Variations in Structural MRI Quality Impact Measures of Brain Anatomy: Relations with Age and Other Sociodemographic Variables

Alysha Gilmore

Nicholas Buser

Jamie L. Hanson

University of Pittsburgh

Author Note

All correspondence related to this work should be sent to: Jamie Hanson, *Learning Research & Development Center, University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260*;

Telephone: 412-383-3250; Email: jamie.hanson@pitt.edu

Potential Conflicts of interest: none to report

Abstract:

In-scanner head movements can introduce artifacts to MRI images and increase errors in brain-behavior studies. The magnitude of in-scanner head movements varies widely across developmental and clinical samples, making it increasingly difficult to parse out “true signal” from motion related noise. Yet, the quantification of structural imaging quality is typically limited to subjective visual assessments and/or proxy measures of motion. It is, however, unknown how direct measures of image quality relate to developmental and behavioral variables, as well as measures of brain morphometrics. To begin to answer this question, we leverage a multi-site dataset of structural MRI images, which includes a range of children and adolescents with varying degrees of psychopathology. We first find that a composite of structural image quality relates to important developmental and behavioral variables (e.g., IQ; clinical diagnoses). Additionally, we demonstrate that even among T1-weighted images which pass visual inspection, variations in image quality impact volumetric derivations of regional gray matter. Image quality was associated with wide-spread variations in gray matter, including in portions of the frontal, parietal, and temporal lobes, as well as the cerebellum. Further, our image quality composite partially mediated the relationship between age and total gray matter volume, explaining 23% of this relationship. Collectively, the effects underscore the need for volumetric studies to model or mitigate the effect of image quality when investigating brain-behavior relations.

Keywords: *Structural MRI; image quality; voxel-based morphometry; gray matter*

1. INTRODUCTION

Neuroimaging methods are increasingly common, but with these advancements, there has been a greater understanding of the potential confounds and limitations of these research techniques. A common limitation of neuroimaging research is that of motion-related artifacts (Alexander-Bloch et al. 2016; Power et al. 2012, 2014; Reuter et al. 2015; Satterthwaite et al. 2013; Van Dijk et al. 2012). For those interested in neurodevelopment and mental health, such noise and bias may be particularly important to address (e.g., Fair et al. 2012; Satterthwaite et al. 2012). Head motion varies considerably among individuals, children typically move more than adults (Satterthwaite et al. 2012; Van Dijk et al. 2012), and patient groups move more than controls (Kong et al. 2014; Power et al. 2012).

Multiple resting state fMRI studies have highlighted the importance of this issue, as small differences in motion have been shown to yield significant differences in estimates of functional connectivity among healthy samples (Power et al. 2012; Satterthwaite et al. 2012; Van Dijk et al. 2012; Zeng et al. 2014). In fact, head movements within fractions of a millimeter distort associations between BOLD-activation time series' in a distant dependent manner, leading to spurious estimates of connectivity within functional networks (Power et al. 2011 & Van Dijk et al. 2014). Further, recent work has shown that head motion is consistent within individual subjects, from one scanning session to the next, raising the potential for motion to confound the exploration of individual differences within the same population (Van Dijk et al. 2012). Particularly challenging, these differences persist even after extensive motion correction procedures (Ciric et al. 2017; Power et al. 2012, 2014; Satterthwaite et al. 2013; Van Dijk et al. 2012). This has thus motivated a methodological subfield focused on robust ways to reduce

motion-related noise in resting-state and other forms of functional MRI (Satterthwaite et al. 2017).

While a great deal of progress has been made in quantifying and addressing the impact of head-motion in functional MRI, less attention has been given to structural MRI. Head motion has been shown to compromise derived measures of volume and thickness in regions of cortical gray matter (Alexander-Bloch et al. 2016; Reuter et al. 2015). These associations are likely not caused by processing failures in software, but rather they reflect systematic bias related to in-scanner motion and this may appear similar to reductions in gray matter (Reuter et al. 2015). However, much of the early work in this area did not include youth and/or clinical populations, and thus, may fail to account for overlapping variance between motion and other psychosocial and sociodemographic factors of interest (e.g., age; psychopathology).

While past work suggests movement is likely to impacts structural MRI, methods of quantifying and addressing motion-related noise in structural MRI have been limited. With particularly noisy structural data, researchers traditionally flag problematic scans and remove these subjects from further analyses. This process involves the visual assessment of each T1-weighted structural image. A limitation of this strategy is that many phenotypes of interest are inherently more prone to head motion (e.g., children under 9; individuals with clinical diagnoses; Alexander et al. 2017). Also, human rating systems are relatively impractical for large scale datasets (Esteban et al. 2017). A further challenge is that visual inspection by human raters is subjective. Numerous studies have showcased this, with high variability in inter- and intra-human-rating systems (e.g., Esteban et al. 2017 & Klapwijk et al. 2018). Further, even for structural scans that pass “visual inspection”, there may still be important variations in data quality which impact morphometric estimates. Put another way, some scans may be “just above”

threshold for raters, while other volumes may be of utmost quality; both types of scans, however, would be simply considered “usable”.

Thinking about functional versus structural MRI, progress on understanding the impact of noise on each modality may be variable due to the measures of movement that are derivable from each scan type. Functional MRI involves the acquisition of dozens, often hundreds, of lower resolution brain volumes; this allows for the calculation of frame-by-frame changes in a volume’s position, and a clear metric of subject movement during fMRI scanning acquisitions (e.g., Framewise Displacement; FD). Structural MRI only involves the acquisition of one, higher resolution volume. Until recently, this has prohibited rich assessments of noise and subject movement in contrast to fMRI. With structural MRI, past work focused on the impact of movement has leveraged indirect proxies of noise (e.g., FD from fMRI scans in the same scanning session), or has binned structural scans based on subjective quality assigned by human-raters. Using these types of approaches, research reports suggest that poor quality data impacts multiple brain-wide measures of morphometry derived from T1w images. For example, Savalia and colleagues (2017) found T1w images from flagged participants exhibited reduced estimates of gray matter thickness and volume in comparison to (non-flagged) age- and gender-matched samples. This caused inflations in the calculated associations between regional anatomical measures and age.

Given that previous studies have used indirect measures of motion and noise, it is possible that elements of past results may be inaccurate, particularly in “motion-prone” populations. Additional work would be necessary to clarify precisely how motion-related bias and noise in T1w images varies and overlaps across distinct study populations. Advancements of novel informatic tools may overcome these issues. Automated quality assessment tools provide

quantitative measures of noise present within structural neuroimaging volumes and are easy-to-implement. Most notably, MRIQC can assess multiple image quality measures (IQMs) that speak to different quality attributes of structural (and other MRI) images (Esteban et al. 2017). Thinking about past research, it is unclear if these IQMs (and structural MRI quality) are related to commonly derived structural measures (e.g., gray matter volume). Additionally, it is unknown if variations in IQMs are impacting associations between structural MRI measures and sociodemographic variables (e.g., age; measures of cognitive functioning; psychopathology). Accounting for such variations may reduce potential spurious associations and increase the replicability of effects.

To make progress on this issue, we completed the following analytic plan: 1) We investigated whether multiple IQMs were uniquely associated with the binary outcome of visual quality checks completed by trained research assistants; 2) We examined if variations in image quality related to demographic and psychosocial variables (e.g., age; measures of cognitive functioning; clinical diagnosis); 3) We evaluated relations between IQMs and a commonly-used structural derivative, gray matter probability (GMP); 4) We probed whether voxel-wise associations between GMP and age were impacted by image quality; 5) We tested if direct measures of structural imaging quality statistically explained the relationship between age and total gray matter volume.

2. MATERIALS AND METHODS

2.1 Participants.

Data from 724 participants between the ages of 5-21 years of age with complete T1-weighted structural images were downloaded from the first two data waves of an ongoing

research initiative, The Healthy Brain Network (HBN), launched by The Child Mind Institute in 2015. Written informed consent was obtained from participants ages 18 or older. For participants younger than 18, written consent was obtained from their legal guardians and written assent was obtained from the participant. Participants with cognitive or behavioral challenges (e.g., being nonverbal, IQ<66), or with medical concerns expected to confound brain-related findings were excluded from the HBN project. The HBN protocol spans four sessions, each approximately three hours in duration. *See Table S1of our supplemental materials for full sample characteristics.* Data and code sharing adopted by the authors (including its licensing) comply with the requirements of the funding body and comply with institutional ethics approval.

2.2 MRI Data Acquisition.

All T1-weighted structural MRIs collected from one of two scanning sites. Scanning for the first wave of data collection ***(n=343)*** took place in Staten Island (SI), near the Child Mind Institute. A mobile 1.5 T Siemens Avanto MRI scanner equipped with 45 mT/m gradients in a mobile trailer (Medical Coaches, Oneonta, NY) parked on 10-inch thick concrete pads. The system was upgraded with 32 RF receive channels, the Siemens 32-channel head coil, and the University of Minnesota Center for Magnetic Resonance Research simultaneous multi-slice echo planar imaging sequence (Moeller *et al. 2010*). T1W acquisition took approximately 6.53 minutes. MPRAGE scanning acquisition parameters included TR=2730ms; TE=1.64ms; TI=1000ms; slices=176; resolution(mm)=1.0x1.0x1.0; percent FOV phase=100; and flip angle=7 degrees.

The second scanning phase ***(n=381)*** occurred at the Rutgers University Brain Imaging Center (RU). A Siemens 3 T Tim Trio MRI scanner was equipped with a Siemens 32-channel

head coil and the CMRR simultaneous multi-slice echo planar imaging sequence. T1W acquisition took approximately 7.00 minutes. MPRAGE scanning acquisition parameters included TR= 2500ms; TE=3.15ms; TI=1060ms; slices=224; resolution(mm)=0.8x0.8x0.8; percent FOV phase=100; and flip angle=8 degrees. *See S1 of our Supplemental Materials for additional information about scanning sites.* All neuroimaging data used in this study are openly available for download with proper data usage agreement via the International Neuroimaging Data-sharing Initiative (Alexander et al., 2017).

2.3 Visual Quality Inspection.

All T1w scans were separated by site and release wave then visually inspected by a series of human raters trained to recognize frequent indications of scan artifacts and motion. This training provided examples and descriptions for artifacts including “ringing”, “ghosting”, “RF-Noise”, “head coverage”, and “susceptibility”. A detailed description of this protocol can be found in our *Supplemental Materials, specifically sections S3A-S3C*. In short, each rater was instructed to give a score between 1 and 10 with 6 being the cutoff for scan inclusion in further research. To minimize effects arising from rater idiosyncrasy, all ratings were z-scored, averaged across raters, and compared to the averaged z-score for the cutoff (6.0) points. Scans for which the averaged z-scored rating was greater than the averaged z-score cutoff point were retained and the rest were removed from further analysis. *Sample characteristics for participants with T1w scans that passed visual inspection can be found in Table S2 of our Supplemental Materials.*

2.4 Image Quality Metrics.

Automated measures of image quality were generated via the MRI Quality Control (*MRIQC*) workflow, which extracts numerous image quality measures (IQMs) based on noise measurements, information theory, or targeting specific artifacts. MRIQC's workflow follows the BIDS standard, and was implemented to estimate the IQMs on the original (T1w) data (Esteban et al. 2017). *See Table S3 of our Supplemental Materials for a summary of each image quality metric extracted using MRIQC toolbox* (as well as this webpage: https://mriqc.readthedocs.io/en/stable/iqms/t1w.html).

2.5 Self-Report Measures of Behavior and Cognition.

Phenotypic (behavioral) data was assessed through the Collaborative Informatics and Neuroimaging Suite (COINS) Data Exchange after completion of appropriate data use agreements (Scott et al. 2011). Of note, neuroimaging data was collected regardless of whether participants completed other portions of the project (e.g., self-report questionnaire battery); therefore, many participants with imaging data did not have phenotypic data available (at present). *Again, See Table S1 of our Supplemental Materials for Sample characteristics.* We chose a number of measures which we predicted would covary with structural MRI quality. These included: age, sex, body mass index (BMI), general cognitive ability/IQ, and clinical diagnoses. We describe information about these measures below. We also examined the effect of scanning site (as there was differences in MRI scanner strength that could likely influence image quality).

2.3.1 Assessment of General Cognitive Ability.

The Wechsler Intelligence Scale for Children (WISC-V) was completed by 464 youth in the sample (age range: 5.99-16.96 years, mean: 10.29, sd: 2.83). The WISC-V is an individually

administered clinical instrument for assessing the intelligence of youth participants 6-16. This measure generates a Full-Scale Intelligence Quotient (FSIQ) score, which represents general cognitive ability (sample range: 50-145, mean: 98.02, sd: 17.31). *Additional information can be found in S4 of our Supplemental Materials.*

2.3.2 Clinical diagnoses.

The presence of psychopathology was assessed by a certified clinician based on review of electronically completed survey items and results generated from the computerized web-based version of the Schedule for Affective Disorders and Schizophrenia for Children (KSADS-COMP). The KSADS-COMP is a semi-structured DSM-5-based psychiatric interview used to derive clinical diagnoses (Kaufman et al. 1997). This data was available for 671 participants, with 554 participants with at least one clinical diagnosis, and 117 subjects with no diagnosis.

3. STATISTICAL ANALYSES

3.1 Image quality composite and bivariate correlations.

MRIQC Toolbox outputs a number of metrics based on noise measurements, information theory, and the targeting of specific MR artifacts. However, many of these measures may be collinear and explaining similar aspects of image quality. Therefore, we sought to determine which quality indices would provide enough information to make valid predictions for the outcome (Pass/Fail) of T1-weighted images subjected to visual inspection. For this reason, seven IQMs were entered into a logistic regression model as independent variables. The dependent variable in this model was the binary outcome (pass/fail) of visual quality assurance checks completed by trained human raters. For all IQMs that uniquely (and significantly) associated

with passing visual inspection, scores from this variable were standardized (z-scored) and then all significant IQMs were averaged to create a composite of image quality. Next, bivariate correlations were used to examine relations between our image quality composite and sociodemographic variables of interest, including age, sex, IQ, MRI field strength, BMI, and clinical diagnosis.

3.2 Image Preprocessing (for Voxel-Based Morphometry).

T1-weighted images were first corrected for field inhomogeneities via N4ITK (Tustison et al. 2010). Brain extraction, tissue segmentation with topological refinements, and diffeomorphic image registration were then completed using Advanced Normalization Tools (ANTs, Avants, Tustison, & Song, 2009; Tustison et al. 2014). ANTs is a volume-based registration/labeling tool (Klein et al. 2009) and we used their antsCorticalThickness.sh script (https://github.com/ANTsX/ANTs/blob/master/Scripts/antsCorticalThickness.sh) for all relevant neuroimaging processing. ~~For~~ brain extraction, the OASIS IXI template was registered to each subject (using an affine transform and symmetric, diffeomorphic normalization) and this yielded a probabilistic brain extraction mask for each file. Next, each extracted brain was segmented using a four-tissue segmentation estimate (Avants et al. 2011). This segmentation algorithm, Atropos, is an open source n-tissue segmentation software tool that uses spatially varying Markov Random Field modeling and prior (tissue segmentation) probability maps. Atropos outputs voxel-wise tissue probability maps, with probabilities at each voxel varying continuously from 0-1 for cerebral spinal fluid, cortical gray matter, “deep” (subcortical) gray matter, and white matter. Cortical and “deep” gray matter were then combined to look at gray matter probability (GMP) in all subsequent analyses. These GMP maps were then smoothed with a

4mm gaussian kernel to lessen any potential effects of misalignment of structural images. GMP maps were not “modulated” as recent work (e.g., Radua et al. 2014) suggests this step may decrease the sensitivity to detect true signal. Of additional note, total tissue volumes (for gray matter, white matter, and cerebral spinal fluid) were also calculated via ANTs. These calculations counted all voxels for a given tissue type above a threshold of 60%.

3.3 Modeling indirect effects of image quality on regional age-GMP associations.

Voxel-wise GMP maps were entered into linear regression models constructed in Analysis of Functional NeuroImages (AFNI; Cox, 1996). In a first set of analyses, we examined the voxel-wise associations between GMP and IQMs. These analyses were completed separately for each neuroimaging site (RU; SI) and were then thresholded at p=.0001 (uncorrected). All values below this threshold were set to zero, and then these thresholded maps were multiplied together (in logical AND conjunction analysis). The resulting maps would indicate brain regions where there was an association between GMP and image quality. This analytic approach has been used in a number of research studies (e.g., Nichols et al. 2006) to find consistent brain differences across datasets. Such an approach was motivated by the differences in acquisition parameters across collection sites. Assuming independence of these tests, these results would be significant at 0.00000001 (.0001x.0001), uncorrected. In these analyses, we also employed a watershed algorithm to divide large clusters into more regionally-specific statistical maps (available for download at: https://www.med.upenn.edu/cmroi/shed.html).

We then constructed two-related sets of voxel-wise models, examining: 1) GMP and age (alone); 2) GMP, age, and our image quality composite. These analyses would allow us to see if associations between GMP and age were influenced after accounting for image quality. For these

analyses, we next subtracted z-scores for age, from Model 1 (age, entered as the only independent variable) from Model 2 (age, as well as our image quality composite, entered as the independent variables). This would yield an estimate of the change in the relation between age and GMP when taking into consideration variations in image quality. Z-score difference maps were then thresholded at p=.01 (uncorrected) for each site independently (RU; SI) and then used in a logical AND conjunction analysis. The resulting maps would indicate brain regions where the association between age and gray matter varied significantly when image quality was accounted for, across data collection sites. Such an approach was, again, motivated by the differences in acquisition parameters across collection sites. Assuming independence of these tests, these results would be significant at 0.0001 (0.01x0.01), uncorrected. To further minimize potential Type I error, we set a minimum cluster of k=90 for these logical AND conjunction maps. Of note, all analyses were completed within a whole-brain gray matter mask, thresholded at 50%.

3.4 Testing the indirect effect of image quality on age to total gray matter relationship.

We tested whether total gray matter volume was predicted by age and structural image quality with a mediation model regression. For this step, path analyses tested whether age (X) was associated with total gray matter volume (Y) and whether the observed association was statistically explained by structural image quality (M). Details about the derivation of total gray matter is noted in Section 3.2, ‘*Image Preprocessing’*. Statistical testing of mediation was done by nonparametric bootstrapping, with 95% confidence interval for indirect (a x b) mediation effects, with standard errors calculated through 5000 bootstrap permutations. Mediation

modeling was completed using lavaan (Rosseel, 2012) in the R statistical package and included total brain volume as a covariate.

4. RESULTS

4.1 Relations between MRIQC measures and visual rejection/acceptance of structural images.

Logistic regression was used to construct a composite of IQMs which accurately predicted the outcome of quality assurance checks completed by trained human raters. Out of seven quality indices, three MRIQC parameters were significant predictors in this model: the signal to noise ratio ($t=4.29$, $p<.005$); entropy focus criterion ($t=5.44$, $p<.005$); and Mortamet's Quality Index 2 ($t=2.92$, $p<.005$). In brief, the signal to noise ratio and index measure for 'goodness of fit' were positively correlated with passing visual inspection (signal to noise ratio, $r=0.40$, $p<.005$; Mortamet's Quality Index 2, $r=0.220$, $p<.005$), while the entropy focus criterion was negatively related to passing visual inspection ($r=-0.234$, $p<.005$). Next, individual scores for each of the significant IQM predictors were standardized among participants (z-scored) and averaged to create a composite for overall image quality. The entropy focus criterion was reverse-scored in this composite. As a form of confirmatory analysis, this composite was then entered into an additional logistic regression model to demonstrate its significance in predicting the dichotomous outcome of human-rater visual inspection. As expected, this composite was a significant predictor of passing visual quality control, with greater numbers indicating a higher likelihood of passing visual quality control ($B=1.11970$, Z-value$=10.213$, $p <.005$).

4.2 Bivariate correlations between image quality and psychosocial variables of interest.

We next examined correlations between this image QC composite; and age, sex, IQ, site/MRI strength, BMI, and clinical diagnosis. Scanning site was the largest source of variation with respect to image quality (r=0.451, t=13.56, p<.005). Image quality was significantly greater in subjects tested at the RU site; this is perhaps not surprising given that scanner strength differed by sites (1.5T at SI; 3T at RU). As expected, image quality increased with IQ (r=0.191, t = 4.19, p<.005), and was reduced among those with a clinical diagnosis (r=-0.082, t=-2.12, p=0.034). In keeping with past work (e.g., Beyer et al. 2017), body mass index was inversely associated with image quality (r=-0.316, t=-8.70, p<.005). Surprisingly, there was a negative association between our QC composite and age (r=-0.126, t=-3.41, p<.005). However, this relation may have been driven by a small number of older participants; when we looked at this association in participants that were less than 16 years of age, this same correlation was not significant (p=0.17). Finally, image quality was not related to sex (r=0.059, t=1.59, p=0.112). *Results also outlined in Table 1.*

Table 1.

Bivariate correlations between image quality and common variables of interest.

Variable	Pearson's correlation coefficient	p-value
Age	-0.1198131	0.002699
Body Mass Index	-0.3221487	$5.587e^{-16}$
IQ	0.1911582	$3.40e^{-5}$
Any Clinical Diagnosis	-0.08164761	0.03447
Site SI (1.5T scanner)	-0.4505629	$< 2.2e^{-16}$
Total grey matter volume	0.1314676	0.01093

4.3 Image quality impacts regional gray matter estimates.

We found multiple large clusters of GMP associated with our image quality composite. In regards to positive associations, there were six large clusters, spanning numerous regions of the brain (max k=22218 voxels) that included portions of the frontal and parietal lobes. *These effects are noted in Table 2 and shown in Figure 1*. There were also significant negative associations between image quality and GMP. In particular, relatively poorer image quality related to higher estimates of gray matter areas of the right Cerebellum and right Parahippocampal Gyrus, *(see Table 3 and Figure 2 below).*

Table 2.

Clusters with increased GMP directly linked to image quality (Cluster size > 90 voxels)

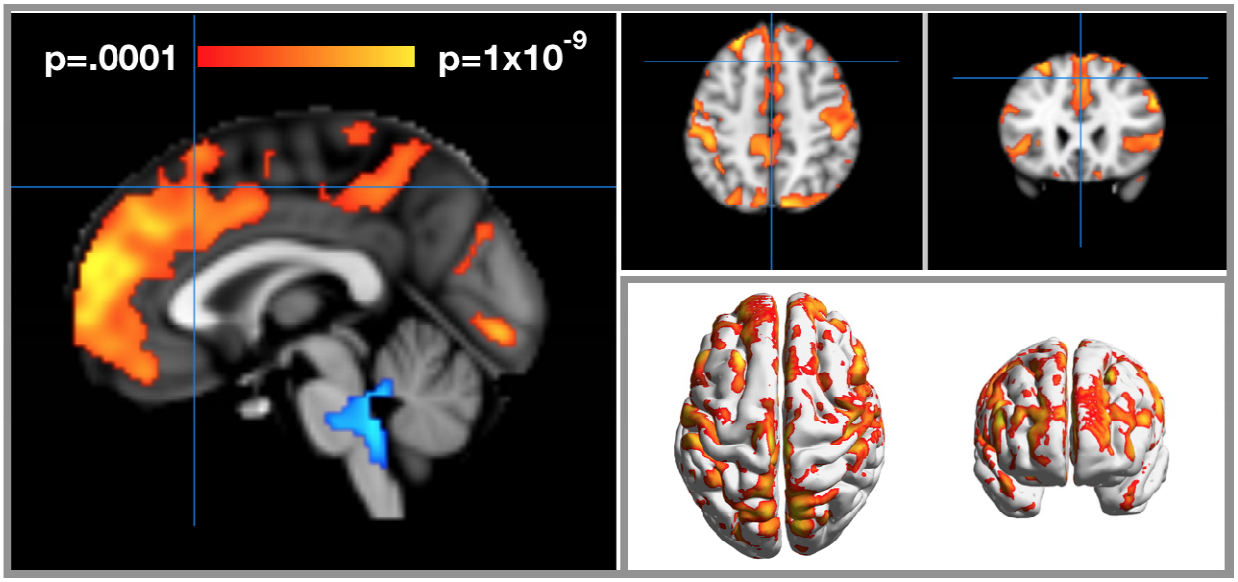
	# Voxels	CM x	CM y	CM z	Region
Cluster 1*	22218	+48.7	+13.9	+21.2	Multiple regions (see table note* below)
Cluster 2	210	50.6	70.7	29.6	Angular Gyrus
Cluster 3	148	-42.1	56.3	-43.7	Right Cerebellar Tonsil
Cluster 4	123	-15.4	76.1	-31.2	Right Uvula
Cluster 5	108	-60.5	27.2	1.4	Right Middle Temporal Gyrus
Cluster 6	93	-47.4	-19.3	-23.8	Right Superior Temporal Gyrus

* Peaks within this cluster (determined by watershed algorithm) included Medial Frontal Gyrus (k=5395, peak XYZ= +1.9, -39.2, +30.6), Precuneus / Superior Parietal Lobule (k=5283, peak XYZ= +0.5, +47.2,

+59.5), the left parietal lobe (k=2171, peak XYZ= +48.7, +13.9, +21.2), and right parietal lobe (k=3572, peak XYZ= -49.2, +6.5, +27.4).

Figure 1.

Regions with Increased GMP directly linked to image quality.



Description: Results of voxel-wise regression for image quality predicting gray matter probability. These statistical maps and surface renderings are the result of conjunction analyses and primarily depict positive associations where with greater image quality, higher gray matter probability is seen. Of note, the threshold of these statistical maps goes from p=.0001 to p=.000000001 (uncorrected).

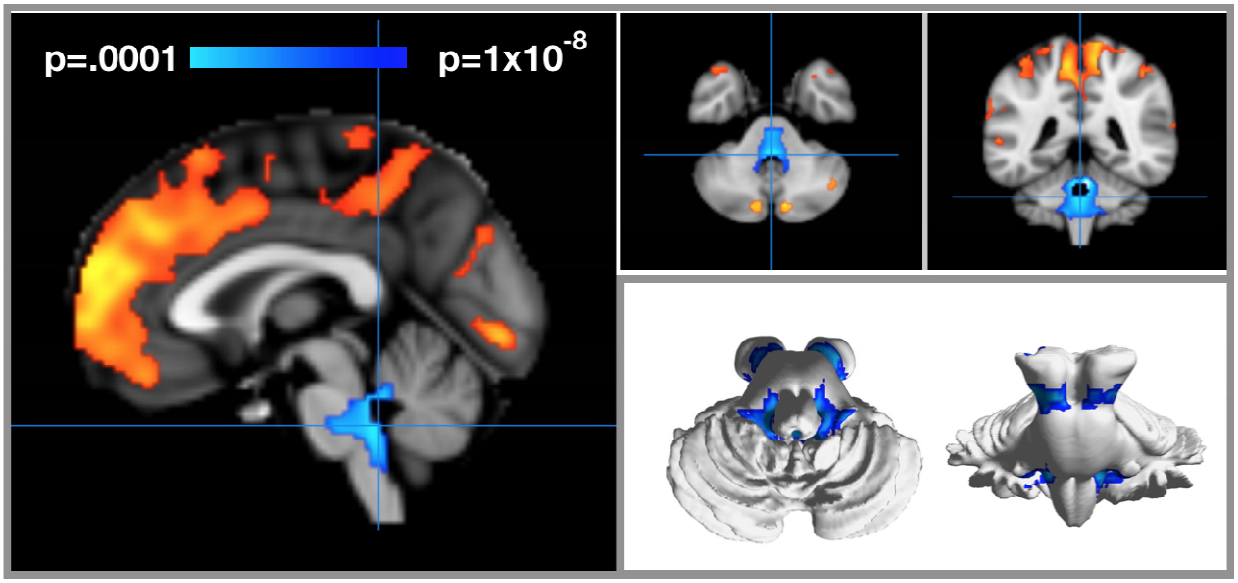
Table 3.

Regions with decreased GMP directly linked to image quality (Cluster size > 90 voxels)

	# of Voxels	CM x	CM y	CM z
Right Cerebellum	980	-1.2	+39.1	-36.5
Right Parahippocampal Gyrus	195	-15.4	19	-10.3

Figure 2.

Regions with decreased GMP directly linked to image quality.



Description: Results of voxel-wise regression for image quality predicting gray matter probability. These statistical maps and surface renderings are the result of conjunction analyses and primarily depict negative associations where with greater image quality, lower gray matter probability is seen. Of note, the threshold of these statistical maps goes from p=.0001 to p=.00000001 (uncorrected).

4.4 Associations between regional gray matter and age, controlling for QC variations.

We then examined if variations in image quality would impact associations between other commonly investigated variables, specifically participant age and gray matter. As noted in our methods section, we computed z-score difference maps for voxel-wise maps correlating age and gray matter (subtracting voxel-wise z-scores for a model including our image quality composite, and one without). We used logical and conjunction analyses to examine differences common across neuroimaging scanning sites, and saw a number of changes in regional associations when image quality was considered in testing the association between age and gray matter. There were significant positive changes in the association between age and gray matter for the Medial Frontal Gyrus, Right Postcentral Gyrus, and two clusters in bilateral cerebellum (Right and Left Tonsils). There were no major clusters (>90 voxels) seen where the association between age and gray matter diminished when controlling for image quality. *Results are outlined in Table 4.*

Table 4.

Increases in associations between age and GMP after controlling for image quality.

	# Voxels	CM x	CM y	CM z
Medial Frontal Gyrus	289	-0.5	-54.2	+18.6
Right Cerebellar Tonsil	222	-31.7	+60.5	-43.1
Left Cerebellar Tonsil	112	+28.1	+62.9	-42.1
Right Postcentral Gyrus	90	-46	+21.0	+57.4

4.5 Image quality partially mediates the relationship between age and total gray matter

Given these voxel-wise differences in our estimates of the effect of age on GMP, we next investigated whether individual differences in image quality might, in part, explain the

association between age and total gray matter. We examined associations between 1) age and total gray matter, 2) age and image quality, and 3) image quality and total gray matter. First, global grey matter volume extracted from the ANTs pipeline increased linearly with image quality ($r=0.131$, $t=2.56$, $p=0.011$). Our non-parametric bootstrapped models then indicated the variations in image quality explained 23.1% of the association between age and total gray matter. The indirect effect (axb; from age to image quality, and image quality to total gray matter) was also significant ($\beta=-0.067$, $Z=6.210$, $p<.005$; 95% confidence interval= -0.095 to -0.049). In line with past reports, we defined this effect as statistical mediation since the confidence intervals did not cross zero. Put another way, in a regression model where total gray matter was entered as the dependent variable, and total brain volume and age were entered as the independent variables, the association between age and total gray matter was significant ($t=15.14$, $p<.005$); however, this association dropped to $t=13.176$ when image quality was entered as an additional independent variable.

5. DISCUSSION

The primary goals of this study were three-fold: 1) to examine if direct measures of structural imaging quality were associated with sociodemographic and behavioral variables of interest; 2) to investigate if there were associations between GMP and image quality; and 3) to evaluate the influence of image quality on associations between structural morphometrics and sociodemographic variables (specifically age). Connected to our first goal, we found significant associations between image quality and age, IQ, and scanner field strength. Image quality increased with general intelligence (IQ) and field strength, and decreased with age and BMI. Next, we show that a commonly derived structural MRI measure, GMP, was strongly related to

image quality. Specifically, we found substantial variations in GMP signal across regions of the Medial Frontal Gyrus, Precuneus, and bilateral Parietal Lobules. With motion likely strongly impacting image quality, this result has significant implications for studies of neurodevelopment, as well as structural MRI work in atypical populations, as motion artifacts are especially problematic for young children and clinical populations, who typically have difficulties remaining still during the time required to collect high-resolution data. Finally, related to our last goal (whether image quality impacts relations between age and structural morphometrics), we report two different critical sets of results. We completed voxel-wise analyses where associations between age and GMP were explored, and then these associations were examined when structural MRI quality was added as an independent variable. Here, we showed that controlling for image quality increased correlations between and GMP and age in the Medial Frontal Cortex, Precuneus, and bilateral Cerebellum. Further cementing the importance of image quality, we showed the associations between age and total gray matter was partially explained by image quality. These mediation analyses suggest a sizable amount of the relation between age and GMP could be explained by variation in image quality.

Contextualizing our results with past research reports, we find significant bivariate associations between image quality and factors such as general intelligence (IQ), and BMI, consistent with prior work (Beyer et al. 2017; Siegel et al. 2017). Furthermore, related to previous studies, we find image quality is related to derived measures of brain anatomy irrespective of typical quality threshold cut-offs. Even in structural scans of high quality (that “pass” visual inspection), in-scanner motion biases volumetric estimations. Indeed, accurate quantification of regional grey matter volume relies on reliable segmentation from high-

resolution MR images. Head motion during an MRI scan can bias segmentation, which in turn can impact morphometric measurements.

Our results have important implications when thinking about structural MRI, especially for studies attempting to center-in on individual differences. To our knowledge, we are one of the first studies to note that accounting for direct measures of structural MRI quality impact relations between variables of interest and the morphometrics typically generated from structural images. For example, other studies have used proxy measures for image quality derived from subject-motion during functional scans (Satterthwaite et al. 2012; Van Dijk et al. 2012; Yan et al. 2013). However, proxy measures for subject motion may be missing true differences obscured by motion (Rosen et al. 2018).

One open question is the best way to deal with variations in data quality and any potential effects on measures of brain anatomy. Data quality (related to subject-motion) may be collinear with other subject-level variables of interest, including general cognition (Siegel et al. 2017), body mass index, as well as group status (Yendiki et al. 2014). As such, group-level corrections (using covariates of no interest) have the potential to remove relevant signals and mask true relationships between variables of interest and brain measures (Power et al 2015; Rosen et al. 2018). Furthermore, here we linearly covary for image quality, but the effects of MRI quality could be nonlinear in nature (or be linear/nonlinear depending on the brain area in question; Despotović, Goossens, & Philips, 2015).

Considering our project, as well as past studies, our results suggest it will be important to consider image quality in future structural MRI analyses. In line with current work, studies interested in individual and/or group differences should flag/exclude scans of extremely poor quality. Furthermore, in the future, research groups may think about accounting for individual

differences in motion-related image quality by using direct measures of image quality as covariates in morphometric analyses. Such a strategy could address indirect effects of motion-related image quality and to confirm main effects for their variables of interest. However, as with any covariate of “no interest”, if motion is collinear with other variables, important variance related to factors of interest may be removed.

Of note, there are many important limitations of our data and our results that must be highlighted. First, we used a composite measure of image quality, informed by a logistic regression and then created through z-scoring. This may be influencing some of the results reported. MRIQC outputs multiple measures of image quality. These multiple measures may be capturing unique aspects of noise relevant for MRI morphometry. Studies in the future could take a more nuanced approach to these different measures, examining associations and influences on similar brain variables to what we reported here. Second, the public access dataset we used here, the Healthy Brain Network, is not a truly random sample. The dataset has a limited age range (5-21 years of age) and also employs a community-referred recruitment model. Study advertisements are specifically targeting families who have concerns about one or more psychiatric symptoms in their child. This fact may explain the puzzling finding of a negative association between age and image quality. Third, voxel-based morphometry is only one approach to deriving measures from structural MRI scans. VBM is highly dependent on tissue segmentation and spatial normalization. These preprocessing steps may be particularly impacted by image quality. However, even more complex structural analytic tools, such as Freesurfer, depend on tissue segmentation and will therefore likely be influenced by image quality. It will be important in future work to potentially see if direct measures of structural MRI quality impact

cortical thickness or other measures derived in Freesurfer (or from other structural analytic software suites).

6. CONCLUSIONS

These limitations, notwithstanding, we demonstrate that direct measures of structural imaging quality are strongly linked to commonly used structural derivatives of gray matter volume, as well as demographics and variables of interest (e.g., age; measures of cognitive functioning; psychopathology). Importantly, we show that variations in image quality are strongly related to measures of brain anatomy. Furthermore, we show that accounting for variations in image quality impacts previous reports of age dependent differences in gray matter morphometrics. Unique to the work, we used direct measures of structural MRI quality rather than proxies of motion and noise. In the future, research groups may consider accounting for such measures in analyses focused on individual differences in age, cognitive functioning, psychopathology, and other factors. This may lead to greater reproducibility in reported effects, as well as a way to minimize any potential spurious associations.

Software and availability:

MRIQC is developed by the Poldrack Lab at Stanford University for use at the Center for Reproducible Neuroscience (CRN), as well as for open-source software distribution. More information can be found at https://mriqc.readthedocs.io/en/stable/index.html

ANTs is open source. Code, binaries and documentation are available at http://stnava.github.io/ANTs/

AFNI It is freely available (both in source code and in precompiled binaries) for research purposes. More information can be found at https://afni.nimh.nih.gov.

Shed – Automated parcellation of functional activation maps is a Matlab program designed to partition statistical activation maps into separate regions based on local peaks and cluster extents. More information can be found at from https://www.med.upenn.edu/cmroi/shed.html.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This work was supported by start-up funds from the University of Pittsburgh.

References

Achenbach, T. (1992). *Manual for the Child Behavior Checklist/2-3 and 1992 Profile*. Burlington, VT: University of Vermont Department of Psychiatry.

Atkinson, D., Hill, D. L., Stoyle, 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

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage, 54*(3), 2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025

Avants BB, Song G, Duda JT, Johnson HJ, Tustison N. ANTs: Advanced Normalization Tools; 2013. Available from: http://www.picsl.upenn.edu/ANTS/

Avants, B. B., Tustison, N., & Song, G. (2009). Advanced normalization tools (ANTS). *The Insight Journal, 2*, 1–35.

Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., & Gee, J. C. (2011). An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics, 9*(4), 381–400. https://doi.org/10.1007/s12021-011-9109-y

[dataset] Alexander, L. M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., . . . Milham, M. P. (2017). An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data, 4*, 170181. https://doi.org/10.1038/sdata.2017.181

Cox, R. W. (1996). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. Computers and Biomedical Research, 29(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Despotović, I., Goossens, B., & Philips, W. (2015). MRI segmentation of the human brain: Challenges, methods, and applications. *Computational and Mathematical Methods in Medicine,2015*, 450341. https://doi.org/10.1155/2015/450341

Dietrich, O., Raya, J. G., Reeder, S. B., Reiser, M. F., & Schoenberg, S. O. (2007). Measurement of signal-to-noise ratios in MR images: Influence of multichannel coils, parallel imaging, and reconstruction filters. *Journal of Magnetic Resonance Imaging, 26*(2), 375–385. https://doi.org/10.1002/jmri.20969

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

Fair, D.A., Posner, J., Nagel, B.J., Bathula, D., Dias, T.G.C., Mills, K.L., Blythe, M.S., Giwa, A., Schmitt, C.F., Nigg, J.T., 2010. Atypical default network connectivity in youth with attention- deficit/hyperactivity disorder. Biological psychiatry 68, 1084–91.

Friedman, L., Glover, G. H., Krenz, D., Magnotta, V., & First, B., & the FIRST BIRN. (2006). Reducing inter-scanner variability of activation in a multicenter fMRI study: Role of smoothness equalization. *NeuroImage, 32*(4), 1656–1668. https://doi.org/10.1016/j.neuroimage.2006.03.062

Hollingshead, A. A. (1975). Four-factor index of social status. Unpublished manuscript, Yale University, New Haven, CT.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., . . . Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*(7), 980–988. https://doi.org/10.1097/00004583-199707000-00021

Kong, X. Z., Zhen, Z., Li, X., Lu, H. H., Wang, R., Liu, L., . . . Liu, J. (2014). Individual differences in impulsivity predict head motion during magnetic resonance imaging. *PLoS One,9*(8), e104989. https://doi.org/10.1371/journal.pone.0104989

Klapwijk, E. T., van de Kamp, F., van der Meulen, M., Peters, S., & Wierenga, L. M. (2019). *Qoala-T: A supervised-learning tool for quality control of FreeSurfer segmented MRI data. NeuroImage, 189, 116–129.* https://doi.org/10.1016/j.neuroimage.2019.01.014

Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M. C., . . . Parsey, R. V. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage, 46*(3), 786–802. https://doi.org/10.1016/j.neuroimage.2008.12.037

Magnotta, V. A., Friedman, L., & First, B., & the FIRST BIRN. (2006). Measurement of signal-to-noise and contrast-to-noise in the fBIRN multicenter imaging study. *Journal of Digital Imaging, 19*(2), 140–147. https://doi.org/10.1007/s10278-006-0264-x

Mennes, M., Biswal, B. B., Castellanos, F. X. & Milham, M. P. Making data sharing work: the FCP/INDI experience. Neuroimage 82, 683–691 (2013).

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., & Uğurbil, K. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel

imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic Resonance in Medicine, 63*(5), 1144–1153. https://doi.org/10.1002/mrm.22361

Mortamet, B., Bernstein, M. A., Jack, C. R., Jr., Gunter, J. L., Ward, C., Britson, P. J., . . . Krueger, G., & the Alzheimer's Disease Neuroimaging Initiative. (2009). Automatic quality assessment in structural brain magnetic resonance imaging. *Magnetic Resonance in Medicine, 62*(2), 365–372. https://doi.org/10.1002/mrm.21992

Nichols, E. A., Kao, Y. C., Verfaellie, M., & Gabrieli, J. D. (2006). Working memory and long-term memory for faces: Evidence from fMRI and global amnesia for involvement of the medial temporal lobes. *Hippocampus, 16*(7), 604–616. https://doi.org/10.1002/hipo.20190

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., . . . Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron, 72*(4), 665–678. https://doi.org/10.1016/j.neuron.2011.09.006

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage, 59*(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage,84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Power, J. D., Schlaggar, B. L., & Petersen, S. E. (2015). Recent progress and outstanding issues in motion correction in resting state fMRI. *NeuroImage, 105*, 536–551. https://doi.org/10.1016/j.neuroimage.2014.10.044

Radua, J., Canales-Rodríguez, E. J., Pomarol-Clotet, E., & Salvador, R. (2014). Validity of modulation and optimal settings for advanced voxel-based morphometry. *NeuroImage, 86*, 81–90. https://doi.org/10.1016/j.neuroimage.2013.07.084

Reuter, M., Tisdall, M. D., Qureshi, A., Buckner, R. L., van der Kouwe, A. J. W., & Fischl, B. (2015). Head motion during MRI acquisition reduces gray matter volume and thickness estimates. *NeuroImage, 107*, 107–115. https://doi.org/10.1016/j.neuroimage.2014.12.006

Rosen, A.F.G., Roalf, D.R., Ruparel, K., Blake, J., Seelaus, K., Villa, L.P., Ciric, R., Cook, P.A., Davatzikos, C., Elliott, M.A., Garcia de La Garza, A., Gennatas, E.D., Quarmley, M., Schmitt, J.E., Shinohara, R.T., Tisdall, M.D., Craddock, R.C., Gur, R.E., Gur, R.C., Satterthwaite, T.D., Data-driven assessment of structural image quality, *NeuroImage* (2018). https://doi.org/10.1016/j.neuroimage.2017.12.059

Rosseel, Y. (2012). lavaan : An *R* Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Savalia, N. K., Agres, P. F., Chan, M. Y., Feczko, E. J., Kennedy, K. M., & Wig, G. S. (2017). Motion-related artifacts in structural brain images revealed with independent estimates of in-scanner head motion. *Human Brain Mapping, 38*(1), 472–492. https://doi.org/10.1002/hbm.23397

Satterthwaite, T. D., Ciric, R., Roalf, D. R., Davatzikos, C., Bassett, D. S., & Wolf, D. H. (2017). *Motion artifact in studies of functional connectivity: Characteristics and mitigation strategies. Human Brain* Mapping. ***https://doi.org/10.1002/hbm.23665***

Satterthwaite TD, Elliott MA, Gerraty RT, Ruparel K, Loughead J, Calkins ME, Eickhoff SB, *Hakonarson H, Gur RC, Gur RE, Wolf DH.* An improved framework for confound

regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data.
Neuroimage. 2013 Jan 1; 64():240-56.

Satterthwaite TD, Wolf DH, Loughead J, Ruparel K, Elliott MA, Hakonarson H, Gur RC, Gur *RE*. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. *Neuroimage. 2012 Mar; 60(1):623-32.*

Scott, A., Courtney, W., Wood, D., de la Garza, R., Lane, S., King, M., … Calhoun, V. D. (2011). COINS: An Innovative Informatics and Neuroimaging Tool Suite Built for Large Heterogeneous Datasets. Frontiers in Neuroinformatics, 5. https://doi.org/10.3389/fninf.2011.00033

Siegel, J. S., Mitra, A., Laumann, T. O., Seitzman, B. A., Raichle, M., Corbetta, M., & Snyder, A. Z. (2017). Data Quality Influences Observed Links Between Functional Connectivity and Behavior. *Cerebral cortex (New York, N.Y.: 1991)*, *27*(9), 4492–4502. https://doi.org/10.1093/cercor/bhw253

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., . . . Avants, B. B. (2014). Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage*, *99*, 166–179. https://doi.org/10.1016/j.neuroimage.2014.05.044

Wechsler, D. Wechsler intelligence scale for children--Fourth Edition (WISC-IV) (The Psychological Corporation, 2003).

Van Dijk KR, Sabuncu MR, Buckner RL. The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage. 2012 Jan 2; 59(1):431-8.*

Yan CG, Cheung B, Kelly C, Colcombe S, Craddock RC, Di Martino A, Li Q, Zuo XN, Castellanos FX, Milham MP. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *Neuroimage. Aug 1; 76:183-201*.

Web References:

Elliott, M. (n.d.). Center for Magnetic Resonance & Optical Imaging. (accessed February 7, 2019, from https://www.med.upenn.edu/cmroi/shed.html).

Oscar Esteban, Krzysztof F. Gorgolewski. Poldrack Lab, Psychology Department, Stanford University, and Center for Reproducible Neuroscience, Stanford University. IQMs for structural images. (n.d.). (accessed February 7, 2019, from https://mriqc.readthedocs.io/en/stable/iqms/t1w.html).

The Scientific and Statistical Computing Core (SSCC) of the NIMH Intramural Research Program. (n.d.). (accessed February 7, 2019, https://afni.nimh.nih.gov).