TECHNICAL REPORT

Validation of the Automated Method VIENA: An Accurate, Precise, and Robust Measure of Ventricular Enlargement

Hugo Vrenken,^{1,2*} Eline K. Vos,¹ W.M. van der Flier,^{3,4,5} Ingrid C. Sluimer,¹
Keith S. Cover,² Dirk L. Knol,⁵ and Frederik Barkhof¹

¹Department of Radiology, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands

²Department of Physics and Medical Technology, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands

³Alzheimer Center, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands

⁴Department of Neurology, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands

⁵Department of Epidemiology and Biostatistics, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands

◆=====◆
Abstract: *Background:* In many retrospective studies and large clinical trials, high-resolution, good-contrast 3DT1 images are unavailable, hampering detailed analysis of brain atrophy. Ventricular enlargement then provides a sensitive indirect measure of ongoing central brain atrophy. Validated automated methods are required that can reliably measure ventricular enlargement and are robust across magnetic resonance (MR) image types. *Aim:* To validate the automated method VIENA for measuring the percentage ventricular volume change (PVVC) between two scans. *Materials and Methods:* Accuracy was assessed using four image types, acquired in 15 elderly patients (five with Alzheimer's disease, five with mild cognitive impairment, and five cognitively normal elderly) and 58 patients with multiple sclerosis (MS), by comparing PVVC values from VIENA to manual outlining. Precision was assessed from data with three imaging time points per MS patient, by measuring the difference between the direct (one-step) and indirect (two-step) measurement of ventricular volume change between the first and last time points. The stringent concordance correlation coefficient (CCC) was used to quantify absolute agreement. *Results:* CCC of VIENA with manual measurement was 0.84, indicating good absolute agreement. The median absolute difference between two-step and one-step measurement with VIENA was 1.01%, while CCC was 0.98. Neither initial ventricular volume nor ventricular volume change affected performance of the method. *Discussion:* VIENA has good accuracy and good precision across four image types. VIENA therefore provides a useful fully automated method for measuring

*Correspondence to: Hugo Vrenken, Departments of Radiology and Physics and Medical Technology, VU University Medical Center, De Boelelaan 1118, 1081 HZ Amsterdam, The Netherlands. E-mail: H.Vrenken@vumc.nl

Received for publication 5 December 2011; Revised 31 August 2012; Accepted 8 November 2012

DOI: 10.1002/hbm.22237

Published online 30 January 2013 in Wiley Online Library (wileyonlinelibrary.com).

ventricular volume change in large datasets. *Conclusion:* VIENA is a robust, accurate, and precise method for measuring ventricular volume change. *Hum Brain Mapp* 35:1101–1110, 2014. © 2013 Wiley Periodicals, Inc.

Key words: brain; atrophy; magnetic resonance imaging; computer-assisted image analysis; 3D imaging; Alzheimer disease; multiple sclerosis

OVERALL DESCRIPTION OF THE PROCEDURE

Rationale and Aim

Neurodegeneration causing local volume loss of brain structures close to the ventricular system occurs in several neurological diseases with large societal impact, such as Alzheimer's disease (AD) and multiple sclerosis (MS). It may affect deep gray matter (GM) such as the thalamus and putamen, as well as white matter (WM) fiber bundles such as the corpus callosum and internal capsule. Direct volumetric measurement of many of these structures from magnetic resonance (MR) images, is now possible through image analysis techniques such as Freesurfer (Fischl et al., 2002) or FIRST (Patenaude et al., 2011), and voxelwise analyses can be done through voxel-based morphometry (Ashburner, 2007; Ashburner and Friston, 2000; Good et al., 2001; Jenkinson et al., 2012). However, such techniques require MR images with good contrast between gray and WM, and good spatial resolution in three dimensions (3D). Typically, they require 3D T_1 -weighted (3DT₁) images with near-isotropic spatial resolution. These requirements cannot always be met: in therapeutic trials, adding a 3DT₁ imaging sequence to the MR protocol at each imaging time point may be too costly. Such 3DT₁ images are generally also unavailable in retrospective studies investigating the long-term clinical predictive value of early atrophic changes from archive images, simply because 3DT₁ imaging was not included in the MR acquisition protocol at the time of the original study.

In such situations, the ventricular enlargement that accompanies the brain volume loss is an important candidate surrogate measure, since it can, in principle, be measured from standard two-dimensional (2D) images, thus allowing a more specific assessment of atrophy than whole-brain measures. Ventricular enlargement has proven to be a sensitive indirect measure of ongoing neurodegeneration in MS, AD, and healthy aging (Dalton et al., 2006; Fjell et al., 2006; Jack et al., 2005).

To understand progression of neurodegeneration in MS, AD, and other diseases, or to develop prediction models that may in future be applied in a clinical setting, retrospective studies on large patient cohorts are warranted. For such studies, validated methods are needed that are fully or almost fully automated, and that do not rely on state of the art imaging.

The Structural Image Evaluation, using Normalization, of Atrophy (SIENA) tool (Smith et al., 2002), provides an automated approach for whole-brain longitudinal atrophy

rate measurements that works well across image types (Smith et al., 2002; Neacsu et al., 2008). As part of the FSL suite, it is freely available to academic researchers and widely used. Finally, the SIENA approach can be easily modified to perform calculations on the ventricles alone instead of the entire brain. For these reasons, we modified the SIENA method to measure ventricular enlargement instead of whole-brain shrinkage. One other group has previously reported using a modified version of SIENA for ventricular measurements in aging (Gunter et al., 2003), although they used an implementation different from standard SIENA. The method has so far not been validated.

The aim of the present study was to validate a modified version of the SIENA method to measure ventricular enlargement, referred to here as VIENA. We used MRI datasets from patients with AD and MS, and included different image types, baseline ventricular volumes, and amounts of ventricular volume change, to assess the method's robustness. Accuracy of the method was assessed by comparing its results to results obtained by manual volumetry. Precision of the method was assessed using a subset of our dataset in which all patients had MRI data at three time points, by comparing the change measured from a two-step analysis with an intermediate time point to that obtained from a one-step analysis. Because VIENA performance depends crucially on SIENA performance, we also calculated precision for SIENA in the same manner to verify correct performance of SIENA in our image data.

VIENA Method

The fully automated method for measuring ventricular enlargement was implemented here directly within the SIENA framework (Smith et al., 2002; <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/SIENA>). Briefly, SIENA linearly registers two brain MRI scans from a single individual together to a "half-way" space, performs tissue type segmentation to find brain-non-brain edge points and then calculates the brain edge displacement between the two time points at each edge point. The average brain edge displacement across all edge points is then calculated, and converted to the percentage brain volume change (PBVC) between the two scans using a scaling factor obtained from a calibration step. These calculations are performed twice, once "forward", i.e., starting from the baseline scan, and once "backward", i.e., starting from the repeat scan.

To obtain the percentage ventricle volume change (PVVC) from these data, three steps are required: (1)

Figure 1.

Illustration of the mask used in the automated VIENA method for excluding nonventricular brain–nonbrain edge points from the analyses. (A) The MR image from a patient with MS with large ventricles, registered to MNI-152 space that was used to draw the mask. (B) MNI-152 image with the mask overlaid.

calculation of the average brain edge displacement for only those brain–nonbrain edge points that are on the ventricular surface; (2) conversion of this average ventricular edge displacement to the PVVC value using the appropriate scaling factor for this ventricular subset of brain–nonbrain edge points; and (3) calculating the final PVVC from the forward and backward results.

Step 1 made use of a mask of the ventricle region, shown in Figure 1, drawn specifically for this purpose by

a trained observer (E.K.V.) on an MRI—registered to MNI-152 space—of a single MS patient with exceptionally large ventricles, thus ensuring that for all expected brains, the ventricular edge points would be included in the analyses. For each new analysis, this single MNI-152 space mask was linearly registered to the individual MRI, and then used to include in the subsequent calculations only those brain/nonbrain edge points that fall within the registered mask.

TABLE I. Demographic characteristics of patients and acquisition parameters of MR images

	Group 1	Group 2	Group 3	Group 4
Disease group	MS	MS	MS	Aging/dementia
Number of subjects	10	10	38	5 Elderly control 5 MCI 5 AD
MR field strength (T)	1.0	1.5	1.0	1.0
2D/3D	2D	2D	2D	3D
Sequence type	T1-weighted spin-echo	T1-weighted spin-echo	T1-weighted spin-echo	MPRAGE
Orientation	Oblique axial	Oblique axial	Oblique axial	Coronal
Voxel size (mm)	$1.01 \times 1.01 \times 5.0$	$0.98 \times 0.98 \times 5.0$	$0.98 \times 0.98 \times 3.0$	$1.0 \times 1.0 \times 1.0$
Inter-slice gap (mm)	0.5	0 (interleaved)	0 (interleaved)	N/A
# MRI time points	2 for 4 patients 3 for 6 patients	3	3	2
Mean (SD) interval between first and last MRIs (y)	4.3 (1.8)	3.7 (1.6)	4.3 (2.0)	1.9 (0.8)

AD, Alzheimer's disease; MCI, mild cognitive impairment; MR, magnetic resonance; MPRAGE, magnetization prepared rapid acquisition gradient echo.

Regarding Step 2, to convert the average ventricular edge displacement to the desired PVVC value, we also repeated SIENA's calibration step for the subset of edge points on the ventricular surface, but with a small difference compared with SIENA. The calibration step applies a known volumetric scaling to one of the scans to obtain the calibration factor necessary to convert average edge displacement to PBVC. This calibration factor is calculated for both the forward and backward analyses separately (from the baseline and follow-up scans, respectively). In whole-brain SIENA, the brain size generally only changes by a limited amount, giving similar numbers of brain-nonbrain edge points for the baseline and follow-up scans, and therefore, the forward and backward calibration factors can be averaged to obtain a more robust result (Smith et al., 2002). However, in contrast to the brain, the size of a patient's ventricles can differ greatly between time points, giving very different numbers of brain-nonbrain edge points. Therefore, we apply separate calibration factors for the forward and backward analyses within VIENA.

Step 3 involved a similar minor deviation from the SIENA processing flow. As for step 2, this difference also was introduced because of the much greater relative volume changes expected for the ventricles when compared with the brain. In standard whole-brain SIENA, the final PBVC value is calculated by simply averaging the PBVC results from the forward and backward analyses, essentially neglecting, in this particular step only, the fact that there is a small volume difference between the baseline and follow-up brain volumes. Because the PBVC is generally only on the order of a few percent, this introduces only a very small error. However, ventricular volume changes in diseases like MS and AD can easily amount to tens of percents volume increase or even more, especially early in the disease when the ventricular volume at baseline is very small. In this case, the percentage volume

increase measured from the forward analysis can be very different from the percentage volume decrease measured from the backward analysis. Simple averaging of forward and backward PVVC results will then produce incorrect answers. Therefore, to obtain the final PVVC value from the forward and backward results, we convert the backward PVVC value to a percentage change relative to the baseline volume before averaging, by applying Eq. (1):

$$PVVC_{final} = 100 \cdot \left(\frac{1}{2} \left(1 + 0.01 \cdot PVVC_{forward} + \frac{1}{1 + 0.01 \cdot PVVC_{backward}} \right) - 1 \right) \quad (1)$$

Note that all PVVC values in Equation 1 are taken as percentages.

All VIENA calculations were performed fully automatically by a set of Linux shell scripts as posthoc operations after running standard SIENA in debug mode (-d option). The only modification to the SIENA software itself, consisted of letting SIENA save a Nifti file to indicate which voxels in the image were identified as brain-nonbrain edge points. Although this information is used by SIENA in its calculations, it is not normally saved to a Nifti file, not even in debug mode. This information is required here to define which set of voxels the brain edge displacement should be averaged over.

SUMMARY OF THE MATERIALS AND EQUIPMENT USED

Data

MR image data previously acquired as part of academic research projects in MS and dementia were analyzed. Data were retrospectively selected from 15 elderly subjects with two imaging time points available: five cognitively normal

TABLE II. Mean values of manual and automated PVVC and PBVC

Patient group ^a	PVVC		PBVC
	Manual	Automated (VIENA)	Automated (SIENA)
	Mean ± SD (median)	Mean ± SD (median)	Mean ± SD (median)
Group 1 (MS 5 mm + gap)	11.7 ± 10.2 (12.1)	12.1 ± 13.8 (8.3)	−1.5 ± 1.0 (−1.4)
Group 2 (MS 5 mm)	16.6 ± 6.9 (16.0)	20.1 ± 17.1 (15.0)	−2.2 ± 0.6 (−2.0)
Group 3 (MS 3 mm)	18.5 ± 18.0 (15.9)	19.0 ± 21.7 (12.8)	−3.2 ± 2.3 (−2.6)
Group 4 (aging 3D)	25.5 ± 30.2 (15.9)	27.7 ± 29.4 (22.2)	−3.5 ± 3.3 (−2.7)

^aFor convenience, disease group, and slice thickness for the 2D sequences is indicated in brackets for each patient group. MS, multiple sclerosis; SD, standard deviation.

elderly controls, five patients with mild cognitive impairment (MCI) (Petersen et al., 1999), and five with probable AD (McKhann et al., 1984), and from 58 patients with MS (McDonald et al., 2001), with either two or three imaging time points. Details of MRI acquisition along with demographic data on these patients are given in Table I.

MANUAL Method for Comparison

To assess the accuracy of the ventricular volume change measured by the VIENA method, we performed manual volumetry of the ventricular system. The ventricular system was defined in this study to consist of both lateral ventricles, third ventricle and fourth ventricle, the foramina of Monro and the aqueduct of Sylvius. The choroid plexus was considered as part of the ventricular volume. The inferior cutoff was taken at the point of the cerebellum where the foramina of Lushkae become visible. The cisterns and the subarachnoid space were excluded. Following this definition, the ventricular system was outlined on the original Dicom images with in-house developed software (Show_Images) using seed growing, combined with additional editing to fill any gaps in the ventricle ROI caused by plexus, and to ensure that hypointense periventricular structures—such as periventricular hypointense lesions—were not included. Figure 2 shows an example of the manual outlining of the ventricular system. All manual measurements were performed by a single trained operator (E.K.V.). Importantly, images were presented for manual measurement in a pseudorandom order, and baseline and follow-up images were analyzed completely separately. To assess reproducibility of the manual measurements, for 25 patients, for two time points, manual measurements were performed twice. To quantify agreement here, as well as further on in this article, we used the concordance correlation coefficient (CCC) (Carrasco and Jover, 2003), which is a stringent measure of absolute agreement. In contrast to the more lenient intraclass correlation coefficient, it punishes systematic offsets or scaling factors between two measures. It is also relatively sensitive to outliers. We use CCC here to be conservative in our estimates of agreement and avoid unwarranted optimism. Reproducibility of the

cross-sectional volume measurements was very good: CCC was 0.9964 for baseline MRI, and 0.9982 for the follow-up MRI. Reproducibility of the manual volume change measurement (PVVC_{manual}), calculated from the two cross-sectional volumes, was also very good, with CCC = 0.92.

STEP-BY-STEP DESCRIPTION OF THE RESEARCH PROTOCOL

The accuracy of the automated method was investigated by assessing, for the longest available interval between scans, the absolute agreement of the automated (VIENA) result PVVC_{VIENA} with the result of the manual method PVVC_{manual}, as measured by CCC. In addition, we calculated the nonparametric chance-corrected agreement coefficient for ordinal data (Kendall's W) (Fagot, 1994) to provide a measure of relative agreement less sensitive to outliers or nonlinearity of relations. The overall and group mean PVVC values were compared between manual and automated methods using Wilcoxon-signed rank test.

To assess the precision of the measurement, we analyzed the subset of 54 MS patients with three imaging time points (Table I). Using VIENA, PVVC_{VIENA,1→3} was calculated directly from time points 1 and 3, and the cumulative

Figure 2.

Example of manual outlining of the ventricles on MR image of an MS patient (see text for details).

Figure 3.

Scatter plot for the comparison of VIENA with manual measurement of percentage ventricle volume change (PVVC). Absolute agreement of VIENA with manual measurement was good, with a concordance correlation coefficient (CCC) of 0.84. Different MR acquisition protocols and patient groups are indicated by different symbols. The solid line indicates perfect agreement.

$PVVC_{VIENA,1 \rightarrow 2 \rightarrow 3}$ was calculated by combining the result obtained from time points 1 and 2 ($PVVC_{VIENA,1 \rightarrow 2}$) with that obtained from time points 2 and 3 ($PVVC_{VIENA,2 \rightarrow 3}$). CCC was used to quantify absolute agreement between $PVVC_{VIENA,1 \rightarrow 3}$ and $PVVC_{VIENA,1 \rightarrow 2 \rightarrow 3}$. The median absolute difference between $PVVC_{VIENA,1 \rightarrow 3}$ and $PVVC_{VIENA,1 \rightarrow 2 \rightarrow 3}$ was also calculated. Finally, to verify correct performance of SIENA, we also calculated the median absolute difference for the whole-brain measure PBVC generated by standard SIENA from the same image data.

RESULTS

Descriptives

Average and median values of manual and automated PVVC and of PBVC are given in Table II and are in the expected range for these patients and follow-up durations.

Accuracy

There was good absolute agreement of the fully automated $PVVC_{VIENA}$ with the manually obtained result $PVVC_{manual}$, with $CCC = 0.84$. Relative agreement was very good with Kendall's $W = 0.88$. Wilcoxon-signed rank test revealed no mean PVVC difference between the two methods, neither analyzed for the entire group ($P = 0.53$) nor analyzed for each patient group separately (P values between 0.10 and 0.96). Figure 3 shows the Bland–Altman plot for the comparison of automated and manual methods, demonstrating that there was good agreement

and no systematic offset. Note that there were some outliers. Figure 4 aims to illustrate whether the inaccuracy of VIENA (compared with manual volumetry) depends on two potentially important factors, namely baseline ventricular volume, and the ventricular volume change between the two scans. It displays on the vertical axis the difference between $PVVC_{VIENA}$ and $PVVC_{manual}$, with on the horizontal axis the baseline ventricular volume in mL in Figure 4A, and the ventricular volume change in mL in Figure 4B. As these plots demonstrate, the accuracy of the automated method did not depend on the initial ventricular volume or the ventricular volume change.

Figure 4.

Potential influence of initial ventricular volume and true mL change in ventricular volume on the accuracy of VIENA. Panel (A) plots the difference between PVVC values obtained from VIENA and from manual measurements as a function of the initial ventricular volume (in mL) obtained from the manual volumetry. Panel (B) plots the difference between PVVC values obtained from VIENA and from manual measurements as a function of the actual change in ventricular volume (measured in mL) as obtained from the manual volumetry.

Figure 5.

Bland–Altman plot illustrating the precision of the VIENA automated method. To determine precision, an MR dataset from MS patients was used in which there were three MR imaging time points (1, 2, and 3) for each patient. The percentage ventricular volumetric change from time point 1 to 3 was measured both directly ($PVVC_{VIENA,1 \rightarrow 3}$), and as a composite ($PVVC_{VIENA,1 \rightarrow 2 \rightarrow 3}$) derived by combining the change measured between time points 1 and 2 ($PVVC_{VIENA,1 \rightarrow 2}$) with that measured between time points 2 and 3 ($PVVC_{VIENA,2 \rightarrow 3}$). The figure shows on the vertical axis the difference between the two measurements, and on the horizontal axis the average of the two measurements. Note the different scales of the two axes. More details are provided in the text.

Precision

There was very good absolute agreement between $PVVC_{VIENA,1 \rightarrow 3}$ (calculated in one step) and $PVVC_{VIENA,1 \rightarrow 2 \rightarrow 3}$ (calculated in two steps), with $CCC = 0.98$. Figure 5A shows the corresponding scatter plot. The median absolute difference between $PVVC_{VIENA,1 \rightarrow 3}$ and $PVVC_{VIENA,1 \rightarrow 2 \rightarrow 3}$ was 1.01%, i.e. 1.01 PVVC. For PBVC, the median absolute difference between $PBVC_{1 \rightarrow 3}$ and $PBVC_{1 \rightarrow 2 \rightarrow 3}$ was 0.18%, i.e. 0.18 PBVC. Figure 5B shows the scatter plot for PBVC.

DISCUSSION

The results in this article demonstrate the validity of the VIENA method for measuring ventricular volume change. Both the accuracy and precision were good, and were robust across image types. The method also performed well across large ranges of both baseline ventricular volume and ventricular volume change.

The VIENA method measures the amount of ventricular volume change in an individual patient over time, thus providing an important indirect measure of central brain atrophy. The results are very comparable with those of labor-intensive manual measurements, as evidenced by the

high CCC value of 0.84, which indicates that VIENA succeeds at determining a good estimate of the amount of ventricular volume change. The CCC is a more conservative measure than the more commonly used intraclass correlation coefficient, punishing both bias and scaling between the two measures being compared, as well as being negatively impacted by outliers. The fact that CCC between VIENA and manual measurements was high even in the presence of a few outliers confirms the good performance of the method. Although indicating that performance of the fully automated VIENA method is good, the CCC value of 0.84 is below common thresholds of at least 0.90 for fully interchangeable measures. Therefore, VIENA and manual measurements are not interchangeable, so results should not be mixed in a single analysis.

The CCC value of 0.98 between the 1-step and 2-step measurements indicates that the method is very precise. The median absolute difference was 1.01% ventricular volume change. Whether this measurement uncertainty is sufficiently small should be decided by investigators by comparison with the expected changes in their patient groups. As a reference, Knopman et al. (2009) report a mean percentage ventricle volume increase of 11.6% over 1 year in patients with frontotemporal lobar degeneration; Nestor et al. (2008) using ADNI data, reported a mean PVVC over 6 months of 3.4% in MCI and 5.7% in AD patients; and Lukas et al. (2010) reported a median PVVC rate of 4.28% per year in early RRMS patients. This fully automated method may therefore be useful for studies of large patient groups, e.g., to investigate the natural development of central atrophy in populations of elderly subjects, or to assess the value of the rate of ventricular volume change as a predictor of future decline in MS. The latter has recently been investigated in an early MS cohort, and that study clearly demonstrated the potential of the VIENA method for predicting subsequent clinical development (Lukas et al., 2010).

To investigate the relation between early atrophy and long-term clinical development, long-term post-MRI follow-up is required. Almost by necessity, this implies analyzing older MR images, which are suboptimal by current standards. It is therefore important to note that the present method works well on such conventional 2D MR images. Even more may be learned from long-term clinical follow-up following more detailed MR atrophy measurements using current state-of-the-art imaging and analysis methods, but this will require waiting several years after the MRI and is therefore currently not yet available. Until then, the ventricular expansion provides a sensible alternative, giving an indirect measure that can already reveal much about disease progression. The VIENA method provides an automated approach for measuring ventricular expansion, which as evidenced by our present results, has good precision and accuracy. Nevertheless, it should be noted that even better results could be obtained through manual volumetry. In many large studies, the cost associated with manual volumetry may be prohibitive, but this may not be true in all cases, and it is important to point

out that expert manual outlining remains the best approach. The current results show that, when manual volumetry is not an option, VIENA provides a good alternative.

The VIENA method was implemented here directly within the FSL SIENA framework, requiring only one modification to the existing SIENA method: the writing to an image file of the map of brain/nonbrain edge points identified by SIENA. The additional calculations were performed by a set of simple Linux shell scripts that use FSL tools to perform registration of the mask and perform calculations for calibration and measurement of PVVC, requiring very little additional calculation time. This implementation ensures that this method can be easily included in the SIENA/FSL framework for use by other investigators. Software will be made available to all interested academic parties.

The absolute median difference for SIENA of 0.18% observed here is comparable with the respective values of 0.15%, obtained in a previous comparison of 1-step and 2-step PBVC analyses for SIENA (Smith et al., 2002), and 0.16% measured from same-day rescans in which the actual atrophy rate should equal zero (Smith et al., 2007); although all three are smaller than the precision estimate of 0.35% obtained in a recent study on a large dataset using two scan-rescan image pairs per patient, from two time points with nonzero actual atrophy rates occurring in between (Cover et al., 2011). Our results indicated that SIENA performed well on our data, which is important since the performance of VIENA depends directly on SIENA. Outliers in our data thus cannot be attributed to poor overall performance of SIENA.

Regarding the accuracy of VIENA, with $CCC = 0.84$ and Kendall's $W = 0.88$, this is at least as good as the accuracy of SIENA. In a recent validation of SIENA against manual segmentations in 10 subjects (de Bresser et al., 2011), de Bresser et al. measured a value of 0.82 for Spearman's rank correlation coefficient. In the absence of ties Kendall's W reduces to Spearman's rank correlation coefficient, indicating that the accuracy of VIENA is comparable with that of SIENA. Nevertheless, the outliers present in our results signify that the accuracy of VIENA could still be improved. In spite of close inspection of intermediate results, we have been unable to identify a clear cause for these outliers. We investigated the presence of misregistration between the subject's scans, misregistration of the standard space mask to the subject's scans—potentially leading to accidental exclusion of parts of the ventricular surface and/or inclusion of peripheral brain-CSF boundaries, and tissue-type segmentation errors, but none of these were observed. Compared with a previous version of the software, the method as described here performed slightly better in part due to improved partial volume estimation as implemented in the FAST segmentation software version 4.1. This yielded slightly better detection of CSF in (relatively thick) 2D slices, which resulted in better accuracy and fewer outliers.

Figure 6.

Scatter plot illustrating the difference between PVVC as measured from forward analysis (baseline $\rightarrow$ follow-up) and PVVC as measured from backward analysis (follow-up $\rightarrow$ baseline), with appropriate conversion for the latter to allow comparison with the forward measurement. The figure plots this difference (on the vertical axis) as a function of PVVC as measured from the manual analyses. The three outliers on the bottom right of the figure suggest that the difference between forward and backward analyses is more likely to be large in the case of large relative volume changes, reflected by large manual PVVC values. In these cases, the large difference stems mainly from an overestimation of the true change by the backward analysis.

Despite MS lesions and age-related WM hyperintensities, both of which may in general hamper tissue-type segmentation, as demonstrated for MS lesions by Nakamura and Fisher (2009) and Gelineau-Morel et al. (2012), and thereby affect the method, VIENA performed well. Nevertheless, inclusion of lesion masks to re-label those voxels in the segmentation, or lesion-inpainting approaches (Chard et al., 2010; Battaglini et al., 2012) may further improve VIENA performance.

The choice to implement VIENA here as a simple extension to existing SIENA also implies that limitations of SIENA apply. Most importantly, the SIENA method favors small displacements of the brain edge, and this has been criticized by Gunter et al. (2003) who used a more sophisticated approach. In their method, the brain edge displacement measured for a given voxel is compared with those observed for its neighboring voxels, and the measurement is repeated several times to obtain an optimum solution. It is likely that in cases with a very large increase in ventricular volume between scans, the brain edge displacement may not be estimated well by the existing SIENA method, and therefore, the VIENA method discussed here may be outperformed in those cases by the approach by Gunter et al. However, the approach chosen in the present work has the advantage that it can be easily implemented within the existing SIENA framework without large modifications. Future work should investigate whether

implementing the method by Gunter et al. leads to the expected improvement in performance of VIENA and reduction of outliers. In trying to explain the outliers in our current study, the inherent punishing of large edge displacements by standard SIENA did not appear to be the cause of the erroneous results. In the major outlier cases, the calculated edge displacements and effective PVVC values were similar, though not identical, between “forward” and “backward” analyses. Figure 6 shows the difference between PVVC values derived from forward and backward analyses, with appropriate conversion for the latter to allow comparison with the forward measurement. The figure plots this difference (on the vertical axis) as a function of PVVC as measured from the manual analyses. The three outliers on the bottom right of the figure suggest that the difference between forward and backward analyses is more likely to be large in the case of large relative volume changes, reflected by large manual PVVC values. In these cases, the large difference stems mainly from an overestimation of the true change by the backward analysis, while the forward analysis gave closer values. Future work should investigate the possibility of improving the PVVC estimate for such cases, perhaps by determining robust criteria for rejecting PVVC as determined by the backward analysis if this deviates too much from that obtained through the forward analysis. In the absence of ground truth knowledge, such criteria may be difficult to formulate, but future work should investigate this.

A further limitation of the method is in the method of selecting the edge points. In the current implementation, the mask used to select the ventricular edge points in individual scans has a large spatial extent to ensure inclusion of ventricular edges also for patients with extremely large ventricles. This implies that, especially in patients with substantial widening of sulci, some peripheral edge points may also be included in the VIENA calculations. The effect of these additional (peripheral) edge points on the PVVC value is difficult to predict. As they cannot a priori be expected to have systematically larger or smaller displacements than the ventricular edge points, and given the variability observed in peripheral brain edge displacement values calculated by SIENA, it would seem reasonable to assume that their effect is mainly to add variability to the edge displacement data rather than introducing any systematic effect. Nevertheless, the method may be further improved by optimizing the ventricle mask for each individual scan, in order to remove peripheral brain edge points and restrict the analysis to the ventricular edge points alone. This may be achieved, e.g., by manual editing, or in an automated way by, e.g., performing additional tissue type segmentation within the registered mask, and/or including morphological operations, or even comparing a range of atlas ventricular outlines to the current scan and either choosing the best match or use some form of voting as has been done in segmentation methods, e.g., of the hippocampus. However, we have chosen in the present implementation to keep the method straightforward,

as well as fast and robust across image types. Any additional steps are likely to compromise especially the latter two points. The aim of the current method was to provide a fast and robust measurement of the overall volume change of the ventricular system. To construct a detailed voxelwise segmentation of the ventricles would require a different approach than a modification of SIENA. It would then seem more logical to us to invest in other techniques that can provide better anatomical precision than can be expected from VIENA, such as methods for deep GM volumetry (Fischl et al., 2002; Patenaude et al., 2011) or modeling methods for ventricular volumetry such as the recent work by Kempton et al. (2011), all of which require better anatomical definition than the type of images typically available in the large and retrospective studies we have designed VIENA for. Investigating these more detailed techniques is beyond the scope of the present work.

For the same reason, we do not include a cross-sectional segmentation of the ventricles for a single scan. The VIENA method presented here, like SIENA, is a solely longitudinal measure, providing the amount of change between scans but not the initial volume at baseline. In principle, it would be possible to obtain a cross-sectional segmentation of the ventricles similar to the cross-sectional extension of the SIENA method, SIENAX, but we have chosen not to do so because a systematic bias can be expected. The crude mask used here to select the ventricular edge points cannot similarly be used in the cross-sectional method SIENAX, because by contrast to the longitudinal method, it can be expected to introduce a systematic bias. The large mask will lead to a systematic overestimation of the ventricular volume, especially in patients with profound sulcal widening. Further adaptations such as those discussed above would be necessary to use this method to obtain a normalized ventricular volume from a single MRI.

In conclusion, this study shows that the fully automated VIENA method for measuring PVVC has good accuracy and precision across different image types. It can therefore be used to measure PVVC in a fully automated fashion with limited calculation time, and is hence an important candidate tool for investigating ventricular volume changes in large datasets. Further improvements may be made to increase its accuracy.

ACKNOWLEDGMENTS

The authors acknowledge Steve Smith of FMRIB, Oxford, UK, for helpful discussion, and Nick Fox of UCL, London, UK, for helpful discussion and for first suggesting the name VIENA.

REFERENCES

- Ashburner J (2007): A fast diffeomorphic image registration algorithm. *Neuroimage* 38:95–113.

- Ashburner J, Friston KJ (2000): Voxel-based morphometry—The methods. *Neuroimage* 11:805–821.
- Babalola KO, Patenaude B, Aljabar P, Schnabel J, Kennedy D, Crum W, Smith S, Cootes T, Jenkinson M, Rueckert D (2009): An evaluation of four automatic methods of segmenting the subcortical structures in the brain. *Neuroimage* 47:1435–1447.
- Battaglini M, Jenkinson M, De Stefano N (2012): Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Hum Brain Mapp* 33:2062–2071.
- Carrasco JL, Jover L (2003): Estimating the generalized concordance correlation coefficient through variance components. *Biometrics* 59:849–858.
- Chard DT, Jackson JS, Miller DH, Wheeler-Kingshott CA (2010): Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. *J Magn Reson Imaging* 32:223–228.
- Cover KS, van Schijndel RA, van Dijk BW, Redolfi A, Knol DL, Frisoni GB, Barkhof F, Vrenken H; neuGRID; Alzheimer's Disease Neuroimaging Initiative (2011): Assessing the reproducibility of the SienaX and Siena brain atrophy measures using the ADNI back-to-back MP-RAGE MRI scans. *Psychiatry Res* 193:182–190.
- Dalton CM, Miszkil KA, O'Connor PW, Plant GT, Rice GP, Miller DH (2006): Ventricular enlargement in MS: One-year change at various stages of disease. *Neurology* 66:693–698.
- De Bresser J, Portegies MP, Leemans A, Biessels GJ, Kappelle LJ, Viergever MA (2011): A comparison of MR based segmentation methods for measuring brain atrophy progression. *Neuroimage* 54:760–768.
- Fagot RF (1994): An ordinal coefficient of relational agreement for multiple judges. *Psychometrika* 59:241–251.
- Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002): Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* 33:341–355.
- Fjell AM, Walhovd KB, Fennema-Notestine C, McEvoy LK, Hagler DJ, Holland D, Brewer JB, Dale AM (2009): One-year brain atrophy evident in healthy aging. *J Neurosci* 29:15223–15231.
- Gelineau-Morel R, Tomassini V, Jenkinson M, Johansen-Berg H, Matthews PM, Palace J (2012): The effect of hypointense white matter lesions on automated gray matter segmentation in multiple sclerosis. *Hum Brain Mapp* 33:2802–2814.
- Gunter JL, Shiung MM, Manduca A, Jack CR Jr (2003): Methodological considerations for measuring rates of brain atrophy. *Magn Reson Imaging* 18:16–24.
- Jack CR Jr, Shiung MM, Weigand SD, O'Brien PC, Gunter JL, Boeve BF, Knopman DS, Smith GE, Ivnik RJ, Tangalos EG, Petersen RC (2005): Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnesic MCI. *Neurology* 65:1227–1231.
- Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012): FSL. *NeuroImage* 62:782–790.
- Knopman DS, Jack CR Jr, Kramer JH, Boeve BF, Caselli RJ, Graff-Radford NR, Mendez MF, Miller BL, Mercaldo ND (2009): Brain and ventricular volumetric changes in frontotemporal lobar degeneration over 1 year. *Neurology* 72:1843–1849.
- Lukas C, Minneboo A, de Groot V, Moraal B, Knol DL, Polman CH, Barkhof F, Vrenken H (2010): Early central atrophy rate predicts 5 year clinical outcome in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 81:1351–1356.
- Nestor SM, Rupsingh R, Borrie M, Smith M, Accomazzi V, Wells JL, Fogarty J, Bartha R; Alzheimer's Disease Neuroimaging Initiative (2008): Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. *Brain* 131:2443–2454.
- McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, McFarland HF, Paty DW, Polman CH, Reingold SC, Sandberg-Wollheim M, Sibley W, Thompson A, van den Noort S, Weinshenker BY, Wolinsky JS (2001): Recommended diagnostic criteria for multiple sclerosis: Guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* 50:121–127.
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984): Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34:939–944.
- Nakamura K, Fisher E (2009): Segmentation of brain magnetic resonance images for measurement of gray matter atrophy in multiple sclerosis patients. *Neuroimage* 44:769–776.
- Neacsu V, Jasperse B, Korteweg T, Knol DL, Valsasina P, Filippi M, Barkhof F, Rovaris M, Vrenken H; MAGNIMS study group (2008): Agreement between different input image types in brain atrophy measurement in multiple sclerosis using SIENAX and SIENA. *J Magn Reson Imaging* 28:559–565.
- Patenaude B, Smith SM, Kennedy DN, Jenkinson M (2011): A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* 56:907–922.
- Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999): Mild cognitive impairment: Clinical characterization and outcome. *Arch Neurol* 56:303–308.
- Smith SM, Zhang Y, Jenkinson M, Chen J, Matthews PM, Federico A, De Stefano N (2002): Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage* 17:479–489.
- Smith SM, Rao A, De Stefano N, Jenkinson M, Schott JM, Matthews PM, Fox NC (2007): Longitudinal and cross-sectional analysis of atrophy in Alzheimer's disease: Cross-validation of BSI, SIENA and SIENAX. *Neuroimage* 36:1200–1206.