NeuroImage 279 (2023) 120306

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Technical Note

FreeSurfer 7 quality control: Key problem areas and importance of manual corrections

Vesa Vahermaa [a,f,#,*], Dogu Baran Aydogan [b,c], Tuukka Raij [c,d], Reetta-Liina Armio [e], Heikki Laurikainen [e], Jari Saramäki [a], Jaana Suvisaari [f]

[a] Department of Computer Science, Aalto School of Science, Aalto University, Finland
[b] A.I. Virtanen Institute for Molecular Science, University of Eastern Finland, Kuopio, Finland
[c] Department of Neuroscience and Biomedical Engineering, Aalto School of Science, Espoo, Finland
[d] University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[e] Department of Psychiatry, University of Turku, Finland
[f] Mental Health Unit, Finnish Institute for Health and Welfare, Finland

ARTICLE INFO

Keywords:
Freesurfer
MRI
Psychosis
Quality control
Neuroimaging

ABSTRACT

We have studied the effects of manual quality control of brain Magnetic Resonance Imaging (MRI) images processed with Freesurfer. T1 images of first episode psychosis patients ($N = 60$) and healthy controls ($N = 41$) were inspected for gray matter boundary errors. The errors were fixed, and the effects of error correction on brain volume, thickness, and surface area were measured.

It is commonplace to apply quality control to Freesurfer MRI recordings to ensure that the edges of gray and white matter are detected properly, as incorrect edge detection leads to changes in variables such as volume, cortical thickness, and cortical surface area. We find that while Freesurfer v7.1.1. does regularly make mistakes in identifying the edges of cortical gray matter, correcting these errors yields limited changes in the commonly measured variables listed above. We further find that the software makes fewer gray matter boundary errors when processing female brains.

The results suggest that manually correcting gray matter boundary errors may not be worthwhile due to its small effect on the measurements, with potential exceptions for studies that focus on the areas that are more commonly affected by errors: the areas around the cerebellar tentorium, paracentral lobule, and the optic nerves, specifically the horizontal segment of the middle cerebral artery.

1. Introduction

Several different automated protocols and pipelines exist for analyzing Magnetic Resonance Imaging (MRI) brain data. Due to the high level of automation involved in these software solutions, ensuring their accuracy is important.

FreeSurfer (FS) is a free software suite that provides a comprehensive processing stream for structural MRI images. The version that was used in this paper, FS v7.1.1, includes automatic functionality for volumetric segmentation of anatomical regions of the brain (Fischl et al., 2002; Fischl et al., 2004). Since the processing of structural MRI images includes several steps including motion correction, removal of non-brain tissue, and gray-white matter segmentation, the ability to automate parts of the process standardizes volumetric quantification between studies and can save time. The latter is especially the case with larger datasets.

Two key features automated by FS are skull stripping and gray-white matter segmentation (Fischl et al., 2001; Ségonne et al., 2004; Segonne et al., 2007). Skull stripping means the removal of the skull and other non-brain matter from the image, and gray-white matter segmentation is a process that automatically outlines and isolates gray and white matter. The developers have made multiple updates to the quality and accuracy

Abbreviations: MRI, Magnetic resonance imaging; fMRI, Functional magnetic resonance imaging; FS, Freesurfer; FEP, First-episode psychosis; QC, Quality control; FD, Framewise displacement.

* Corresponding author.
E-mail address: vesa.vahermaa@thl.fi (V. Vahermaa).
Present/permanent address: Haitinkuja 3B 62, Helsinki 00220, Finland.

https://doi.org/10.1016/j.neuroimage.2023.120306
Received 30 March 2023; Received in revised form 30 July 2023; Accepted 1 August 2023
Available online 2 August 2023

of the processing stream dating back to at least 2006 (FreeSurfer, 2023). However, manual quality control of the results has been shown to be useful in past versions of FS, as the software suite does not always accurately trace the edges of gray and white matter (Waters et al., 2018).

Striving to achieve optimally accurate structural MRI estimates is paramount; for instance, volumes of specific brain regions have been found to be linked to diseases such as various mood and psychotic disorders, though the effects are generally small (De Peri et al., 2012; Armio et al., 2020; Salokangas et al., 2021; Cheon et al., 2022). However, while accuracy is important, recent research with an older version of FS, such as FS v5, has shown that manual editing and quality control of reading-related brain regions in pediatric populations offer limited gains (Beelen et al., 2020). Other recent research has also pointed out that the utility of correcting processing errors made by FS v5 is incremental, and it has been noted that despite the errors made by the software, researchers with limited resources may not find manual intervention worthwhile (Waters et al., 2018). Thus, even though errors can be identified and corrected with manual quality control, the process may not result in meaningful changes to the data, while still being time-intensive.

It should be noted that different versions of FS may produce different results. In particular, two recent studies exploring compatibility of brain metrics across different FS versions from 5.3 to 7.1 found that between-version compatibility and correlation of measurements such as volume can vary significantly depending on the examined brain region and which versions of FS are being compared (Bigler et al., 2020; Haddad, 2022). FS calculates metrics such as surface area and volume based on the gray-white matter borders, which are the areas targeted by manual QC of MRI images. As such, the usefulness of manual QC may also be dependent on the version of FS being used.

In this paper, we find support for the notion that the benefits of manual quality assurance are limited in FS v7.1.1 with regard to gray matter boundary identification, and research resources are often better spent elsewhere. While accurately measured brain surface areas and volumes can be considered important for various anatomical studies, the degree of accuracy exhibited by FS v7.1.1 is already very high. However, in cases where uncompromised accuracy is vital, having knowledge of the regions most prone to errors can increase the speed of the manual quality control process. Here, we also show gray matter regions that were most commonly affected with automatic processing errors.

In this work, the structural brain data of 101 subjects was automatically processed by FS v7.1.1. The images then underwent manual gray matter boundary quality control by a single researcher who analyzed and corrected errors related to skull stripping and gray matter surface.

The location of each error was recorded to determine the areas most prone to these errors, and the automatic segmentation step was done again for each image after manually correcting the recorded errors. The surface area of the whole brain and its subsections were recorded both before and after the manual quality control of the structural images, and the differences in surface area were calculated to determine the impact of manual quality control on the surface areas of the automatically segmented regions. Besides the surface area, the volumes and cortical thicknesses were also calculated for the subjects, and compared between groups based on sex and patient status.

2. Methods

2.1. Participants

The dataset is from the Helsinki Early Psychosis Study, which was conducted between December 2010 and July 2016. The FEP patients in the dataset are 18–40 years old, and were recruited from the in- and outpatient units of the University Hospital District of Helsinki and Uusimaa as well as the City of Helsinki. The inclusion criterion for FEP patients was a score of 4 or greater in the items assessing delusions or hallucinations in the Brief Psychiatric Rating Scale, Expanded version 4.0, BPRS et al., 1993). The diagnoses were confirmed using the research version of the Structured Clinical Interview for DSM-IV (First, 2007) complemented by a review of all medical records by a senior psychiatrist (JS). Psychotic disorders inarguably substance-induced or caused by a general medical condition were excluded (Karpov et al., 2020, Keinänen et al., 2018, Lindgren et al., 2017, Mäntylä, 2018).

Control subjects were matched from the Population Register to be comparable to the patients based on age, sex, and region of residence. Exclusion criteria for controls consisted of a lifetime history of psychotic disorders, chronic neurological, endocrinological, and cardiovascular disease, and any condition that would prevent MRI (Mäntylä et al., 2015). MRI was acquired once at the beginning of the study as baseline, and once at a 1 year follow-up point.

The dataset included 101 total subjects, of which 60 were FEP patients (19 females) and 41 healthy controls (15 females).

2.2. Image acquisition

The MRI data was acquired with a 3T MAGNETOM Skyra whole-body scanner (Siemens Healthcare, Erlangen, Germany) at Aalto AMI centre, Aalto NeuroImaging, Aalto University School of Science, using a standard 32 -channel head-neck coil and established sequences. A T1 weighted magnetization-prepared rapid gradient echo (MPRAGE) sequence was used for 176 sagittal/192 transversal slices with $1 \times 1 \times 1$ mm voxels.

The sequence had a TR of 2530 ms, TE of 3.3–3.75 ms, flip angle was 7°, matrix size 256×256, and the field of view was 25.6 cm.

2.3. Image processing

Cortical reconstruction and volumetric segmentation was performed with the FS image analysis suite v7.1.1, which is documented and freely available for download online (http://surfer.nmr.mgh.harvard.edu/). All subject images were processed with the *recon-all* function of FS v7.1.1 prior to manual error correction. The technical details of these procedures are described in prior publications (Dale et al., 1999; Dale and Sereno, 1993; Fischl and Dale, 2000; Fischl et al., 2001, 2002, 2004; Fischl et al., 1999a, 1999b, 2004b; Han et al., 2006; Jovicich et al., 2006; Segonne et al., 2004; Reuter et al., 2010; Reuter et al. 2012). Image processing was done on the Aalto Triton high-performance computing cluster using FS v7.1.1 on Dell Poweredge C6420 computers (2×20 core Intel Xeon Gold 6148 2.40 GHz, 192GB DDR4–2667).

2.4. Manual error correction

After image pre-processing, one researcher (VV) examined all images for gray matter boundary errors using FreeSurfer's Freeview application, and marked any significant and visible boundary errors for correction, along with the area of the location where applicable (such as the cerebellar tentorium or the middle cerebral artery). As selection criteria, errors were required to be clearly identifiable on at least three consecutive slices. The images were then processed again with the FS pipeline, resulting in images with corrected boundaries. Appendix A describes the details of this manual intervention process.

Written guidelines were established to ensure consistency in the process, and the error correction for the first subjects was done together under the supervision of RLA and HL. VV was also in touch with RLA and BA later during the process regarding some of the perceived errors to ensure consistency in interpretation.

The process was carried out using FS v7.1.1, running on a Macbook Pro 2019 16″ computer (2,4 GHz, 32GB, macOS Big Sur). All the images were processed under similar conditions in the same office, using the same monitor with identical brightness and contrast levels. Subjects were processed in numerical order based on subject ID number. 65

subjects had two anatomical images to analyze, one taken at baseline and one at 1-year follow-up. In the case where a subject had two images to process, the second image was processed immediately after the first during the same session.

2.5. Heatmap generation

A heatmap visualization was generated to illustrate the distribution of the boundary errors in MNI305 space. The process started with the use of FreeSurfer's *mri_convert* function to convert the coordinates of each recorded error from Freesurfer's own coordinate space to MNI305 space. Errors were pooled for each voxel in the MNI305 image space.

To improve the visibility of the errors and show how the errors were clustered, Gaussian smoothing with Sigma of 3 voxels was applied to the combined error image. Finally, the combined error image with Gaussian smoothing was applied as an overlay on the MNI305 brain. Heatmap visualizations of the results were generated with *ITK-SNAP* (Yushkevich et al., 2006).

The dimensions of the MNI305 brain used for the coordinates in the caption of Fig. 1 are 172 × 220 × 156. The MNI305 brain data was obtained in *NIFTI* file format from the NIST laboratory of McGill University (NIST, 2023).

2.6. Statistical analysis

The locations of the recorded errors were labeled according to labels obtained from individual images that contain the parcellation and segmentation data (*aparc+aseg.mgz* files). In case the recorded error was not part of the parcellated cerebral gray or white matter, such as when the error was located in the meninges, the closest labeled voxel was chosen.

Comparison of brain surface areas, volumes, and thicknesses was done by generating csv files with detailed information of each subject by using the FS *aparcstats2table* function after having run the *recon-all* function on all subjects. This information was generated from the subjects both prior and after manual quality control. The measurements from both runs were then subtracted from each other, and the percentage change was calculated by comparing the result of the subtraction to the measurements prior to manual quality control. The mean change to each brain area was calculated by averaging the measurement changes of all subjects.

The number of identified errors was analyzed on a group level, investigating possible differences in identified errors related to sex, age, and patient status. These analyses were done using only baseline images, which were available for every subject of every group. When comparing the distributions of errors between different groups, Mann-Whitney *U test* was used.

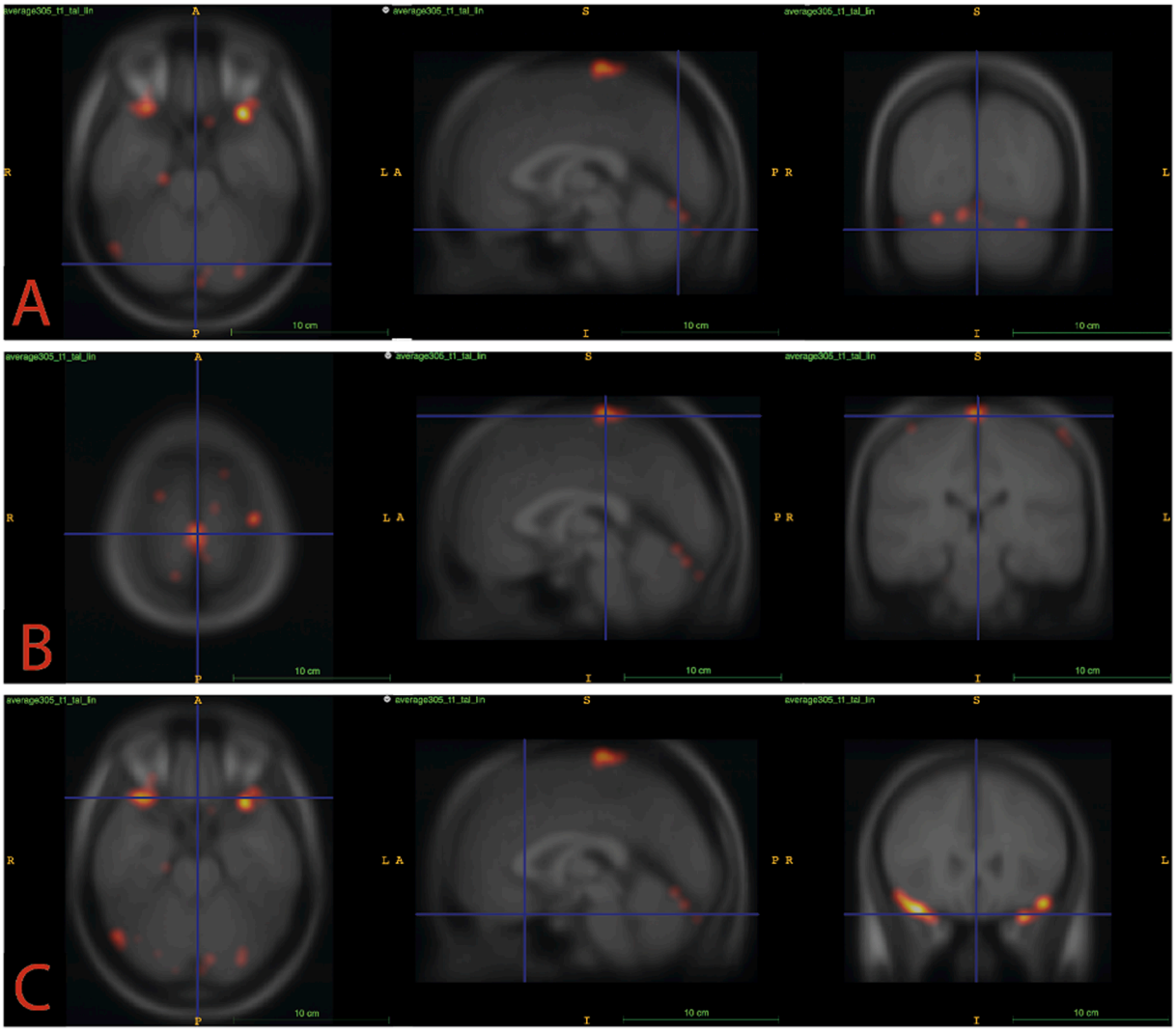

Fig. 1. Main error clusters and coordinates near the cerebellar tentorium (A - 87,52,42); paracentral lobule (B – 87,100,143); and optic nerves (C – 86,150,44).

The correlation regarding the number of errors between baseline and follow-up images of the same subject was investigated for those subjects for whom both a baseline and a 1-year follow-up image were available. This was done using the Pearson correlation coefficient.

All reported difference percentages are measured in terms of changes to the pre-QC values. As such, a change of +0.1% should be interpreted as an increase of 0.1% in the post-QC round of measurements. The differences between patients and healthy controls are measured in terms of absolute differences between the means of the two groups.

Mean framewise displacement (FD) data recorded during functional magnetic resonance imaging (fMRI) sessions of the same study was available for analysis (Rikandi, 2022). This data was used to investigate the significant differences in number of errors between male and female subjects. Data was available from 209 recordings (70 females).

FD data was compared between males and females, and between the 5% of recordings with the highest number of recorded errors and the 95% of the recordings with the lowest number of errors. When comparing the distributions of mean framewise displacement, Student's T-test was used.

2.7. Ethics statement

The study protocol was approved by the Ethics Committee of the Hospital District of Helsinki and Uusimaa (257/12/03/03/2009 and 226/13/03/03/2013) and by the institutional review board of the Finnish National Institute for Health and Welfare, Helsinki, Finland. All participants gave written informed consent. The treating psychiatrist assessed the patient's capacity to give informed consent. The work described has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

2.8. Data availability

The participants of this study, the Helsinki Early Psychosis Study, did not give written consent for their data to be shared publicly. As such, due to the sensitive nature of the research the data is not openly available as it is considered clinical data. However, anonymized data that support the findings of this study are available from the corresponding author, VV, upon reasonable request including clarification on how the anonymized data would be used.

2.9. Code availability

The software used for this study is Freesurfer, which is publicly available online at https://www.freesurfer.net. The code used for the data analysis and heatmap generation in this study is available online at https://github.com/Schwerbelastung/freesurfer-qc-paper

3. Results

3.1. Cortical volume changes

The mean local brain volume changes, as read from the *aparc.stats* file after corrections of the overextension of gray matter boundary, were very small across all regions (−0.01% +- 0.37). The largest mean volume changes across all subjects occurred in the left caudal anterior cingulate (+1.12% +- 3.72) and the left entorhinal cortex (−1.12% +−5.73).

Between patients and controls, the average difference in regional cortical volumes when comparing pre- and post-QC volumes was 0.37% +- 0.40. Full details of the volume changes can be found in Appendices B and C.

3.2. Cortical thickness changes

The mean change in cortical thickness across all regions and all subjects was −0.06% +- 0.13. The largest mean changes occurred in the left insula (−0.29% +- 1.55) and left caudal anterior cingulate (+0.28% +- 2.18).

Between patients and controls, the average difference in regional cortical thickness changes when comparing pre- and post-QC volumes was 0.21% +- 0.13. Full details of the cortical thickness changes can be found in Appendices D and E.

3.3. Cortical surface area changes

The mean local surface area change across all regions was very small (+0.03% +- 0.36). The largest mean surface area changes were in the left entorhinal cortex (+1.09% +- 5.26) and in the left caudal anterior cingulate (+0.90% +- 3.84).

Between patients and controls, the average difference in regional cortical surface area changes when comparing pre- and post-QC volumes was 0.36% +- 0.37. Full details of the cortical surface area changes can be found in Appendices F and G.

3.4. Error locations

The errors were not uniformly distributed in the brain. The most affected areas were the areas in proximity to the middle cerebral artery (132 errors, or 25.2%) and the vicinity of the cerebellar tentorium (also 132 errors, or 25.2%). The errors were relatively evenly distributed across hemispheres, with 251 errors (48%) in the left hemisphere, and 272 (52%) in the right hemisphere.

Fig. 1 visualizes the errors close to the three most affected areas. No images had obvious and significant processing errors such as large areas of the brain excluded from the brain mask.

A video has been recorded of the three-dimensional heatmap with the recorded errors. It can be found linked in the GitHub repository mentioned in Section 2.9, and in the following location: https://figshare.com/articles/media/Video_of_the_3d_heatmap_of_errors_mov/22341454

3.5. Number of errors

A total of 523 errors related to the overextension of the gray matter boundary were recorded from the 167 MRI images.

The number of errors between the baseline and follow-up images of the same subject was significantly correlated ($r = 0.63$, $p < 0.01$).

The number of errors in FEP subjects (3.27 +- 2.8, $n = 96$) did not differ significantly from those in healthy controls (2.89 +- 2.35, $n = 70$), Mann-Whitney U test ($U = 3144.5$, $p = 0.48$).

Spearman's rank correlation was computed to assess the relationship between subject age and number of errors. There was no significant correlation between the two variables, $r(99) = -0.012$, $p = 0.91$).

However, there was a significant effect for sex in the number of recorded errors per image. On average, images of male brains had more errors per image (3.96 +- 2.78, $n = 67$) than images of female brains (1.68 +- 1.84, $n = 34$), based on a Mann-Whitney U test ($U = 538$, $p < 0.001$). The images with the most errors were exclusively male images. The distribution can be seen in Fig. 2.

3.6. Subject movement

FD data from 209 recordings (139 male, 70 female) was analyzed to assess potential differences in subject movement.

The mean FD of male subjects did not differ significantly from the mean FD of female subjects ($p = 0.75$). Likewise, the mean FD of the top 5% of recordings did not differ from the mean FD of the rest of the sample ($p = 0.17$). The data used for the analysis can be found in

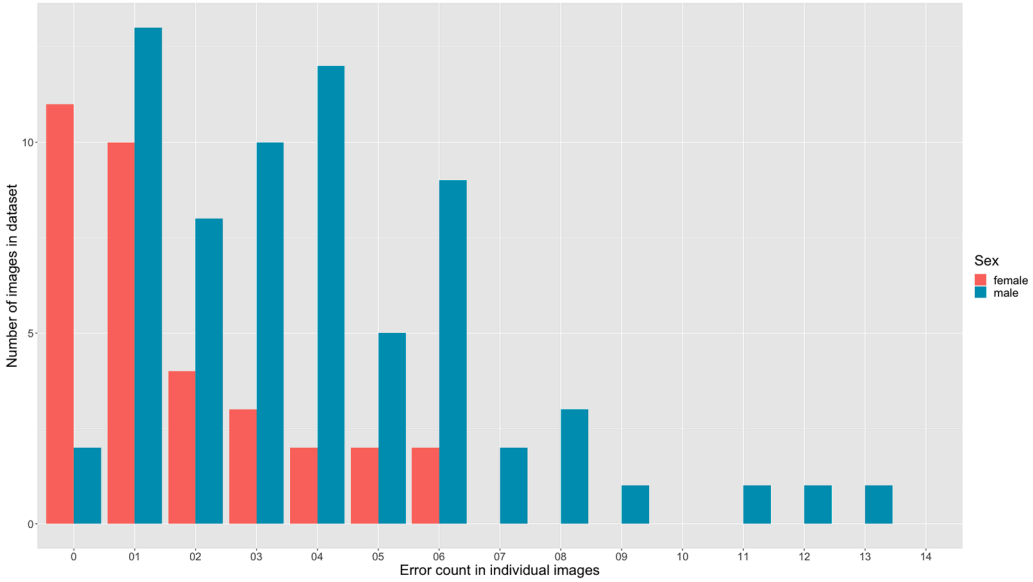

Fig. 2. Number of errors in inspected images between men and women.

Appendix H.

4. Discussion

Our results suggest that cortical surface gray matter QC of structural images results only in minor differences in surface area, volume, and thickness across all subjects. Most errors are located near the cerebellar tentorium, the optic chiasm, and the paracentral lobule. Thus, unless the research project requires significant precision of the most affected areas, manual error correction may not be necessary.

Further, when comparing the FEP patient and healthy control populations, the differences in terms of pre- and post-QC measurements were small. When investigating these differences using a paired T-test, various brain regions showed statistically significant differences. However, these typically would not survive correction for multiple testing.

These results are in line with previous studies that investigated the effects of QC with earlier versions of FS and found that benefits of manual QC may be limited (Beelen et al., 2020; Waters et al., 2018).

The correlation found in the number of errors between the baseline and follow-up images of the same subjects suggests that the number of errors made by FS is related to the individual subject's brain anatomy, as the number of errors does not seem to strongly fluctuate despite a 1 year difference between measurements.

The dataset of 167 individual MRI images used in this study was large for a single-site study where quality control was done by a single researcher. The dataset consisted of both first episode psychosis patients as well as healthy controls, all adults 18–40 years of age, which allowed us to investigate whether recent psychosis affects the need for manual quality control. Still, it is important to note that the dataset did not include children, adolescents, or adults over 40 years of age, so the results may not be directly applicable to those populations. For instance, movement artifacts caused due to subject movement can cause problems in structural MRI interpretation (Havsteen et al., 2017). If there is reason to believe that the investigated population has a higher propensity for movement within the scanner, manual QC may be appropriate.

Future research may also investigate whether errors fixed by manual QC could feasibly affect other sequences, such as fMRI or Diffusion-Weighted Imaging, if the T1 images are used for co-registration.

We did not find any obvious cause for the sex differences in the number of fixed cortical surface errors found in the study. While there are sex differences in brain anatomy (Luders and Toga, 2010), it is not immediately clear how they would translate to differences in gray matter segmentation accuracy. It might also be that the differences are not directly related to physical sex-based differences, but instead to differences in behavior during the study. For instance, a recent study suggests that males exhibit more motion inside the scanner than females, which might lead to differences in cortical surface voxel clarity even after correcting for errors due to motion (Alexander-Bloch et al., 2016). However, FD data recorded from this dataset in fMRI context does not show significant differences in motion between female and male subjects. Framewise displacement analysis can be found in Appendix H.

In addition to the relatively small effects in volume, thickness, and surface area of the brain when comparing pre- and post-QC measurements, it should also be noted that manual QC of FS analyses can suffer from multiple problems. For instance, manual QC protocols of different academic and clinical groups using FS are rarely compared, they can be time consuming, and they may exhibit both inter-rater and intra-rater variability. There are automated software solutions such as Qoala-T that aim to address this by automating parts of the QC process to a large degree (Klapwijk et al., 2019).

It should also be noted that re-running the *recon-all* command after fixing any boundary errors results in re-calculation of all the brain regions. Correcting even a single gray matter boundary error results in changes to the *aparcstats2table* output regarding surface area, volume, and thickness in every brain region, not just the ones where errors were corrected.

In conclusion, these findings suggest that manual gray matter QC and error corrections of MRI images processed with FS v7.1.1 have only limited effect on the surface, volume, and cortical thickness of the processed images, and this should be taken into account when deciding whether to invest researcher work time into manual QC. While there are significant differences in how many errors male and female brains have in terms of gray matter boundary errors, these do not translate into significant differences in the surface area, volume, or cortical thickness in either males or females pre- or post-QC.

Data and code availability statement

The participants of this study, the Helsinki Early Psychosis Study, did not give written consent for their data to be shared publicly. As such, due to the sensitive nature of the research the data is not openly available as it is considered clinical data. However, anonymized data that support the findings of this study are available from the corresponding author, VV, upon reasonable request including clarification on how the

anonymized data would be used.

The software used for this study is Freesurfer, which is publicly available online at https://www.freesurfer.net. The code used for the data analysis and heatmap generation in this study is available online at https://github.com/Schwerbelastung/freesurfer-qc-paper

CRediT authorship contribution statement

Vesa Vahermaa: Conceptualization, Methodology, Software, Writing – original draft, Visualization, Funding acquisition, Investigation, Formal analysis. **Dogu Baran Aydogan:** Conceptualization, Methodology, Software, Writing – review & editing. **Tuukka Raij:** Investigation, Resources, Writing – review & editing, Project administration. **Reetta-Liina Armio:** Methodology, Writing – review & editing. **Heikki Laurikainen:** Methodology, Writing – review & editing. **Jari Saramäki:** Writing – review & editing, Supervision. **Jaana Suvisaari:** Writing – review & editing, Supervision, Resources, Investigation, Funding acquisition, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The original clinical data is confidential, as the participants did not give written consent for their data to be shared publicly. Code and anonymized data are available as described in 2.8 and 2.9.

Acknowledgements and funding

The project received funding from the Finnish Foundation for Psychiatric Research, Niilo Helander foundation, Waldemar von Frenckell foundation, Emil Aaltonen foundation, and the Foundation for Aalto University Science and Technology for VV. The project also received funding from Academy of Finland grants #348631 and #353798 to BA. Finally, the project received funding from Academy of Finland grants #278171 and #323035 and the Sigrid Juselius foundation for JS.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.120306.

References

Alexander-Bloch, A., et al., 2016. Subtle in-scanner motion biases automated measurement of brain anatomy from in vivo MRI. Hum. Brain Mapp. 37 (7), 2385–2397. https://doi.org/10.1002/hbm.23180.

Armio, R.L., et al., 2020. Amygdala subnucleus volumes in psychosis high-risk state and first-episode psychosis. Schizophr. Res. 215, 284–292. https://doi.org/10.1016/j.schres.2019.10.014.

Atlases (2023) NIST. Available at: https://nist.mni.mcgill.ca/atlases/(Accessed: 28 July 2023).

Beelen, C., et al., 2020. Investigating the added value of FreeSurfer's manual editing procedure for the study of the reading network in a pediatric population. Front. Hum. Neurosci. 14 https://doi.org/10.3389/fnhum.2020.00143.

Bigler, E.D., et al., 2020. Freesurfer 5.3 versus 6.0: are volumes comparable? A chronic effects of neurotrauma consortium study. Brain Imaging Behav. 14 (5), 1318–1327. https://doi.org/10.1007/s11682-018-9994-x.

Cheon, E.J., et al., 2022. Cross disorder comparisons of brain structure in schizophrenia, bipolar disorder, major depressive disorder, and 22q11.2 deletion syndrome: a review of enigma findings. Psychiatry Clin. Neurosci. 76 (5), 140–161. https://doi.org/10.1111/pcn.13337.

Dale, A.M., Sereno, M.I., 1993. Improved localizadon of cortical activity by combining EEG and Meg with MRI cortical surface reconstruction: a linear approach. J. Cogn. Neurosci. 5 (2), 162–176. https://doi.org/10.1162/jocn.1993.5.2.162.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. Neuroimage 9 (2), 179–194. https://doi.org/10.1006/nimg.1998.0395.

De Peri, L., et al., 2012. Brain structural abnormalities at the onset of schizophrenia and bipolar disorder: a meta-analysis of controlled magnetic resonance imaging studies. Current Pharmaceutical Design 18 (4), 486–494. https://doi.org/10.2174/138161212799316253.

First, M.B., 2007. Structured Clinical Interview For DSM-IV-TR Axis I disorders: SCID-I. New York State Psychiatric Institutute, New York, NY. Biometrics Research Dept.

Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. 97 (20), 11050–11055. https://doi.org/10.1073/pnas.200033797.

Fischl, B., Sereno, M.I., Dale, A.M., 1999a. Cortical surface-based analysis. Neuroimage 9 (2), 195–207. https://doi.org/10.1006/nimg.1998.0396.

Fischl, B., et al., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8 (4), 272–284. https://doi.org/10.1002/(sici)1097-0193(1999)8:4<272::aid−hbm10>3.0.co;2-4.

Fischl, B., Liu, A., Dale, A.M., 2001. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Trans. Med. Imaging 20 (1), 70–80. https://doi.org/10.1109/42.906426.

Fischl, B., et al., 2002. Automated labeling of neuroanatomical structures in the human brain neuron. Neuron 33 (3), 341–355. https://doi.org/10.1016/s0896-6273(02)00569-x.

Fischl, B., et al., 2004. Sequence-independent segmentation of magnetic resonance images. Neuroimage 23. https://doi.org/10.1016/j.neuroimage.2004.07.016.

Fischl, B., et al., 2004b. Automatically parcellating the human cerebral cortex. Cerebral Cortex 14 (1), 11–22. https://doi.org/10.1093/cercor/bhg087.

Freesurfer. Previous Release Notes (no date) Previousreleasenotes - Free Surfer Wiki. Available at: https://surfer.nmr.mgh.harvard.edu/fswiki/PreviousReleaseNotes (Accessed: March 22, 2023).

Haddad, E., et al., 2022. Multisite test–retest reliability and compatibility of brain metrics derived from freesurfer versions 7.1, 6.0, and 5.3. Hum. Brain Mapp. 44 (4), 1515–1532. https://doi.org/10.1002/hbm.26147.

Han, X., et al., 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage 32 (1), 180–194. https://doi.org/10.1016/j.neuroimage.2006.02.051.

Havsteen, I., et al., 2017. Are movement artifacts in magnetic resonance imaging a real problem?—A narrative review. Front. Neurol. 8 https://doi.org/10.3389/fneur.2017.00232.

Jovicich, J., et al., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on Phantom and human data. Neuroimage 30 (2), 436–443. https://doi.org/10.1016/j.neuroimage.2005.09.046.

Karpov, B., et al., 2020. Anxiety symptoms in first-episode psychosis. Early Interv. Psychiatry 15 (3), 569–576. https://doi.org/10.1111/eip.12986.

Klapwijk, E.T., et al., 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.

Lindgren, M., et al., 2017. Childhood adversities and clinical symptomatology in first-episode psychosis. Psychiatry Res. 258, 374–381. https://doi.org/10.1016/j.psychres.2017.08.070.

Luders, E. and Toga, A.W. (2010) "Sex differences in brain anatomy," Sex Differences in the Human Brain, Their Underpinnings and Implications, pp. 2–12. Available at: 10.1016/b978-0-444-53630-3.00001-4.

Mäntylä, T., et al., 2015. Altered activation of innate immunity associates with white matter volume and diffusion in first-episode psychosis. PLoS One 10 (5). https://doi.org/10.1371/journal.pone.0125112.

Mäntylä, T., et al., 2018. Aberrant cortical integration in first-episode psychosis during natural audiovisual processing. Biol. Psychiatry 84 (9), 655–664. https://doi.org/10.1016/j.biopsych.2018.04.014.

Reuter, M., Rosas, H.D., Fischl, B., 2010. Highly accurate inverse consistent registration: a robust approach. Neuroimage 53 (4), 1181–1196. https://doi.org/10.1016/j.neuroimage.2010.07.020.

Reuter, M., et al., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61 (4), 1402–1418. https://doi.org/10.1016/j.neuroimage.2012.02.084.

Rikandi, E., et al., 2022. Functional network connectivity and topology during naturalistic stimulus is altered in first-episode psychosis. Schizophr. Res. 241, 83–91. https://doi.org/10.1016/j.schres.2022.01.006.

Ségonne, F., et al., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage 22 (3), 1060–1075. https://doi.org/10.1016/j.neuroimage.2004.03.032.

Salokangas, R.K.R., et al., 2021. Effect of childhood physical abuse on social anxiety is mediated via reduced frontal lobe and amygdala-hippocampus complex volume in adult clinical high-risk subjects. Schizophr. Res. 227, 101–109. https://doi.org/10.1016/j.schres.2020.05.041.

Segonne, F., Pacheco, J., Fischl, B., 2007. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE Trans. Med. Imaging 26 (4), 518–529. https://doi.org/10.1109/tmi.2006.887364.

Waters, A.B., et al., 2018. Identifying errors in freesurfer automated skull stripping and the incremental utility of manual intervention. Brain Imaging Behav. 13 (5), 1281–1291. https://doi.org/10.1007/s11682-018-9951-8.

Yushkevich, P.A., et al., 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31 (3), 1116–1128. https://doi.org/10.1016/j.neuroimage.2006.01.015.