Variations in Structural MRI Quality Significantly Impact Commonly-Used Measures of Brain Anatomy

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

Abstract:

Subject motion can introduce noise into neuroimaging data and result in biased estimations of brain structure. In-scanner motion can compromise data quality in a number of ways and varies widely across developmental and clinical populations. However, quantification of structural image quality is often limited to proxy and indirect measures gathered from functional scans; this may be missing true differences related to these potential artifacts. In this study, we take advantage of novel informatic tools, the CAT12 toolbox, to directly measure image quality from T1-weighted images to understand if these measures of image quality: 1) relate to rigorous quality-control checks visually completed by human raters; 2) are associated with sociodemographic variables of interest; 3) influence regional estimates of cortical thickness and subcortical volumes from the commonly-used Freesurfer tool suite. We leverage public-access data that includes a community-based sample of children and adolescents, spanning a large age-range (*n=388; ages 5-21*). Interestingly, even after visually inspecting our data, we find image quality significantly impacts derived cortical thickness and subcortical volumes from multiple regions across the brain (~44% of the regions investigated). We believe these results our important for research groups completing structural MRI studies using Freesurfer or other morphometric tools. As such, future studies should consider using direct measures of image quality to minimize the influence of this potential confound in group comparisons or studies focused on individual differences.

Keywords: *Structural MRI; image quality; Freesurfer; 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. One of the most common limitations of neuroimaging research is that of motion-related artifacts. This type of noise is caused by participant movement during a neuroimaging session and may impact assessment of brain structure and function (Power et al., 2014, 2012; Satterthwaite et al., 2012; van Dijk et al., 2012). For those interested in neurodevelopment and mental health, such noise and bias may be particularly important to address. While head motion varies considerably among individuals, children typically move more than adults and patient groups move on average more than controls (Kong et al., 2014; Yendiki et al., 2014).

Multiple resting state fMRI studies have highlighted the importance of this issue, as incredibly small differences in motion have been shown to yield significant differences in estimates of functional connectivity among healthy samples (Power et al., 2012; van Dijk et al., 2012). In fact, head movements within fractions of a millimeter have been shown to significantly bias correlations 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., 2012). 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 (Hodgson et al., 2017). Particularly challenging, these differences persist even after extensive motion correction procedures (Ciric et al., 2017; Power et al., 2015). This has thus motivated a methodological sub-

field focused on effective ways to reduce motion-related noise in resting-state and other forms of functional MRI.

While a great deal of progress has been made in quantifying and addressing the impact of head-motion in functional analyses, less attention has been given to structural MRI. It is, however, clear that head motion has been shown to compromise derived measures of volume and thickness in regions of cortical gray matter (Alexander-Bloch et al., 2016; Pardoe et al., 2016; Reuter et al., 2015). Such effects remain after automated correction, suggesting that in-scanner motion induces spurious effects that do not reflect a processing failure in software; rather, they reflect systematic bias (e.g., such as motion-induced blurring) and this may appear similar to gray matter atrophy (Reuter et al., 2015). These may be particularly important issues to examine in youth and/or clinical populations.

While the impact of movement on structural MRI is clear, 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 raters visually assessing 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-Bloch et al., 2016; Pardoe et al., 2016). Also, human rating systems are relatively impractical for large scale datasets (Esteban et al. 2017). A further challenge under discussed in the research literature, is that visual inspection by human raters is relatively subjective. Numerous studies have showcased this, with moderately concerning inter- and intra- related variability among human-rating

systems (Klapwijk et al., 2019). 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” (Alexander-Bloch et al., 2016).

Thinking holistically, these multiple problems are in part due to the limited information about noise typically available for structural MRI scans. Structural MRI involves the acquisition of only one, higher resolution volume. To date, this has prohibited rich assessments of noise and subject movement in contrast to fMRI. 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. The ease in collection of this sort of data has led some to advocate for the use of fMRI-derived motion parameters, such as mean Framewise Displacement (FD), to identify structural brain scans that contain motion-related bias. Recent work has showed that by additionally removing FD outliers from a sample of visually inspected T1-weighted images, the effect sizes of age and gray matter thickness were attenuated across a majority of the cortex (Savalia et al., 2017). It is, therefore, possible that some past results of associations between participant variables and brain morphometry may be inaccurate, likely particularly inflated 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.

While past structural MRI studies have suffered from the limitations noted above, advancements of novel informatic tools may overcome these issues. Quality assessment tools have been recently introduced that provide easy-to-implement, automated, quantitative measures of structural neuroimaging volumes. For example, the MRI Quality Control tool (MRIQC) has recently been introduced and can speak to different quality attributes of structural (and other MRI) images (Esteban et al., 2017). Similarly, the Computational Anatomy Toolbox for SPM (CAT12) assesses multiple image quality metrics and provides an aggregate “grade” for a given structural MRI scan (Dahnke and Gaser, 2016). Thinking about past research, it is unclear if structural MRI quality is related to commonly derived structural measures (e.g., cortical thickness; regional subcortical volumes). Thoughtful work by Rosen and colleagues (Rosen et al., 2018) began to investigate this idea. These researchers found that metrics from Freesurfer, specifically Euler number, were consistently correlated with human raters’ assessments of image quality. Furthermore, Euler number, a summary statistic of the topological complexity of a reconstructed brain surface, was significantly related to cortical thickness.

While important, one of Rosen and colleagues’ (Rosen et al., 2018) major results could be described as “circular” in nature-- a measure of Freesurfer re-construction (Euler number) is related to measures output by Freesurfer (cortical thickness). In theory, inaccuracy or variability of Freesurfer re-construction could be due to MR quality and/or algorithmic issues. The use of an independent measure of quality in relation to Freesurfer outputs would provide stronger evidence of the potential impact of MRI quality on morphometric measures. In addition, Rosen and colleagues did not

investigate if Euler number, their measure of MR quality, was related to subcortical (e.g., amygdala) volumes. Given the major interest from cognitive and affective neuroscientists in these areas (e.g., (Caldwell et al., 2015; Hanson et al., 2015), it will be important to know if MRI quality impacts volumetric variations in these structures. Accounting for such variations may be important in reducing potential spurious associations and increasing the replicability of effects.

To these ends, we investigated three key questions: 1) We sought to look if an integrated measure of image quality, output by the CAT12 toolbox, uniquely related to visual rater judgement (retain/exclude) of structural MRI images; 2) We examined if variations in image quality related to sociodemographic and psychosocial variables (e.g., age; sex; clinical diagnosis); 3) we evaluated if CAT12 image quality was associated with differences in commonly-used structural measures derived from Freesurfer (both cortical thickness and subcortical volume).

2. MATERIALS AND METHODS

2.1 Participants.

Data from 388 participants between the ages of 5-21 years of age with T1-weighted structural images were downloaded from two data waves of an ongoing research initiative, The Healthy Brain Network (HBN), launched by The Child Mind Institute in 2015. For sample characteristics, see our Table 1. 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. For

additional information about the full HBN sample and measures, please see the HBN data-descriptor (Alexander et al., 2017).

2.2 MRI Data Acquisition.

MRI acquisition included structural MRI (T1, T2-space), magnetization transfer imaging, and quantitative T1- and T2-weighted mapping. Here, we focused on only T1-weighted structural MRI scans. A Siemens 3-Telsa Tim Trio MRI scanner located at the Rutgers University Brain Imaging Center (RU) was equipped with a Siemens 32-channel head coil. T1-weighted scans were acquired with a Magnetization Prepared - RApid Gradient Echo (MPRAGE) sequence with the following parameters: 224 Slices, 0.8×0.8×0.8 mm resolution, TR=2500 ms, TE=3.15 ms, and Flip Angle=8°. All neuroimaging data used in this study are openly available for download with proper data usage agreement via the International Neuroimaging Data-sharing Initiative (http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/). Again, please see the HBN data descriptor for more information for additional information (Alexander et al., 2017).

2.3 Visual Quality Inspection.

All T1-weighted scans were separated by release wave then visually inspected by a series of human raters that were 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”. Examples of this protocol are detailed in our *Supplemental Materials*. 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 any rater idiosyncrasy, all ratings were z-scored (within rater), 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 (passing n=209) and the rest were removed from further analysis.

2.4 Image Quality Metrics.

The CAT12 toolbox (Computational Anatomy Toolbox 12) from the Structural Brain Mapping group, implemented in Statistical Parametric Mapping, was used to generate a quantitative metric indicating the quality of each collected MR-image (Dahnke and Gaser, 2016; Gaser and Kurth, 2017). The method employed considers four summary measures of image quality: 1) noise to contrast ratio, 2) coefficient of joint variation, 3) inhomogeneity to contrast ratio, and 4) Root mean squared voxel resolution. To produce a single aggregate metric that serves as an indicator of overall quality, this toolbox normalizes each measure and combines them using a kappa statistic-based framework, for optimizing a generalized linear model through solving least squares (Dahnke et al., 2015). This measure ranged from 0-1, with higher values indicating better image quality. Additional information is available at: http://www.neuro.uni-jena.de/cat/index.html#QA. Quality assessment for one MRI scan could not be completed through the CAT12 toolbox due to excessive noise.

2.5 Sociodemographic, Cognitive, and Psychiatric Measures

Sociodemographic (self-report), cognitive, and psychiatric data was assessed through the COllaborative Informatics and Neuroimaging Suite (COINS) Data Exchange after completion of appropriate data use agreements. We chose a number of measures relevant to research groups that we believed may covary with structural MRI quality. Motivated by past studies, these included: age, sex, body mass index (BMI), general cognitive ability (IQ), and clinical diagnoses. The Wechsler Intelligence Scale for Children (WISC-V) was used as a measure of general cognitive ability (IQ) and was completed on 336 participants in the sample; the WISC-V is an individually administered clinical instrument for assessing the intelligence of youth participants 6-16 and generates a general cognitive ability score (Full-Scale Intelligence Quotient; FSIQ). Related to clinical diagnoses, the presence of psychopathology was assessed by a certified clinician using semi-structured DSM-5-based psychiatric interview (i.e., the Schedule for Affective Disorders and Schizophrenia for Children; KSADS-COMP). This data was available for 367 participants in our sample. Mean, standard deviation, and ranges for all the sociodemographic, cognitive, and psychiatric measures are noted in Table 1. Additional information about these measures are noted in our *Supplemental Materials*.

3.1 Image Pre/processing (Freesurfer)

Standard-processing approaches from Freesurfer (e.g., cortical reconstruction; volumetric segmentation) were performed in version 6.0. Freesurfer is a widely-documented and freely available morphometric processing tool suite (http://surfer.nmr.mgh.harvard.edu/). The technical details of these procedures are

described in prior publications (Dale et al., 1999; Fischl et al., 2004a, 2004b, 2002, 1999a, 1999b). Briefly, this processing includes motion correction and intensity normalization of T1 weighted images, removal of non-brain tissue using a hybrid watershed/surface deformation procedure (Ségonne et al., 2004), automated Talairach transformation, segmentation of the subcortical white matter and deep gray matter volumetric structures (including hippocampus, amygdala, caudate, putamen, ventricles), tessellation of the gray matter white matter boundary, and derivation of cortical thickness. Of note, the "recon-all" pipeline with the default set of parameters (no flag options) was used and no manual editing was conducted. After successful processing, we extracted volumes from subcortical structures, as well as mean cortical thickness for the 34 bilateral Desikan-Killiany (DK) atlas regions (Desikan et al., 2006). Scans from four participants did not complete processing in Freesurfer due to technical issues; this brought the total sample size (passed visual inspection, and with Freesurfer processing completed) to n=205.

3.2 Statistical Modeling

We first constructed logistic regression models that used an aggregated measure of image quality from the CAT12 toolbox and the outcome of passing or failing visual quality assurance checks completed by trained human raters. Receiver operating characteristic curves were computed to understand true positive (sensitivity) and false positive rates. For these receiver operating characteristic measures, the area under the curve (AUC) was computed to show classification performance at all classification thresholds (and distinguishing between classes of passing or failing visual quality

assurance checks). Bayesian logistic models were also constructed to probe potential over-fitting and biases common to Frequentist logistic models (Christmann and Rousseeuw, 2001). Next, bivariate correlations were calculated to examine relations between our image quality and sociodemographic variables of interest, including age, sex, IQ, BMI, and clinical diagnosis. Finally, we computed 84 bivariate correlations between image quality and Freesurfer outputs (68 mean cortical thickness estimates from the DK atlas; 16 subcortical regions). Of note, cerebral spinal fluid Freesurfer subcortical outputs (e.g., Lateral ventricle; Left-choroid-plexus) were excluded from analyses. Given the number of statistical tests conducted and to further reproducibility, we adjusted all p-values of this last step based on the Benjamini & Hochberg False Discovery Rate Correction (Benjamini and Hochberg, 1995). This commonly-used approach has been shown to have appropriate power to detect true positives, while still controlling the proportion of type I errors at a specified level (α=.05).

4. RESULTS

4.1 Relations Between Structural MRI Quality and Visual Rejection/Acceptance of Structural Images.

Logistic regression was used to examine relationships between our aggregated MRI quality measure and the outcome of passing or failing quality assurance checks completed by trained human raters. Logistic regression models indicated that structural MRI quality, derived by the CAT12 toolbox) was significantly related to passing or failing quality assurance checks completed by trained human raters (*$z=7.877$, $p<.005$; Nagelkerke's $R^2=0.8951$).* This indicated that greater CAT12 MRI quality was related to

a higher likelihood of passing visual inspection. Bayesian GLM modeling suggested a similar relation, with higher MRI quality significantly relating to passing visual checks (*z=8.141, p<.005).* Receiver operating characteristic analyses indicated a mean AUC of 98.9% (with 95% confidence intervals spanning 98.2-99.6%, as shown in Figure 1).

---- Figure 1 Approximately Here ---

4.2 Bivariate Correlations Between Image Quality and Sociodemographic Variables of Interest.

We next examined correlations between image quality, sociodemographic variables of interest (e.g., age, sex, BMI, and clinical diagnosis). As expected and in line with other reports, image quality was related to age (r=0.321, p<.005; as shown in Figure 2). Older subjects typically had better quality scans. Interestingly, no other sociodemographic factors were significantly related to image quality (Sex p=0.196; BMI p=.227; Clinical Diagnosis [binary indicator] p=.189). The BMI finding is in contrast to past results reported in adults (e.g., (Hodgson et al., 2017; Siegel et al., 2017). There was a trend association for image quality and IQ (r=.101, p= 0.06), with high IQ relating to better image quality. Of note, this is for all participants (not only those passing human rater visual inspection). If associations are investigated in only those passing visual inspection, the association with age and image quality remains significant (p= 0.036). All other associations remained non-significant (all p's >.3).

---- Figure 2 Approximately Here ---

4.3 Associations Between Freesurfer Outputs and Structural MRI Quality.

We next examined correlations between structural MRI quality and 84 morphometric outputs from Freesurfer (68 mean cortical thickness estimates from the DK atlas; 16 subcortical regions). Surprisingly, structural MRI quality was related to estimates for 37 of the regions (44.05%, as listed in Table 2). Of note, this is after multiple comparison correction. Thirty-one cortical sub-divisions related to image quality, all showing positive relations ranging from β=0.161 to β=0.523. For these areas, with greater MRI image quality, higher mean cortical thickness was found. These multiple cortical regions are shown in Figure 3. The volumes of our subcortical regions were also related to MRI image quality. Six had positive associations (3 subdivisions of the corpus callosum, Left Amygdala, and the Left and Right Putamen), ranging from β=0.160 to β=0.243. These are shown in Figure 4. Example scatterplots depict associations between image quality and cortical thickness and subcortical volume in Figure 5A and 5B.

---- Table 2 Approximately Here ---

---- Figures 3-5 Approximately Here ---

5. DISCUSSION

The primary goals of this study were two-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 commonly-

used Freesurfer outputs and MRI quality. Connected to this first goal, we first found significant associations between image quality and age; there were, however, no relations between IQ, BMI, sex, or clinical diagnosis. Next, we demonstrated commonly-derived structural MRI measures were strongly related to image quality. Even after correcting for multiple comparisons, numerous measurements of cortical thickness and subcortical volumes were connected to image quality. This was for a large percentage (44%) of the brain regions investigated, suggesting diffuse, but significant, impacts of image quality on structural morphometric measures. This result has significant implications for studies of neurodevelopment and other applied work using structural MRI, as motion artifacts are especially problematic for young children and clinical populations; these groups may have difficulties remaining still during the time required to collect high-resolution neuroimaging data.

Contextualizing our results with past research reports, we find significant bivariate associations between image quality and age. However, we did not find associations between image quality and factors such as general intelligence (IQ), and BMI. Such findings are in contrast to a few prior publications (Hodgson et al., 2017; Siegel et al., 2017). This may be due to the age-range of our sample (5-21 years of age), while those relevant past studies have been primarily completed in adult samples . Building off of previous studies, we find image quality is related to derived measures of brain anatomy, irrespective of typical (binary) quality threshold cut-offs. Even in structural scans of high quality (that “pass” visual inspection), in-scanner motion appears to influence 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 that has used direct measures of structural MRI quality. Other past studies (e.g., (Pardoe et al., 2016; Savalia et al., 2017) have used measures derived from resting state; while this is a measure of participant movement, it is not specifically during the (structural) MRI scan. Furthermore, this type of information may not be available for all studies, but the measure we employ here could be derived for any T1-weighted scan. Using this more direct measure of MRI quality, we found impacts on morphometric variables typically generated from structural images. For example, other studies have used proxy measures for image quality derived from subject-motion during functional scans (Alexander-Bloch et al., 2016; Savalia et al., 2017). However, proxy measures for subject motion may be missing true differences obscured by motion (Rosen et al. 2018). Our findings build off of past work by Rosen and colleagues' that found Freesurfer Euler number was related to Freesurfer cortical thickness volumes. Here, however, we used an independent metric of image quality (derived from the CAT12 toolbox) and examined correlations with this measure and commonly-used Freesurfer outputs. This use of an independent image quality metric provides stronger evidence of the impact of image quality on volume and cortical thickness. Furthermore, we detail associations between image quality and subcortical volumes. Of particular interest to those studying emotion, we find that volumetric

measures of amygdala and putamen were related to image quality; specifically, greater image volume is related to higher volumes in these regions.

One open question is how best to deal with variations in data quality and the 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, body mass index, as well as group status (Hodgson et al., 2017; Siegel et al., 2017; 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 (Satterthwaite et al., 2012).While the inclusion of image quality in statistical models as a variable of no-interest could be a useful starting approach, the effects of MRI quality could be non-linear in nature (or be linear/non-linear depending on the brain area in question; (Despotović et al., 2015). Additional studies are needed to explore this idea.

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, constructed in the CAT12 toolbox. This may be influencing some of the results reported. There are many metrics of image quality, each potentially capturing unique aspects of noise relevant for MRI morphometry. We relied on an aggregated metric that combined noise to contrast ratio, coefficient of joint variation, inhomogeneity to contrast ratio, and root mean squared voxel resolution. Studies in the future could take a more nuanced approach to these different measures (for additional information, see https://mriqc.readthedocs.io/en/stable/iqms/t1w.html), 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. Third, Freesurfer is only one approach to deriving measures from structural MRI scans. Other metrics, such as voxel-based morphometry or region of interest drawing, may be similarly impacted by image quality. These approaches, however, often depend on tissue segmentation and would likely also be influenced by image quality.

6. CONCLUSIONS

These limitations, notwithstanding, we demonstrate that direct measures of structural imaging quality are strongly linked to commonly-used structural MRI measures, as well as participant age. Importantly, we show that variations in image

quality are strongly related to derivation of brain anatomy. Accounting for variations in image quality could impact results from applied studies (focused on age, clinical status, etc.). 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.

Acknowledgements

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

Competing interests: The authors declare that they have no competing interests.

References

Alexander-Bloch, A., Clasen, L., Stockman, M., Ronan, L., Lalonde, F., Giedd, J., Raznahan, A., 2016. Subtle in-scanner motion biases automated measurement of brain anatomy from in vivo MRI. Hum. Brain Mapp. 37, 2385–2397. https://doi.org/10.1002/hbm.23180

Alexander, L.M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega-Potler, N., Langer, N., Alexander, A., Kovacs, M., Litke, S., O'Hagan, B., Andersen, J., Bronstein, B., Bui, A., Bushey, M., Butler, H., Castagna, V., Camacho, N., Chan, E., Citera, D., Clucas, J., Cohen, S., Dufek, S., Eaves, M., Fradera, B., Gardner, J., Grant-Villegas, N., Green, G., Gregory, C., Hart, E., Harris, S., Horton, M., Kahn, D., Kabotyanski, K., Karmel, B., Kelly, S.P., Kleinman, K., Koo, B., Kramer, E., Lennon, E., Lord, C., Mantello, G., Margolis, A., Merikangas, K.R., Milham, J., Minniti, G., Neuhaus, R., Levine, A., Osman, Y., Parra, L.C., Pugh, K.R., Racanello, A., Restrepo, A., Saltzman, T., Septimus, B., Tobe, R., Waltz, R., Williams, A., Yeo, A., Castellanos, F.X., Klein, A., Paus, T., Leventhal, B.L., Craddock, R.C., Koplewicz, H.S., Milham, M.P., 2017. Data Descriptor: An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci. Data 4. https://doi.org/10.1038/sdata.2017.181

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Caldwell, J.Z.K., Armstrong, J.M., Hanson, J.L., Sutterer, M.J., Stodola, D.E., Koenigs, M., Kalin, N.H., Essex, M.J., Davidson, R.J., 2015. Preschool externalizing

behavior predicts gender-specific variation in adolescent neural structure. PLoS One 10. https://doi.org/10.1371/journal.pone.0117453

Christmann, A., Rousseeuw, P.J., 2001. Measuring overlap in binary regression. Comput. Stat. Data Anal. 37, 65–75. https://doi.org/10.1016/S0167-9473(00)00063-3

Ciric, R., Wolf, D.H., Power, J.D., Roalf, D.R., Baum, G.L., Ruparel, K., Shinohara, R.T., Elliott, M.A., Eickhoff, S.B., Davatzikos, C., Gur, R.C., Gur, R.E., Bassett, D.S., Satterthwaite, T.D., 2017. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. Neuroimage 154, 174–187. https://doi.org/10.1016/j.neuroimage.2017.03.020

Dahnke, R., Gaser, C., 2016. CAT-A Computational Anatomy Toolbox for the Analysis of Structural MRI Data, in: 22nd Annual Meeting of the Organization For Human Brain Mapping.

Dahnke, R., Ziegler, G., Grosskreutz, J., Gaser, C., 2015. Quality Assurance in Structural MRI. 21st Annu. Meet. Organ. Hum. Brain Mapp. 1556.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 9, 179–194. https://doi.org/10.1006/nimg.1998.0395

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31, 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Despotović, I., Goossens, B., Philips, W., 2015. MRI segmentation of the human brain: Challenges, methods, and applications. Comput. Math. Methods Med. https://doi.org/10.1155/2015/450341

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, e0184661. https://doi.org/10.1371/journal.pone.0184661

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355. https://doi.org/10.1016/S0896-6273(02)00569-X

Fischl, B., Salat, D.H., Van Der Kouwe, A.J.W., Makris, N., Ségonne, F., Quinn, B.T., Dale, A.M., 2004a. Sequence-independent segmentation of magnetic resonance images. Neuroimage 23. https://doi.org/10.1016/j.neuroimage.2004.07.016

Fischl, B., Sereno, M.I., Dale, A.M., 1999a. Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207. https://doi.org/10.1006/nimg.1998.0396

Fischl, B., Sereno, M.I., Tootell, R.B.H., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284. https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4

Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H.,

Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004b. Automatically Parcellating the Human Cerebral Cortex. Cereb. Cortex 14, 11–22. https://doi.org/10.1093/cercor/bhg087

Gaser, C., Kurth, F., 2017. Manual Computational Anatomy Toolbox-CAT12, in: Structural Brain Mapping Group at the Departments of Psychiatry and Neurology, University of Jena.

Hanson, J.L., Nacewicz, B.M., Sutterer, M.J., Cayo, A.A., Schaefer, S.M., Rudolph, K.D., Shirtcliff, E.A., Pollak, S.D., Davidson, R.J., 2015. Behavioral problems after early life stress: Contributions of the hippocampus and amygdala. Biol. Psychiatry 77, 314–323. https://doi.org/10.1016/j.biopsych.2014.04.020

Hodgson, K., Poldrack, R.A., Curran, J.E., Knowles, E.E., Mathias, S., Göring, H.H.H., Yao, N., Olvera, R.L., Fox, P.T., Almasy, L., Duggirala, R., Barch, D.M., Blangero, J., Glahn, D.C., 2017. Shared Genetic Factors Influence Head Motion During MRI and Body Mass Index. Cereb. Cortex 27, 5539–5546. https://doi.org/10.1093/cercor/bhw321

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

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

Pardoe, H.R., Kucharsky Hiess, R., Kuzniecky, R., 2016. Motion and morphometry in

clinical and nonclinical populations. Neuroimage. https://doi.org/10.1016/j.neuroimage.2016.05.005

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, 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2011. Functional Network Organization of the Human Brain. Neuron 72, 665–678. https://doi.org/10.1016/j.neuron.2011.09.006

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

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., 2018. Quantitative assessment of structural image quality. Neuroimage 169, 407–418. https://doi.org/10.1016/j.neuroimage.2017.12.059

Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., Gur, R.E., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. Neuroimage 60, 623–632. https://doi.org/10.1016/j.neuroimage.2011.12.063

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. Hum. Brain Mapp. 38, 472–492. https://doi.org/10.1002/hbm.23397

Ségonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage 22, 1060–1075. https://doi.org/10.1016/j.neuroimage.2004.03.032

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. Cereb. Cortex 27, 4492–4502. https://doi.org/10.1093/cercor/bhw253

van Dijk, K.R.A., Sabuncu, M.R., Buckner, R.L., 2012. The influence of head motion on intrinsic functional connectivity MRI. Neuroimage 59, 431–438. https://doi.org/10.1016/j.neuroimage.2011.07.044

Yendiki, A., Koldewyn, K., Kakunoori, S., Kanwisher, N., Fischl, B., 2014. Spurious group differences due to head motion in a diffusion MRI study. Neuroimage 88, 79–

90. https://doi.org/10.1016/j.neuroimage.2013.11.027

Table 1.

Demographic Table

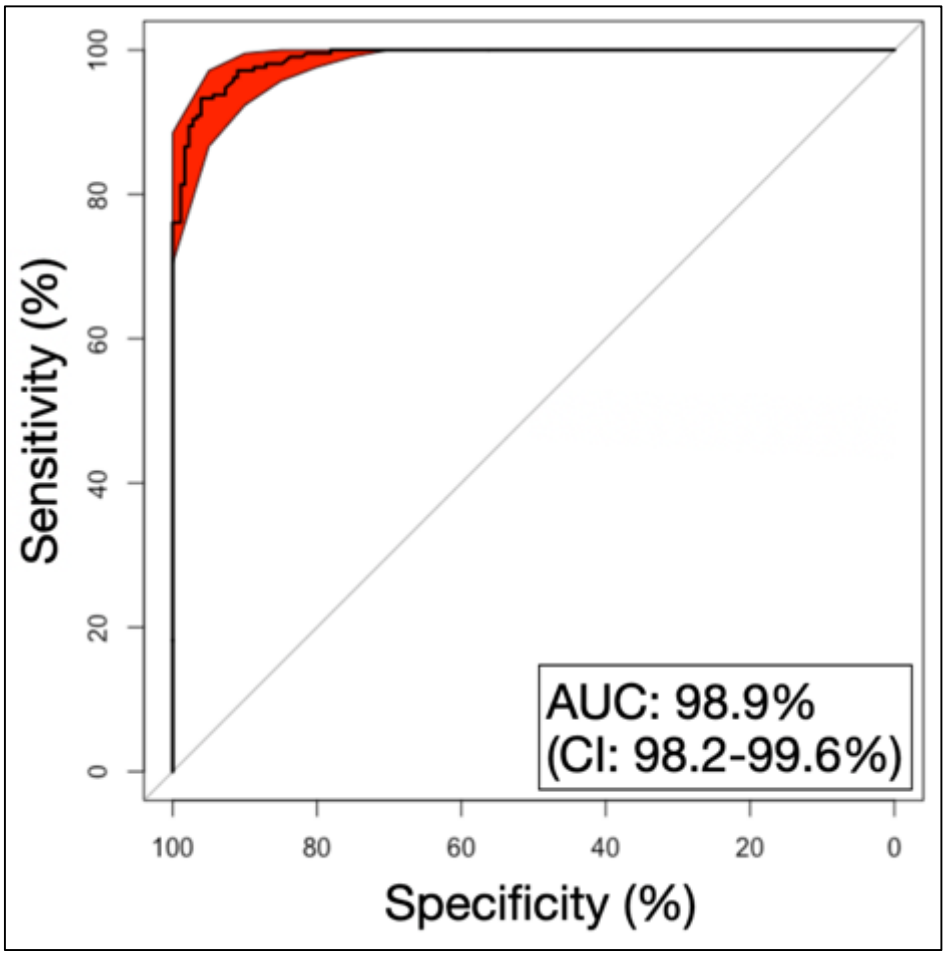
	Overall (N=388)
Sex	
Female	142 (36.6%)
Male	246 (63.4%)
Age	
Mean (SD)	10.1 (3.37)
Median [Min, Max]	9.46 [5.02, 20.8]
Diagnosis	
No History	60 (15.5%)
One or more Disorder	307 (79.1%)
Missing	21 (5.4%)
IQ	
Mean (SD)	98.4 (16.9)
Median [Min, Max]	98.0 [51.0, 145]
Missing	52 (13.4%)
BMI	
Mean (SD)	19.5 (5.20)
Median [Min, Max]	18.0 [11.7, 45.7]
Missing	8 (2.1%)
StructuralMRIQuality	
Mean (SD)	0.832 (0.0668)
Median [Min, Max]	0.855 [0.577, 0.900]
Missing	1 (0.3%)

Table 2.

Freesurfer Outputs Significantly Related to Structural MRI Quality.

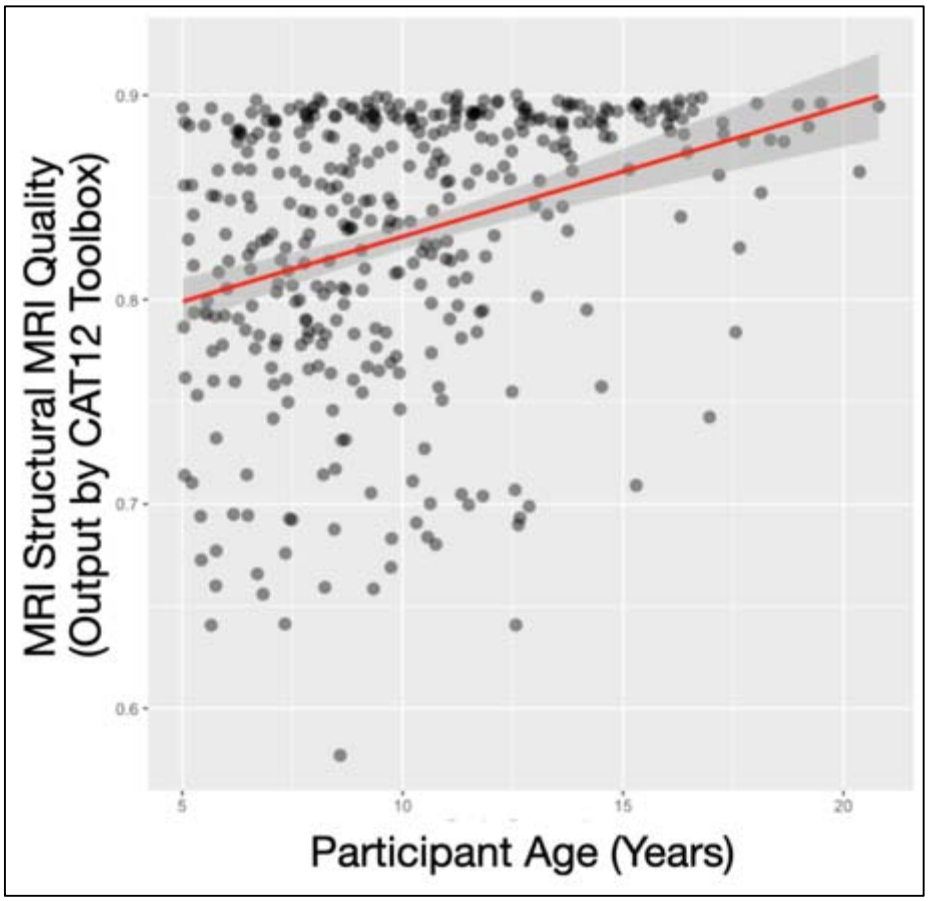
Brain Areas	t stats	p value adjusted
cc.anterior	2.136817	0.0387
cc.mid.anterior	2.313882	0.0196
cc.posterior	3.577070	0.0019
Left.Amygdala	2.817391	0.0186
Left.Putamen	2.195752	0.0310
Right.Putamen	2.506151	0.0310
lh caudalmiddlefrontal	7.172163	0.0000
lh entorhinal	1.770691	0.0000
lh fusiform	2.578228	0.0293
lh inferiorparietal	3.803395	0.0011
lh inferiortemporal	1.792572	0.0000
lh insula	3.034712	0.0101
lh lateraloccipital	2.659558	0.0253
lh lateralorbitofrontal	3.711327	0.0013
lh middletemporal	2.705017	0.0239
lh paracentral	2.331749	0.0196
lh parsopercularis	2.310613	0.0196
lh postcentral	3.293811	0.0019
lh precentral	8.758671	0.0000
lh superiorfrontal	1.003513	0.0005
lh superiorparietal	1.134813	0.0003
lh superiortemporal	5.969925	0.0000
lh supramarginal	1.854834	0.0000
rh caudalmiddlefrontal	5.751109	0.0000
rh entorhinal	5.567120	0.0000
rh inferiortemporal	3.780885	0.0011
rh lateraloccipital	2.608773	0.0283
rh middletemporal	2.572109	0.0293
rh paracentral	2.871558	0.0165
rh parsopercularis	2.670209	0.0253
rh postcentral	3.583014	0.0019
rh precentral	7.778537	0.0000
rh superiorfrontal	2.764778	0.0209
rh superiorparietal	1.218265	0.0003
rh superiortemporal	1.143510	0.0001
rh supramarginal	3.191515	0.0066
rh transversetemporal	1.158138	0.0003

Figure 1.



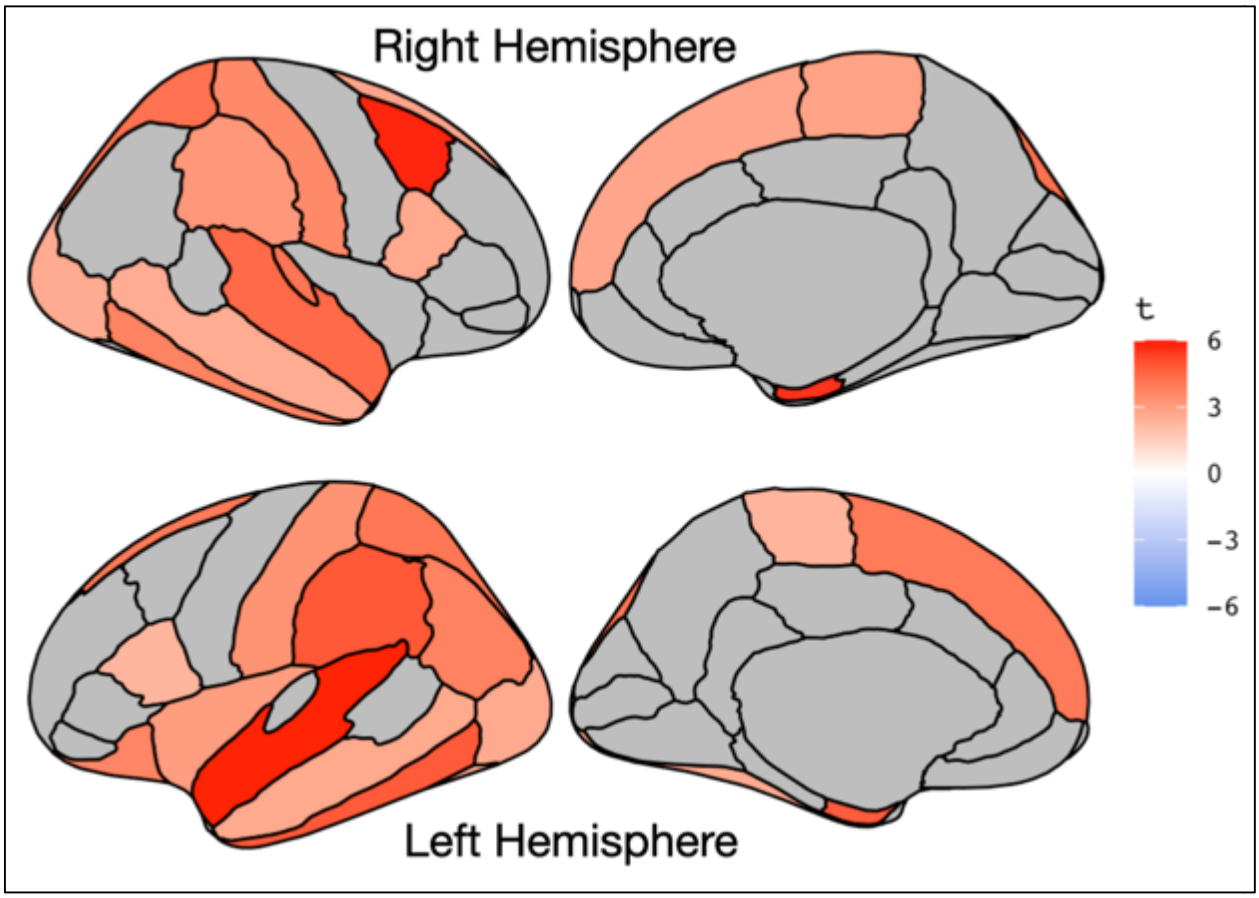
Caption: ROC curves showing the validity of image quality (derived from the CAT12 toolbox) for discriminating passing (versus failing) human rater visual checks of quality. Sensitivity and specificity were both high, suggesting image quality was able to robustly parse this binary categorization.

Figure 2.



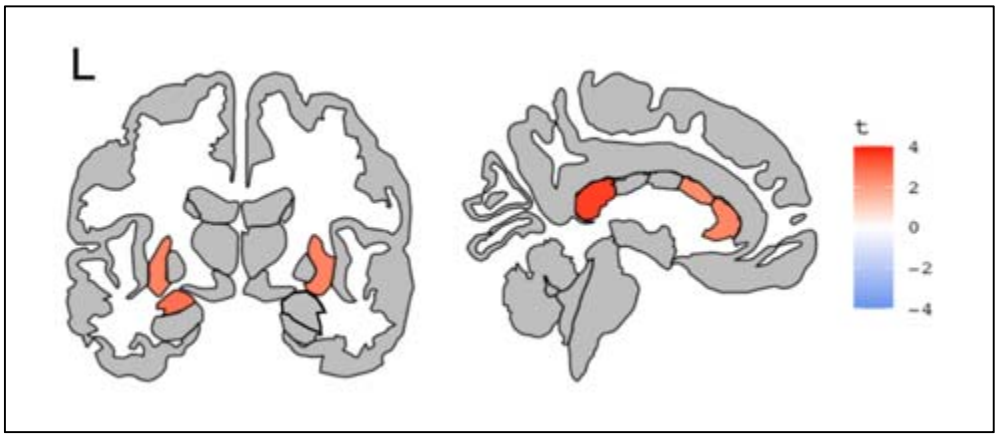
Caption: Scatterplot showing participant age (in years; horizontal axis) and image quality (an aggregated measure of noise to contrast ratio, coefficient of joint variation, inhomogeneity to contrast ratio, and Root mean squared voxel resolution, ranging from 0-1; vertical axis).

Figure 3.



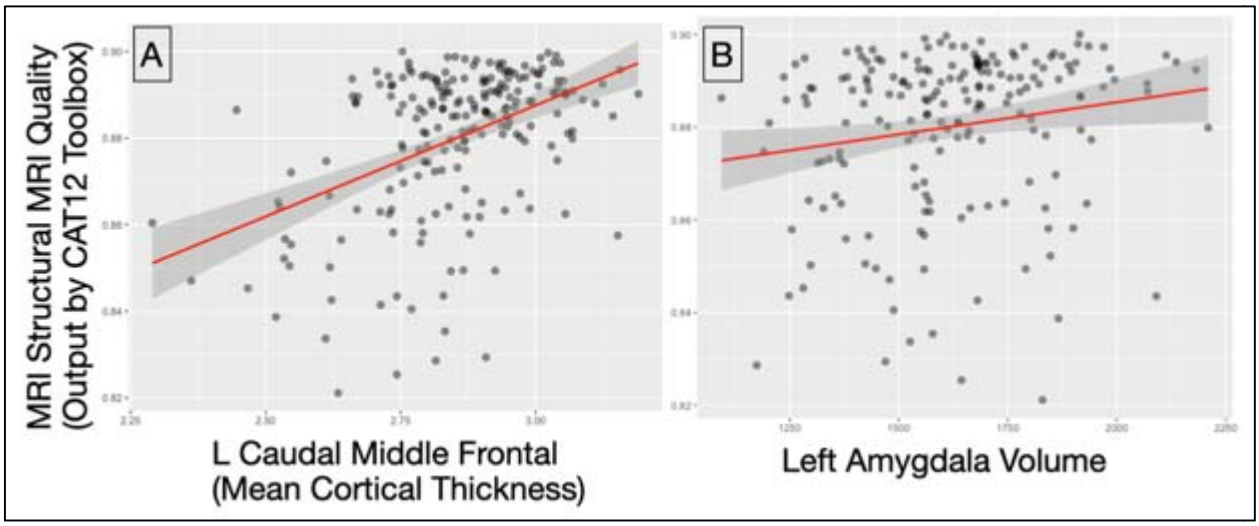
Caption: A graphic depiction (from the R library ggseg) showing associations between image quality and derived (mean) cortical thickness. This is shown for the Desikan atlas commonly used in Freesurfer. Lateral and medial views are shown for the right (top) and left (bottom) hemispheres. Only regions passing multiple comparison correction are depicted, with t-statistics of these associations ranging from 2.31 to 8.75.

Figure 4.



Caption: A graphic depiction (from the R library ggseg) showing associations between image quality and subcortical volume. This is shown for the Freesurfer ‘aseg’ atlas. Coronal (left) and sagittal (right) views are shown. Only regions passing multiple comparison correction are depicted, with t-statistics of these associations ranging from 2.31 to 3.57

Figure 5.



Caption: Example scatterplots showing associations between image quality (vertical axis, panels A & B) and Freesurfer outputs. In panel A, left caudal middle frontal cortical thickness is depicted on the horizontal axis, while left amygdala volume is shown on the horizontal axis in panel B.

Figure 1.

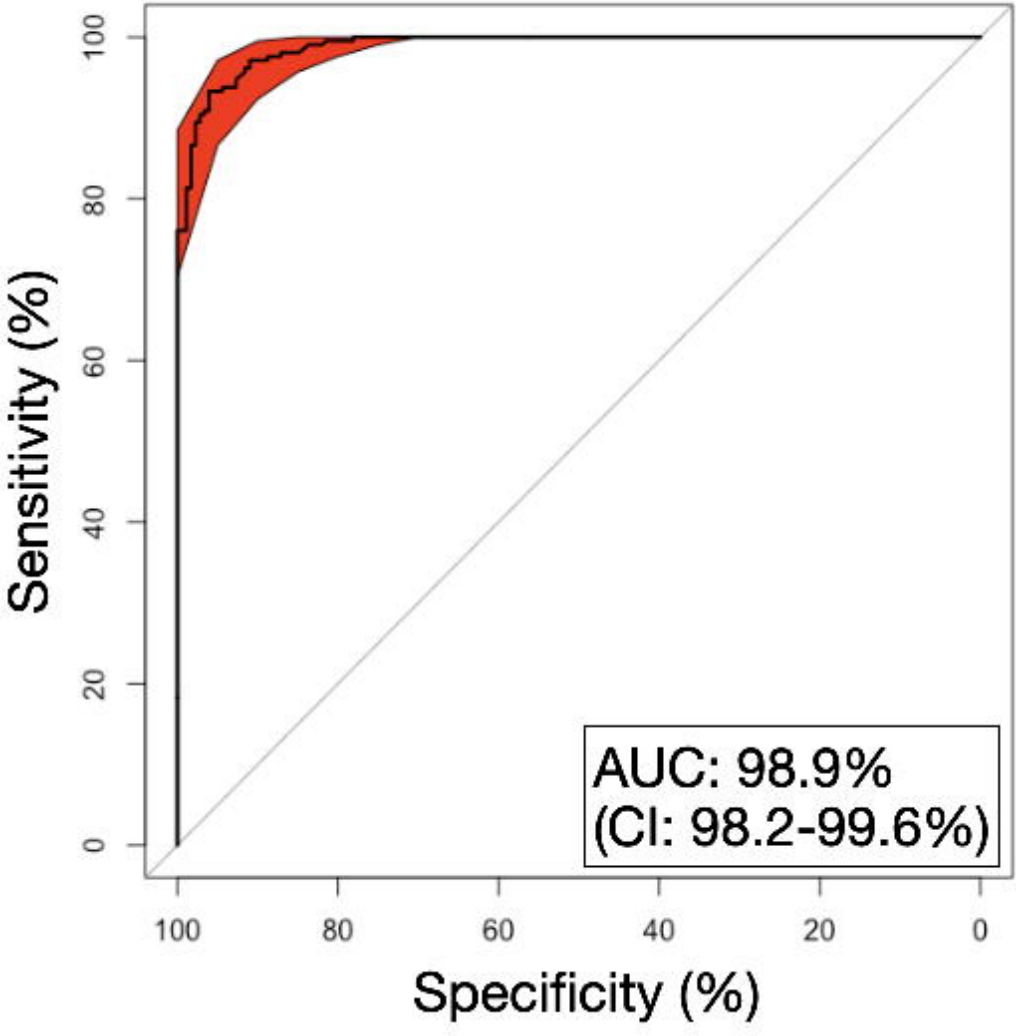

Sensitivity (%)
100
80
60
40
20
0
AUC: 98.9%
(CI: 98.2-99.6%)
100
80
60
40
20
0
Specificity (%)

Figure 2.

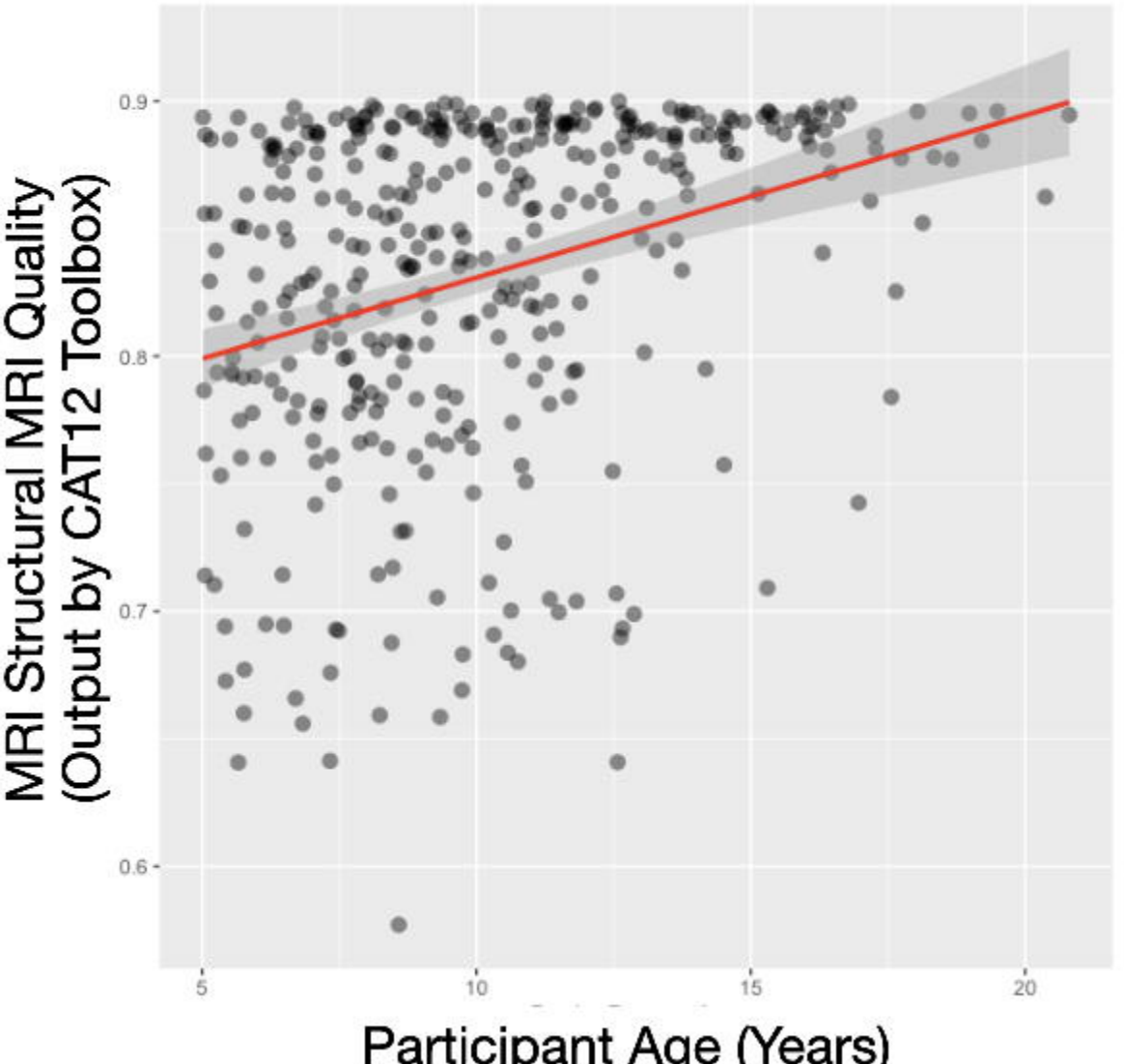

MRI Structural MRI Quality
(Output by CAT12 Toolbox)
0.9
0.8
0.7
0.6
5
10
15
20
Participant Age (Years)

Figure 3.

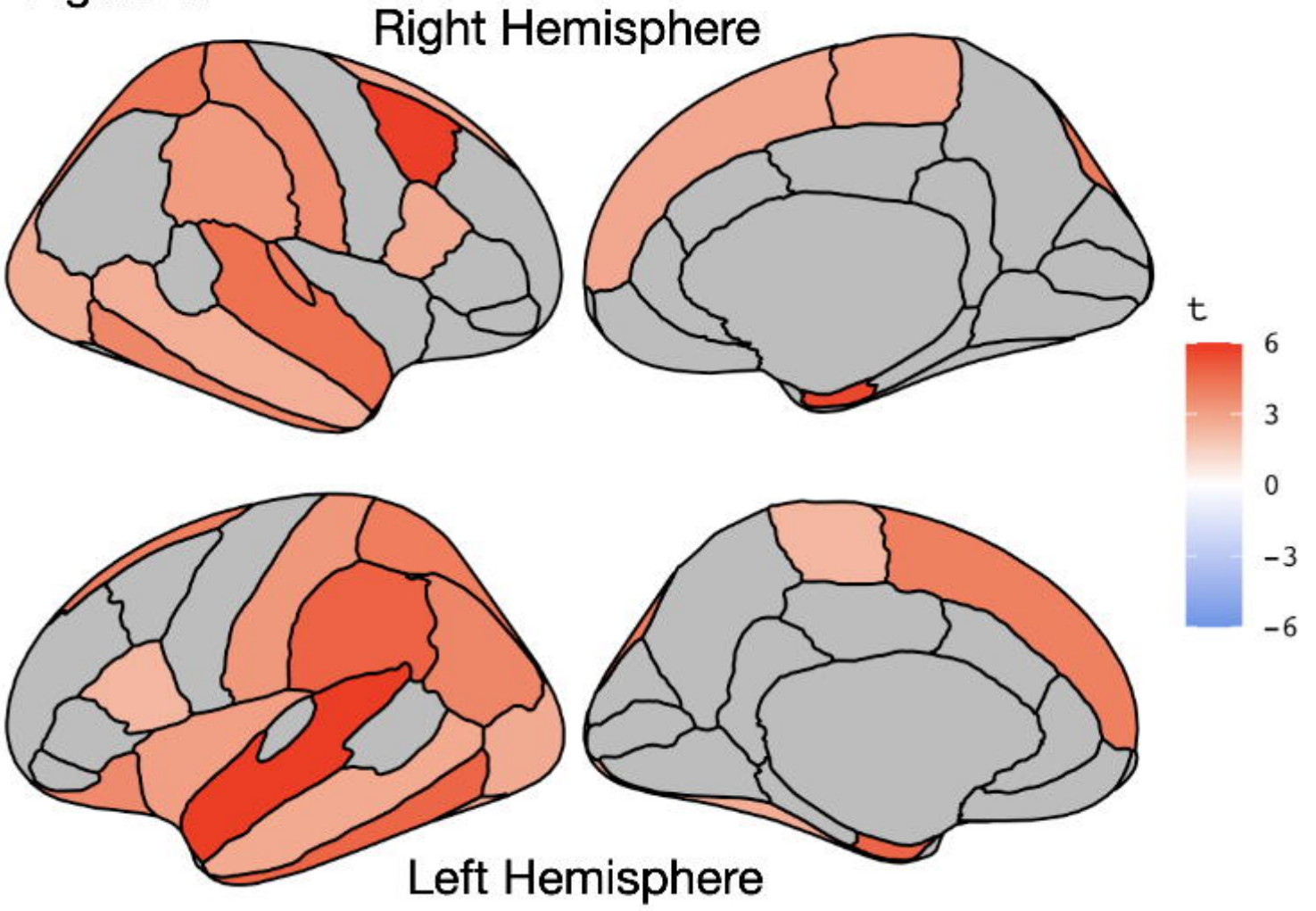


Figure 4.

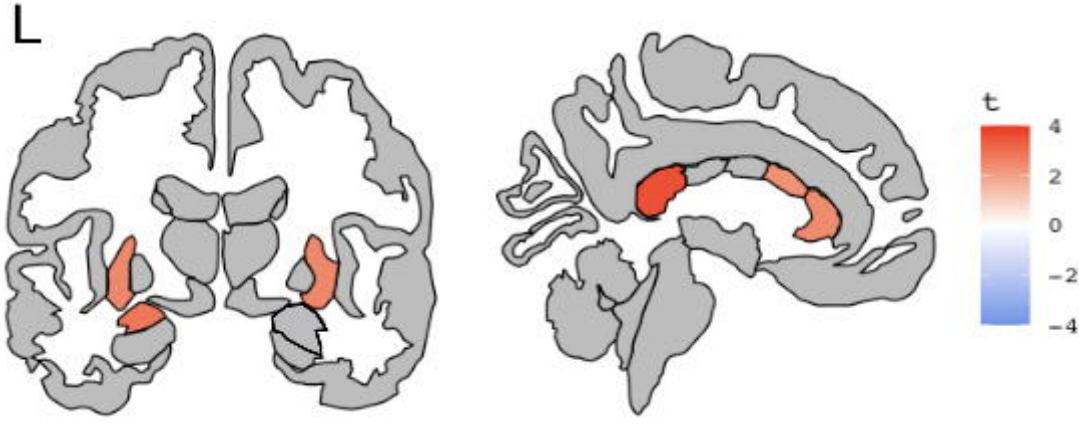
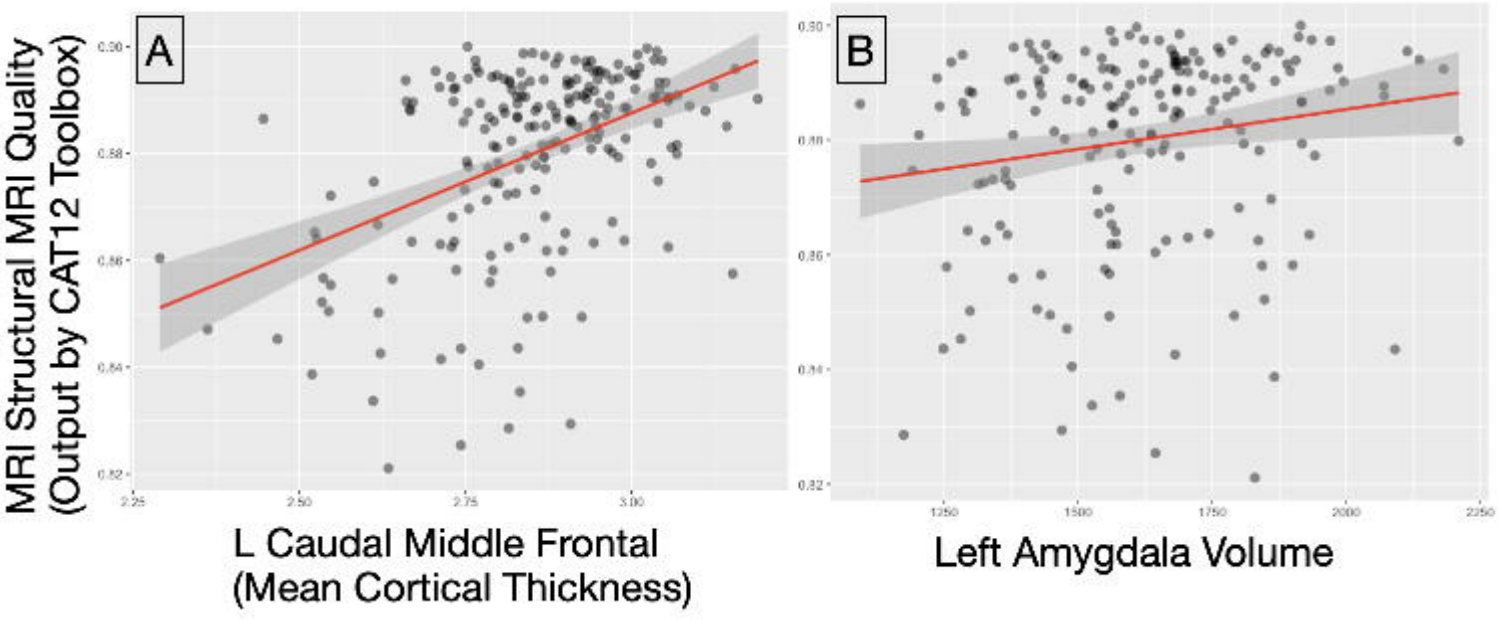


Figure 5.


A
B
MRI Structural MRI Quality
(Output by CAT12 Toolbox)
0.90
0.88
0.86
0.84
0.82
2.25
2.50
2.75
3.00
1250
1500
1750
2000
2250
L Caudal Middle Frontal
(Mean Cortical Thickness)
Left Amygdala Volume