Comparison of Automated Brain Volume Measures obtained with NeuroQuant® and FreeSurfer

Alfred L. Ochs, David E. Ross, Megan D. Zannoni, Tracy J. Abildskov, Erin D. Bigler, For the Alzheimer's Disease Neuroimaging Initiative*

From the Virginia Institute of Neuropsychiatry, Midlothian, VA (ALO, DER, MDZ); Department of Biomedical Engineering, Virginia Commonwealth University, Richmond, VA (ALO); Department of Psychiatry, Virginia Commonwealth University, Richmond, VA (DER); Department of Psychology and Neuroscience Center, Brigham Young University, Provo, UT (TJA, EDB); and Department of Psychiatry and The Brain Institute of Utah, University of Utah, Salt Lake City, UT (EDB).

ABSTRACT

PURPOSE: **To examine intermethod reliabilities and differences between FreeSurfer and the FDA-cleared congener, NeuroQuant®, both fully automated methods for structural brain MRI measurements.**

MATERIALS AND METHODS: **MRI scans from 20 normal control subjects, 20 Alzheimer's disease patients, and 20 mild traumatically brain-injured patients were analyzed with NeuroQuant® and with FreeSurfer. Intermethod reliability was evaluated.**

RESULTS: **Pairwise correlation coefficients, intraclass correlation coefficients, and effect size differences were computed. NeuroQuant® versus FreeSurfer measures showed excellent to good intermethod reliability for the 21 regions evaluated (*r*: .63 to .99/ICC: .62 to .99/ES: −.33 to 2.08) except for the pallidum (*r*/ICC/ES = .31/.29/−2.2) and cerebellar white matter (*r*/ICC/ES = .31/.31/.08). Volumes reported by NeuroQuant were generally larger than those reported by FreeSurfer with the whole brain parenchyma volume reported by NeuroQuant 6.50% larger than the volume reported by FreeSurfer. There was no systematic difference in results between the 3 subgroups.**

CONCLUSION: **NeuroQuant® and FreeSurfer showed good to excellent intermethod reliability in volumetric measurements for all brain regions examined with the only exceptions being the pallidum and cerebellar white matter. This finding was robust for normal individuals, patients with Alzheimer's disease, and patients with mild traumatic brain injury.**

Keywords: Brain morphometry, freesurfer, neuroquant®, reliability, pallidum.

Acceptance: Received March 31, 2014, and in revised form January 4, 2015. Accepted for publication January 16, 2015.

Correspondence: Address correspondence to Alfred L. Ochs, Virginia Institute of Neuropsychiatry, 364 Browns Hill Court, Midlothian, VA 23114, 804-502-7331; E-mail: aochs@VaNeuropsychiatry.org.

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Funding was internal for work done at the Virginia Institute of Neuropsychiatry and the Brigham Young University.

J Neuroimaging 2015;25:721-727.
DOI: 10.1111/jon.12229

Introduction

Over the past few decades, magnetic resonance imaging (MRI) of the brain has proven to be useful for detecting brain volume abnormalities in many neuropsychiatric disorders. Currently, several software packages for brain segmentation and morphometry are available. Notable among these are FSL,[1,2] Voxel-Based Morphometry,[3] FreeSurfer (FS),[4] and NeuroQuant® (NQ).[5] The first 3 of these techniques have undergone well-documented development, have been used in diverse research environments, and are accepted by the imaging community. To a limited extent, these 4 techniques have been cross-validated with each other and all with expert manual tracing.[6–14] At this time, a majority of current investigations involving brain segment morphometry, and in particular subcortical nuclei, seemingly utilize either FSL or FS (eg, see Stein et al, Table A2[12]). The purpose of this paper is to show the reliability of NQ (http://www.cortechs.net) in relation to its well-studied progenitor, FS (https://surfer.nmr.mgh.harvard.edu), now especially significant because NQ is being marketed as a U.S. Food and Drug Administration (FDA) 510k cleared clinical device.

In 2006, CorTechs Labs, San Diego, CA, introduced NQ as a commercial subset of FS.[15] Although CorTechs Labs initially marketed NQ for Alzheimer's disease, its FDA clearance for clinical use does not preclude NQ's use for other conditions. NQ has been reported to be useful in evaluating patients with progression of mild cognitive impairment,[16–19] Alzheimer's disease,[20,21] mild traumatic brain injury (TBI),[22–28] and epilepsy.[29] Although FS is widely used for research purposes, the developers have not applied to the FDA for FS to be cleared or approved for clinical use.

The NQ segmentation algorithm and validation results were described in Brewer et al.[20] "NeuroQuant relies on similar segmentation methods to those used by FreeSurfer, but it utilizes a different probabilistic atlas and independent code base, and it includes separate methods for intensity normalization and gradient distortion correction to accommodate scanner specific acquisition-level differences" (personal communication, CorTechs Labs, 2013).

Both NQ and FS have been validated with comparisons to manually segmented volumetric measurements.[7,17,20,30] However, with the exception of hippocampal volume measures,[17]

they have not been compared with each other. In that study of hippocampal volume, NQ was found to be reliable when compared with FS. Other studies compared different versions of FS with each other and found that differences occur with different FS versions, workstations, and operating systems.[31,32]

Methods

Subjects

Some of the data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu[33–35]). The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. For this study, we analyzed MRI scans from 20 ADNI1 normal subjects, 10 male, and 10 female, chosen from the larger elderly group to be as young as possible (mean age = 68.0, std. dev. = 3.75; mean years education = 16.0, std. dev. = 3.06). Scans from 20 ADNI1 Alzheimer's disease patients matched in sex, age, and education also were chosen (mean age = 68.0, std. dev. = 4.31; mean years education = 16.0, std. dev. = 2.95).

For these ADNI scans, high-resolution, T1, sagittal, MRI scans were performed without the use of contrast media. The diversity among scanning conditions is described in Table A1. To compensate for the inherent diversity among scanner manufactures, models, software, and machine settings, stringent quality control was employed by ADNI. Each scanning session included a scan of a custom-designed brain phantom. All scans were then submitted to an ADNI site at the Mayo Clinic for inspection and approval before entry into the public database. The use of consistent sequence protocols, consistent phantoms, and quality assurance criteria have been shown to allow for fully automated evaluation and interpretation of data collected across many different sites and different brand MRI machines.[36] A FS and NQ analyses requires that the scanning center follow acquisition protocols published by ADNI (http://adni.loni.usc.edu/methods/documents/mri-protocols), the Martinos Center (http://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferBeginnersGuide), or CorTechs Labs (http://www.cortechslabs.com/wp-content/uploads/2014/06/ScannerSetup.pdf).

In addition to the above-described ADNI data, 20 scans were performed on patients with nonpenetrating mild TBI, 10 male and 10 female (mean age at the time of the injury = 44.3, std. dev. = 11.42; mean years education = 14.3, std. dev. = 2.63). Fifteen of the TBIs resulted from motor vehicle accidents and the remaining 5 from other impacts to the head. These patients presented to the Virginia Institute of Neuropsychiatry with various neuropsychiatric conditions attributed to their brain injury and were determined to have suffered a mild TBI according to the definitions of Menon et al[37] and Kay et al.[38] Initial MRIs were performed at this time (median interval from date of TBI = 18 months, std. dev. = 13 months). MRIs were performed at various community hospitals using NQ-recommended protocols derived from ADNI recommendations. ADNI phantoms were not employed.

Informed Consent

For each ADNI subject, written informed consent, approved by an Institutional Review Board of each ADNI study site, was obtained[34] The written informed consent obtained from TBI patients included in this study was approved by the New England Institutional Review Board and satisfied the requirements of the Code of Ethics of the World Medical Association (Declaration of Helsinki) for human research.

Volumetric Analyses

NQ analyses were performed by CorTechs Labs, San Diego, CA, using a CentOS 5 UNIX operating system running on a Dell PowerEdge 1950 platform utilizing NQ version 1.4. FS analyses were performed on a MacBook Pro computer utilizing FS version 5.3.0. The Macintosh computer had a 2.4 GHz Intel Quad-Core i7 running a Mac OS 10.8.5 operating system (Mountain Lion) with a 64-bit Mac UNIX terminal using a tsch shell.

NQ returns volumetric data in a variety of ways. The NQ General Report returns volumes from 22 regions with data for 32 regions available in a spreadsheet format. The FS aseg.stats file contains measurements of subcortical regions, and it is this file that corresponds most directly to the output of NQ. Twenty-eight NQ regions corresponded directly to regions measured by FS (Table A2A). Three regions (whole brain parenchyma, cerebellum and total CSF) were summated from components for convenience (Table A2B). NQ nomenclature is used throughout this paper. Since the purpose of this study is to assess intermethod reliability under diverse condition, scans were not edited for misidentified regions.

Statistics

Intermethod reliabilities were measured by calculating intraclass correlation coefficients (ICC; 3, 1) using the terminology of Shrout and Fleis,[39] and were performed using SPSS version 22 (Model = 2-Way Mixed; Type = Consistency; Confidence Interval = 95%). To interpret ICC values, the following guidelines were used[40]:

$$\text{ICC} > .75 \quad \text{excellent reliability}$$

$$.4 \leq \text{ICC} \leq .75 \quad \text{fair to good reliability}$$

$$\text{ICC} < .4 \quad \text{poor reliability}$$

Pearson's r correlations were computed using the CORREL function in Microsoft Excel. Effect size (Cohen's d) was used to document the magnitude of differences between the two techniques without any implication of causality as the word "effect" is commonly understood. With the mean of each group, designated as $\bar{x}$, and the standard deviation, designated by σ, the effect size, ie, the standardized mean difference between 2 groups, is given as

$$d = (\bar{x}_1 - \bar{x}_2)/\sigma$$

Since neither Group 1 (NQ) nor Group 2 (FS) could be considered *a priori* to be the more accurate, a root mean square average of $\sigma_1 + \sigma_2$ was used.[41,42]

$$\sigma = \left(\left[\sigma_1^2 + \sigma_2^2\right]/2\right)^{1/2}$$

To interpret effect size values, the following guidelines were used[43]:

Small: d = .2
Medium: d = .5
Large: d = .8

Table 1. Correlations for 60 Subjects. Effect Size, Pearson's Product Moment (r), and ICC (3, 1) Used to Assess NQ and FS Intermethod Reliability in 21 Brain Segments. A positive effect size indicates that the NQ volume measure is larger than the corresponding FS volume measure.

	Effect Size and Correlation Coefficients (Pearson's *r* & ICC)								
	Left			Right			Total (Left + Right)		
Region	E.S.	*r*	ICC	E.S.	*r*	ICC	E.S.	*r*	ICC
Total intracranial volume	*	*	*	*	*	*	.35	.93	.92
Whole brain parenchyma	*	*	*	*	*	*	.61	.96	.96
Forebrain parenchyma	*	*	*	*	*	*	.59	.96	.95
Cortical gray matter	*	*	*	*	*	*	−.33	.89	.88
Cerebral white matter	−.29	.87	.87	−.25	.86	.86	−.27	.86	.87
Lateral ventricle	.23	1.00	1.00	.32	.97	.97	.28	.99	.99
Inferior lateral ventricle	.97	.96	.96	.87	.97	.97	.97	.97	.97
3rd ventricle	*	*	*	*	*	*	−.01	.98	.97
4th ventricle	*	*	*	*	*	*	.64	.89	.89
Total cerebrospinal fluid	*	*	*	*	*	*	.12	.99	.99
Caudate	.71	.63	.62	.79	.58	.56	.77	.63	.62
Putamen	.05	.80	.79	.02	.79	.77	.03	.84	.83
Pallidum	−2.05	.21	.21	−2.04	.36	.34	−2.20	.31	.29
Thalamus	1.56	.67	.67	2.41	.79	.74	2.08	.79	.77
Amygdala	1.10	.86	.85	.87	.85	.85	1.02	.89	.89
Hippocampus	.20	.82	.78	.40	.81	.79	.31	.84	.80
Ventral diencephalon	.62	.77	.77	.87	.75	.75	.77	.81	.81
Cerebellar white matter	.18	.34	.34	−.02	.28	.28	.08	.31	.31
Cerebellar gray matter	.60	.93	.93	.33	.95	.94	.46	.94	.94
Cerebellum	.55	.94	.94	.27	.94	.94	.41	.94	.94
Brainstem	*	*	*	*	*	*	1.71	.95	.95

*Indicates that a direct match could not be determined.

Results

Intermethod Reliability

Across all groups, ICCs between NQ and the FS results were generally excellent to good as were correlations as measured by Pearson's r, except for the pallidum and cerebellar white matter for which the ICCs were poor (Table 1). Although the effect size between FS and NQ measures were very large for the pallidum, they were small for the cerebellar white matter among all 60 subjects (Table 1). Volumes reported by NQ were generally larger than those reported by FS. Overall, the whole brain parenchyma volume reported by NQ was 6.50% larger than the volume reported by FS. A mean effect size difference of +0.40 was determined for individually measured regions.

The same pattern emerged when the 3 subgroups: normal subjects, Alzheimer's disease patients, and TBI patients, were analyzed individually. As documented in Table 2, the pallidum appeared small with NQ measures and showed a poor ICC reliability in all groupings. This large discrepancy warranted further investigation. Figure 1 shows a plot of NQ and FS determinations of pallidum volume for all subjects. For the pallidum, the NQ volume was generally smaller than the corresponding FS value. To provide perspective, volume measures for the amygdala are included in Figure 1. The most extreme difference for the pallidum occurred in the normal subject 002_S_1261. No important difference was seen in a similar plot of the volume of these structures divided by the total intracranial volume. Also for perspective, the left and right pallidum volumes of normal subject 098_S_0172 are designated. This comparison subject was chosen for illustration because the FS and NQ determinations of the right pallidum volume were virtually equal and the

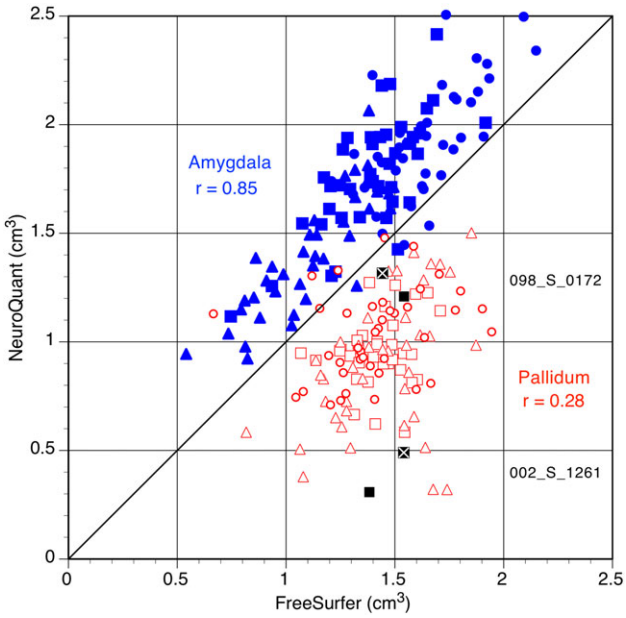

Fig 1. NeuroQuant versus FreeSurfer. Shown are NQ and FS determined pallidum volumes (red open symbols) and the similar-sized amygdala (blue filled symbols) for all subjects. Squares are measurements from normal subjects, triangles are from Alzheimer's disease patients and circles from TBI patients. Pearson's r is shown for each region. The diagonal line indicates equal volumes for both measurement systems. A filled black square indicates the left and a black square with a diagonal cross indicates the right pallidum volumes for normal subjects 002_S_1261 and 098_S_0172. The largest discrepancy between the two measurement systems was for subject 002_S_1261, having among the smallest NQ measured volumes.

Table 2. Correlations for Each Group. Effect Size, Pearson's Product Moment (r), and ICC (3,1) Used to Assess NQ and FS Intermethod Reliability in 21 Brain Segments (Left + Right Hemispheres) Computed for Normal Subjects, Alzheimer's Disease Patients, and Mild TBI Patients.

	Effect Size and Correlation Coefficients (Pearson's *r* & ICC)								
	Normal Subjects			Alzheimer's Disease			TBI Patients		
Region	E.S.	*r*	ICC	E.S.	*r*	ICC	E.S.	*r*	ICC
Total intracranial volume	.33	.94	.93	.04	.99	.99	.74	.96	.96
Whole brain parenchyma	.66	.93	.93	.56	.97	.97	.74	.96	.94
Forebrain parenchyma	.63	.93	.92	.53	.97	.97	.74	.95	.93
Cortical gray matter	−.61	.83	.80	−.45	.82	.82	−.15	.90	.90
Cerebral white matter	−.09	.88	.88	−.23	.89	.89	−.54	.86	.86
Lateral ventricle	.33	1.00	1.00	.39	1.00	1.00	.35	.95	.94
Inferior lateral ventricle	2.13	.91	.91	1.01	.97	.97	3.11	.71	.66
3rd ventricle	−.10	.97	.97	−.04	.96	.95	.11	.95	.94
4th ventricle	.68	.92	.92	.99	.97	.96	.45	.70	.70
Total cerebrospinal fluid	.15	1.00	1.00	.17	1.00	1.00	.17	.95	.95
Caudate	.78	.79	.79	1.31	.87	.85	.17	.46	.42
Putamen	.16	.84	.80	.17	.92	.91	−.27	.63	.63
Pallidum	−2.97	.16	.14	−2.08	.47	.44	−1.91	.25	.25
Thalamus	2.28	.63	.59	2.05	.90	.89	2.30	.71	.69
Amygdala	1.61	.75	.74	1.21	.91	.90	1.34	.80	.76
Hippocampus	.56	.72	.72	.70	.79	.74	−.19	.85	.85
Ventral diencephalon	.85	.70	.70	.65	.95	.95	.90	.67	.67
Cerebellar white matter	.57	.21	.16	.33	.38	.38	−.77	.70	.69
Cerebellar gray matter	.68	.85	.84	.38	.96	.95	.49	.95	.95
Cerebellum	.77	.91	.90	.40	.94	.95	.25	.96	.96
Brainstem	2.12	.98	.97	1.21	.99	.99	2.49	.91	.91

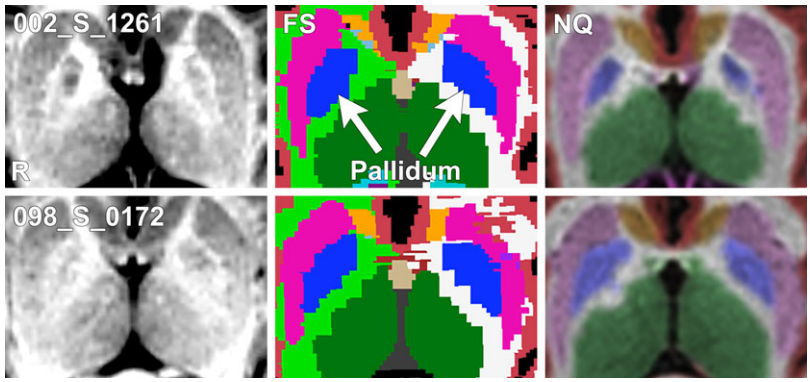

Fig 2. MRI SLICES through the pallidum show axial views of a grayscale MRI, an FS segmentation (FS), and an NQ segmentation (NQ) for subjects 002_S_1261 and 098_S_0172. FreeSurfer routinely returns images with the right white matter shown in green and the left in white. Sagittal, T1-weighted MRIs were obtained on 1.5 Tesla GE Excite scanners described in Table A1 according to the ADNI1 acquisition protocol (see Ref 44). The grayscale and FS images were manually coregistered with the closest image match possible to a NQ slice through the long axis of the pallidum with 4 degrees of freedom (rotation and global scaling). 3D Slicer imaging software (http://www.slicer.org) displayed the grayscale and FS segmented images. OsiriX imaging software (http://www.osirix-viewer.com) displayed the NQ segmented images.

scans for the two subjects were performed on identical model scanners with virtually identical scanning protocols.

Figure 2 shows the original grayscale MRI, an FS segmentation, and an NQ segmentation manually registered through the long dimension of the pallidum for normal subjects 002_S_1261 and 098_S_0172. Visually, the FS identification of the pallidum matches that discernible in the original grayscale image whereas the NQ segmentation appears smaller for subject 002_S_1261. For both subjects shown, FS identification appears to include white matter between the pallidum and the neighboring putamen, an observation consistent with the large effect size difference between FS and NQ. Other deep brain segments imaged in Figure 2 show a good correspondence, reflecting the statistical results presented here. The lack of jagged edges in the NQ image is due to how NQ presents its data. In NQ, the colored segmentation maps are smoothed and overlaid onto the original grayscale image. For FS, the colored map is neither smoothed nor overlaid onto a grayscale image.

Discussion

An advantage of NQ over FS is speed. NQ accomplishes this by abandoning the extremely computationally intensive FS routines for parcellation of the cerebral cortex. That is, NQ returns overall cerebral gray and white matter volumes whereas FS returns individual volume and thickness measures of virtually every cerebral gyrus. Both return volumes of deep brain nuclei,

ventricles, cerebellum, and brain stem. NQ results are available in about 20 minutes using NQ over a web-based link. Analysis of the same scan using FS on a 2.4 GHz Macintosh takes 8 hours. Other practical advantages of NQ include providing segmented, color-identified brain images in a DICOM formatted file, and in its relative ease of use, in contrast to FS which requires knowledge of UNIX programming. However, NQ is unable to provide measures of individual cortical gyri or lobes.

ICCs between NQ and FS evaluations for the 60 subjects evaluated (Table 1) were generally excellent (ICC > .75), good for the caudate and thalamus (.4 ≤ ICC ≤ .75), and poor for the pallidum and cerebellar white matter (ICC < .4). In this regard, Stein et al, in table A4,[12] noted that excellent ICC reliability in the range of .73 to .85 is obtained among human raters and "is arguably a reasonable upper bound on the accuracy of automated segmentation."

As illustrated in Table A1, 60 MRI scans were performed on a wide variety of machines. Overall, volume measures were made on MRI machines from 3 companies and performed on 29 different scanners with both 1.5 and 3.0 Tesla magnets. This diversity of measuring conditions represents what may be obtained across clinical practices and is made possible by scanning protocols developed by ADNI.

Effect sizes differences between NQ and FS (Tables 1 and 2) were consistently small across all groups for lateral ventricles, third ventricles, total cerebrospinal fluid and putamen. They were consistently large for the inferior lateral ventricles, pallidum, thalamus, amygdala, and the brainstem. Effect sizes for other regions were medium or variable.

The poor reliability and very large effect size for NQ versus FS measurement of the pallidum appears to arise from the similar intensities of the pallidum and white matter in T1-weighted MRIs.[7] T1-weighted MRIs are required for NQ because this timing sequence provides good contrast between gray matter, white matter, and cerebrospinal fluid. The anterior portions of the 4 pallidal regions shown in Fig 2 are identifiable visually but a drop in the MRI intensity pattern along the pallidum results in misidentification by the NQ algorithm. In particular, the left pallidum for subject 002_S_1261 is poorly segmented by NQ, with the posterior tail not identified. In contrast, even though the grayscale image appears indistinct, the NQ and FS segmentation of the right pallidum for subject 098_S_0172 match and the 2 segmentation algorithms yielded an almost equal volume. We note that Brewer et al[20] found a borderline excellent ICC (.76) for the comparison of the pallidum between NQ volume determinations and expert computer-aided manual segmentation in 40 elderly subjects. However, Brewer et al used Cronbach's alpha to define ICCs, which returns a higher estimate of reliability than the ICC definition recommended by Shrout and Fleiss.[39] For example, the ICC (3, 1) obtained here for the pallidum of combined subjects is .29 as compared to .45 using Cronbach's alpha (Table 1).

Fischl et al[7] compared volume measures of the same deep brain structures, ventricles and brain stem described here (but not the cortical white matter, cortical gray matter, and the cerebellum). T1-weighted scans from 7 healthy young subjects were analyzed with automated segmentation (FS) and separately by 5 persons expert in performing manual segmentation. For each segment, differences between automated and manual segmentation were "statistically indistinguishable." Given the careful validation of the FS algorithm by Fischl and coworkers and the data presented in Figures 1 and 2, we suggest that determinations of pallidum volumes by NQ should be interpreted with caution.

Even when volume measurements from automated segmentation are shown to be statistically indistinguishable from careful and repeated manual segmentation, one cannot say that they are identical to the actual size of brain segments *in situ* because acquisition factors affect measured volumes.[45–47] Moreover, repeated scans on the same individual performed within a few days show test-retest variability.[13]

Especially for clinical use, as is the case with virtually all medical laboratory tests (clinical chemistry, electrophysiology, etc.), results from each laboratory must be interpreted with standards developed for their own protocols. Here, we see that this is true for computer determination of brain segmentation. FS installation instructions warn "it is essential to process all your subjects with the same version of FS, on the same OS platform and vendor, and for safety, even the same version of the OS."[48] Beyond this, CorTechs Labs and our clinical experience with TBI patients strongly suggests that when using NQ clinically, each scan be inspected and that misidentified regions be excluded in forming a clinical impression. Inspection of NQ DICOM images with color-identified regions overlaid on corresponding grayscale images facilitates this inspection.

Conclusion

NQ and FS showed good to excellent intermethod reliability in volumetric measurements for all brain regions examined with the only exceptions being the pallidum and cerebellar white matter. This finding was robust for normal individuals, patients with Alzheimer's disease, and patients with mild TBI.

We thank Doug Greve of the Athinoula A. Martinos Center for Biomedical Imaging, Harvard University, for his help in defining the relationship between NQ and FS measurements described in Tables A2A and A2B. We are grateful to CorTechs Labs for providing the NQ analyses used in this paper without charge or editorial involvement.

Support

Funding was internal for work done at the Virginia Institute of Neuropsychiatry and the Brigham Young University.

Data collection and sharing for this project was funded by the ADNI (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org). The grantee

organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuroimaging at the University of Southern California.

Conflicts of Interests

There are no financial or other conflicts of interest to disclose. The authors have no financial interest in CorTechs Labs or NQ.

References

1. Jenkinson M, Beckmann CF, Behrens TE, et al. FSL. NeuroImage 2012;62:782-90.
2. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 2004;23(Suppl 1):S208-19.
3. Ashburner J, Friston KJ. Voxel-based morphometry–the methods. NeuroImage 2000;11:805-21.
4. Fischl B. FreeSurfer. NeuroImage 2012;62:774-81.
5. Brewer JB. Fully-automated volumetric MRI with normative ranges: translation to clinical practice. Behav Neurol 2009;21:21-8.
6. Carmichael OT, Aizenstein HA, Davis SW, et al. Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment. NeuroImage 2005;27:979-90.
7. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002;33:341-55.
8. Jatzko A, Rothenhofer S, Schmitt A, et al. Hippocampal volume in chronic posttraumatic stress disorder (PTSD): MRI study using two different evaluation methods. J Affect Disorders 2006;94:121-6.
9. Morey RA, Petty CM, Xu Y, et al. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. NeuroImage 2009;45:855-66.
10. Testa C, Laakso MP, Sabattoli F, et al. A comparison between the accuracy of voxel-based morphometry and hippocampal volumetry in Alzheimer's disease. JMRI-J Magn Reson Im 2004;19:274-82.
11. Voets NL, Hough MG, Douaud G, et al. Evidence for abnormalities of cortical development in adolescent-onset schizophrenia. NeuroImage 2008;43:665-75.
12. Stein JL, Medland SE, Vasquez AA, et al. Identification of common variants associated with human hippocampal and intracranial volumes. Nat Genet 2012;44:552-61.
13. Morey RA, Selgrade ES, Wagner HR 2nd, et al. Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. Hum Brain Mapp 2010;31:1751-62.
14. Mulder ER, de Jong RA, Knol DL, et al. Alzheimer's Disease Neuroimaging I. Hippocampal volume change measurement: quantitative assessment of the reproducibility of expert manual outlining and the automated methods FreeSurfer and FIRST. NeuroImage 2014;92:169-81.
15. Birk S. Hippocampal atrophy: biomarker for early AD? Intern Med News 2009;9:12.
16. Heister D, Brewer JB, Magda S, et al. Predicting MCI outcome with clinically available MRI and CSF biomarkers. Neurology 2011;77:1619-28.
17. Kovacevic S, Rafii MS, Brewer JB. Alzheimer's Disease Neuroimaging I. High-throughput, fully automated volumetry for prediction of MMSE and CDR decline in mild cognitive impairment. Alzheimer Dis Assoc Disord 2009;23:139-45.
18. Desikan RS, Rafii MS, Brewer JB, et al. An expanded role for neuroimaging in the evaluation of memory impairment. Am J Neuroradiol 2013; 34:2075–82.
19. McEvoy LK, Brewer JB. Biomarkers for the clinical evaluation of the cognitively impaired elderly: amyloid is not enough. Imaging in Medicine 2012;4:343-57.
20. Brewer JB, Magda S, Airriess C, et al. Fully-automated quantification of regional brain volumes for improved detection of focal atrophy in Alzheimer disease. Am J Neuroradiol 2009;30: 578-80.
21. McEvoy LK, Brewer JB. Quantitative structural MRI for early detection of Alzheimer's disease. Expert Rev Neurotherapeutics 2010;10:1675-88.
22. Ross DE, Graham TJ, Ochs AL. Review of the evidence supporting the medical and legal use of neuroquant® in patients with traumatic brain injury. Psychol Inj Law 2012;6:75-80.
23. Ross DE, Castelvecchi C, Ochs AL. Brain MRI volumetry in a single patient with mild traumatic brain injury. Brain Injury 2013;27:634-6.
24. Ross DE, Ochs AL, Seabaugh JM, et al. Progressive brain atrophy in patients with chronic neuropsychiatric symptoms after mild traumatic brain injury: a preliminary study. Brain Injury 2012;26: 1500-9.
25. Ross DE, Ochs AL, Seabaugh JM, et al. Man versus machine: comparison of radiologists' interpretations and NeuroQuant® volumetric analysis of brain MRIs in patients with traumatic brain injury. J Neuropsychiatry Clin Neurosci 2013;25:1-8.
26. Ross DE, Ochs AL, DeSmit ME, et al. Man vs. machine Part 2: comparison of radiologists' interpretations and NeuroQuant measures of brain asymmetry and progressive atrophy in patients with traumatic brain injury. J Neuropsychiatry Clin Neurosci 2015;27:147-52.
27. Ross DE, Ochs AL, Zannoni MD, et al. Back to the future: estimating pre-injury brain volume in patients with traumatic brain injury. NeuroImage 2014;102P2:565-78.
28. Brezova V, Moen KG, Skandsen T, et al. Prospective longitudinal MRI study of brain volumes and diffusion changes during the first year after moderate to severe traumatic brain injury. NeuroImage Clin 2014;5:128-40.
29. Farid N, Girard HM, Kemmotsu N, et al. Temporal lobe epilepsy: quantitative MR volumetry in detection of hippocampal atrophy. Radiology 2012;264:542-50.
30. Bigler ED, Abildskov TJ, Wilde EA, et al. Diffuse damage in pediatric traumatic brain injury: a comparison of automated versus operator-controlled quantification methods. NeuroImage 2010;50:1017-26.
31. Gronenschild EH, Habets P, Jacobs HI, 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.
32. Jovicich J, Czanner S, Han X, et al. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. NeuroImage 2009;46:177-92.
33. Jack CR Jr, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. JMRI-J Magn Reson Im 2008;27:685-91.
34. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. Neurology 2010;74:201-9.
35. Weiner MW, Aisen PS, Jack CR Jr, et al. The Alzheimer's disease neuroimaging initiative: progress report and future plans. Alzheimers Dement 2010;6:202-11.
36. Davids M, Zollner FG, Ruttorf M, et al. Fully-automated quality assurance in multi-center studies using MRI phantom measurements. Magn Reson Imaging 2014;32:771-80.
37. Menon DK, Schwab K, Wright DW, et al. Demographics, Clinical Assessment Working Group of the I, Interagency Initiative toward Common Data Elements for Research on Traumatic Brain I, Psychological H. Position statement: definition of traumatic brain injury. Arch Phys Med Rehab 2010;91:1637-40.
38. Kay T, Harrington DE, Adams R, et al. Definition of mild traumatic brain injury. J Head Trauma Rehab 1993;8:86-7.
39. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420-8.
40. Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York: Wiley, 1986.

41. Coe R. It's the effect size, stupid: what effect size is and why it is important. 2002. URL: http://www.leeds.ac.uk/educol/documents/00002182.htm
42. Olejnik S, Algina J. Measures of effect size for comparative studies: applications, interpretations, and limitations. Contemp Educ Psychol 2000;25:241-86.
43. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.
44. MRI Scanner Protocols. URL: http://adni.loni.usc.edu/methods/documents/mri-protocols/
45. Han X, Jovicich J, Salat D, et al. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. NeuroImage 2006;32:180-94.
46. Jovicich J, Czanner S, Greve D, et al. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. NeuroImage 2006;30:436-43.
47. Wonderlick JS, Ziegler DA, Hosseini-Varnamkhasti P, et al. Reliability of MRI-derived cortical and subcortical morphometric measures: effects of pulse sequence, voxel geometry, and parallel imaging. NeuroImage 2009;44:1324-33.
48. FreeSurfer Download. 2013. URL: http://surfer.nmr.mgh.harvard.edu/fswiki/Download

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table A1. Characteristics of MRI Scanners. Sixty subjects were scanned at 29 centers using the ADNI1 protocol or NQ protocol for TBI patients. All centers used a 3D MP-RAGE T1-weighted sequence, except those performed at MCT that used FSPGR. All used a flip angle of 8° or 9° and a 1.2-mm slice thickness without contrast media. Centers are identified by the first 3 alphameric characters.

Table A2A. Key for Corresponding NeuroQuant® and FreeSurfer Names of Brain Regions. NQ and FS use similar names for several brain regions. For the regions in this table, there is an exact one-to-one correspondence between the names. That is, NQ and FS here purport to measure the same brain region.

Table A2B. Key relating Noncorresponding NeuroQuant® and FreeSurfer Names of Brain Regions.