of successive difference images: The rate of change of voxel intensity across the entire brain at each volume (Afyouni & Nichols, 2018). Lower is better.	fMRI
FD	Frame-wise displacement: An index of the amount of frame-to-frame displacement during scanning, calculated as the root mean square of the six translation and rotation parameters.	fMRI, DWI
Outliers	Percentage of total number of slices classified by EDDY as outliers due to motion-related signal drop-out. Calculated for each b-shell separately.	DWI
SSE	Sum of squared errors of the diffusion tensor fit: a measure of the accuracy of the tensor fit. The median value within an eroded WM mask is calculated separately for each b-shell. Lower values are better.	DWI
WMH	Temporal variation of WM heterogeneity: Variation within time-series of SD/median of signal intensity in WM (after linear detrending).	fMRI

Note: For more information about the image quality metrics derived from MRIqc or eddyqc, see: mriqc.readthedocs.io/en/latest/measures.html or Bastiani et al. (Bastiani et al., 2019), respectively.

thickness (Dale, Fischl, & Sereno, 1999). We determined both raw volumes and volumes normalized for estimated total intracranial volume by multiplying with 1948 ml/eTIV (<http://www.freesurfer.net/fswiki/eTIV> (accessed May 20, 2022)).

Independent component analysis (ICA) was performed to identify the spatial consistency of resting-state networks between sites. We conducted group ICA (FSL) on the denoised fMRI scans, separately for each site, excluding the traveling volunteers. We applied a low model order ICA (6 components) to identify large 'standard' intrinsic resting-state networks and prevent them from breaking down into smaller subnetworks. For each site we identified the default mode network (DMN), somatomotor network (SMN) and visual network (VN) based on their spatial likeness to the Yeo 7-Network parcellation (Yeo et al., 2011). We quantified the spatial overlap (cross-correlation) between these resting-state networks across sites and with the Yeo parcellation.

Using DTIFIT, we calculated fractional anisotropy (FA), MD, axial diffusivity (AD) and radial diffusivity (RD) for each b-shell (b1000, b2000, b3000), and compared these between sites using tract-based spatial statistics (TBSS) (Smith et al., 2006). To register the DWI scans to a common space, we used DTI-TK (Zhang et al., 2006), which utilizes the full tensor orientation information (Bach et al., 2014; Wang et al., 2011). We determined median diffusion measures of the whole brain skeleton (thresholded at FA 0.2), and of the skeleton voxels within the forceps major (through the splenium of the corpus callosum) and forceps minor (genu corpus callosum). We selected the forceps ROIs as examples of robust tracts with mostly parallel fibers. These ROIs were taken from the JHU-ICBM tracts (Hua et al., 2008) with 25% threshold.

2.8 | Statistics

We calculated means and standard deviation (SD) of IQMs and outcome measures for the 28 participants per site. For between-site comparisons, Cohen's f was computed as a standardized effect size. Conventionally, we consider $f = 0.1$, $f = 0.25$, $f = 0.4$ as small, medium, large effect, respectively. Inter- and intra-site variability of IQMs and

outcome measures were visualized using raincloud plots (Allen et al., 2019). Data from the two traveling volunteers were displayed on the same plots.

3 | RESULTS

3.1 | Human participants

Mean age of the 140 healthy participants (73 females) was 27.3 (SD = 6.0) years, with on average 15.7 (SD = 2.1) years of education. Demographics per site are shown in Table 3. Between sites, participants were well matched on age, sex and education.

3.2 | Human IQMs

Human IQMs are shown in Table 4. For most IQMs of structural MRI, within-site variation was smaller than between-site variation (Figure 2). Effect sizes were large, and between-site variation was also reflected as intra-subject variation of the traveling volunteers.

IQMs of rsfMRI were variable (Figure 3) and showed medium to large effect sizes. Temporal WM heterogeneity was low in sites 3, 4 and 5 (median values below 0.5%, with an incidental outlier corresponding to a higher value of framewise displacement). Temporal WM heterogeneity was higher in sites 1 and 2 (median 0.77% and 0.75%, respectively). Scans with very high values could also be identified in maps provided by MRIQC, displaying the temporal SD (Figure 1). Temporal WM heterogeneity was not clearly related to other human IQMs, although static SNR values were notably higher for the Philips scanners (sites 1 and 2).

Like for rsfMRI, motion-related DWI IQMs (i.e., framewise displacement) were similar across sites (Figure 4). At high b-values, lower signal intensity makes outlier detection more difficult, resulting in a lower percentage outliers and smaller effect sizes than at b1000. Effect sizes of CNR for all b-values were large and within-site vari-

TABLE 3 Demographics

site	1	2	3	4	5	Statistics
Age (years)	28.4 ± 5.2 [21-38]	26.9 ± 4.9 [19-37]	27.0 ± 6.3 [18-45]	27.3 ± 6.9 [19-42]	27.1 ± 6.8 [20-49]	H (4) = 1.90 P = 0.75
Sex (M/F)	14/14	15/13	14/14	12/16	16/12	X ² (4) = 1.03 P = 0.91
Education level (years)	15.8 ± 2.9 [7-20]	15.9 ± 1.7 [11-18]	16.0 ± 2.4 [11-21]	15.1 ± 1.9 [12-21]	16.1 ± 1.6 [13-20]	H (4) = 4.072 P = 0.396

Note: Age and education level are shown as mean ± SD and [range].

TABLE 4 Human image quality measures (IQMs) (mean \pm SD) for 28 subjects per site for T1w structural scans, rsfMRI and diffusion-weighted imaging (DWI)

T1	1	2	3	4	5	Cohen's f
CJV	0.35 \pm 0.03	0.29 \pm 0.02	0.39 \pm 0.04	0.33 \pm 0.02	0.34 \pm 0.02	1.41
EFC	0.68 \pm 0.01	0.72 \pm 0.01	0.71 \pm 0.01	0.62 \pm 0.02	0.68 \pm 0.02	2.59
FWHM	3.85 \pm 0.13	4.22 \pm 0.15	3.21 \pm 0.1	3.66 \pm 0.09	3.65 \pm 0.1	2.96
INU	11.74 \pm 0.75	14.16 \pm 0.76	10.86 \pm 0.75	10.53 \pm 0.42	12.37 \pm 0.84	1.89
SNR GM	1.06 \pm 0.03	1.04 \pm 0.03	1.03 \pm 0.05	1.19 \pm 0.05	0.96 \pm 0.05	1.82
rsfMRI						
EFC	0.44 \pm 0.03	0.43 \pm 0.03	0.49 \pm 0.03	0.50 \pm 0.03	0.48 \pm 0.03	1.40
FWHM	2.40 \pm 0.18	2.18 \pm 0.34	2.37 \pm 0.11	2.64 \pm 0.15	2.39 \pm 0.11	0.76
FD (mm)	0.15 \pm 0.09	0.12 \pm 0.04	0.13 \pm 0.05	0.14 \pm 0.07	0.14 \pm 0.07	0.20
DVARS	27.7 \pm 6.2	23.0 \pm 3.6	25.4 \pm 2.3	23.8 \pm 3.8	24.7 \pm 6.1	0.37
SNR	4.62 \pm 0.44	5.45 \pm 0.84	3.06 \pm 0.41	2.31 \pm 0.4	3.26 \pm 0.38	2.26
tSNR	51.4 \pm 13.3	62.3 \pm 9.6	48.5 \pm 8.6	60.4 \pm 7.0	47.6 \pm 10.3	0.64
WMH	1.85 \pm 2.65	0.83 \pm 0.32	0.51 \pm 0.48	0.29 \pm 0.09	0.58 \pm 0.25	0.46
DWI						
CNR b0	31.6 \pm 5.95	29.0 \pm 4.2	36.6 \pm 5.1	42.2 \pm 6.8	36.8 \pm 6.3	0.83
CNR b1000	2.27 \pm 0.57	1.93 \pm 0.32	1.88 \pm 0.33	3.30 \pm 0.96	3.27 \pm 0.96	0.96
CNR b2000	5.6 \pm 2.32	4.08 \pm 1.64	2.67 \pm 0.48	6.22 \pm 2.43	6.14 \pm 3.52	0.62
CNR b3000	3.65 \pm 0.73	2.74 \pm 0.63	2.16 \pm 0.31	4.24 \pm 1.19	5.33 \pm 2.06	1.01
Outliers b1000 (%)	1.71 \pm 0.68	1.52 \pm 0.91	0.57 \pm 0.36	1.90 \pm 0.77	1.02 \pm 0.47	0.76
Outliers b2000 (%)	0.12 \pm 0.26	0.12 \pm 0.17	0.14 \pm 0.30	0.10 \pm 0.23	0.06 \pm 0.09	0.13
Outliers b3000 (%)	0.14 \pm 0.23	0.12 \pm 0.16	0.07 \pm 0.10	0.14 \pm 0.17	0.07 \pm 0.10	0.19
SSE b1000	0.06 \pm 0.01	0.06 \pm 0.01	0.09 \pm 0.02	0.05 \pm 0.01	0.04 \pm 0.01	1.43
SSE b2000	0.18 \pm 0.02	0.21 \pm 0.04	0.29 \pm 0.06	0.17 \pm 0.02	0.14 \pm 0.03	1.44
SSE b3000	0.38 \pm 0.04	0.46 \pm 0.09	0.64 \pm 0.13	0.39 \pm 0.04	0.31 \pm 0.07	1.48
FD (mm)	0.32 \pm 0.07	0.34 \pm 0.08	0.31 \pm 0.09	0.36 \pm 0.10	0.28 \pm 0.08	0.33

Note: Effect sizes are given with Cohen's f. IQMs are described in more detail in Table 2.

Abbreviations: (t)SNR, (temporal) signal-to-noise ratio; CJV, coefficient of joint variation; CNR, contrast-to-noise ratio; DVARS, spatial standard deviation of successive difference images; EFC, entropy-focus criterion; FD, framewise displacement; FWHM, full-width half maximum; INU, intensity non-uniformity; SSE, sum-of-squared error; WMH, temporal variation of WM heterogeneity.

ation was larger in sites with a high median CNR. High values of CNR were partly reflected by lower SSE values.

3.3 | Human outcome measures

Human outcome measures are shown in Table 5. Effect sizes for all metrics were still medium to large, but were markedly smaller than for the IQMs.

Normalized volumes were more comparable within sites than raw volumes (Figure 5), but effect sizes were similar. Importantly, normalization based on eTIV may introduce additional variability (see traveling volunteer B). Further, eTIV was initially underestimated for

7 participants, and overestimated in one; this could be solved (see Supporting Methods). Cortical thickness showed some variation between sites. Between-site variation in raw volumes of the traveling volunteers was lower than within- or between-site variation of the 140 healthy participants.

Spatial correlations on the fMRI-derived resting-state networks showed high consistency across sites 3, 4, and 5, while overlap between sites 1 and 2 and other sites was slightly lower (Figure 6). This might reflect differences in fMRI IQMs between sites. Generally, overlap with the Yeo parcellation was smaller.

Diffusion measures are shown in Figure 7, Figure 8, and Figure 9. For each b-value separately, we observed only a small variation of MD values within and between sites, although effect sizes were still

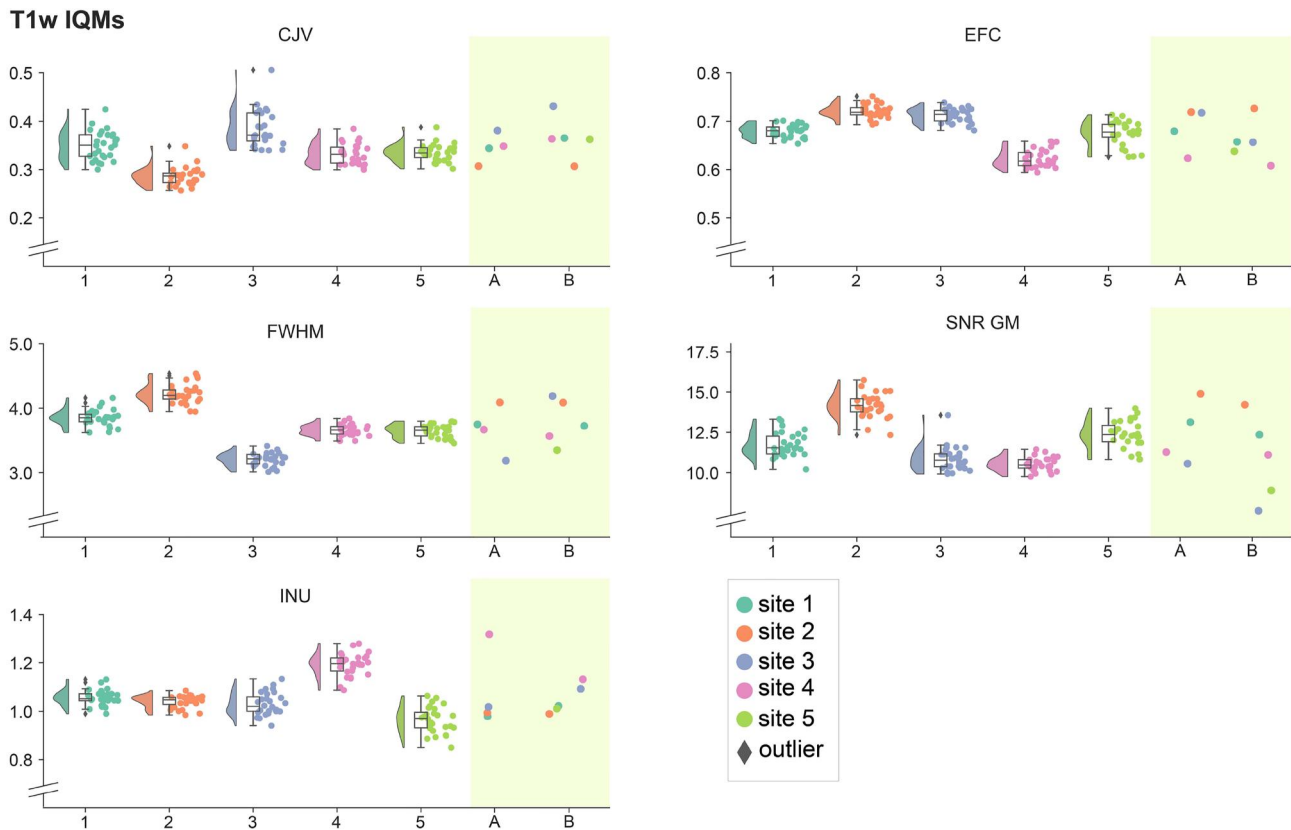


FIGURE 2 Image Quality Metrics (IQMs) of structural T1w MRI. The variation per site is indicated by the raincloud plots. The traveling volunteers A and B are indicated with the color corresponding to each of the sites. CJV, coefficient of joint variation; EFC, entropy focus criterion; FWHM, full width at half maximum; SNR GM, signal-to-noise ratio of gray matter; INU, intensity non-uniformity

large. Effect sizes for ROIs (skeletonized Forceps Major and Forceps Minor, indicated in Table 5) were smaller than for the full WM skeleton. MD, AD, and RD values decreased with increasing b-value, as expected for non-Gaussian diffusion. The traveling volunteers showed little between-site variation in diffusion measures.

3.4 | Phantoms

Deviations in the geometry phantom varied per site and direction, but remained below 0.50%, and were systematic, as shown by small SD's, although small distinct changes in deviations at site 3 and 5 could be explained by gradient re-calibration, and a software upgrade, respectively (Table 6).

Within sites, normalized SNR values obtained from the agar phantom were stable over time, with COV between 2.8% and 13.9%. We detected outliers in SNR at site 2, due to a temporary change of head coil (that needed repair), and an SNR increase at site 5, due to an upgrade in reconstruction software (Figure S1).

Compared to these standard NEMA SNR, for most sites fMRI IQMs of the agar phantom yielded higher COVs for SNR, between 4.5% and 12.4%. FFNR, the ratio between SNR and SFNR, had mean values per site between 0.99 and 1.08. Fluctuation was

below 0.1% for 4 sites, and 0.34% for site 1. Whether this relates to the larger temporal WM heterogeneity in human IQMs needs future investigation. Mean trend varied between 0.39% and 1.07%, with relatively large SD, showing variation both within and between sites.

DWI IQMs of the agar phantom showed variation of MD over time (Figure S2). Fluctuations were present in all sites, likely due to temperature fluctuations between sessions, and typically random. Only in site 2 we observed a sudden jump in May 2019, coinciding with a software update of the scanner. Within sessions, MD was comparable for the three b-values, with mean within-session COV between 0.34% and 1.13% (Table 6). We noticed a small systematic effect of gradient direction, which differed by site (Figure S2), with mean within-session COV between directions between 1.11% and 1.36%.

4 | DISCUSSION

We described the image quality monitoring pipeline of our global multi-center multi-vendor study on OCD (Simpson et al., 2020). We showed the importance of collecting regular phantom scans to monitor stability, and noticed relationships between some IQMs of

rsfMRI IQMs

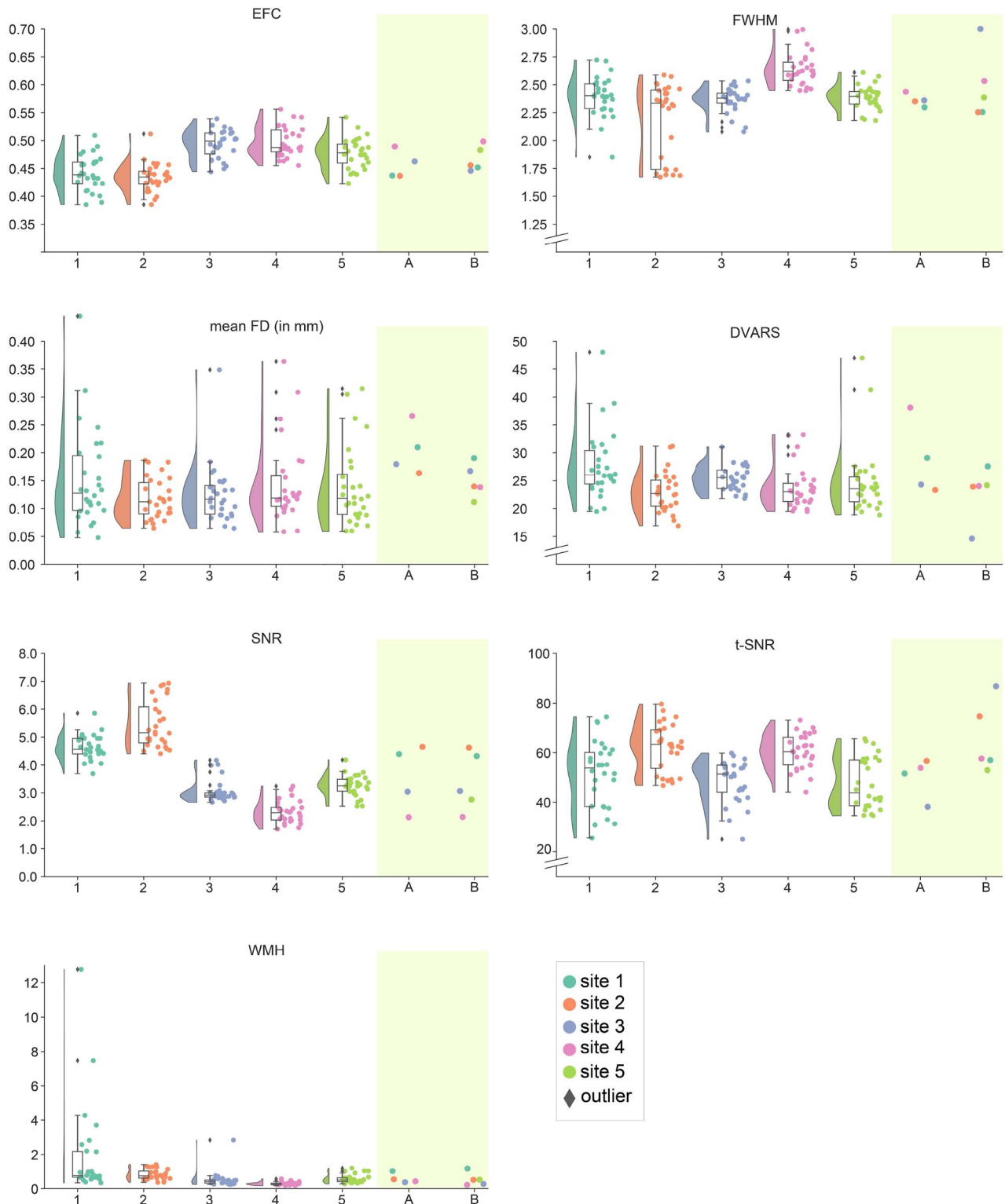


FIGURE 3 Image Quality Metrics (IQMs) of resting-state functional MRI (fMRI). The variation per site is indicated by the raincloud plots. The traveling volunteers A and B are indicated with the color corresponding to each of the sites. DVARS, rate of change of voxel intensity across the entire brain at each volume (t-); EFC, entropy focus criterion; FD, framewise displacement; FWHM, full width at half maximum; SNR, (temporal) signal-to-noise ratio; WMH, temporal variation in WM heterogeneity expressed in %

DWI IQMs

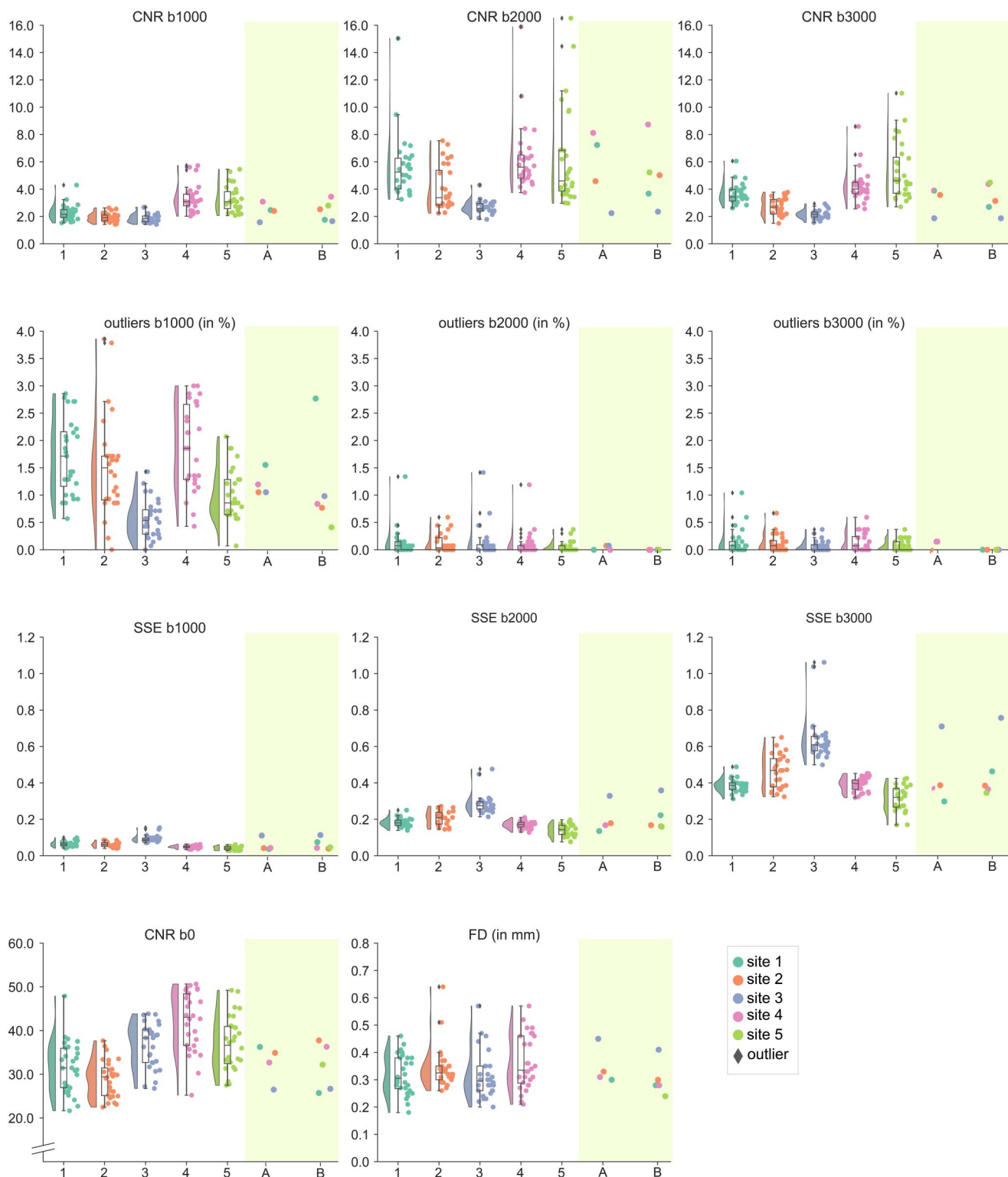


FIGURE 4 Image Quality Metrics (IQMs) of diffusion-weighted imaging (DWI). The variation per site is indicated by the raincloud plots. The traveling volunteers A and B are indicated with the color corresponding to each of the sites. CNR: contrast-to-noise-ratio for b0 and per b-shell, SSE: sum-of-squared error of diffusion tensor fit, FD: framewise displacement. For site 5 FD was not determined, because volumes consist of slices that are acquired at different times. On this scanner the design of the DWI sequence used maximal interleaving of b-values for subsequent slices, which were reconstructed by the scanner into volumes with one b-value

TABLE 5 Human outcome measures (mean \pm SD) for 28 subjects per site for T1w structural scans and diffusion-weighted imaging (DWI)

T1	1	2	3	4	5	Cohen's f
Brain volume (ml)	1134 \pm 108	1110 \pm 114	1204 \pm 122	1149 \pm 124	1154 \pm 125	0.32
GM volume (ml)	649 \pm 58	647 \pm 60	708 \pm 57	655 \pm 66	678 \pm 66	0.47
WM volume (ml)	458 \pm 55	438 \pm 57	470 \pm 68	467 \pm 60	449 \pm 60	0.23
NBV (ml)	1479 \pm 59	1518 \pm 59	1469 \pm 41	1559 \pm 63	1488 \pm 55	0.47
Norm. GM (ml)	847 \pm 47	885 \pm 40	866 \pm 36	889 \pm 42	876 \pm 35	0.38
Norm. WM (ml)	596 \pm 33	597 \pm 39	571 \pm 38	633 \pm 37	577 \pm 37	0.52
Cort. Thickness (mm)	2.526 \pm 0.066	2.582 \pm 0.103	2.509 \pm 0.075	2.471 \pm 0.07	2.541 \pm 0.069	0.49
DWI WM skeleton						
FA b1000	0.509 \pm 0.018	0.469 \pm 0.018	0.520 \pm 0.017	0.500 \pm 0.02	0.488 \pm 0.015	1.04
FA b2000	0.497 \pm 0.016	0.472 \pm 0.016	0.507 \pm 0.015	0.489 \pm 0.019	0.479 \pm 0.012	0.84
FA b3000	0.477 \pm 0.015	0.46 \pm 0.014	0.495 \pm 0.015	0.470 \pm 0.018	0.465 \pm 0.012	0.87
MD b1000 ^a	0.703 \pm 0.021	0.717 \pm 0.016	0.706 \pm 0.018	0.720 \pm 0.021	0.724 \pm 0.020	0.43
MD b2000	0.589 \pm 0.018	0.602 \pm 0.014	0.602 \pm 0.017	0.609 \pm 0.018	0.611 \pm 0.016	0.46
MD b3000	0.495 \pm 0.015	0.505 \pm 0.012	0.511 \pm 0.014	0.512 \pm 0.016	0.515 \pm 0.014	0.53
RD b1000	0.486 \pm 0.023	0.517 \pm 0.020	0.482 \pm 0.020	0.503 \pm 0.026	0.512 \pm 0.020	0.66
RD b2000	0.409 \pm 0.019	0.428 \pm 0.016	0.413 \pm 0.017	0.426 \pm 0.021	0.432 \pm 0.016	0.53
RD b3000	0.349 \pm 0.016	0.362 \pm 0.013	0.353 \pm 0.014	0.363 \pm 0.017	0.367 \pm 0.013	0.50
AD b1000	1.140 \pm 0.022	1.127 \pm 0.016	1.164 \pm 0.024	1.161 \pm 0.021	1.157 \pm 0.024	0.66
AD b2000	0.954 \pm 0.019	0.953 \pm 0.014	0.985 \pm 0.021	0.977 \pm 0.019	0.970 \pm 0.020	0.69
AD b3000	0.793 \pm 0.016	0.796 \pm 0.014	0.834 \pm 0.017	0.814 \pm 0.015	0.814 \pm 0.017	0.97
DWI forceps major skeleton						
FA b1000	0.661 \pm 0.044	0.648 \pm 0.052	0.658 \pm 0.043	0.669 \pm 0.050	0.622 \pm 0.036	0.38
FA b2000	0.663 \pm 0.041	0.664 \pm 0.045	0.654 \pm 0.043	0.675 \pm 0.046	0.621 \pm 0.033	0.47
FA b3000	0.650 \pm 0.043	0.651 \pm 0.042	0.649 \pm 0.045	0.656 \pm 0.042	0.617 \pm 0.034	0.36
MD b1000	0.753 \pm 0.032	0.744 \pm 0.026	0.75 \pm 0.028	0.777 \pm 0.029	0.781 \pm 0.024	0.56
MD b2000	0.634 \pm 0.028	0.624 \pm 0.021	0.638 \pm 0.025	0.655 \pm 0.023	0.656 \pm 0.024	0.54
MD b3000	0.532 \pm 0.023	0.524 \pm 0.018	0.54 \pm 0.021	0.549 \pm 0.018	0.551 \pm 0.022	0.52
RD b1000	0.416 \pm 0.05	0.424 \pm 0.051	0.423 \pm 0.039	0.429 \pm 0.050	0.464 \pm 0.031	0.38
RD b2000	0.352 \pm 0.038	0.349 \pm 0.038	0.361 \pm 0.033	0.361 \pm 0.036	0.390 \pm 0.024	0.44
RD b3000	0.302 \pm 0.032	0.299 \pm 0.031	0.307 \pm 0.029	0.312 \pm 0.028	0.330 \pm 0.021	0.40
AD b1000	1.455 \pm 0.058	1.419 \pm 0.048	1.463 \pm 0.055	1.516 \pm 0.066	1.468 \pm 0.065	0.56
AD b2000	1.23 \pm 0.052	1.212 \pm 0.037	1.245 \pm 0.055	1.287 \pm 0.056	1.239 \pm 0.067	0.50
AD b3000	1.004 \pm 0.035	0.99 \pm 0.028	1.045 \pm 0.045	1.045 \pm 0.037	1.034 \pm 0.053	0.61
DWI forceps minor skeleton						
FA b1000	0.555 \pm 0.029	0.507 \pm 0.031	0.573 \pm 0.029	0.552 \pm 0.034	0.541 \pm 0.027	0.75
FA b2000	0.545 \pm 0.027	0.506 \pm 0.029	0.552 \pm 0.025	0.543 \pm 0.032	0.529 \pm 0.024	0.63
FA b3000	0.526 \pm 0.026	0.489 \pm 0.027	0.537 \pm 0.026	0.523 \pm 0.033	0.511 \pm 0.025	0.63
MD b1000	0.726 \pm 0.029	0.740 \pm 0.026	0.731 \pm 0.029	0.732 \pm 0.031	0.749 \pm 0.031	0.29
MD b2000	0.606 \pm 0.023	0.620 \pm 0.020	0.628 \pm 0.022	0.623 \pm 0.025	0.631 \pm 0.026	0.37
MD b3000	0.505 \pm 0.020	0.516 \pm 0.015	0.532 \pm 0.019	0.524 \pm 0.02	0.530 \pm 0.021	0.53

(Continues)

TABLE 5 (Continued)

T1	1	2	3	4	5	Cohen's f
RD b1000	0.474 ± 0.032	0.510 ± 0.034	0.468 ± 0.031	0.482 ± 0.038	0.500 ± 0.033	0.49
RD b2000	0.399 ± 0.025	0.426 ± 0.026	0.406 ± 0.022	0.411 ± 0.029	0.422 ± 0.024	0.40
RD b3000	0.339 ± 0.020	0.361 ± 0.021	0.349 ± 0.019	0.352 ± 0.023	0.360 ± 0.019	0.40
AD b1000	1.227 ± 0.039	1.203 ± 0.031	1.274 ± 0.053	1.236 ± 0.044	1.265 ± 0.049	0.60
AD b2000	1.020 ± 0.037	1.008 ± 0.027	1.065 ± 0.041	1.041 ± 0.038	1.046 ± 0.045	0.54
AD b3000	0.842 ± 0.029	0.829 ± 0.022	0.887 ± 0.035	0.861 ± 0.031	0.862 ± 0.037	0.66

Note: Effect sizes are given with Cohen's f.

^aMD, RD and AD are given in units of 10^{-3} mm²/s.

human and phantom scans, but these need to be confirmed in the final cohort. Overall, within-site variations were smaller than between-site variations, and effect sizes of IQMs were typically larger than effect sizes of outcome measures. Thus, despite efforts to harmonize scan protocol across sites, between-site variations persist that can stem from variation between participants other than age, sex, and education or from differences between MRI scanners. Indeed, outcome measures from the traveling volunteers still showed variation, although limited. This highlights the importance of collecting data of matched control subjects at each site, which is part of this study design.

Despite protocol standardization with respect to spatial and timing parameters, between-site variation in IQMs of the human structural, rsfMRI and DWI images persisted, likely due to factors such as general scanner performance, pulse sequence implementation, and image reconstruction. For rsfMRI the cause of higher temporal WM heterogeneity in two sites is unknown. For DWI, the high between-site variation in CNR at higher b-values is unclear, as is the observation that sites with high mean CNR also showed a large within-site variation; both observations need further investigation in the final cohort. The variation in IQMs had no direct relationship, however, with outcome measures. Outcome measures of human structural scans showed similar between-site variation for raw volumes compared with normalized volumes. This was especially clear for the two traveling volunteers, who demonstrated consistent raw volumes at all sites, while normalized volumes were more variable. It has been noted that scaling with eTIV might lead to over-correction (Klasson et al., 2018), and we noticed that eTIV values may depend on whether or not the neck has been removed from the input images. Differences in cortical thickness between sites, also for the two traveling volunteers, are possibly due to subtle tissue contrast differences. DWI outcome metrics were also more comparable between sites, despite the large variation in the corresponding IQMs. Of note, motion-related IQMs will play only a minor role in DWI outcome measures, because of efficient outlier replacement (Bastiani et al., 2019). Similar as for structural outcome measures, DWI metrics of the traveling subjects were highly comparable between sites, although different from most control participants. Finally, the between-site variation in functional IQMs

might have had a small effect on the spatial consistency of the identified resting-state networks. Statistical techniques may be employed to reduce site effects further. ComBat, for example, has been shown to be an effective multi-site harmonization method for DWI (Fortin et al., 2017), fMRI (Yu et al., 2018), and structural MRI data (Fortin et al., 2018; Radua et al., 2020), without removing true effects. Although such techniques may further improve the quality of the data, our study is based on the assumption that prospective harmonization of input data remains preferable. In future analyses we will be able to compare the power of our study with studies based on retrospectively collected imaging data, such as ENIGMA. Geometry phantom IQMs showed systematic but small deviations (0.50% or < 1 mm over 180 mm distance). Relative SNR, an IQM of the agar phantom, remained stable over time for most sites, but was sensitive in detecting the temporary coil replacement at site 2, and a change in reconstruction software at site 5. The timing of this change, and of the gradient re-calibration at site 3, could be included in structural analyses of the final cohort. IQMs of the agar phantom showed between-site variation in fMRI fluctuation. Whether this relates to the larger temporal WM heterogeneity in human IQMs in two sites needs to be investigated in the future. Within- and between-site variations in fMRI trend may depend on vendor-specific implementation of stabilization, and on timing of the experiment (e.g., measuring after scanner start-up or shortly after extensive gradient use). DWI IQMs of the agar phantom showed good correspondence between MD values at different b-values, and only small systematic differences along different gradient orientations. These differences are well below the variations observed in a previous large multi-center study (Belli et al., 2016).

Limitations of this study include the fact that a traveling volunteer could not be scanned at one site due to COVID-19 restrictions. We only assessed DWI outcome measures based on multiple single-shell data, while the multi-shell dataset is valuable for tractography and analyses like NODDI or estimation of free water content (Pasternak et al., 2012; Zhang et al., 2012). We also used different agar phantoms, not necessarily from the same batch, and measured DWI at room temperature for pragmatic reasons. Only with a phantom in ice-water can the temperature be held stable between sessions and sites (Chenevert et al., 2011). Instead, we choose a short pragmatic

Global brain morphometry

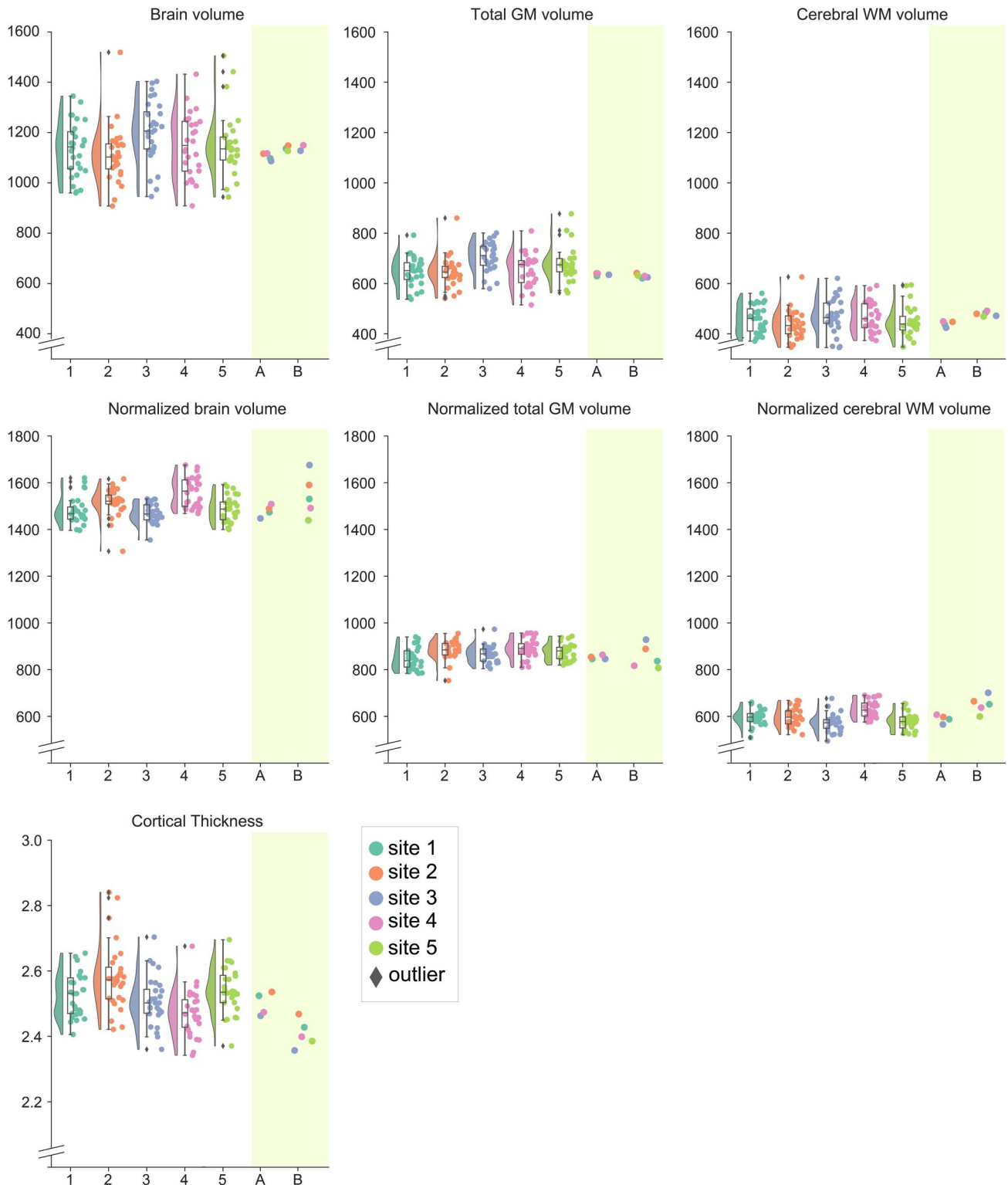


FIGURE 5 Structural outcome measures from T1w scans. The variation per site is indicated by the raincloud plots. The traveling volunteers A and B are indicated with the color corresponding to each of the sites

agar protocol, which can regularly be repeated on clinical scanners, while still providing information about MD stability at different b-values and gradient directions.

In conclusion, this paper described how we perform the image processing and monitor scan quality in our worldwide multi-center multi-vendor study that utilizes clinical scanners and an a priori

Spatial correlations RSNs

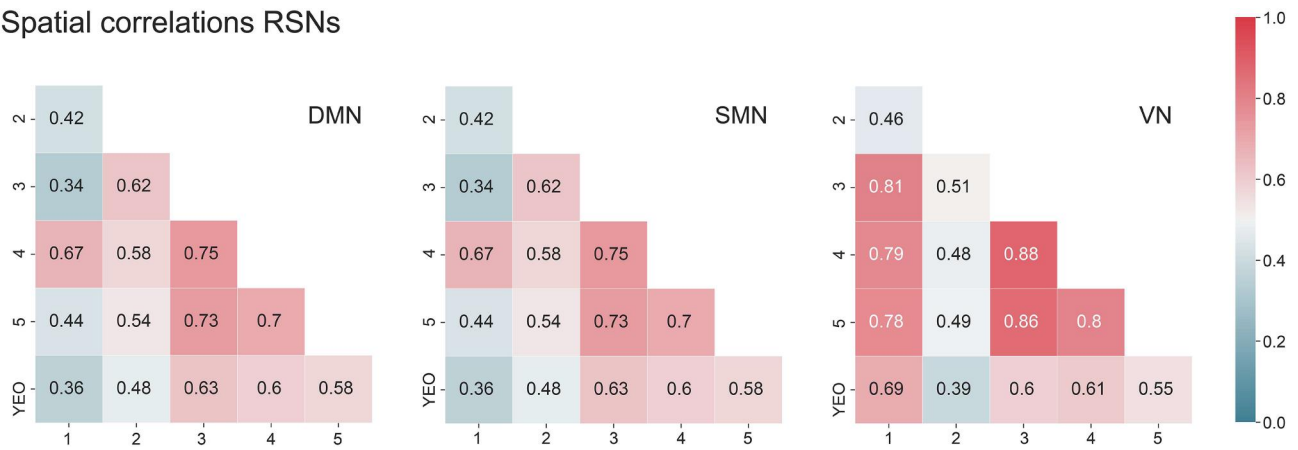


FIGURE 6 Spatial correlations between sites for three selected resting-state networks: the default mode network (DMN), the somatomotor network (SMN) and the visual network (VN). Correlations are shown between each of the five sites and between all sites and the YEO parcellation (lower row)

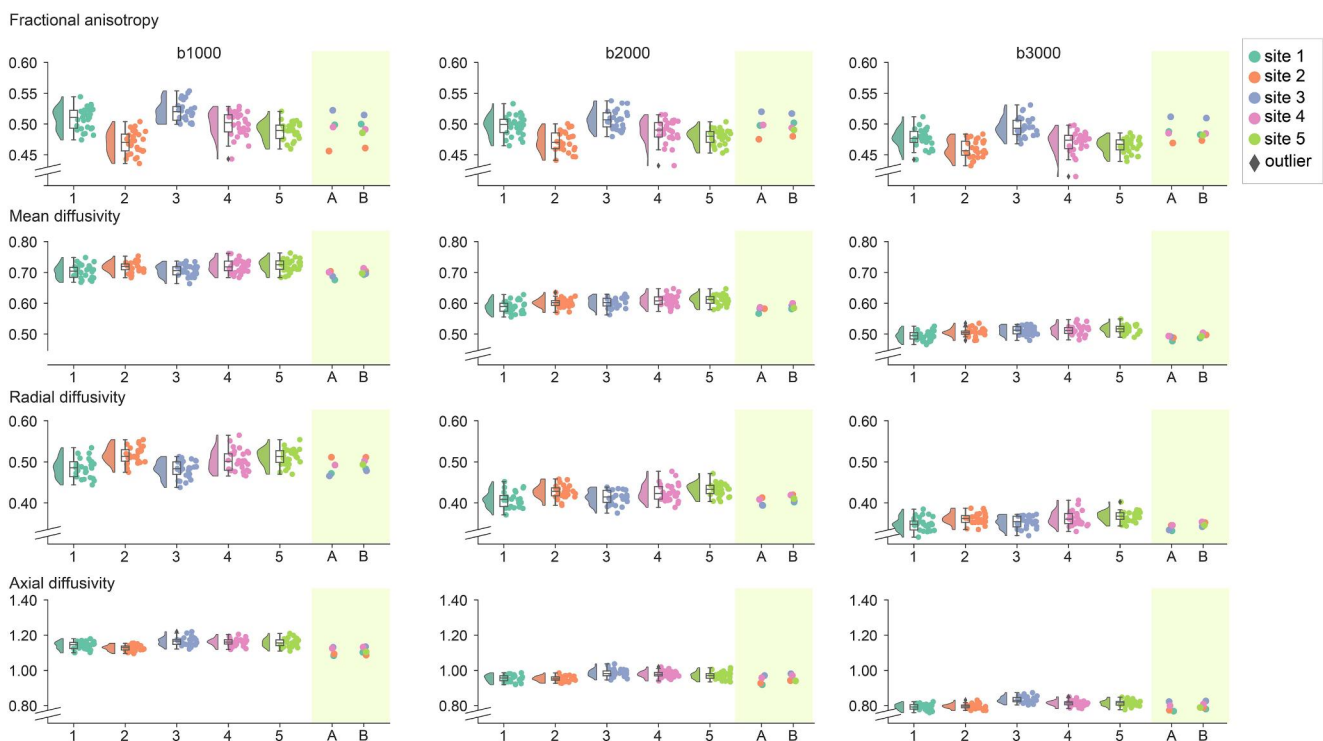


FIGURE 7 Fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), and axial diffusivity (AD) of the WM skeleton. The variation per site is indicated by the raincloud plots. The traveling volunteers A and B are indicated with the color corresponding to each of the sites. The unit of diffusivity is $10^{-3} \text{ mm}^2/\text{s}$. MD, AD, and RD values decrease with increasing b-value, as expected for non-Gaussian diffusion. For instance, MD of the skeleton (mean over all 140 subjects) decreased from 0.71 (b1000) to 0.60 (b2000) to $0.51 \times 10^{-3} \text{ mm}^2/\text{s}$ (b3000)

harmonized data acquisition protocol. The results suggest that phantom and human IQMs can be used to monitor scan quality within sites across the period of participant recruitment. Still, the field may benefit from additional benchmark studies to recommend limits for (in)sufficient data quality (similar to the cut-off for framewise displacement during rsfMRI). The between-site variability in outcome

measures emphasizes the importance of including patients and matched control subjects at each site. Although it will remain important to reduce remaining site effects in the final dataset using statistical techniques, the development of standardized acquisition and analysis protocols could help the neuroimaging field in psychiatry move toward greater reproducibility and rigor.

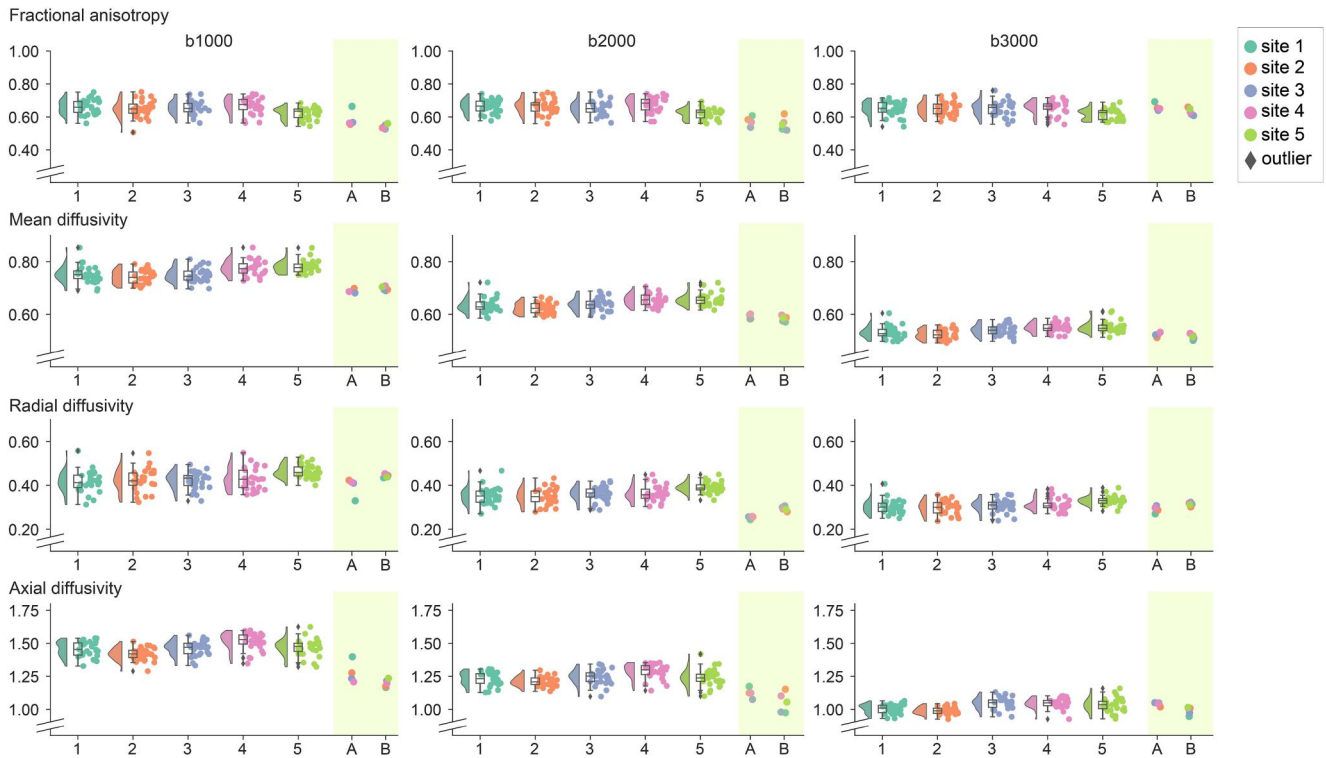


FIGURE 8 Diffusion measures of skeleton voxels within forceps major. The variation per site is indicated by the raincloud plots. The traveling volunteers A and B are indicated with the color corresponding to each of the sites. The unit of diffusivity is $10^{-3} \text{ mm}^2/\text{s}$

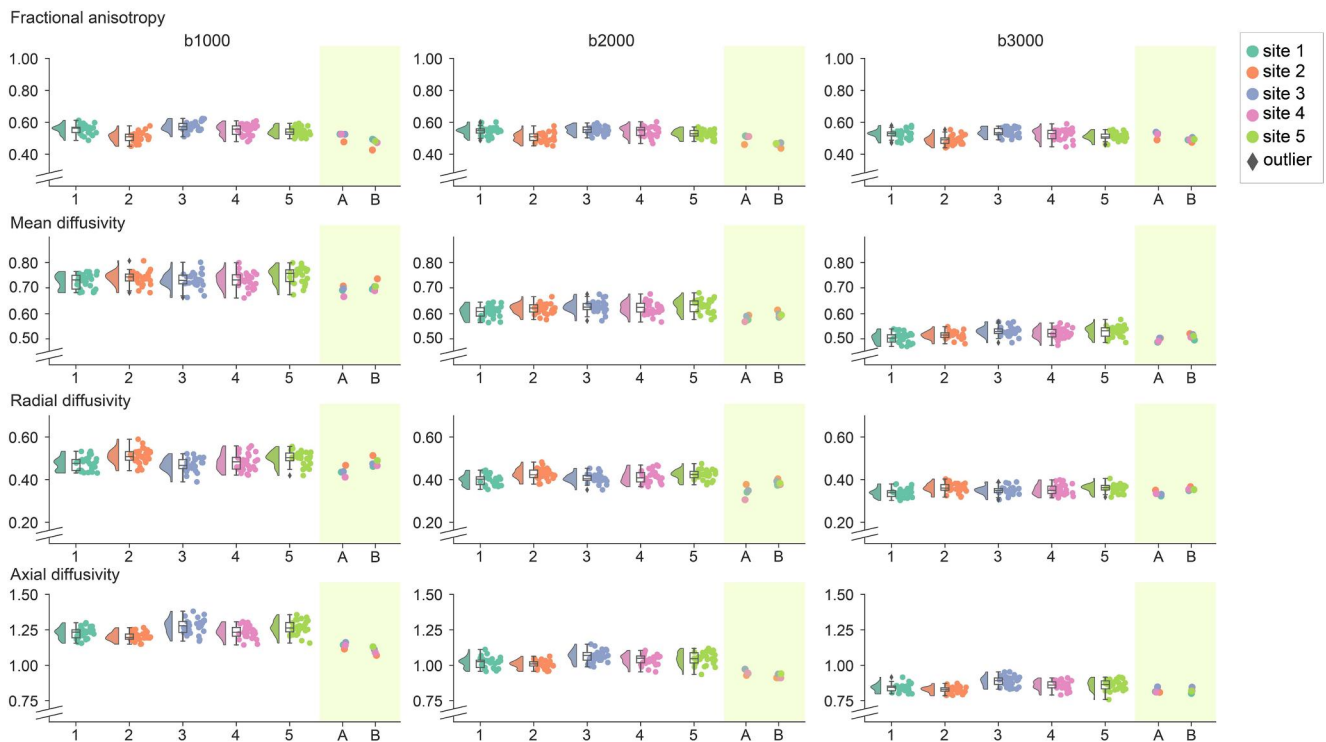


FIGURE 9 Diffusion measures of skeleton voxels within forceps minor. The variation per site is indicated by the raincloud plots. The traveling volunteers A and B are indicated with the color corresponding to each of the sites. The unit of diffusivity is $10^{-3} \text{ mm}^2/\text{s}$

TABLE 6 Image quality measures (IQMs) phantom data

Site	1	2	3	4	5
Geometry					
Deviation x (%)	-0.46 ± 0.03	-0.24 ± 0.07	-0.26 ± 0.16	-0.14 ± 0.12	-0.14 ± 0.27
Deviation y (%)	-0.50 ± 0.03	-0.33 ± 0.05	-0.01 ± 0.02	-0.23 ± 0.13	0.35 ± 0.06
Deviation z (%)	-0.32 ± 0.03	-0.23 ± 0.05	-0.45 ± 0.14	-0.17 ± 0.16	-0.13 ± 0.37
Agar SNR					
NEMA SNR COV (%)	6.5	7.0 (14.5) ^a	3.9	2.8	13.9 ^a
Agar fMRI					
ROI SNR COV (%)	11.5	10.9	4.5	12.4	7.2
ROI FFNR	0.99 ± 0.08	1.04 ± 0.07	1.02 ± 0.03	1.08 ± 0.04	1.01 ± 0.03
Trend (%)	0.87 ± 0.43	0.62 ± 0.51	0.50 ± 0.20	1.07 ± 0.45	0.39 ± 0.22
Fluctuation (%)	0.34 ± 0.06	0.10 ± 0.02	0.05 ± 0.02	0.03 ± 0.00	0.08 ± 0.01
Agar DWI					
MD COV x-y-z ^b (%)	1.36	1.29	1.28	1.25	1.11
MD COV b-shells ^b (%)	1.04	1.13	1.13	0.34	1.05

^aAt site 2 higher COV was related to temporary change of head coil, and at site 5 higher COV was related to an upgrade of reconstruction software (see Figure S1).

^bWithin-session COV of MD (mean diffusivity) between directions x-y-z, and between b-shells b400-800-1200.

ACKNOWLEDGMENTS

This paper uses data from a study funded by the National Institute of Mental Health (NIMH; R01 MH113250) that is a collaboration between five global sites (sites (Principal Investigators): Brazil (Drs. Euripides Miguel & Roseli G. Shavitt); India (Dr. Janardhan Reddy YC); Netherlands (Dr. Odile A. van den Heuvel); South Africa (Drs. Dan J. Stein & Christine Lochner); USA (Drs. Helen Blair Simpson & Melanie Wall).

In addition to NIMH funding, we acknowledge the infrastructural and imaging support provided by the New York State-Office of Mental Health at the New York site and the "Accelerator Program for Discovery in Brain disorders using Stem Cells (ADBS)" funded by the Department of Biotechnology, Government of India. Ganesan Venkatasubramanian acknowledges the support of Department of Biotechnology (DBT) - Wellcome Trust India Alliance (IA/CRC/19/1/610005) and Department of Biotechnology, Government of India (BT/HRD-NBA-NWB/38/2019-20 (6)).

We thank all study team members who worked on this study across the global sites, including: Neeltje M. Batelaan, Anish Cherian, Daniel Lucas Conceição Costa, Dianne M. Hezel, Marinês Joaquim, Martha Katechis, Roberto Lewis-Fernandez, Loche Manuel, Karen Mare, Clara Marincowitz, Maria Alice de Mathis, Gabrielle R. Messner, Rachel Middleton, Madhuri Narayan, Nienke Pannekoek, Jamila Rocha, Sarah Rose, Deise Ruiz, Petty Samuels, Yael R. Stovezky, Page Van Meter, Shivakumar Venkataram.

CONFLICT OF INTEREST

Dr. van den Heuvel received consultancy fee from Lundbeck and a stipend from Elsevier for serving as associate editor of Journal of Obsessive-Compulsive and Related Disorders.

Dr. Simpson has received research support from Biohaven Pharmaceuticals, royalties from UpToDate Inc and Cambridge University Press, and a stipend from the American Medical Association for serving as Associate Editor of JAMA Psychiatry.

Dr. Stein has received research grants and/or consultancy honoraria from Johnson & Johnson, Lundbeck, Servier, and Takeda.

Dr. Vriend is listed as an inventor on a patent licensed to General Electronic.

(WO2018115148A1).

None of the remaining authors has any disclosures.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Petra J. W. Pouwels  <https://orcid.org/0000-0001-8430-0606>

Ganesan Venkatasubramanian  <https://orcid.org/0000-0002-0949-898X>

REFERENCES

- Afyouni, S., & Nichols, T. E. (2018). Insight and inference for DVARS. *NeuroImage*, 172, 291–312. <https://doi.org/10.1016/j.neuroimage.2017.12.098>
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, 4, 63. <https://doi.org/10.12688/wellcomeopenres.15191.1>
- Andersson, J. L. R., & Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, 125, 1063–1078. <https://doi.org/10.1016/j.neuroimage.2015.10.019>
- Atkinson, D., Hill, D. L., Stoye, P. N., Summers, P. E., & Keevil, S. F. (1997). Automatic correction of motion artifacts in magnetic resonance

- images using an entropy focus criterion. *IEEE Transactions on Medical Imaging*, 16(6), 903–910. <https://doi.org/10.1109/42.650886>
- Bach, M., Laun, F. B., Leemans, A., Tax, C. M., Biessels, G. J., Stieltjes, B., & Maier-Hein, K. H. (2014). Methodological considerations on tract-based spatial statistics (TBSS). *NeuroImage*, 100, 358–369. <https://doi.org/10.1016/j.neuroimage.2014.06.021>
- Bastiani, M., Cottaar, M., Fitzgibbon, S. P., Suri, S., Alfaro-Almagro, F., Sotiropoulos, S. N., Jbabdi, S., & Andersson, J. L. R. (2019). Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *NeuroImage*, 184, 801–812. <https://doi.org/10.1016/j.neuroimage.2018.09.073>
- Belli, G., Busoni, S., Ciccarone, A., Coniglio, A., Esposito, M., Giannelli, M., Mazzoni, L. N., Nocetti, L., Sghedoni, R., Tarducci, R., Zatelli, G., Anoja, R. A., Belmonte, G., Bertolino, N., Betti, M., Biagini, C., Ciarmatori, A., Cretti, F., Fabbri, E., ... Gobbi, G. (2016). Quality assurance multicenter comparison of different MR scanners for quantitative diffusion-weighted imaging. *Journal of Magnetic Resonance Imaging*, 43(1), 213–219. <https://doi.org/10.1002/jmri.24956>
- Caruyer, E., Lenglet, C., Sapiro, G., & Deriche, R. (2013). Design of multi-shell sampling schemes with uniform coverage in diffusion MRI. *Magnetic Resonance in Medicine*, 69(6), 1534–1540. <https://doi.org/10.1002/mrm.24736>
- Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., Thomas, K. M., ... Dale, A. M. (2018). The adolescent brain cognitive development (ABCD) study: Imaging acquisition across 21 sites. *Developmental cognitive neuroscience*, 32, 43–54. <https://doi.org/10.1016/j.dcn.2018.03.001>
- Chenevert, T. L., Galbán, C. J., Ivancevic, M. K., Rohrer, S. E., Londy, F. J., Kwee, T. C., Johnson, T. D., Rehemtulla, A., & Ross, B. D. (2011). Diffusion coefficient measurement using a temperature-controlled fluid for quality control in multicenter studies. *Journal of Magnetic Resonance Imaging*, 34(4), 983–987. <https://doi.org/10.1002/jmri.22363>
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, 9(2), 179–194. <https://doi.org/10.1006/nimg.1998.0395>
- Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS One*, 12(9), e0184661. <https://doi.org/10.1371/journal.pone.0184661>
- Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, 16(1), 111–116. <https://doi.org/10.1038/s41592-018-0235-4>
- Fortin, J. P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., Adams, P., Cooper, C., Fava, M., McGrath, P. J., McInnis, M., Phillips, M. L., Trivedi, M. H., Weissman, M. M., & Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, 167, 104–120. <https://doi.org/10.1016/j.neuroimage.2017.11.024>
- Fortin, J. P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., Roalf, D. R., Satterthwaite, T. D., Gur, R. C., Gur, R. E., Schultz, R. T., Verma, R., & Shinohara, R. T. (2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, 161, 149–170. <https://doi.org/10.1016/j.neuroimage.2017.08.047>
- Friedman, L., & Glover, G. H. (2006). Report on a multicenter fMRI quality assurance protocol. *Journal of Magnetic Resonance Imaging*, 23(6), 827–839. <https://doi.org/10.1002/jmri.20583>
- Ganzetti, M., Wenderoth, N., & Mantini, D. (2016). Intensity inhomogeneity correction of structural MR images: A data-driven approach to define input algorithm parameters. *Frontiers in Neuroinformatics*, 10, 10. <https://doi.org/10.3389/fninf.2016.00010>
- Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., Calabresi, P. A., Pekar, J. J., van Zijl, P. C., & Mori, S. (2008). Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. *NeuroImage*, 39(1), 336–347. <https://doi.org/10.1016/j.neuroimage.2007.07.053>
- Klasson, N., Olsson, E., Eckerström, C., Malmgren, H., & Wallin, A. (2018). Estimated intracranial volume from FreeSurfer is biased by total brain volume. *European Radiology Experimental*, 2(1), 24. <https://doi.org/10.1186/s41747-018-0055-4>
- Magnotta, V. A., Friedman, L., & First, B. (2006). Measurement of signal-to-noise and contrast-to-noise in the fBIRN multicenter imaging study. *Journal of Digital Imaging*, 19(2), 140–147. <https://doi.org/10.1007/s10278-006-0264-x>
- Pasternak, O., Shenton, M. E., & Westin, C. F. (2012). Estimation of extracellular volume from regularized multi-shell diffusion MRI. *Medical Image Computing and Computer Assisted Intervention*, 15(Pt 2), 305–312. https://doi.org/10.1007/978-3-642-33418-4_38
- Radua, J., Vieta, E., Shinohara, R., Kochunov, P., Quidé, Y., Green, M. J., Weickert, C. S., Weickert, T., Bruggemann, J., Kircher, T., Nenadic, I., Cairns, M. J., Seal, M., Schall, U., Henskens, F., Fullerton, J. M., Mowry, B., Pantelis, C., Lenroot, R., ... van Erp, T. (2020). Increased power by harmonizing structural MRI site differences with the ComBat batch adjustment method in ENIGMA. *NeuroImage*, 218, 116956. <https://doi.org/10.1016/j.neuroimage.2020.116956>
- Simpson, H. B., van den Heuvel, O. A., Miguel, E. C., Reddy, Y. C. J., Stein, D. J., Lewis-Fernández, R., Shavitt, R. G., Lochner, C., Pouwels, P. J. W., Narayanawamy, J. C., Venkatasubramanian, G., Hezel, D. M., Vriend, C., Batistuzzo, M. C., Hoexter, M. Q., de Jooze, N. T., Costa, D. L., de Mathis, M. A., Sheshachala, K., ... Wall, M. (2020). Toward identifying reproducible brain signatures of obsessive-compulsive profiles: Rationale and methods for a new global initiative. *BMC Psychiatry*, 20(1), 68. <https://doi.org/10.1186/s12888-020-2439-2>
- Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., & Behrens, T. E. (2006). Tract-based spatial statistics: Voxel-wise analysis of multi-subject diffusion data. *NeuroImage*, 31(4), 1487–1505. <https://doi.org/10.1016/j.neuroimage.2006.02.024>
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23(Suppl 1), S208–S219. <https://doi.org/10.1016/j.neuroimage.2004.07.051>
- Sotiropoulos, S. N., Jbabdi, S., Xu, J., Andersson, J. L., Moeller, S., Auerbach, E. J., Glasser, M. F., Hernandez, M., Sapiro, G., Jenkinson, M., Feinberg, D. A., Yacoub, E., Lenglet, C., Van Essen, D. C., Ugurbil, K., & Behrens, T. E. (2013). Advances in diffusion MRI acquisition and processing in the human connectome project. *NeuroImage*, 80, 125–143. <https://doi.org/10.1016/j.neuroimage.2013.05.057>
- Stein, D. J., Costa, D. L. C., Lochner, C., Miguel, E. C., Reddy, Y. C. J., Shavitt, R. G., van den Heuvel, O. A., & Simpson, H. B. (2019). Obsessive-compulsive disorder. *Nature Reviews Disease Primers*, 5(1), 52. <https://doi.org/10.1038/s41572-019-0102-3>
- Szekér, S., & Vathy-Fogarassy, Á. (2020). Weighted nearest neighbours-based control group selection method for observational studies. *PLoS One*, 15(7), e0236531. <https://doi.org/10.1371/journal.pone.0236531>
- Thompson, P. M., Jahanshad, N., Ching, C. R. K., Salminen, L. E., Thomopoulos, S. I., Bright, J., Baune, B. T., Bertolin, S., Bralten, J., Bruin, W. B., Bulow, R., Chen, J., Chye, Y., Dannlowski, U., de Kovel, C. G. F., Donohoe, G., Eyer, L. T., Faraone, S. V., Favre, P., ... Zelman, V. (2020). ENIGMA and global neuroscience: A decade of large-scale

- studies of the brain in health and disease across more than 40 countries. *Translational Psychiatry*, 10(1), 100. <https://doi.org/10.1038/s41398-020-0705-1>
- van den Heuvel, O. A., Boedhoe, P. S. W., Bertolin, S., Bruin, W. B., Francks, C., Ivanov, I., Jahanshad, N., Kong, X., Kwon, J. S., O'Neill, J., Paus, T., Patel, Y., Piras, F., Schmaal, L., Soriano-Mas, C., Spalletta, G., Wingen, G. A., Yun, J., Vriend, C., ... Stein, D. J. (2020). An overview of the first 5 years of the ENIGMA obsessive-compulsive disorder working group: The power of worldwide collaboration. *Human Brain Mapping*, 43(1), 23–36. <https://doi.org/10.1002/hbm.24972>
- Vovk, U., Pernus, F., & Likar, B. (2007). A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Transactions on Medical Imaging*, 26(3), 405–421. <https://doi.org/10.1109/TMI.2006.891486>
- Wang, Y., Gupta, A., Liu, Z., Zhang, H., Escobar, M. L., Gilmore, J. H., Gouttard, S., Fillard, P., Maltbie, E., Gerig, G., & Styner, M. (2011). DTI registration in atlas based fiber analysis of infantile Krabbe disease. *NeuroImage*, 55(4), 1577–1586. <https://doi.org/10.1016/j.neuroimage.2011.01.038>
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Morris, J. C., Petersen, R. C., Salazar, J., Saykin, A. J., Shaw, L. M., Toga, A. W., Trojanowski, J. Q., & Alzheimer's Disease Neuroimaging, I. (2017). The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. *Alzheimers Dement*, 13(5), 561–571. <https://doi.org/10.1016/j.jalz.2016.10.006>
- Weisskoff, R. M. (1996). Simple measurement of scanner stability for functional NMR imaging of activation in the brain. *Magnetic Resonance in Medicine*, 36(4), 643–645. <https://doi.org/10.1002/mrm.1910360422>
- Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hol-linshead, M., Roffman, J. L., Smoller, J. W., Zollei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3), 1125–1165. <https://doi.org/10.1152/jn.00338.2011>
- Yu, M., Linn, K. A., Cook, P. A., Phillips, M. L., McInnis, M., Fava, M., Trivedi, M. H., Weissman, M. M., Shinohara, R. T., & Sheline, Y. I. (2018). Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Human Brain Mapping*, 39(11), 4213–4227. <https://doi.org/10.1002/hbm.24241>
- Zhang, H., Schneider, T., Wheeler-Kingshott, C. A., & Alexander, D. C. (2012). NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage*, 61(4), 1000–1016. <https://doi.org/10.1016/j.neuroimage.2012.03.072>
- Zhang, H., Yushkevich, P. A., Alexander, D. C., & Gee, J. C. (2006). Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Medical Image Analysis*, 10(5), 764–785. <https://doi.org/10.1016/j.media.2006.06.004>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pouwels, P. J. W., Vriend, C., Liu, F., de Joode, N. T., Otaduy, M. C. G., Pastorello, B., Robertson, F. C., Venkatasubramanian, G., Ipser, J., Lee, S., Batistuzzo, M. C., Hoexter, M. Q., Lochner, C., Miguel, E. C., Narayanaswamy, J. C., Rao, R., Janardhan Reddy, Y. C., Shavitt, R. G., Sheshachala, K., ... van den Heuvel, O. A. (2023). Global multi-center and multi-modal magnetic resonance imaging study of obsessive-compulsive disorder: Harmonization and monitoring of protocols in healthy volunteers and phantoms. *International Journal of Methods in Psychiatric Research*, 32(1), e1931. <https://doi.org/10.1002/mpr.1931>