Reliability and Reproducibility of T_1-weighted images for Development Studies: comparing MPRAGE and Prospective Motion Correction with Volumetric Navigators

Lei Ai[1], R. Cameron Craddock[2], Nim Tottenham[3], Jonathan P Dyke[4], Stanley Colcombe[5], Michael Milham[1,5], Alexandre R. Franco[1,5*]

[1]Center for the Developing Brain, Child Mind Institute, New York, New York
[2]Department of Diagnostic Medicine, Dell Medical School, Austin, Texas
[3]Department of Psychology, Columbia University, New York, New York
[4] Citigroup Biomedical Imaging Center, Weill Cornell Medicine, New York, New York
[5]Center for Biomedical Imaging and Neuromodulation, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, New York
* Corresponding author: alexandre.franco@nki.rfmh.org or alexandre.franco@childmind.org

Abstract

New large neuroimaging studies, such as the ABCD and HCP Development studies are adopting a new T_1-weighted imaging sequence with volumetric navigators (vNav) in favor of the more traditional 3-Dimensional Magnetization-Prepared Rapid Gradient-Echo Imaging (MPRAGE) sequence. In this study, we used a dataset of children (ages 5-21, N=366) from the Healthy Brain Network Initiative and directly compared the MPRAGE and vNav sequences to determine if the morphometric measurements obtained from both protocols are equivalent. Reliability and reproducibility were assessed with a subset of subjects (N=61) that performed two MPRAGE and two vNav sequences within the same imaging session, with one MPRAGE (MPRAGE1) and vNav (vNav1) pair at the beginning of the session and another pair (MPRAGE2 and vNav2) at the end of the session. We directly compared the two sequences in 6 quality control indexes from the Quality Assessment Protocol (QAP) toolbox. The MPRAGE sequence showed superior scores compared to the vNav sequence in 5 out of the 6 measurements. With morphometric measurements such as volume and cortical thickness, Intraclass correlation coefficients (ICC) scores show that reliability is the highest within the vNav sequences and lowest within the MPRAGE sequences. ICC scores were also higher for the MPRAGE1-vNav1 pair at the beginning of the session then the MPRAGE1-MPRAGE2 pair, possibly due to the higher motion in the MPRAGE2 run. Results also show that the vNav sequence is robust to a high head motion population, such as children. In conclusion, morphometric measurements evaluated here for the MPRAGE and vNav sequences can be directly compared, even though the MPRAGE showed superior scores in the quality control indexes.

Introduction

New technologies are constantly being developed to improve the quality of Magnetic Resonance Images (MRI). At the same time, this presents a large challenge to longitudinal studies that might be wary of changing methods mid-study. One of the most widely used MRI pulse sequences is the T_1 weighted 3-Dimensional Magnetization-Prepared Rapid Gradient-Echo Imaging (3D MPRAGE or MPR) (Mugler and Brookeman 1990). This ultrafast gradient echo 3D pulse sequence is used by a large fraction of neuroimaging researchers with Siemens MR equipment[1] because of its excellent contrast properties and capacity to collect reliable structural images (Wonderlick et al. 2009). MPRAGE can be considered as the basic acquisition protocol for brain morphometry studies[2]. As such, large neuroimaging studies such as the WU-Minn Human Connectome Project (HCP) (David C. Van Essen et al. 2013), which is more recently referred to as the HCP Young-Adult (HCP-YA), the Rockland Sample (Nooner et al. 2012), the UK BioBank (Sudlow et al. 2015), and the Alzheimer's Disease Neuroimaging Initiative - ADNI (Jack et al. 2008), all use the MPRAGE sequence to collect structural T_1 images of the brain. There are some slight differences between sequence parameters, such as voxel size and TR/TI (see Table 1), but parameters are very similar across studies. In these large neuroimaging studies, researchers are fixated on maintaining the same sequence parameters across sites and throughout the duration of the study. Also, researchers with studies of smaller samples sizes will also prefer to use the same pulse sequences, so that their results can be compared to the larger studies. However, there are technological advancements that can occur during the data collection phase of the study such as the use prospective motion correction for structural scans (Tisdall et al. 2012) and even using multiband acquisition for fMRI scans (Feinberg et al. 2010).

Even though the concept of prospective motion correction for MRIs is not something new (Thesen et al. 2000), more recently several prospective motion correction methods have been proposed for acquiring structural and functional images (Godenschweger et al. 2016). Prospective motion correction performs an adaptive update of the MR gradients during data acquisition by using navigators that can be used as references regarding the direction of the head motion. In particular, one technique developed by Tisdall et al. is a pulse sequence based on multi-echo MPRAGE (MEPRAGE) that collects Short 3D EPI volumetric navigators (vNavs) are that are embedded in a long 3D sequence (Tisdall et al. 2012, 2016). Beyond using the EPI navigators to adjust the gradients, TRs that are detected to have a large quantity of motion are reacquired. The number of TRs that can be reacquired can be set by the operator. The vNav sequence has been widely adopted by research groups and more specifically, by large imaging studies (see Table 1). Other T_1-weighted pulse sequences with prospective motion correction are also available in other platforms, such as PROMO on GE (N. White et al. 2010).

The use of prospective motion correction pulse sequences can be used complementary to several other methods, such as training the subject in a mock scanner to get acclimated to the environment (de Bie et al. 2010), movie watching can reduce motion (Vanderwal et al. 2015;

---

[1] the equivalent of this sequence for GE machines is the 3-D Fast SPGR and for Philips is its 3D TFE

[2] MPRAGE is the recommended sequence to be used by Freesurfer

Greene et al. 2018) and other methods such as using head molds to reduce motion have also been proposed (Power et al. 2019). Methods that quickly quantify the quality of the structural images have also been proposed (T. White et al. 2018), so if necessary a structural scan can be repeated within the same session.

Head motion is a significant concern for neuroimaging studies, especially in pediatric studies (S. Y. Bookheimer 2000). Head motion has been shown to significantly reduce gray matter volume and thickness estimates (Reuter et al. 2015). Due to this concern, the longitudinal Adolescent Brain Cognitive Development (ABCD) Study (Casey et al. 2018) adopted the vNav sequence as its standard T1-weighted structural sequence. The Healthy Brain Network (HBN) study (Alexander et al. 2017) also adopted the new vNav sequence, while also maintaining the HCP style MPRAGE sequence due to concerns regarding reproducibility across sequences. The original HCP study (HCP-Young Adult) is a study that has already concluded its data acquisition, but the study now is being prolonged and continued through the HCP Lifespan Studies. The Lifespan Studies have all converted to collect structural imagen with vNav protocol. This includes the HCP Aging (HCP-A) (Susan Y. Bookheimer et al. 2019) for ages 36-100+ years old, the HCP Development (HCP-D) (Somerville et al. 2018) for ages 5-21 years old, and the Lifespan Baby Connectome Project (BCP) for children aged 0-5 years old (Howell et al. 2019). For the HCP, the impact of the change from the MPRAGE to the vNav sequence has not been fully quantified (Harms et al. 2018), in part, because few datasets contain a large enough sample size with MPRAGE and vNav in the same subjects. See Table 1 for details regarding T_1-weighted structural sequences used in large neuroimaging studies.

Study	Pulse sequence	Voxel size (mm)	Matrix Size	Num Slices	TI (ms)	TR (ms)	Scanner
Rockland Sample	MPRAGE	1.0x1.0x1.0	256x256	176	900	1900	Tim-Trio
UK Biobank	MPRAGE	1.0x1.0x1.0	256x256	208	880	2000	Siemens 3T Skyra
ADNI	MPRAGE	1.0x1.0x1.0	256x240	176	900	2300	Many
WU-Minn HCP (D. C. Van Essen et al. 2012)	MPRAGE	0.7x0.7x0.7	320x320	256	1000	2400	Custom HCP Skyra
HCP Aging	vNav	0.8x0.8x0.8	320x300	208	1000	2500	Prisma

(HCP-A)							
HCP Development (HCP-D)	vNav	0.8x0.8x0.8	320x300	208	1000	2500	Prisma
HCP Baby (BCP)	vNav	0.8x0.8x0.8	320x320	208	1060	2400	Prisma
ABCD	vNav	1.0x1.0x1.0	256x256	176	1060	2500	Prisma
Healthy Brain Network (HBN) **	MPRAGE	0.8x0.8x0.8	320x320	224	1060	2500	Prisma
	vNav	1.0x1.0x1.0	256x256	176	1060	2500	Prisma

Table 1. T_1-weighted structural imaging parameters across large imaging studies.
* Only reporting imaging sequences performed on 3T Siemens MRI's
** For the HBN study, currently only one imaging site (CBIC) is collecting two structural scans for all subjects. Data is also currently being collected at the Rutgers University Brain Imaging Center. At this site, only the MPRAGE sequence is being collected since the vNav sequence is currently unavailable for software version VB19.

Given that new large studies (i.e. HCP Lifespan and ABCD) are using the vNav sequence to collect their structural data, we raise a question; are the derivatives from the newer vNav sequence comparable to MPRAGE? Also, is there any difference in brain development measurements between the two sequences? In this study, we present quantifiable differences between the HCP style MPRAGE to the ABCD style vNav sequence.

Methods

Neuroimaging Data

All neuroimaging data used in this study were collected as part of the Healthy Brain Network (HBM) Project (Alexander et al. 2017) and were acquired on a Siemens Prisma Fit located at the Citigroup Biomedical Imaging Center (CBIC) at Weill Cornell Medicine. A total of 465 imaging sessions were analyzed. Of these 465 participants, 366 completed the full HBN protocol and are included in this study, with an age range of 5 to 21 years old (mean=11.3±3.6) which included 134 females and 331 males. The HBN protocol at CBIC includes two structural T1 weighted sequences, one based on the Human Connectome Project -YA (here referred to as the "MPRAGE" sequence), and another based on the ABCD study with prospective volumetric navigators (here referred to as the "vNav" sequence).

A subset of these participants (N=61) performed a test-retest protocol. As part of the protocol, these participants performed two MPRAGE structural scans and two vNav scans within the same imaging session. Specifically, one vNav (vNav1) and then one MPRAGE (MPRAGE1) sequence were performed at the beginning of the imaging session and the other two sequences were repeated at the end of the imaging session (MPRAGE2 and then vNav2). It was expected that there would be more head motion at the end of the imaging session. The test-retest group had an age range of 5 to 20 years old (mean=11.9±3.6) with 23 females and 38 males. The HBN protocol and timing of sequences are provided in Supplementary Tables 1 and 2.

Imaging Parameters

All imaging data reported in this study were collected on a Siemens Prisma Fit Scanner with a 32 channel head coil. The MR Protocol Guidance from the Human Connectome Project (Glasser et al. 2016) was followed to define the 3D MPRAGE HCP imaging sequence. For the structural sequences with prospective navigators (vNav), we used the protocol from the ABCD study (Casey et al. 2018). The imaging sequence protocol parameters used in this study are shown in Table 1. Additionally, for vNav, we configured the sequence with a reacquisition threshold of 0.5 (see equation [3] in (Tisdall et al. 2012)) and up to 24 TRs could be remeasured. The HCP sequence has a duration of 7 minutes and 19 seconds, while the vNav can take up to 7 minutes and 12 seconds to be acquired. During the structural runs the participants were shown the Inscapes movie (Vanderwal et al. 2015), a video developed to improve compliance related to motion and wakefulness.

Quality Control

Six measures of quality control for the structural images were performed by using the Quality Assessment Protocol (QAP) toolbox (Zarrar et al. 2015). Specifically for each subject and structural image, the Contrast to Noise Ratio (CNR, measures the mean of the gray matter intensity values minus the mean of the white matter intensity values divided by the standard deviation of the values outside the brain) (Magnotta, Friedman, and FIRST BIRN 2006), the Signal to Noise Ratio (SNR, measures the mean intensity within the gray matter divided by the standard deviation of the values outside the brain)(Magnotta, Friedman, and FIRST BIRN 2006), the Foreground to Background Energy Ratio (FBER, measures the variance of voxels inside the brain divided by the variance of voxels outside the brain), the Percent Artifact Voxels (Qi1, measures the proportion of voxels outside the brain with artifacts to the total number of voxels outside the brain) (Mortamet et al. 2009), the Smoothness of Voxels (FWHM, measures the full-width half maximum of the spatial distribution of the image intensity values in voxel units) (Magnotta, Friedman, and FIRST BIRN 2006), and the Entropy Focus Criterion (EFC, measures the Shannon entropy of voxel intensities proportional to the maximum possible entropy for a similarly sized image; Indicates ghosting and head motion-induced blurring) (Atkinson et al. 1997) were calculated. For the CNR, SNR, and FBER, a higher score means a better image, while for the FWHM, Qi1, and EFC, a lower score is better.

Motion estimation

With the vNav pulse sequence, it is possible to obtain EPI navigators. These EPI volumes with relatively low resolution (8 mm x 8 mm x 8 mm) are used as navigators to estimate the head

motion during the scan. At each TR an EPI volume is collected. The total number of volumes are acquired at each vNav sequence range from 143 to 168, depending on the number of TRs that need to be required based on subject motion (Tisdall et al. 2012). A framewise displacement (FD) (Jenkinson et al. 2002) was calculated with these EPIs. The FD was then converted to FD per minute (FDpm) by:

$$FD_{pm} = \frac{\sum_{i=2}^{N} FD(i-1,i)}{N \cdot TR} \cdot 60$$

where *N* is the number of TRs, *TR* is the repetition time in seconds, *i* is the volume, and *FD(i-1,i)* is the FD between two subsequent volumes.

With the MPRAGE sequence, we cannot directly estimate motion, hence we investigated if the average motion across all functional scans can be used as an estimate of how much a participant moves during a structural scan. Specifically, the FDpm for all functional MRI scans of the protocol were also calculated.

Structural Quantitative Measurements

We extracted morphometry measurements from the images using Mindboggle (Klein et al. 2017). For each Freesurfer label, measurements include volume, area, median travel and geodesic depth, and the median measurement of Freesurfer's cortical thickness, curvature, and convexity of the sulcus (Fischl 2012). Geodesic depth is the shortest distance along the surface of the brain from the point to where the brain surface makes contact with the outer reference surface (Klein et al. 2017), whereas travel depth is the shortest distance from a point to the outer reference surface without penetrating any surface (Giard et al. 2011; Klein et al. 2017). Intraclass correlation coefficients (ICC) was used to calculate the reproducibility and test-retest reliability of these measures (Shrout and Fleiss 1979). Reproducibility is measured between the different imaging sequence, MPRAGE, and vNav. Test-retest reliability is measured across different imaging runs for the same pulse sequence.

The gray matter volume across different structural runs was measured with FSL's SIENAX (Smith et al. 2002) package.

Intensity Bias Normalization

Two T_1 weighted images are generated when collecting vNav images, one without and one with intensity bias normalization. Both these images were tested for quality control. The image with the intensity bias correction is referred to as "vNavNorm", while "vNav" is without the correction.

Age-Related Changes

We estimated age-related changes to compare the two imaging sequences. Age-related curves were separated by sex and by the quantity of motion during the functional sequences.

Results

Quality Control Measurements

There were some slight key differences in image intensity (Figure 1) and image quality (Figure 2) when comparing the two sequences for the test-retest group. From Figure 1 it can be observed that there were large intensity inhomogeneities for the vNav images that were not normalized. The vNavNorm images in the bottom row were the same images as the vNav in the middle row but with the intensity normalization applied. Through visual inspection of the structural images, we identified two participants, one with a low amount of motion (non-mover) and one with a high amount of motion (mover). For the participant data on the left (non-mover), visually, the data quality appears to be of excellent quality. This subject has a low amount of motion during the data collection of the vNav sequence (FDpm =6.04). For the functional scans of the protocol, the same participant has a low FDpm=7.24. The images on the right represent a participant with a large amount of motion for the MPRAGE and vNav sequence. Even though there was a large amount of head motion during the acquisition in the vNav sequence (FDminute = 62.2), the quality of the T1's is still acceptable. That is not the case for the MPRAGE images, where the ringing artifacts are strikingly pronounced.

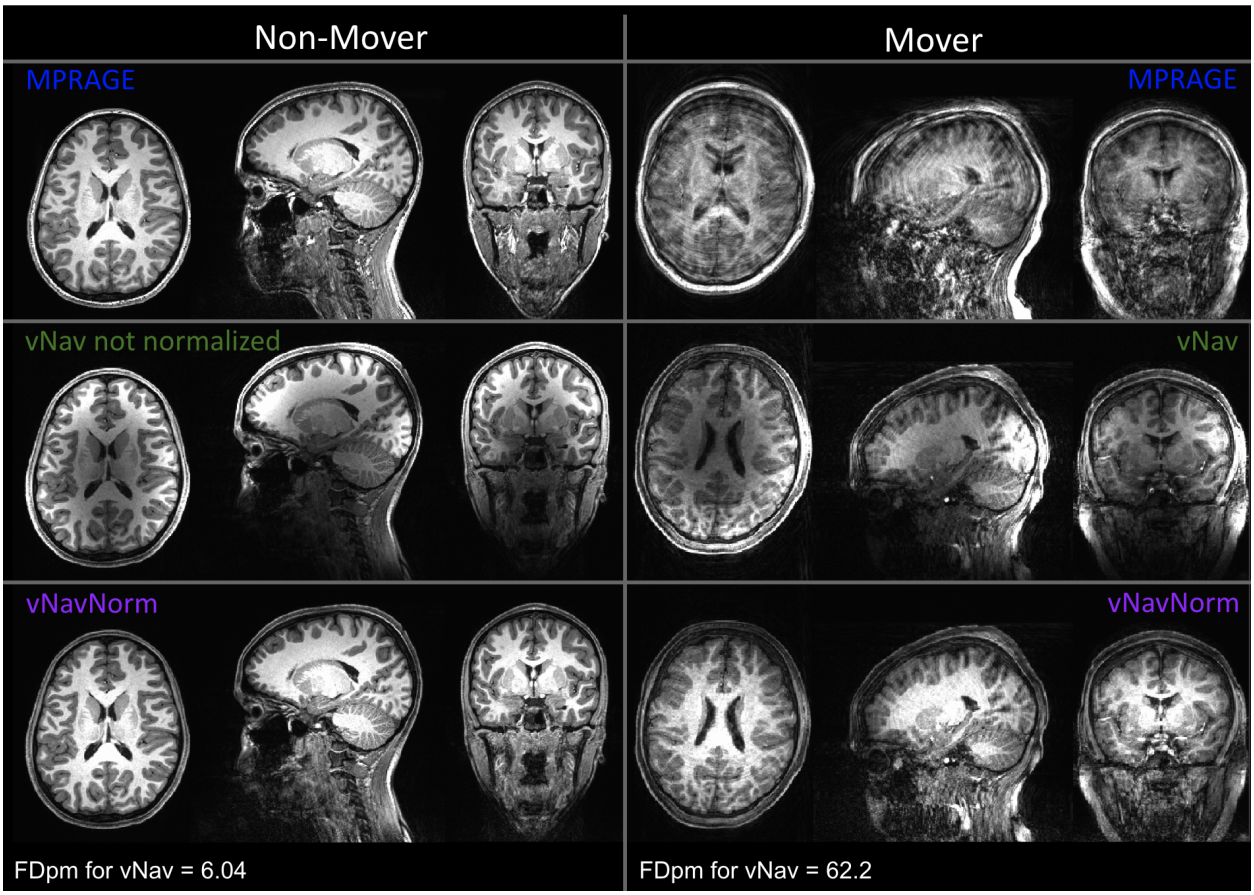


Figure 1. T_1 structural images for the two sequences MPRAGE and vNav. The top row shows the MPRAGE sequence, while the middle and bottom rows show the two images that were generated from the vNav sequence, one without (vNav) and one with (vNavNorm) intensity normalization applied. Columns represent two different participants, one with minimal head motion (left) and another with a large quantity of motion (right).

Figure 2 shows the quality control metrics from QAP for the 366 participants. Quality control metrics for the test-retest group are shown in Supplementary Figure S1. One-way repeated measures ANOVAs were calculated to compare the metrics across the three structural scans for each QAP measure. Results were significant ($p<0.05$) for all measures. Results from follow up paired t-tests are shown in Table 2. When comparing vNav and vNavNorm, vNavNorm was significantly better ($p<0.05$) in SNR and FBER, which are metrics that compare the signal within the brain versus the signal outside the brain. For CNR, the score for vNav was higher ($p<0.05$) then vNavNorm. VNav was also better then vNavNorm for QI1, FWHM and EFC (lower score is better). However, given that CNR, SNR, and FBER are the quality control measures of most importance, and by visual inspection (Figure 1), the remaining results presented in this manuscript will only be with the intensity normalized images (vNavNorm). When comparing MPRAGE and vNavNorm, vNavNorm had a better score for CNR.

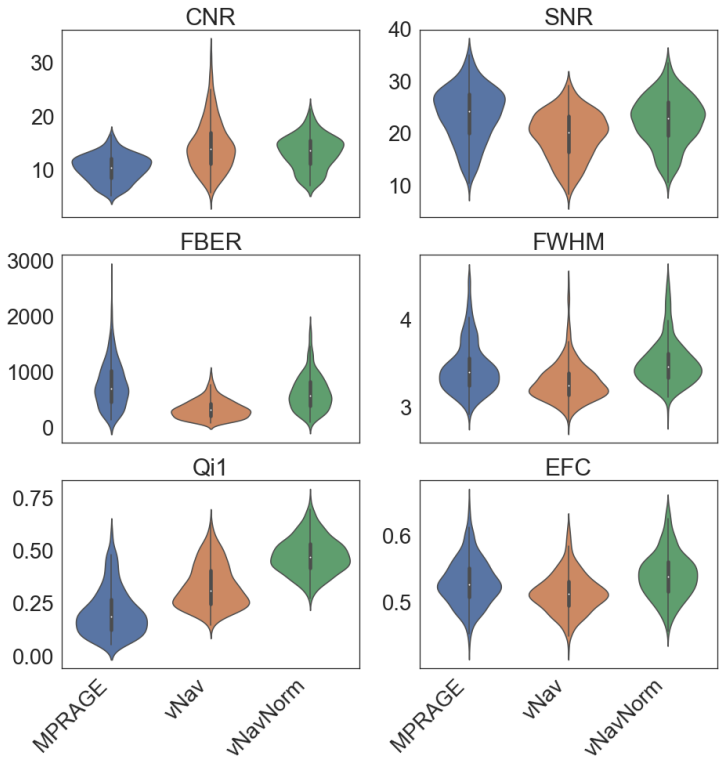


Figure 2. QAP metrics for the MPRAGE, vNav and vNavNorm images across 465 participants. Quality control metrics include; Contrast to Noise Ratio (CNR), Signal to Noise Ratio (SNR), (FBER), Smoothness of Voxels (FWHM), Percent Artifact Voxels(Qi1), Entropy Focus Criterion (EFC).

QAP Measure	Contrast	t-score	P-value (uncorrected)
CNR	MPRAGE-vNav	-15.91	7.000e-43
	MPRAGE-vNavNorm	-19.67	9.436e-58
	vNav-vNavNorm	2.09	**0.039**
SNR	MPRAGE-vNav	15.17	6.496e-40
	MPRAGE-vNavNorm	4.67	4.408e-06
	vNav-vNavNorm	-2.62	**0.011**

FBER	MPRAGE-vNav	24.96	2.499e-78
	MPRAGE-vNavNorm	8.58	3.662e-16
	vNav-vNavNorm	-9.23	6.038e-14
FWHM	MPRAGE-vNav	9.48	4.810e-19
	MPRAGE-vNavNorm	-4.96	1.144e-06
	vNav-vNavNorm	-12.65	3.543e-20
QI1	MPRAGE-vNav	-20.55	2.980e-61
	MPRAGE-vNavNorm	-47.25	6.012e-150
	vNav-vNavNorm	-18.37	2.755e-29
EFC	MPRAGE-vNav	10.79	1.696e-23
	MPRAGE-vNavNorm	-9.91	1.761e-20
	vNav-vNavNorm	-10.87	5.469e-17

Table 2. Paired t-tests comparing the QAP metrics.

Gray Matter Volume

For the test-retest group, we calculated the absolute difference between the gray matter volume measured through MindBoggle and SIENAX (Figure 3). Pairwise comparisons were made for each imaging run, MPRAGE 1 and 2, and vNavNorm 1 and 2. The comparison between vNavNorm1 and vNavNorm2 in gray matter volume presented the smallest difference in gray matter volume for results from both toolboxes. When examining the results for MPRAGE1, the difference in gray matter volume for vNavNorm1 and vNavNorm2 was smaller than for the MPRAGE1 and MPRAGE2 pair. This result was possibly due to the higher motion in the MPRAGE2 run since it was collected at the end of the session when there tends to be a higher incidence of participant head motion (see section below on motion estimation). Importantly, the vNavNorm2 sequence was collected after the MPRAGE2 sequence. These results indicate that the prospective motion correction sequence is robust for measuring gray matter volume, and does not depend on when the structural sequence is performed within the session.

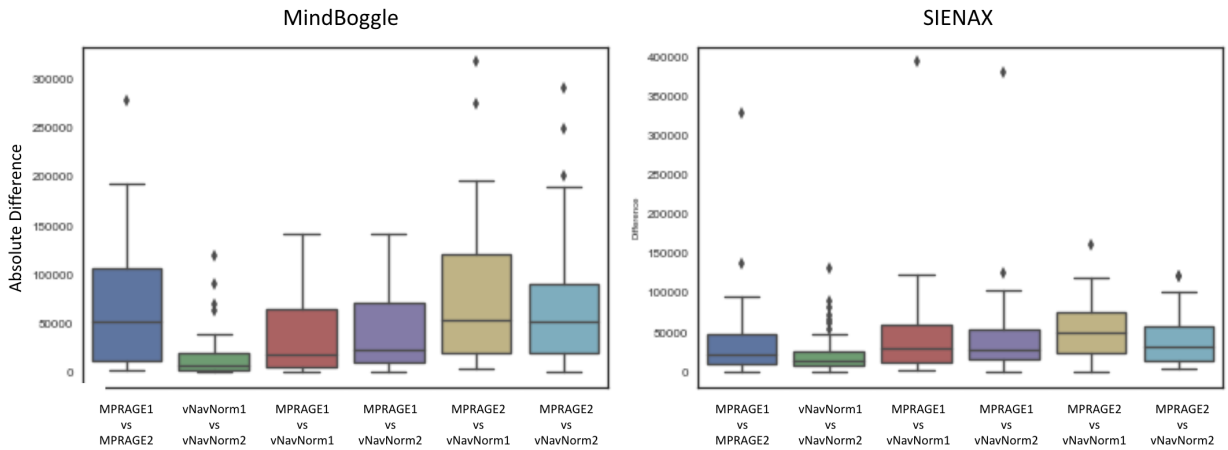


Figure 3. The absolute difference in gray matter volume within the test-retest group. Gray matter was measured using MindBoggle and SIENAX.

Motion Estimation

Since head motion cannot be directly estimated for the MPRAGE sequence, we tested if there was another means to estimate head motion during this structural scan. With the test-retest group, we calculated the FDpm for the vNav and functional sequences across the HBN full protocol. The mean FDpm across the functional runs was also calculated (See Supplementary Tables S1 and S2 and Alexander (Alexander et al. 2017) for details on the full imaging session.). We then calculated the pairwise correlation between each run and the average FDpm (Figure 4) and assessed order effects (since scans collected closer together in time should have more similar motion values). The runs are organized in the order that there were collected (Top to bottom and left to right). Runs that were adjacent to each other showed higher correlations relative to runs that were distant in time. For example, rest1 has a much higher correlation with peer1 (r=0.68) which are sequences collected one immediately after the other, than it did to movieTP (r=0.38) which occurred at the very end of the scanning session and thus farthest away in time from rest1. The exception to this order effect was rest1 and rest2, which were separated in time, with a correlation of r=0.79. Nonetheless, the average FDpm of the functional runs did show a high correlation with all the runs, with scores ranging from r=[0.54, 0.83]. The correlation scores with the two structural vNav runs were r=0.54 and r=0.63 for vNav1 and vNav2, respectively, even though they were not included in the calculation of the mean FDpm. Hence, the average FDpm of the functional runs seemed to be a good unbiased estimator of how much a participant moved during the MPRAGE structural runs.

On the leftmost column of Figure 4, the average FPpm is shown. As can be seen, there was a slight tendency for the motion to increase as the session went on. There were two runs that were repeated at the beginning and at the end of the run, vNav1-2, and peer1-3[3]. For the vNav runs, there was an increase in average FDpm from 7.51mm/min to 14.61, for vNav1 and vNav2 respectively. A large increase in motion also occurs for the peer runs, from 10.49 in peer1 to 17.09

[3] Peer (Peer Eye Estimation Regression) is a short (<2 minutes) functional run to calibrate an fMRI-based eye tracking algorithm. See (Son et al. 2019) for more details. In the initial HBN protocol, there was a peer2 run which was later dropped due to time constraints and no need to have 3 calibration runs.

in peer3. An exception was MovieTP, which was a short animated and engaging movie, which could have been the reason for lower head motion during that run.

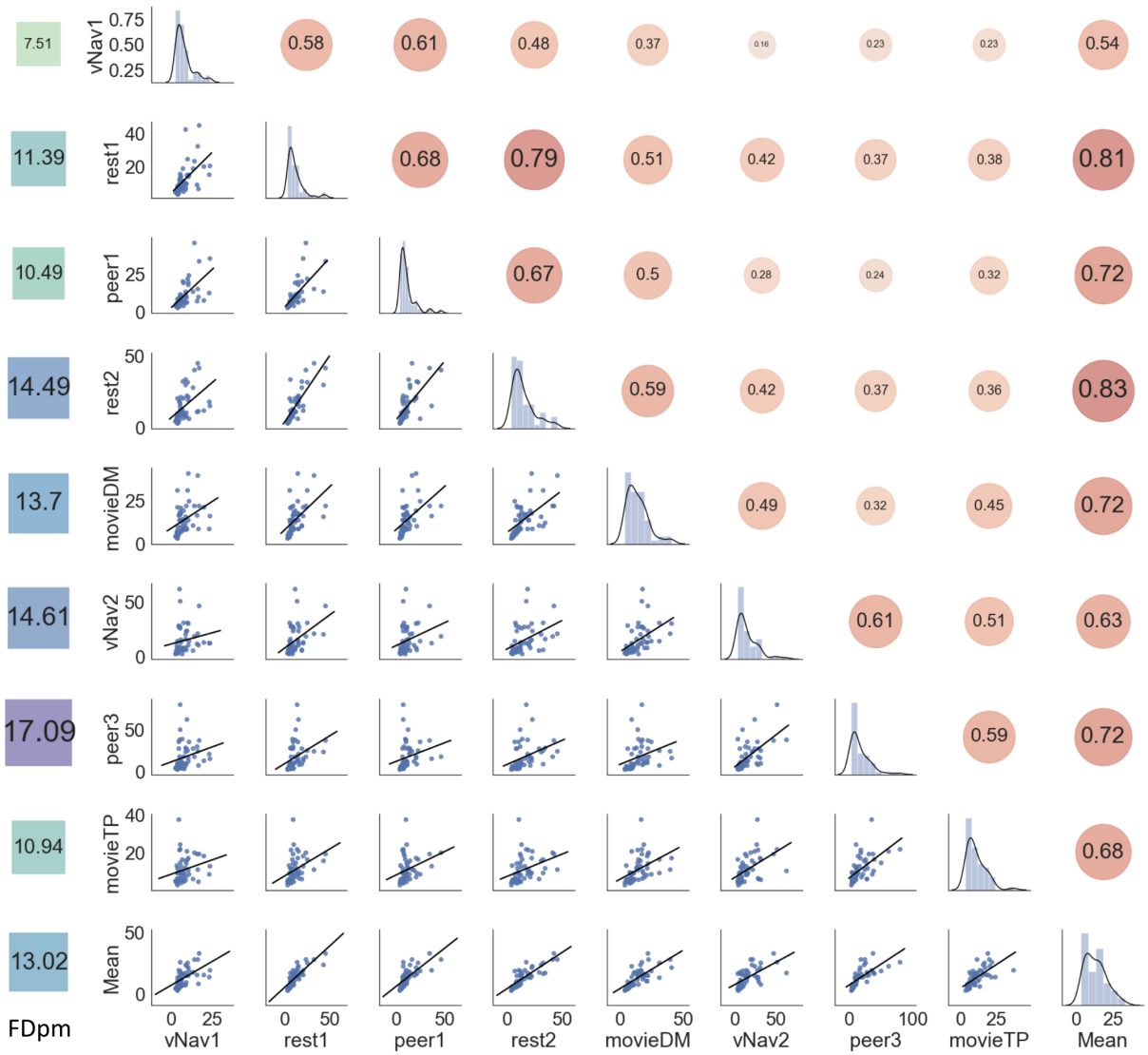


Figure 4. Correlation of the FDpm of all the runs. The main diagonal shows the distribution of FDpm each run and the average of the functional runs (Mean). The values on the left show the average FDpm for each run.

<u>Mindboggle measurements</u>

Results for the test-retest reliability and reproducibility for the Mindboggle measurements are shown in Figure 5. ICC results are shown for each of Freesurfer 62 cortical regions and for the following measurements, (1) area, (2) Freesurfer median cortical thickness, (3) travel depth, (4) geodesic distance, (5) curvature, and (6) convexity. The first row shows results for all subjects. Subjects were then divided into two groups, through a median split of the average FDpm of the functional scans. What is most noticeable for the test-retest reliability results is that for the

vNavNorm images, the ICC was not as sensitive to motion as the MPRAGE images. Another noticeable result was that for the low motion subjects, the ICC is higher for the MPRAGE1 x vNavNorm1 pair then the MPRAGE1 x MPRAGE2 pair. Again, this is possibly due to the higher motion in the MPRAGE2 runs, even though these were the subjects with lower motion estimation scores.

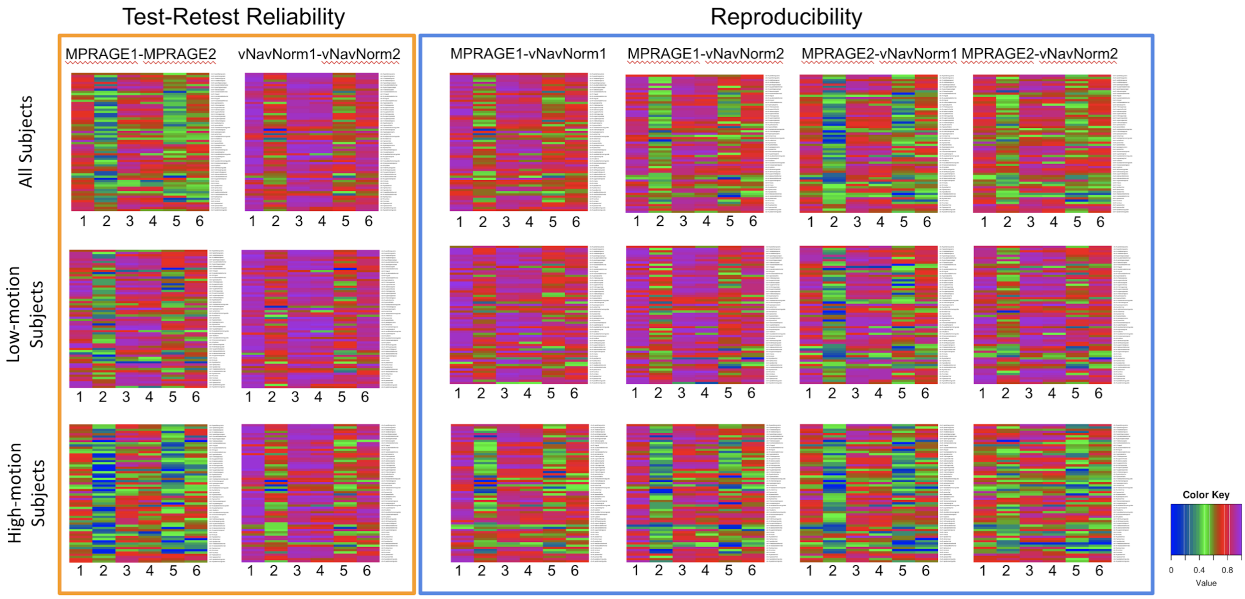


Figure 5. Test-retest reliability and reproducibility ICC results for Mindboggle measurements within each of Freesurfer 62 cortical regions. Measurements tested were (1) area, (2) Freesurfer median cortical thickness, (3) travel depth, (4) geodesic distance, (5) curvature, and (6) convexity.

Histogram plots of the area, volume, and Freesurfer median cortical thickness across all brain regions are shown in Figure 6. As can be seen in the histogram, the pair vNavNorm1 x vNavNorm2 outperformed in all other pairs (green line). The pair MPRAGE1 x vNavNorm1 typically showed the second best performance. The decrease in ICC scores between the low-motion and high-motion subjects for any pair that contains the MPRAGE2 run is highly noticeable, especially for Area. Again, this was presumably affected by the increase in motion at the end of the scan. ICC of Cortical Thickness was smaller for all pairs when compared to the ICC of Area and Volume. This shows how sensitive the measurement of Cortical Thickness is on the head motion. This results is unanticipated considering that the MPRAGE has a voxel resolution almost twice as of the vNav sequence. (voxel volume is 0.513mm^3 for MPRAGE and 1.0mm^3 for vNav).

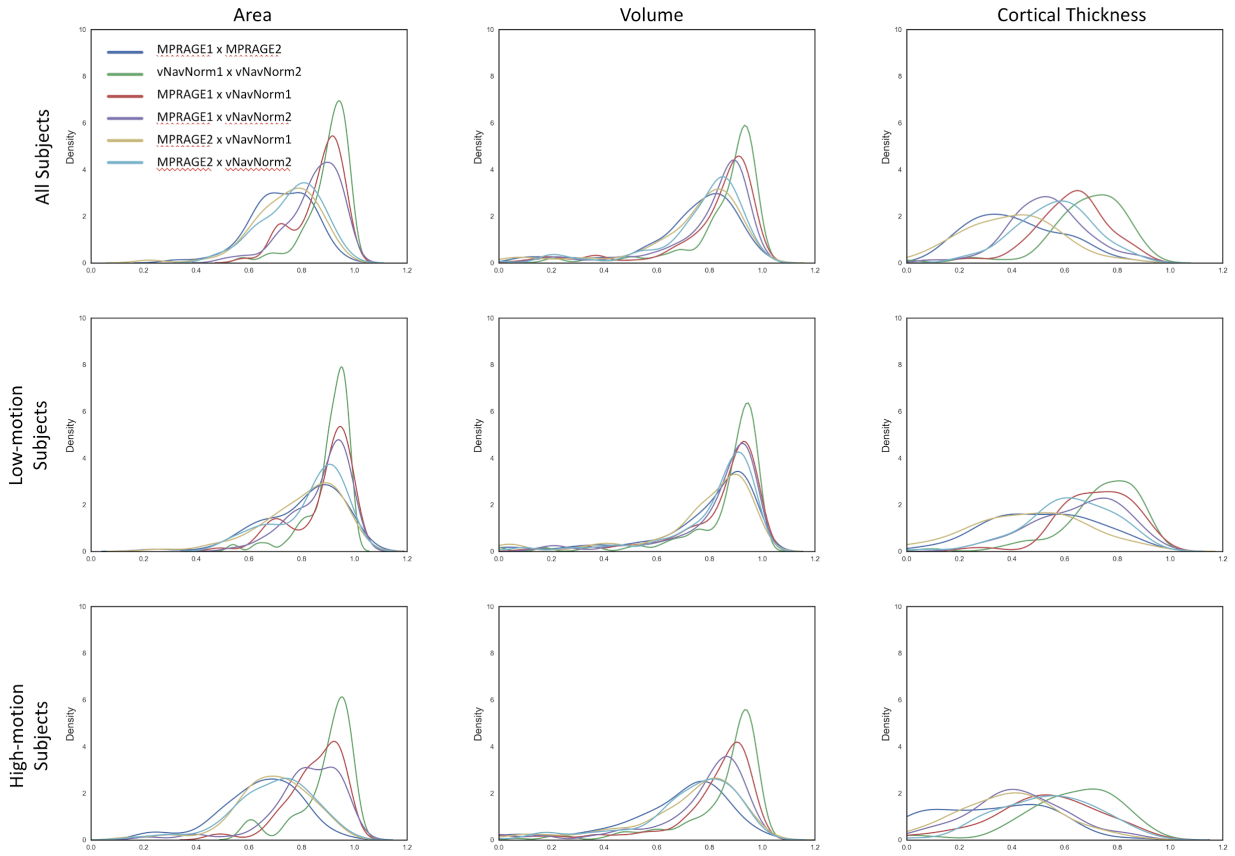


Figure 6. Histogram plots of ICC for the test-retest group that performed two MPRAGE scans and two vNav scans within the same session. ICC is calculated for Area, Volume and Cortical Thickness.

<u>Age-Related Measures</u>

Figure 7 shows development curves for Total Volume, Gray and White Matter Volume, and Ventricle Volume for male and female participants. A smaller subset of subjects (N=319) was used to calculate the development curves since there were not enough subjects older than 16 to obtain a good estimate. Hence, development curves are shown for ages 6 to 16. Black dots represent volumes calculated with the MPRAGE sequence while red dots represent the vNav sequence. For each sequence, a quadratic curve was fit for estimating development growth. As can be observed, the graphs are similar for both imaging sequences. With a median split, participants were grouped by low and high motion. Even for the subjects with large motion, the development curves are similar, just deviating a bit at the higher ages for the female group. This is possibly due to the low amount of female subjects that are older with a higher amount of motion. In Supplementary Figures S2 and S3, development curves are shown for cortical thickness measurements for males and females, respectively.

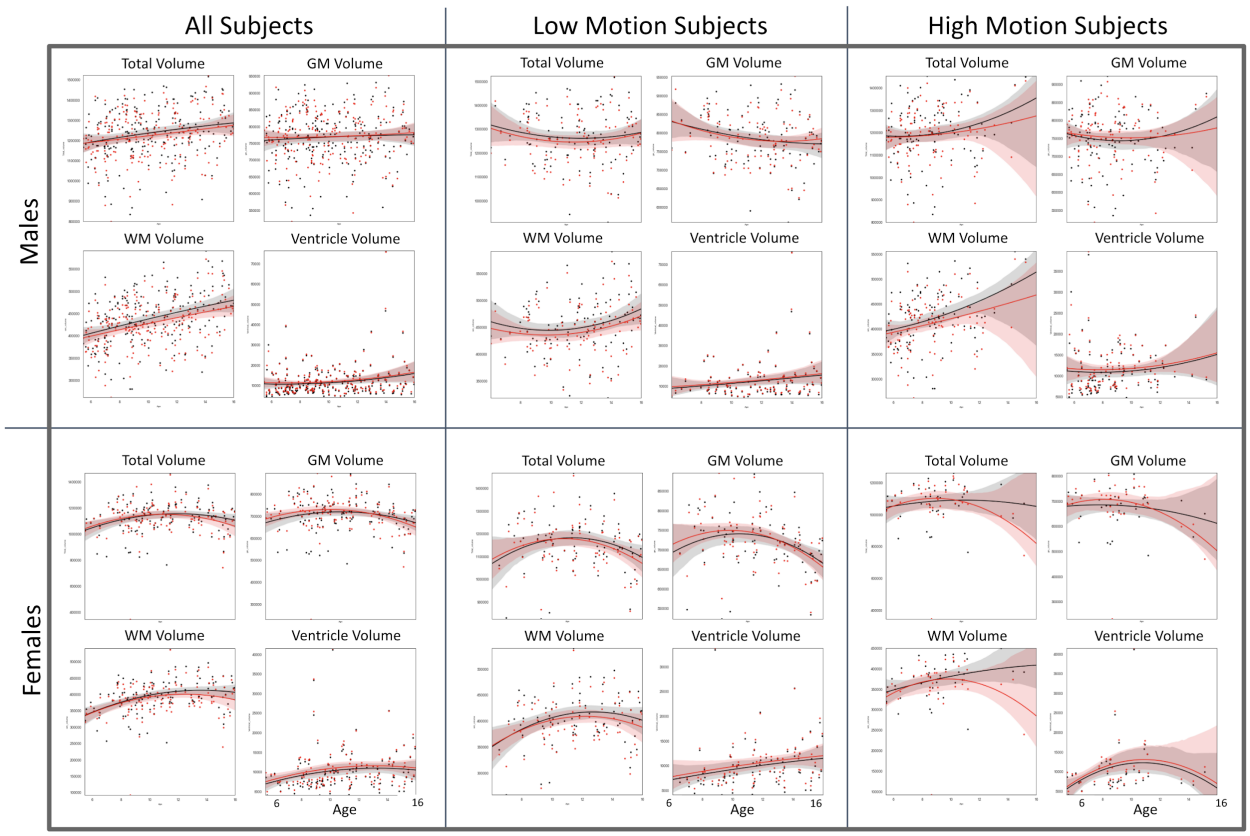


Figure 7. Black dots and lines (with 95% confidence intervals) are developmental measures for the MPRAGE sequence, while the red dots and lines are for the vNav sequence.

Discussion

Choosing the optimal pulse sequences for a study is always a challenging task for a neuroimaging researcher. Since its development in 1992(Brant-Zawadzki, Gillan, and Nitz 1992), MPRAGE has been the pulse sequence of choice by researchers collecting T_1 structural scans. This sequence, or similar sequences from other manufacturers, has been widely adopted for studies with large or small sample sizes. However, as all MRI sequences, it is susceptible to head motion which can significantly alter the quality of the morphometry measurements(Reuter et al. 2015; Pardoe, Kucharsky Hiess, and Kuzniecky 2016; Alexander-Bloch et al. 2016). More recently, the vNav sequence and its derivative vNavNorm have been shown to be more robust to head motion(Tisdall et al. 2016). This new pulse sequence is has the potential to be ideal for studies involving populations that tend to have more head motion, such as children or patients with movement disorders.

In this study, we directly compared MPRAGE and vNav pulse sequences, through a variety of morphometric measurements and quality control indexes.

Regarding quality control measures, the MPRAGE image is superior to vNavNorm in 5 of the 6 quality control indexes. These scores are possibly due to the higher resolution of the MPRAGE sequences. The only exception is CNR, for which vNavNorm has better scores. CNR is essential

for programs that are implementing automatic segmentation algorithms. This better score in CNR might help explain why the vNavNorm is showing better reliability scores.

Compared to other studies, we tested the vNav sequence in a “real world” scenario, where we were not explicitly asking the subjects to move or maintain still during the acquisition of the structural images (Tisdall et al. 2016). With the test-retest group, we have shown that the highest reliability across several morphometric measurements is between the two vNav runs. This is confirmed through the higher ICC scores and smaller difference between gray matter volume. In contrast, the lowest ICC scores are typically between the two MPRAGE runs. This is due to the higher motion of the second MPRAGE (MPRAGE2) run, which is collected towards the end of the session where participants typically move their head more. However, as can be seen in Supplementary Table S1, vNavNorm2 is collected after MPRAGE2 where even more motion should be present. But still, the reliability of the vNav sequence is not affected, showing the robustness to motion of the sequence with the prospective motion correction with volumetric navigators. This is due not only to the adaptation of the gradients to motion, but the recollection of TRs with large motion. Typically, the ICC scores that include the MPRAGE2 sequence are the lowest.

By observing ICC scores in Figure 6, it is clear that the reproducibility between MPRAGE1 and vNavNorm1 is high, with an average ICC score above 0.8 for Area and Volume and above 0.6 for cortical thickness. Those ICC scores were only lower than the ICC scores between two vNavNorm sequences which indicates reliability. Even though the voxel size in the MPRAGE and vNavNorm are different ($0.8x0.8x0.8mm^3$ vs $1.0x1.0x1.0mm^3$), there is still high reproducibility between the sequences. The results obtained in this study corroborate that the MPRAGE and vNav sequences are equivalent, i.e, morphometric results obtained with the MPRAGE sequence should be replicated with the vNav sequences, with an advantage of that vNav sequences are more robust to head motion.

The last set of tests that we performed was comparing development trajectories of the total brain, white matter, gray matter, and ventricle volumes between the two sequences. Within the full set of subjects and with the low motion subject, the trajectories are very similar. There is some larger deviation in trajectories for the participants with higher motion, but only for the subjects with higher age. This is possibly due to the lower number of subjects with high motion in the larger age group, hence larger error in trajectories. These results reinforce the idea that both sequences are comparable and that switching from the MPRAGE sequence to the vNav sequence will not compromise your results.

The only downside that we can observe for adopting the vNav sequence is the potential increase in acquisition time. However, this increase in acquisition time is mostly due to the repetition of TRs that surpass a motion threshold. If you are studying a population with high motion, on average this is actually a reduction in session time, since you have to repeat a full acquisition much fewer times. In our HBN initiative with children, we have adopted a maximum repeat of 24 TRs. If time is of the essence and the study is with a low moving population, the maximum number of repeated TRs can be reduced.

This study is limited in the sense that we do not have a direct measurement for motion during the MPRAGE sequence. However, we have attempted to estimate the motion by using the average motion across the functional runs. We have also not performed any rigorous visual inspection(Iscan et al. 2015) or post-processing quality control on segmentation measurements (Ducharme et al. 2016). We did not want to discard any data due to poor image quality, hence directly comparing the two sequences. As can be seen in Figure 1, though a visual inspection for the high motion participant, we would probably discard the MPRAGE image but not the vNav image. Another limitation of this study is that the voxel size of the sequences that we are testing have different voxel sizes. However, the objective of this paper is to compare two T_1-weighted MRI sequences that are used by a broad amount of researchers and by large imaging studies, such as the ABCD study and HCP.

Conclusions

Our results indicate that researchers should adopt the vNav sequence in their new studies, especially if there are studying populations with high levels of head motion. Morphometric results obtained from the vNav sequences are comparable to MPRAGE with low motion and better than MPRAGE with high motion.. T_1s obtained with volumetric navigators show much higher reliability compared to the MPRAGE sequence.

References

Alexander-Bloch, Aaron, Liv Clasen, Michael Stockman, Lisa Ronan, Francois Lalonde, Jay Giedd, and Armin Raznahan. 2016. “Subtle in-Scanner Motion Biases Automated Measurement of Brain Anatomy from in Vivo MRI.” *Human Brain Mapping* 37 (7): 2385–97.

Alexander, Lindsay M., Jasmine Escalera, Lei Ai, Charissa Andreotti, Karina Febre, Alexander Mangone, Natan Vega-Potler, et al. 2017. “An Open Resource for Transdiagnostic Research in Pediatric Mental Health and Learning Disorders.” *Scientific Data* 4 (December): 170181.

Atkinson, D., D. L. Hill, P. N. Stoyle, P. E. Summers, and S. F. Keevil. 1997. “Automatic Correction of Motion Artifacts in Magnetic Resonance Images Using an Entropy Focus Criterion.” *IEEE Transactions on Medical Imaging* 16 (6): 903–10.

Bie, Henrica M. A. de, Maria Boersma, Mike P. Wattjes, Sofie Adriaanse, R. Jeroen Vermeulen, Kim J. Oostrom, Jaap Huisman, Dick J. Veltman, and Henriette A. Delemarre-Van de Waal. 2010. “Preparing Children with a Mock Scanner Training Protocol Results in High Quality Structural and Functional MRI Scans.” *European Journal of Pediatrics* 169 (9): 1079–85.

Bookheimer, Susan Y., David H. Salat, Melissa Terpstra, Beau M. Ances, Deanna M. Barch, Randy L. Buckner, Gregory C. Burgess, et al. 2019. “The Lifespan Human Connectome Project in Aging: An Overview.” *NeuroImage* 185 (January): 335–48.

Bookheimer, S. Y. 2000. “Methodological Issues in Pediatric Neuroimaging.” *Mental Retardation and Developmental Disabilities Research Reviews* 6 (3): 161–65.

Brant-Zawadzki, M., G. D. Gillan, and W. R. Nitz. 1992. “MP RAGE: A Three-Dimensional, T1-Weighted, Gradient-Echo Sequence--Initial Experience in the Brain.” *Radiology* 182 (3): 769–75.

Casey, B. J., Tariq Cannonier, May I. Conley, Alexandra O. Cohen, Deanna M. Barch, Mary M.

Heitzeg, Mary E. Soules, et al. 2018. “The Adolescent Brain Cognitive Development (ABCD) Study: Imaging Acquisition across 21 Sites.” *Developmental Cognitive Neuroscience*, March. https://doi.org/10.1016/j.dcn.2018.03.001.

Ducharme, Simon, Matthew D. Albaugh, Tuong-Vi Nguyen, James J. Hudziak, J. M. Mateos-Pérez, Aurelie Labbe, Alan C. Evans, Sherif Karama, Brain Development Cooperative Group, and Others. 2016. “Trajectories of Cortical Thickness Maturation in Normal Brain Development—the Importance of Quality Control Procedures.” *NeuroImage* 125: 267–79.

Feinberg, David A., Steen Moeller, Stephen M. Smith, Edward Auerbach, Sudhir Ramanna, Matthias Gunther, Matt F. Glasser, Karla L. Miller, Kamil Ugurbil, and Essa Yacoub. 2010. “Multiplexed Echo Planar Imaging for Sub-Second Whole Brain FMRI and Fast Diffusion Imaging.” *PloS One* 5 (12): e15710.

Fischl, Bruce. 2012. “FreeSurfer.” *NeuroImage* 62 (2): 774–81.

Giard, Joachim, Patrice Rondao Alface, Jean-Luc Gala, and Benoît Macq. 2011. “Fast Surface-Based Travel Depth Estimation Algorithm for Macromolecule Surface Shape Description.” *IEEE/ACM Transactions on Computational Biology and Bioinformatics / IEEE, ACM* 8 (1): 59–68.

Glasser, Matthew F., Stephen M. Smith, Daniel S. Marcus, Jesper L. R. Andersson, Edward J. Auerbach, Timothy E. J. Behrens, Timothy S. Coalson, et al. 2016. “The Human Connectome Project’s Neuroimaging Approach.” *Nature Neuroscience* 19 (9): 1175–87.

Godenschweger, F., U. Kägebein, D. Stucht, U. Yarach, A. Sciarra, R. Yakupov, F. Lüsebrink, P. Schulze, and O. Speck. 2016. “Motion Correction in MRI of the Brain.” *Physics in Medicine and Biology* 61 (5): R32–56.

Greene, Deanna J., Jonathan M. Koller, Jacqueline M. Hampton, Victoria Wesevich, Andrew N. Van, Annie L. Nguyen, Catherine R. Hoyt, et al. 2018. “Behavioral Interventions for Reducing Head Motion during MRI Scans in Children.” *NeuroImage* 171 (May): 234–45.

Harms, Michael P., Leah H. Somerville, Beau M. Ances, Jesper Andersson, Deanna M. Barch, Matteo Bastiani, Susan Y. Bookheimer, et al. 2018. “Extending the Human Connectome Project across Ages: Imaging Protocols for the Lifespan Development and Aging Projects.” *NeuroImage* 183 (December): 972–84.

Howell, Brittany R., Martin A. Styner, Wei Gao, Pew-Thian Yap, Li Wang, Kristine Baluyot, Essa Yacoub, et al. 2019. “The UNC/UMN Baby Connectome Project (BCP): An Overview of the Study Design and Protocol Development.” *NeuroImage* 185 (January): 891–905.

Iscan, Zafer, Tony B. Jin, Alexandria Kendrick, Bryan Szeglin, Hanzhang Lu, Madhukar Trivedi, Maurizio Fava, et al. 2015. “Test-Retest Reliability of Freesurfer Measurements within and between Sites: Effects of Visual Approval Process.” *Human Brain Mapping* 36 (9): 3472–85.

Jack, Clifford R., Jr, Matt A. Bernstein, Nick C. Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, et al. 2008. “The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI Methods.” *Journal of Magnetic Resonance Imaging: JMRI* 27 (4): 685–91.

Jenkinson, Mark, Peter Bannister, Michael Brady, and Stephen Smith. 2002. “Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images.” *NeuroImage* 17 (2): 825–41.

Klein, Arno, Satrajit S. Ghosh, Forrest S. Bao, Joachim Giard, Yrjö Häme, Eliezer Stavsky, Noah Lee, et al. 2017. “Mindboggling Morphometry of Human Brains.” *PLoS Computational Biology* 13 (2): e1005350.

Magnotta, Vincent A., Lee Friedman, and FIRST BIRN. 2006. “Measurement of Signal-to-Noise and Contrast-to-Noise in the fBIRN Multicenter Imaging Study.” *Journal of Digital Imaging* 19 (2): 140–47.

Mortamet, Bénédicte, Matt A. Bernstein, Clifford R. Jack Jr, Jeffrey L. Gunter, Chadwick Ward, Paula J. Britson, Reto Meuli, Jean-Philippe Thiran, Gunnar Krueger, and Alzheimer’s Disease Neuroimaging Initiative. 2009. “Automatic Quality Assessment in Structural Brain

Magnetic Resonance Imaging." *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 62 (2): 365–72.

Mugler, J. P., 3rd, and J. R. Brookeman. 1990. "Three-Dimensional Magnetization-Prepared Rapid Gradient-Echo Imaging (3D MP RAGE)." *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 15 (1): 152–57.

Nooner, Kate Brody, Stanley J. Colcombe, Russell H. Tobe, Maarten Mennes, Melissa M. Benedict, Alexis L. Moreno, Laura J. Panek, et al. 2012. "The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry." *Frontiers in Neuroscience* 6 (October): 152.

Pardoe, Heath R., Rebecca Kucharsky Hiess, and Ruben Kuzniecky. 2016. "Motion and Morphometry in Clinical and Nonclinical Populations." *NeuroImage* 135 (July): 177–85.

Power, Jonathan D., Benjamin M. Silver, Melanie R. Silverman, Eliana L. Ajodan, Dienke J. Bos, and Rebecca M. Jones. 2019. "Customized Head Molds Reduce Motion during Resting State fMRI Scans." *NeuroImage* 189 (April): 141–49.

Reuter, Martin, M. Dylan Tisdall, Abid Qureshi, Randy L. Buckner, André J. W. van der Kouwe, and Bruce Fischl. 2015. "Head Motion during MRI Acquisition Reduces Gray Matter Volume and Thickness Estimates." *NeuroImage* 107 (February): 107–15.

Shrout, P. E., and J. L. Fleiss. 1979. "Intraclass Correlations: Uses in Assessing Rater Reliability." *Psychological Bulletin* 86 (2): 420–28.

Smith, Stephen M., Yongyue Zhang, Mark Jenkinson, Jacqueline Chen, P. M. Matthews, Antonio Federico, and Nicola De Stefano. 2002. "Accurate, Robust, and Automated Longitudinal and Cross-Sectional Brain Change Analysis." *NeuroImage* 17 (1): 479–89.

Somerville, Leah H., Susan Y. Bookheimer, Randy L. Buckner, Gregory C. Burgess, Sandra W. Curtiss, Mirella Dapretto, Jennifer Stine Elam, et al. 2018. "The Lifespan Human Connectome Project in Development: A Large-Scale Study of Brain Connectivity Development in 5-21 Year Olds." *NeuroImage* 183 (December): 456–68.

Son, J., L. Ai, R. Lim, T. Xu, S. Colcombe, and A. Franco. 2019. "Evaluating fMRI-Based Estimation of Eye Gaze during Naturalistic Viewing." *BioRxiv*. https://www.biorxiv.org/content/10.1101/347765v3.abstract.

Sudlow, Cathie, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, et al. 2015. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age." *PLoS Medicine* 12 (3): e1001779.

Thesen, Stefan, Oliver Heid, Edgar Mueller, and Lothar R. Schad. 2000. "Prospective Acquisition Correction for Head Motion with Image-Based Tracking for Real-Time fMRI." *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 44 (3): 457–65.

Tisdall, M. Dylan, Aaron T. Hess, Martin Reuter, Ernesta M. Meintjes, Bruce Fischl, and André J. W. van der Kouwe. 2012. "Volumetric Navigators for Prospective Motion Correction and Selective Reacquisition in Neuroanatomical MRI." *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 68 (2): 389–99.

Tisdall, M. Dylan, Martin Reuter, Abid Qureshi, Randy L. Buckner, Bruce Fischl, and André J. W. van der Kouwe. 2016. "Prospective Motion Correction with Volumetric Navigators (vNavs) Reduces the Bias and Variance in Brain Morphometry Induced by Subject Motion." *NeuroImage* 127 (February): 11–22.

Vanderwal, Tamara, Clare Kelly, Jeffrey Eilbott, Linda C. Mayes, and F. Xavier Castellanos. 2015. "Inscapes: A Movie Paradigm to Improve Compliance in Functional Magnetic Resonance Imaging." *NeuroImage* 122 (November): 222–32.

Van Essen, David C., Stephen M. Smith, Deanna M. Barch, Timothy E. J. Behrens, Essa Yacoub, Kamil Ugurbil, and WU-Minn HCP Consortium. 2013. “The WU-Minn Human Connectome Project: An Overview.” *NeuroImage* 80 (October): 62–79.

Van Essen, D. C., K. Ugurbil, E. Auerbach, D. Barch, T. E. J. Behrens, R. Bucholz, A. Chang, et al. 2012. “The Human Connectome Project: A Data Acquisition Perspective.” *NeuroImage* 62 (4): 2222–31.

White, Nathan, Cooper Roddey, Ajit Shankaranarayanan, Eric Han, Dan Rettmann, Juan Santos, Josh Kuperman, and Anders Dale. 2010. “PROMO: Real-Time Prospective Motion Correction in MRI Using Image-Based Tracking.” *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 63 (1): 91–105.

White, Tonya, Philip R. Jansen, Ryan L. Muetzel, Gustavo Sudre, Hanan El Marroun, Henning Tiemeier, Anqi Qiu, Philip Shaw, Andrew M. Michael, and Frank C. Verhulst. 2018. “Automated Quality Assessment of Structural Magnetic Resonance Images in Children: Comparison with Visual Inspection and Surface-Based Reconstruction.” *Human Brain Mapping* 39 (3): 1218–31.

Wonderlick, J. S., D. A. Ziegler, P. Hosseini-Varnamkhasti, J. J. Locascio, A. Bakkour, A. van der Kouwe, C. Triantafyllou, S. Corkin, and B. C. Dickerson. 2009. “Reliability of MRI-Derived Cortical and Subcortical Morphometric Measures: Effects of Pulse Sequence, Voxel Geometry, and Parallel Imaging.” *NeuroImage* 44 (4): 1324–33.

Zarrar, Shehzad, Giavasis Steven, Li Qingyang, Benhajali Yassine, Yan Chaogan, Yang Zhen, Milham Michael, Bellec Pierre, and Craddock Cameron. 2015. “The Preprocessed Connectomes Project Quality Assessment Protocol - a Resource for Measuring the Quality of MRI Data.” *Frontiers in Neuroscience*. https://doi.org/10.3389/conf.fnins.2015.91.00047.

Supplementary Material

Supplementary Table 1 - Imaging Session Sequence for the Participant that conducted the regular HBN protocol

Order	Sequence	Time / Run	Cumulative time
1	localizer_32ch	0:00:47	-
2	ABCD_T1w_MPR_vNAV_setter	0:00:00	0:00:47
3	ABCD_T1w_MPR_vNAV **(vNAV)**	0:07:12	0:00:47

4	ANAT_T1W-RU **(MPRAGE)**	0:07:19	0:07:59
5	cmrr_fMRI_DistortionMap_AP	0:00:05	0:15:18
6	cmrr_fMRI_DistortionMap_PA	0:00:05	0:15:23
7	cmrr_REST1	0:05:08	0:15:28
8	cmrr_PEER1	0:01:56	0:20:36
9	cmrr_REST2	0:05:08	0:22:32
10	cmrr_MOVIE1	0:10:08	0:27:40
11	cmrr__dMRI_DistortionMap_AP	0:00:50	0:37:48
12	cmrr__dMRI_DistortionMap_PA	0:00:50	0:38:38
13	cmrr_DKI_018	0:07:55	0:39:28
14	ABCD_T2w_SPC_vNAV_setter	0:00:00	0:47:23
15	ABCD_T2w_SPC_vNav	0:06:35	0:47:23
16	cmrr_PEER3	0:01:56	0:53:58
17	cmrr_MOVIE2	0:03:28	0:55:54

Supplementary Table 2 - Imaging Session Sequence for Participants that conducted the Structural Imaging test-retest protocol.

Order	Sequence	Time / Run	Cumulative time

1	localizer_32ch	0:00:47	-
2	ABCD_T1w_MPR_vNAV_setter	0:00:00	0:00:47
3	ABCD_T1w_MPR_vNAV **(vNAV_1)**	0:07:12	0:00:47
4	ANAT_T1W-RU **(MPRAGE_1)**	0:07:19	0:07:59
5	cmrr_fMRI_DistortionMap_AP	0:00:05	0:15:18
6	cmrr_fMRI_DistortionMap_PA	0:00:05	0:15:23
7	cmrr_REST1	0:05:08	0:15:28
8	cmrr_PEER1	0:01:56	0:20:36
9	cmrr_REST2	0:05:08	0:22:32
10	cmrr_MOVIE1	0:10:08	0:27:40
11	cmrr__dMRI_DistortionMap_AP	0:00:50	0:37:48
12	cmrr__dMRI_DistortionMap_PA	0:00:50	0:38:38
13	cmrr_DKI_018	0:07:55	0:39:28
14	ABCD_T2w_SPC_vNAV_setter	0:00:00	0:47:23
15	ABCD_T2w_SPC_vNav	0:06:35	0:47:23
16	ANAT_T1W-RU **(MPRAGE_2)**	0:07:19	0:53:58
17	ABCD_T1w_MPR_vNAV_setter	0:00:00	1:01:17
18	ABCD_T1w_MPR_vNAV **(vNAV_2)**	0:07:12	1:01:17

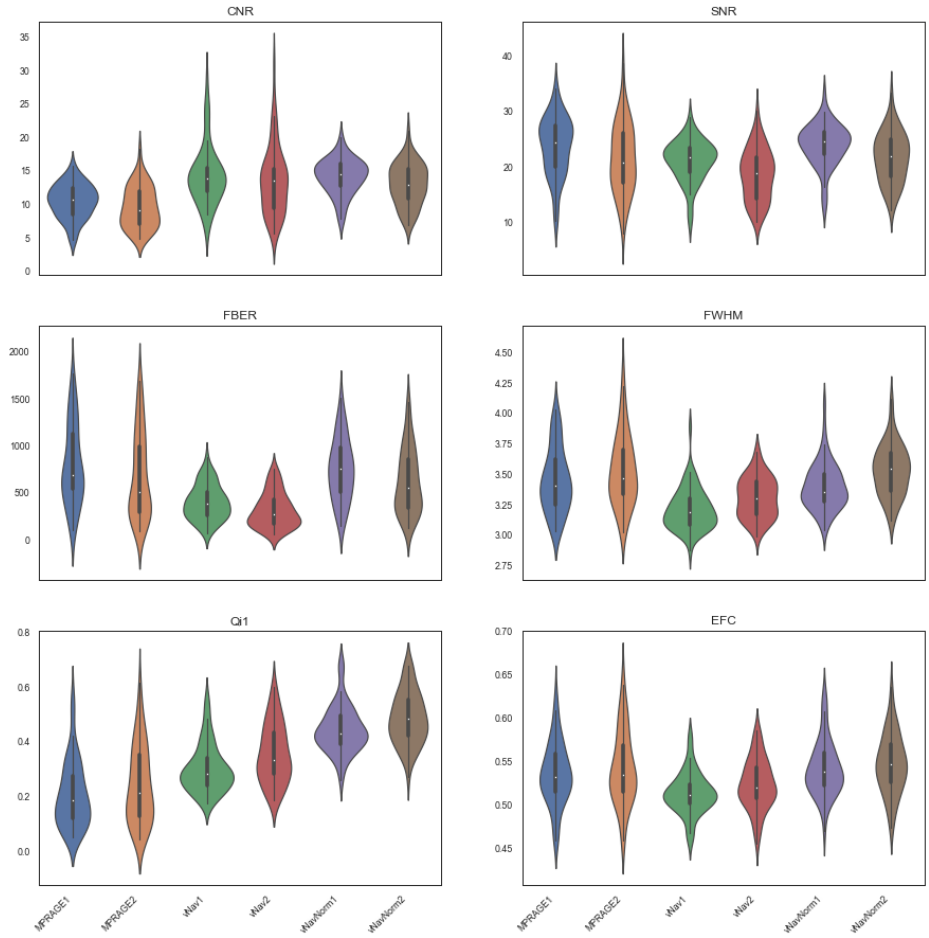
19	cmrr_PEER3	0:01:56	1:08:29
20	cmrr_MOVIE2	0:03:28	1:10:25

Time in between the end of HCP1 and start of HCP2 sequences is 39 minutes
Time in between the end of vNav1 and start of vNav2 sequences is 53 minutes

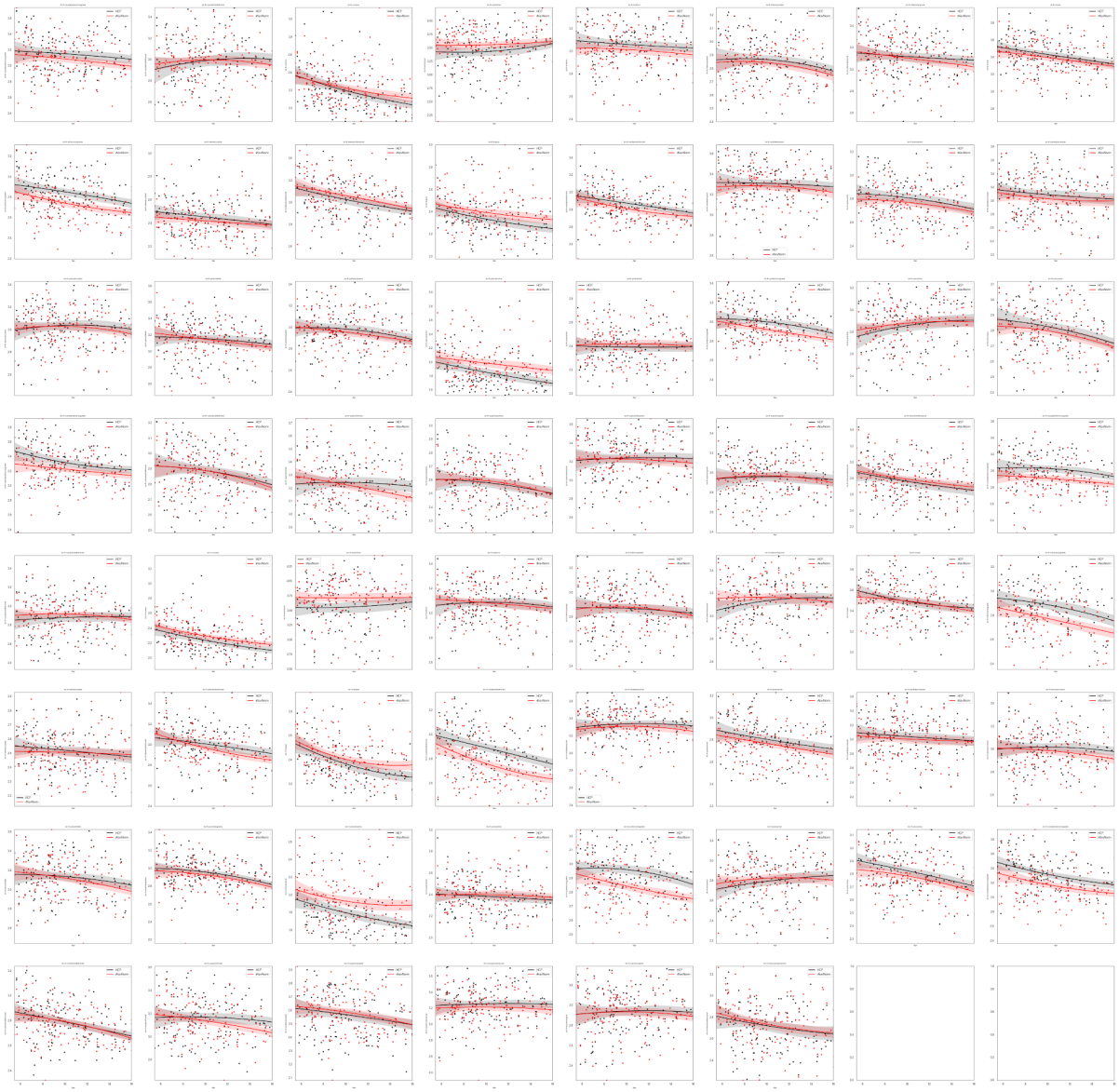
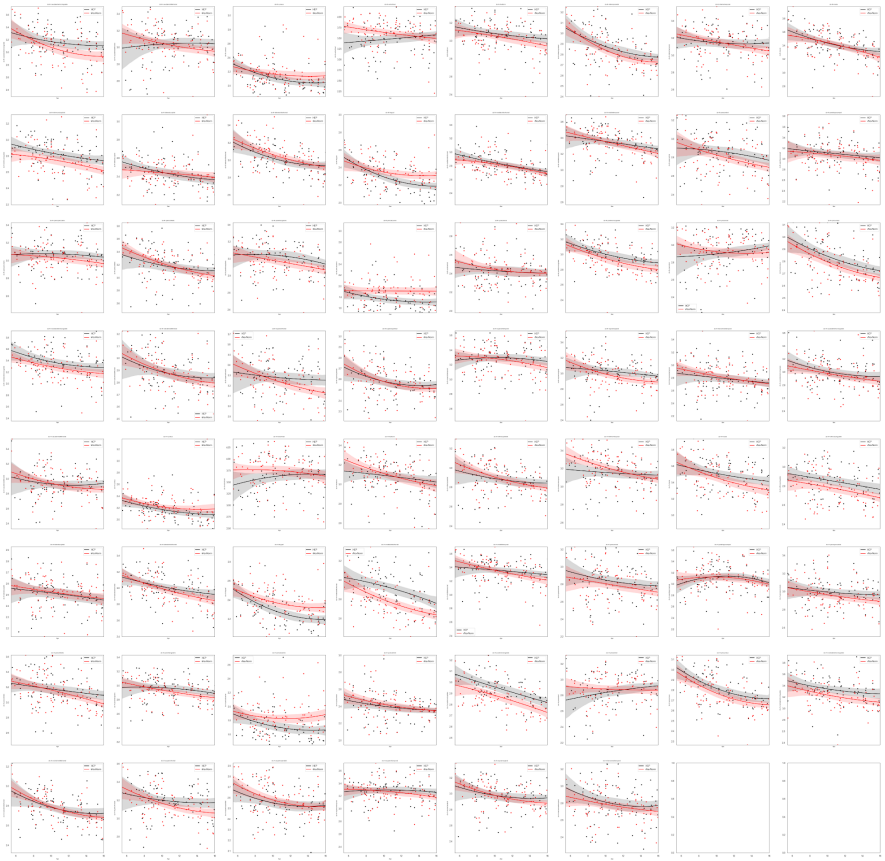
The time between the start of vNav1 and vNav2 is at least 1:00:30
Time between the start of HCP1 and HCP2 is at least 0:45:49
Time between the start of vNav1 and HCP2 is at least 0:53:11
Time between HCP1 and vNav2 0:53:18

Supplementary Figures

Supplementary Figure S1 - Quality Control Metrics for the test-retest group.


CNR
SNR
FBER
FWHM
Qi1
EFC
MPRAGE1
MPRAGE2
vNav1
vNav2
vNavNorm1
vNavNorm2

Supplementary Figure S2 - Cortical Thickness Development curves for Males.



Supplementary Figure S3 - Cortical Thickness Development curves for Females.