Original Article
Yonsei Med J 2021 Mar;62(3):255-261
https://doi.org/10.3349/ymj.2021.62.3.255
Yonsei Medical Journal
YMJ
pISSN: 0513-5796 · eISSN: 1976-2437

Comparison of Automated Brain Volume Measures by NeuroQuant vs. Freesurfer in Patients with Mild Cognitive Impairment: Effect of Slice Thickness

Younghee Yim[1,2], Ji Young Lee[3], Se Won Oh[4,5], Mi Sun Chung[2], Ji Eun Park[6], Yeonsil Moon[7], Hong Jun Jeon[8], and Won-Jin Moon[1]

Departments of [1]Radiology, [7]Neurology, and [8]Psychiatry, Konkuk University Medical Center, Konkuk University School of Medicine, Seoul;
[2]Department of Radiology, Chung-Ang University Hospital, Seoul;
[3]Department of Radiology, Hanyang University Medical Center, Seoul;
[4]Department of Radiology, Soonchunhyang University Cheonan Hospital, Cheonan;
[5]Department of Radiology, Eunpyeong St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul;
[6]Department of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea.

Purpose: This study aimed to examine the inter-method reliability and volumetric differences between NeuroQuant (NQ) and Freesurfer (FS) using T1 volume imaging sequence with different slice thicknesses in patients with mild cognitive impairment (MCI).

Materials and Methods: This retrospective study enrolled 80 patients diagnosed with MCI at our memory clinic. NQ and FS were used for volumetric analysis of three-dimensional T1-weighted images with slice thickness of 1 and 1.2 mm. Inter-method reliability was measured with Pearson correlation coefficient (r), intraclass correlation coefficient (ICC), and effect size (ES).

Results: Overall, NQ volumes were larger than FS volumes in several locations: whole brain (0.78%), cortical gray matter (5.34%), and white matter (2.68%). Volume measures by NQ and FS showed good-to-excellent ICCs with both 1 and 1.2 mm slice thickness (ICC=0.75–0.97, ES=-1.0–0.73 vs. ICC=0.78–0.96, ES=-0.9–0.77, respectively), except for putamen, pallidum, thalamus, and total intracranial volumes. The ICCs in all locations, except the putamen and cerebellum, were slightly higher with a slice thickness of 1 mm compared to those of 1.2 mm.

Conclusion: Inter-method reliability between NQ and FS was good-to-excellent in most regions with improvement with a 1-mm slice thickness. This finding indicates that the potential effects of slice thickness should be considered when performing volumetric measurements for cognitive impairment.

Key Words: Brain volume, FreeSurfer, mild cognitive impairment, NeuroQuant

INTRODUCTION

Alzheimer's disease (AD) and other neurodegenerative disorders are associated with brain atrophy.[1] Consistent with histologic findings,[2] significant medial temporal lobe or hippocampal atrophy can be detected by structural magnetic resonance imaging (MRI).[3-5] MRI measurements of brain volumes assess regional brain atrophy and quantify neurodegeneration, thereby enabling the detection of patients with the risk of rapid clinical deterioration.[1]

Software packages available for volumetric brain analysis include the FSL,[6,7] voxel-based morphometry,[8] FreeSurfer (FS),[9] and NeuroQuant (NQ).[10] NQ[11] was originally designed for the quantification of brain atrophy in AD,[12] and is a fully-automated software approved by the United States Federal Drug Administration for cross-sectional brain volume measurement.

Received: November 11, 2020 **Revised:** December 29, 2020
Accepted: January 5, 2021
Corresponding author: Won-Jin Moon, MD, PhD, Department of Radiology, Konkuk University Medical Center, Konkuk University School of Medicine, 120-1 Neungdong-ro, Gwangjin-gu, Seoul 05030, Korea.
Tel: 82-2-2030-5544, Fax: 82-2-2030-5549, E-mail: mdmoonwj@kuh.ac.kr

•The authors have no potential conflicts of interest to disclose.

NQ uses high-resolution three-dimensional (3D) T1-weighted volumetric images to automatically provide segmentation-based measurements of cortical and subcortical volumes. NQ also provides normative percentiles of regional brain atrophy by comparing the measured volumes to a normative database adjusted for age, sex, and intracranial volume (ICV). NQ and FS use similar segmentation methods; however, NQ utilizes a different probabilistic atlas, independent codebase, intensity normalization, and gradient distortion correction method to accommodate the scanner-specific acquisition-level differences.[13] Although NQ was introduced clinically for brain atrophy measurement, FS is still regarded as a reference standard for brain volumetry and has been used extensively in research.

In the era of big data, both clinicians and researchers are becoming increasingly aware of the reproducibility problems between different software.[14,15] Volumetric results can be affected not only by the use of different software,[16] but also by image acquisition conditions such as slice thickness. Previous studies have shown good inter-method reliability of volumetric measurements by NQ and FS for most brain regions, including the hippocampus.[1,12,17,18] However, the effect of 3D T1 slice thickness on brain volumetry has not yet been investigated.

We hypothesized that different slice thicknesses in T1 volume imaging sequence might affect the inter-method reliabilities of different software depending on the structure. We reasoned that this evaluation would be more appropriately tested in participants with clinical mild cognitive impairment (MCI), as both normal healthy controls and persons with advanced stage of AD are the extremes of the spectrum in the context of clinical practice. Accordingly, this study aimed to examine the effect of slice thickness in T1 volume imaging sequence on the inter-method reliability and volumetry of NQ and FS in patients with MCI.

MATERIALS AND METHODS

Study population

The Institutional Review Board approved this study and waived the requirement for informed consent (IRB number : 2019-08-034, IRB institution : Konkuk University Medical Cente) due to the retrospective nature of the study. Patients who underwent brain MRI and were subsequently diagnosed with MCI (n=102) between September 2016 and December 2017 at our memory clinic were considered. MCI was diagnosed according to the operational criteria of Petersen, et al.[19] Patients with insufficient MRI or clinical data were excluded. Two groups were identified based on the type of MR protocol. The case-control matching procedure was used to select 40 age-matched patients from each group. Patients in Group 1 (n=40; female:male=26:14; mean age=71.6±7.0 years; age range=57–85 years) had 1.2 mm thick sagittal T1-weighted MRI, and those in Group 2 (n=40; female:male=25:15; mean age=72.2±6.8 years; age range=57–81 years) had 1 mm thick sagittal T1-weighted MRI.

Image acquisition

Routine MRI protocols were obtained with a 3T MR scanner (Discovery MR750+; GE Healthcare, Waukesha, WI, USA) for the following sequences: axial and sagittal T1-weighted inversion recovery imaging [repetition time (TR)/echo time (TE)=2468/12; inversion time=920 ms; section thickness 5 mm; matrix 512×224]; axial T2-weighted fast spin-echo imaging (TR/effective TE=4000/106; section thickness 5 mm; matrix 384×384); axial fluid-attenuated inversion recovery imaging (TR/TE=11000/105; inversion time=2600 ms; section thickness 5 mm; matrix 384×224); and axial T2-weighted gradient-recalled echo imaging (TR/TE=550/17; section thickness 5 mm; matrix 384×224; flip angle 15°). The sagittal T1-weighted volumetric fast spoiled gradient-recalled echo was either TR/TE=5.692/2.36; section thickness 1.2 mm; matrix 192×192; flip angle 8°; field of view (FOV) 240×240 mm for Group 1, or TR/TE=8.224/3.192; section thickness 1 mm; matrix 256×256; flip angle 12°; FOV 250×250 mm for Group 2.

Volumetric analyses

The sagittal T1-weighted volumetric images were analyzed with automated segmentation methods. The brain MRI data of each MCI patient were uploaded to the tool's server. The steps of NQ image processing were as follows: stripping the brain of scalp, skull, and meninges; inflating the brain to a spherical shape; mapping the spherical brain to a common spherical space shared with the Talairach atlas coordinates; identification of segmented brain regions; and deflation of the brain back to its original shape. The brain volume was corrected for the head size difference using division by ICV, and was expressed as a percentage. The results were saved in the NQ database. When a patient's brain region fell below the 5th normative percentile, it was classified as abnormally small. The automated tool also provided an age-related atrophy report, with absolute and relative volumes as a percentage of the ICV for hippocampi, lateral ventricles, and inferior lateral ventricles. The total duration for NQ image processing ranged from 10 to 15 minutes.

The FreeSurfer 6.0.0 (http://surfer.nmr.mgh.harvard.edu, Harvard University, Boston, MA, USA) uses a template-driven approach for volumetric- and surface-based segmentation, as previously described.[9,20] The steps involved in FS image processing were as follows: motion correction, removal of non-brain tissue, automated Talairach transformation, segmentation of the subcortical structures, intensity normalization, tessellation of the gray matter (GM) white matter (WM) boundary, automated topology correction, surface deformation following intensity gradients to optimally place the gray/white and gray/cerebrospinal fluid borders at the location where the greatest shift in intensity defines the transition to the other tissue class, surface inflation, parcellation of the cerebral cortex into units

https://doi.org/10.3349/ymj.2021.62.3.255

based on gyrus and sulcus structure, and finally, creation of a variety of surface-based data. The variable "Brain Segmentation Volume Without Ventricles from Surf," which excludes the brainstem, was used as the FS estimate for brain volume. The variable "Total Gray Matter Volume" was used as the estimate GM volume. The WM volume was obtained by summing "cerebral WM," "cerebellar WM," "brainstem," and "corpus callosum" FS variables. It is notable that FS specifically segments WM hypointensities. The brain, GM, and WM volumes were divided by the "Estimated Total Intracranial Volume" for normalization.[21]

Statistical analysis

The inter-method reliability was assessed by calculating the intraclass correlation coefficient (ICCs) between NQ volumes and FS volumes using MedCalc, version 19.0.5 (MedCalc Software, Ostend, Belgium). The ICC model was based on two-way mixed effects, absolute agreement, and average measures. The guidelines used for ICC interpretation were as follows:[22] ICC >0.9 excellent reliability, 0.75≤ICC≤0.9 good reliability, 0.5≤ICC<0.75 moderate reliability, and ICC<0.5 poor reliability.

As a secondary approach, Pearson's r values were also calculated. Effect size (ESCohen's d) was used to document the magnitude of differences between the two techniques without any implication of causality. The guidelines used to interpret effect size (ES) values were as follows: small, d=0.2; medium, d=0.5; and large, d=0.8.[23]

Ethical approval

All procedures performed in the studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent

The requirement for informed consent was waived due to the retrospective nature of this study.

RESULTS

Basic characteristics of the patients

Table 1 shows the clinicodemographic characteristics of the study patients. Age, sex ratio, WM hyperintensities, and medial temporal lobar atrophy were not significantly different between the two groups. Clinical Dementia Rating and Mini-Mental State Examination (MMSE) scores of the available subjects were significantly different between the groups (p=0.032, <0.001, respectively).

Inter-method reliability

Overall, volume measures by NQ and FS showed a strong positive correlation (r=0.988; 95% confidence interval: 0.986–0.989; p<0.001) (Fig. 1). For both groups, NQ and FS volume measurements showed good-to-excellent inter-method reliability for 20 regions (r=0.78 to 0.94; ICC=0.72–0.96; ES=-1.0–0.75) with the exception of total ICV (ICC=0.45), putamen (ICC: left=0.26, right=0.29), pallidum (ICC: left=-0.02, right=-0.02), and thalamus (ICC: left=0.58, right=0.62) (Table 2).

In both groups, the volumes reported by NQ were larger than those reported by FS in several locations, including the whole brain volume (0.78%), cortical GM (5.34%), and WM (2.68%) (Fig. 2). A mean ES difference of +0.40 was determined for individually measured regions. The same pattern was evident in data using slices with 1.2 and 1 mm thickness. Volume measurements using 1 and 1.2 mm thick slices showed good-to-excellent ICCs (r=0.84–0.96; ICC=0.75–0.97; ES=-1.0–0.73 vs. r=0.66–0.95, ICC=0.68–0.96; ES=-0.9–0.77, respectively), with the exception of the putamen, pallidum, thalamus, and total ICV (Table 3). The ICCs of 1 mm thick slices were slightly higher than the ICCs of 1.2 mm thick slices in every brain region, except the putamen and cerebellum.

Table 1. Clinical Characteristics of the Study Patients

Variables	Group 1 (n=40)	Group 2 (n=40)	p value
Sex male	14 (35.0)	15 (37.5)	0.817
Age (yr)	71.6±7.0	72.2±6.8	0.699
MMSE	25.8±3.1	21.8±4.2	<0.0001
CDR	0.46±0.46	0.67±0.34	0.032
WMH	1.7±0.8	1.6±0.7	0.554
MTA	1.5±0.8	1.6±0.9	0.601

CDR, Clinical Dementia Rating; WMH, white matter hyperintensity; MMSE, Mini-Mental State Examination; MTA, medial temporal lobe atrophy.
Data are expressed as mean±standard deviation or n (%) for continuous variables, unless otherwise specified.

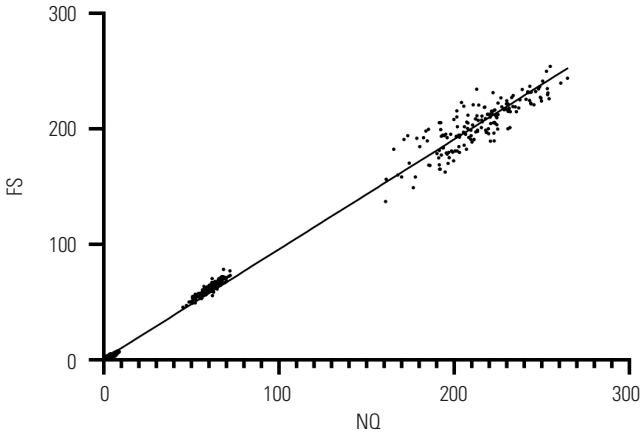

Fig. 1. Scatterplot showing the correlation between NQ and FS. A Pearson correlation coefficient was computed to assess the relationship between volume measurements obtained by NQ and FS. There was a strong positive correlation between the two variables; r=0.988, n=80, 95% CI: 0.986–0.989, p<0.001. NQ, NeuroQuant; FS, Freesurfer.

Table 2. Correlation of Volume Measurements between NeuroQuant and Freesurfer

Region	Left				Right				Total			
	ES	r	ICC	95% CI	ES	r	ICC	95% CI	ES	r	ICC	95% CI
Total intracranial volume									0.05	0.36	0.45	0.15–0.65
Whole brain volume									-0.08	0.90	0.94	0.91–0.96
Cortical gray matter	-0.46	0.83	0.86	0.53–0.94	-0.56	0.87	0.86	0.22–0.95	-0.52	0.85	0.86	0.37–0.95
White matter									-0.20	0.78	0.86	0.77–0.91
Hippocampus	0.51	0.83	0.84	0.43–0.93	0.53	0.89	0.87	0.25–0.96				
Amygdala	-0.96	0.82	0.72	-0.20–0.91	0.13	0.84	0.90	0.85–0.94				
Caudate	0.75	0.84	0.79	-0.08–0.93	0.71	0.83	0.79	0.02–0.93				
Putamen	-2.56	0.68	0.26	-0.10–0.63	-2.25	0.72	0.29	-0.16–0.57				
Pallidum	6.31	0.19	-0.02	-0.06–0.07	6.35	-0.25	-0.02	-0.06–0.05				
Thalamus	-1.21	0.71	0.58	-0.22–0.84	-1.30	0.84	0.62	-0.17–0.88				
Cerebellum	0.26	0.94	0.96	0.84–0.98	0.28	0.92	0.94	0.81–0.97				

ES, effect size; r, Pearson correlation coefficient; ICC, intraclass correlation coefficient; CI, confidence interval.

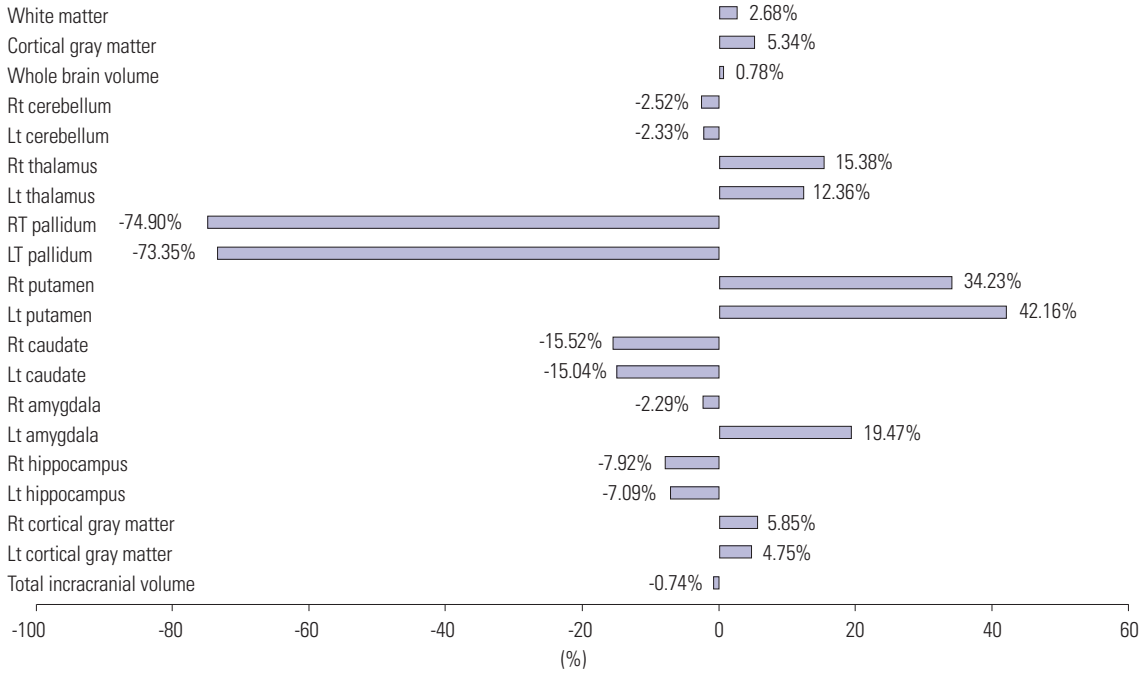

Fig. 2. Percentages of volumes measured by NQ and FS, (NQ–FS)/FS×100%. NQ, NeuroQuant; FS, Freesurfer.

DISCUSSION

We found that the volume measurements between NQ and FS were excellently correlated, regardless of the slice thickness used. Furthermore, our results showed that systematically measured volumes by NQ were slightly higher than those by FS, and the inter-method reliability was slightly higher for 1 mm thick slices compared to 1.2 mm thick slices.

Neurodegenerative disorders are reliably associated with the patterns of progressive neural atrophy that can be quantified by MRI post-processing techniques. In clinical practice, brain volumetry can characterize disease processes, identify the risk for rapid clinical deterioration, and predict prognosis by providing objective and quantitative evidence.[1] However, different MRI parameters influence the results of volumetric measures. Therefore, in many clinical trials using volumetry as the outcome measure, it is recommended to use the suggested MR sequence and to require careful consideration while interpreting the data using existing methods.

In this study, the volume measurements by NQ and FS showed excellent inter-method reliability for 20 brain regions (ICC=0.72–0.96), except for the total ICV, putamen, pallidum, and thalamus. Our finding was consistent with the findings of Ochs, et al.[13] which reported good-to-excellent inter-method reliability between NQ and FS in 60 subjects for all brain regions (caudate and thalamus 0.4≤ICC≤ 0.75, others ICC>0.75), except the pallidum and cerebellar WM (ICC<0.4). Our ICCs between the two volumetric analyses were comparable to the

https://doi.org/10.3349/ymj.2021.62.3.255

Table 3. Correlation of Volume Measurements between NeuroQuant and Freesurfer with Different Slice Thickness

Region	With 1 mm slice thickness				With 1.2 mm slice thickness			
	ES	r	ICC	95% CI	ES	r	ICC	95% CI
Total intracranial volume	-0.14	0.43	0.61	0.26–0.79	0.15	0.37	0.41	-0.12–0.69
Whole brain volume	-0.19	0.96	0.97	0.91–0.99	0.00	0.84	0.91	0.83–0.95
Lt cortical gray matter	-0.37	0.91	0.92	0.64–0.97	-0.57	0.75	0.79	0.33–0.91
Rt cortical gray matter	-0.56	0.92	0.89	0.00–0.97	-0.57	0.81	0.82	0.27–0.94
Cortical gray matter	-0.47	0.92	0.91	0.27–0.97	-0.57	0.78	0.81	0.30–0.92
White matter	-0.51	0.94	0.90	0.10–0.97	-0.02	0.66	0.78	0.59–0.89
Lt hippocampus	0.40	0.88	0.87	0.62–0.95	0.64	0.80	0.80	0.16–0.93
Rt hippocampus	0.45	0.92	0.91	0.38–0.97	0.64	0.84	0.82	0.09–0.94
Lt amygdala	-1.00	0.89	0.75	-0.16–0.93	-0.93	0.79	0.68	-0.16–0.88
Rt amygdala	0.04	0.85	0.90	0.82–0.95	0.25	0.85	0.90	0.79–0.95
Lt caudate	0.73	0.84	0.78	-0.18–0.94	0.77	0.85	0.78	-0.11–0.93
Rt caudate	0.64	0.87	0.84	0.03–0.95	0.82	0.75	0.69	-0.06–0.88
Lt putamen	-2.34	0.73	0.26	-0.07–0.64	-2.57	0.69	0.27	-0.09–0.65
Rt putamen	-2.30	0.79	0.32	-0.10–0.70	-2.27	0.78	0.36	-0.08–0.74
Lt pallidum	7.14	-0.16	-0.01	-0.04–0.04	6.25	-0.31	-0.02	-0.06–0.06
Rt pallidum	7.28	0.06	0.00	-0.02–0.04	6.28	-0.33	-0.02	-0.06–0.07
Lt thalamus	-1.05	0.76	0.66	-0.21–0.89	-1.42	0.65	0.47	-0.23–0.78
Rt thalamus	-1.34	0.84	0.60	-0.16–0.88	-1.35	0.82	0.59	-0.17–0.87
Lt cerebellum	0.23	0.94	0.96	0.86–0.98	0.27	0.95	0.96	0.78–0.98
Rt cerebellum	0.23	0.90	0.93	0.83–0.97	0.33	0.94	0.93	0.67–0.98

ES, effect size; r, Pearson correlation coefficient; ICC, intraclass correlation coefficient; CI, confidence interval; Lt, left; Rt, right.

reliability between human raters (0.73 to 0.85). [24]

The poor reliability and very large ES by NQ versus FS measurement of the pallidum was notable and could be explained by the similar intensities of the pallidum and WM in T1-weighted MRIs.[17] This finding also corroborated the previous observation by Ochs, et al.[13] It is known that FS calculates the volume by including WM between the pallidum and the neighboring putamen; in contrast, NQ uses color mapping images.[13] Furthermore, colored segmentation maps of NQ are smoothed and overlaid onto the original grayscale image, whereas the colored map of FS is neither smoothed nor overlaid onto a grayscale image, which might further exacerbate differences in the volume measurements. Fischl, et al.[17] reported statistically indistinguishable results between automated FS segmentation and manual segmentation of deep brain structures, which has made FS a status of the gold standard for volumetric measurement. Therefore, we recommend careful interpretation while determining the pallidum volumes by NQ.

We also observed larger volumes by NQ compared to FS in several brain locations, including the whole brain (0.78%), cortical GM (5.34%), and WM (2.68%). A mean ES difference of +0.40 was determined for the individually measured regions. These results were similar to the reports of Ochs, et al.,[13] where the whole brain parenchyma volume by NQ was 6.5% larger than that reported by FS, with a mean ES difference of +0.40 for individually measured regions. However, the whole brain volume showed excellent ICC; however, the total ICV showed weaker ICC (0.4–0.6), which was not in line with the previous study.[13] We speculate this may have originated from the fundamental errors owned by the software, as ICV is calculated and estimated based on each calculation formula.

Interestingly, different slice thickness (1.2 mm vs. 1 mm) did not affect the final volumetric results, although the ICCs improved slightly with thinner image slices. Furthermore, regardless of the slice thickness used, the volume measurements were consistently higher with NQ as compared to FS.

We opine that our observation of slightly improved ICCs with thinner image slices would have many implications in the near future. Currently, many software venders recommend the use of rather thick slices (1.2 mm) for clinical practice, instead of using 1 mm, which is a norm of research community.[25] To clarify these recommendations, we used the ES to compare the mean difference between the two groups in a standardized manner.[26] For instance, while the overall reliability was excellent, the ES of hippocampal measurement with 1.2 mm thick slices was larger (ES=0.64) compared to using 1 mm thick slices (ES=0.40–0.45), which means that the use of 1.2 mm slice thickness is prone to a bigger difference in volume measurement when a volumetry software is switched to another software. Therefore, we recommend careful interpretation of the results of volume measurements using a slice thickness of 1.2 mm, instead of 1.0 mm, in both FS and NQ.

In the context of clinical practice, speed is a major advantage of NQ over FS. NQ saves time by abandoning the intensive com-

putation process of FS for parcellation of the cerebral cortex. As a result, NQ outputs the overall volumes of cerebral GM and WM, whereas FS calculates the individual volume and thickness measurements of virtually every cerebral gyrus.[13] NQ processing takes approximately 10 minutes in a conventional desktop computer, and the only input is the study to be segmented. NQ can interact directly with a PACS server, or can be configured as a remotely hosted web server. The final report provides the volumes of structures in cubic centimeters and the ICVs as percentage. A normative range is provided for the hippocampus, lateral ventricle, and temporal horn of the lateral ventricle, based on previously segmented healthy subjects aged 50–95 years.[1] In contrast, FS analysis takes approximately 8 hours with a 2.4 GHz Macintosh computer, and requires knowledge of UNIX programming for analysis.[13] In our study, NQ took 5–10 minutes while FS took 4–6 hours for analysis.

Since the results of the analysis can be affected by the MRI scanner setting, MRI software, NQ and FS software, and computer hardware, it is ideal to use the same hardware and software for comparison purposes.[27] In this study, the effect of hardware or software was controlled by using the same volumetric software programs and computer hardware for all patients.

The current study had several limitations. First, as the study used a small sample of patient data from a single tertiary referral hospital, there is potential for selection bias. Second, the MMSE results between two groups differ significantly, which might affect the difference in volume. Third, the repeatability of different MR sequences in the same scanner was not considered. Moreover, different MR scanning parameters might affect the volume measurements in different ways. Lastly, this study lacked biomarkers or neuropsychological assessments; therefore, patient factors, such as disease severity, could have affected the results between groups.

In conclusion, NQ and FS showed excellent inter-method reliability in volumetric measurements of all brain regions, except pallidum, in patients with MCI. The slice thickness might affect the inter-method reliability of volumetric measures, albeit to a very small degree, with thinner slices providing better reliability than thicker slices. The study outcomes could improve the precise interpretation of automated volume measurements in clinical practice. Future studies are warranted to examine specific measures as biological markers in patients with cognitive impairment.

ACKNOWLEDGEMENTS

This study was supported by a 2017 Clinical Practice Guideline Research Fund grant from the Korean Society of Radiology.

AUTHOR CONTRIBUTIONS

Conceptualization: Younghee Yim and Won-Jin Moon. Data curation: Yeonsil Moon and Hong Jun Jeon. Formal analysis: Younghee Yim, Yeonsil Moon, and Hong Jun Jeon. Funding acquisition: Won-Jin Moon. Investigation: all authors. Methodology: Younghee Yim, Won-Jin Moon, Ji Young Lee, and Ji Eun Park. Project administration: Younghee Yim. Resources: Yeonsil Moon and Hong Jun Jeon. Software: Younghee Yim. Supervision: Won-Jin Moon. Validation: Ji Young Lee, Se Won Oh, Mi Sun Chung, and Ji Eun Park. Visualization: Younghee Yim and Won-Jin Moon. Writing—original draft: Younghee Yim and Won-Jin Moon. Writing—review & editing: Younghee Yim, Ji Young Lee, Se Won Oh, Mi Sun Chung, Ji Eun Park, and Yeonsil Moon. Approval of final manuscript: all authors.

ORCID iDs

Younghee Yim https://orcid.org/0000-0002-4224-7832
Ji Young Lee https://orcid.org/0000-0003-1181-8070
Se Won Oh https://orcid.org/0000-0003-1336-4498
Mi Sun Chung https://orcid.org/0000-0003-1141-9555
Ji Eun Park https://orcid.org/0000-0002-4419-4682
Yeonsil Moon https://orcid.org/0000-0001-7770-4127
Hong Jun Jeon https://orcid.org/0000-0002-0260-0494
Won-Jin Moon https://orcid.org/0000-0002-8925-7376

REFERENCES

1. Brewer JB, Magda S, Airriess C, Smith ME. Fully-automated quantification of regional brain volumes for improved detection of focal atrophy in Alzheimer disease. AJNR Am J Neuroradiol 2009; 30:578-80.
2. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 1991;82:239-59.
3. Jack CR Jr, Petersen RC, Xu YC, Waring SC, O'brien PC, Tangalos EG, et al. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. Neurology 1997;49:786-94.
4. Park M, Moon WJ. Structural MR imaging in the diagnosis of Alzheimer's disease and other neurodegenerative dementia: current imaging approach and future perspectives. Korean J Radiol 2016; 17:827-45.
5. Min J, Moon WJ, Jeon JY, Choi JW, Moon YS, Han SH. Diagnostic efficacy of structural MRI in patients with mild-to-moderate Alzheimer disease: automated volumetric assessment versus visual assessment. AJR Am J Roentgenol 2017;208:617-23.
6. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. Neuroimage 2012;62:782-90.
7. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004; 23 Suppl 1:S208-19.
8. Ashburner J, Friston KJ. Voxel-based morphometry--the methods. Neuroimage 2000;11(6 Pt 1):805-21.
9. Fischl B. FreeSurfer. Neuroimage 2012;62:774-81.
10. Brewer JB. Fully-automated volumetric MRI with normative ranges: translation to clinical practice. Behav Neurol 2009;21:21-8.
11. Birk S. Hippocampal atrophy: biomarker for early AD? Hippocampal volume in patients with AD is typically two standard deviations below normal. Internal Medicine News 2009;42:12.
12. Kovacevic S, Rafii MS, Brewer JB; Alzheimer's Disease Neuroimaging Initiative. High-throughput, fully automated volumetry for prediction of MMSE and CDR decline in mild cognitive impairment. Alzheimer Dis Assoc Disord 2009;23:139-45.
13. Ochs AL, Ross DE, Zannoni MD, Abildskov TJ, Bigler ED; Alzheimer's Disease Neuroimaging Initiative. Comparison of automated brain volume measures obtained with NeuroQuant® and FreeSurfer. J Neuroimaging 2015;25:721-7.

https://doi.org/10.3349/ymj.2021.62.3.255

14. Baker M. Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the'crisis rocking science and what they think will help. Nature 2016;533:452-5.
15. Ross DE, Ochs AL, Tate DF, Tokac U, Seabaugh J, Abildskov TJ, et al. High correlations between MRI brain volume measurements based on NeuroQuant® and FreeSurfer. Psychiatry Res Neuroimaging 2018;278:69-76.
16. Storelli L, Rocca MA, Pagani E, Van Hecke W, Horsfield MA, De Stefano N, et al. Measurement of whole-brain and gray matter atrophy in multiple sclerosis: assessment with MR imaging. Radiology 2018;288:554-64.
17. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002;33:341-55.
18. Bigler ED, Abildskov TJ, Wilde EA, McCauley SR, Li X, Merkley TL, et al. Diffuse damage in pediatric traumatic brain injury: a comparison of automated versus operator-controlled quantification methods. Neuroimage 2010;50:1017-26.
19. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56:303-8.
20. Ferreira D, Voevodskaya O, Imrell K, Stawiarz L, Spulber G, Wahlund LO, et al. Multiple sclerosis patients lacking oligoclonal bands in the cerebrospinal fluid have less global and regional brain atrophy. J Neuroimmunol 2014;274:149-54.
21. Guo C, Ferreira D, Fink K, Westman E, Granberg T. Repeatability and reproducibility of FreeSurfer, FSL-SIENAX and SPM brain volumetric measurements and the effect of lesion filling in multiple sclerosis. Eur Radiol 2019;29:1355-64.
22. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155-63.
23. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988.
24. Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM, et al. Identification of common variants associated with human hippocampal and intracranial volumes. Nat Genet 2012;44:552-61.
25. CorTechs Labs Inc. NeuroQuant 3.0: scanner parameters [accessed on 2020 October 30]. Available at: https://www.cortechs.ai/wp-content/uploads/2019/04/NeuroQuant-Scanner-Parameters_NQ3.0.pdf.
26. Olejnik S, Algina J. Measures of effect size for comparative studies: applications, interpretations, and limitations. Contemp Educ Psychol 2000;25:241-86.
27. Gronenschild EH, Habets P, Jacobs HI, Mengelers R, Rozendaal N, Van Os J, et al. The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. PLoS One 2012;7:e38234.