NeuroImage 54 (2011) 760–768

Contents lists available at ScienceDirect

NeuroImage

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

A comparison of MR based segmentation methods for measuring brain atrophy progression

Jeroen de Bresser [a,b,*], Marileen P. Portegies [b], Alexander Leemans [a], Geert Jan Biessels [b], L. Jaap Kappelle [b], Max A. Viergever [a]

[a] Image Sciences Institute, University Medical Center Utrecht, The Netherlands
[b] Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 28 May 2010
Revised 23 August 2010
Accepted 24 September 2010
Available online 1 October 2010

Keywords:
Brain segmentation
Brain MRI
Brain atrophy
Image processing
Cross-sectional studies
Longitudinal studies

ABSTRACT

Automated brain segmentation methods with a good precision and accuracy are required to detect subtle changes in brain volumes over time in clinical applications. However, the ability of established methods such as SIENA, US and kNN to estimate brain volume change have not been compared on the same data, nor been evaluated with ground-truth manual segmentations. We compared measurements of brain volume change between SIENA, US and kNN in terms of precision (repeatability) and accuracy (ground-truth) using one baseline and two repeated follow-up 1.5 T MRI scans after 4 years of 10 subjects. The coefficient of repeatability (brain volume/volume change) was larger for US (29.6 cc/2.84%) than for kNN (4.9 cc/0.31%) and SIENA (−/0.92%). In terms of absolute brain volume measurements US and kNN showed good correlation with the manual segmentations and with each other (all Spearman's correlation coefficients $\rho \geq 0.96$; all $p<0.001$). Concerning brain volume changes, SIENA showed a good ($\rho = 0.82$; $p = 0.004$), kNN a moderate ($\rho = 0.60$; $p = 0.067$) and US a weak ($\rho = 0.50$; $p = 0.138$) correlation with the manual segmentations. For measurements of volume change, SIENA–US (mean correlation coefficient and p-value: $\rho = 0.28$; $p = 0.442$) and US–kNN ($\rho = 0.17$; $p = 0.641$) showed a weak correlation, but correlation was fairly good for kNN–SIENA ($\rho = 0.65$; $p = 0.048$). In conclusion, US and kNN showed a good precision, accuracy and comparability for brain volume measurements. For measurements of volume change, SIENA showed the best performance. kNN is a good alternative if volume change measurements of other brain structures are required.

Introduction

Brain volumes can be measured to assess atrophy due to normal ageing, or as a marker of disease progression in clinical studies of different pathologic conditions, like dementia (Fox et al., 1996; Karas et al., 2003), or in other diseases with a less marked loss of brain volume, such as multiple sclerosis and type 2 diabetes mellitus (de Bresser et al., 2010; Filippi et al., 2004; Jasperse et al., 2007b; Jongen et al., 2007).

There are a number of automated methods that can estimate brain atrophy from magnetic resonance images; they can be subdivided in methods measuring brain volume and in methods measuring brain volume change over time. Methods measuring brain volume require only a single scan of a subject and generally measure the absolute brain volume. By comparing two scans such methods can also assess brain volume change. By contrast, methods measuring changes in brain volume typically measure a percentage difference in brain volume between two or more scans of the same subject, through a direct comparison of these scans without the need of calculating the absolute brain volume.

Thus far, methods measuring brain volume showed a lower precision for assessment of volume change than methods specifically designed to measure volume change (Smith et al., 2007). High precision and accuracy are important to detect subtle differences in brain volume changes between patient groups. Studies comparing different methods are few in number (i.e. Lee and Prohovnik, 2008; Smith et al., 2007), and only few methods have been evaluated with the reference standard of manual segmentation to test accuracy in addition to precision (i.e. Anbeek et al., 2005).

Structural Image Evaluation, using Normalization, of Atrophy (SIENA), Unified Segmentation (US) and k-Nearest Neighbor-based probabilistic segmentation (kNN) are well-established and widely used methods to measure brain atrophy (Anbeek et al., 2005; Ashburner and Friston, 2005; Smith et al., 2002). Although precision of SIENA has been assessed by repeated imaging with patient repositioning, accuracy has not been evaluated with manual segmentations (Smith et al., 2001, 2002). In addition, accuracy of brain volume measurements by US was tested only on simulated data (Ashburner and Friston, 2005). Furthermore, kNN has been compared

* Corresponding author. University Medical Center Utrecht, Department of Neurology G03.228, P.O. Box 85500, Heidelberglaan 100, 3508 GA Utrecht, The Netherlands. Fax: +31 30 2513399.
E-mail address: J.deBresser@umcutrecht.nl (J. de Bresser).

1053-8119/$ – see front matter
doi:10.1016/j.neuroimage.2010.09.060

with manual segmentations, but only evaluating accuracy for absolute brain volume measurements, not for brain volume change assessment (Anbeek et al., 2005). Finally, no studies have compared these methods on the same data.

In this study, we compared measurements of brain volume change between SIENA, US and kNN, and assessed potential differences in precision (via repeatability) and accuracy (compared with ground-truth manual segmentations). The methods were applied as described in the respective validation studies in order not to favor one method (Anbeek et al., 2005; Ashburner and Friston, 2005; Smith et al., 2002).

Materials and methods

Data

Ten subjects were included in this study (mean age ± SD = 70 years ± 6). Subjects were part of the Utrecht Diabetic Encephalopathy Study (de Bresser et al., 2010; van den Berg et al., 2010). This study was approved by the local ethics committee of the University Medical Center Utrecht and all participants signed an informed consent form. Baseline (BL) magnetic resonance imaging (MRI) scans and follow-up scans 4 years later were acquired on the same 1.5 T Philips MR system. At follow-up, scanning was performed twice (FU1/FU2) with patient repositioning in between. All three scans were made by a standardized scanning protocol which consisted of an axial T1 (TR/TE = 234/2 ms), T2 (TR/TE = 2200/100 ms), proton density (PD) (TR/TE = 2200/11 ms), inversion recovery (IR) (TR/TE/TI = 2919/22/410 ms) and fluid attenuated inversion recovery (FLAIR) (TR/TE/TI = 6000/100/2000 ms); all scans with 38 contiguous slices and 0.90 × 0.90 × 4.00 mm voxels.

SIENA

SIENA is a fully automated method that calculates brain volume change between two scans (Smith et al., 2001, 2002). SIENA as part of the FMRIB Software Library (FSL) 4.0 was used on the T1 images to calculate differences between BL and FU1, BL and FU2, and FU1 and FU2 (Smith et al., 2004). Brain masks were created with the Brain Extraction Tool (BET) (Smith, 2002), the resulting masked brain images were aligned to each other (Jenkinson et al., 2002; Jenkinson and Smith, 2001) and both data sets were resampled into the space halfway between the two. Brain/non-brain edge points were detected by tissue type segmentation (Zhang et al., 2001), and used to calculate perpendicular edge displacement between the two scans. The resulting mean edge displacement was converted into a global estimate of percentage brain volume change between the scans. BET parameters and other parameters were optimized for our data by qualitatively evaluating output. Resulting BET masks were of good quality and no problems due to non-uniformities were detected. All resulting output images were visually inspected and all output was considered to be of good quality.

US

The US algorithm combines tissue classification, bias correction, and image registration in the same generative model and is able to calculate absolute volumes of gray and white matter and cerebrospinal fluid (CSF) (Ashburner and Friston, 2000, 2005). US as part of Statistical Parametric Mapping (SPM) 8 was performed using the T1 images from the BL, FU1 and FU2 scans. The gray matter volume was added to the white matter volume for all scans to calculate total brain volume. From these absolute volumes, brain volume changes were also calculated. Parameters were optimized for our data by qualitative evaluation of the output. All resulting segmentations of brain volumes were visually inspected and considered to be of good quality.

kNN

kNN is a fully automated method that can calculate absolute brain volume (Anbeek et al., 2005). This method is based on manually segmented training data, which consisted of IR and FLAIR images of 10 subjects from another cohort without known neurodegenerative disease, comparable in age to the subjects in our study and made by an identical scanning protocol to the images in our study (Anbeek et al., 2005; de Bresser et al., 2010). A percentage of 40% training samples was first randomly selected from the training data, and then fixed for use in this study. For one scan, the T1, T2, PD and IR were rigidly registered to the FLAIR image with the Elastix tool (Klein et al., 2010). Scan inhomogeneities were corrected by a shading correction algorithm (Likar et al., 2001). A BL brain mask was created by using a k-means clustering algorithm with eight clusters (Jongen et al., 2007). FU masks were created by rigidly registering the BL to the FU FLAIR of the same subject and by using the resulting transform parameters to transform the BL brain mask (de Bresser et al., 2010). The uncorrected FLAIR images were multiplied voxelwise by the binary brain mask, followed by a shading correction (Likar et al., 2001). IR and FLAIR images were used for measurements of brain volume, by applying the k-nearest neighbor classification technique that builds a feature space from spatial information and voxel intensities of manually segmented training data (Anbeek et al., 2005). For each voxel to be classified, it determines the k (= 100) nearest neighbors from the training data and calculates a probability that a voxel is of a certain tissue type. Parameters were optimized for our data by qualitative evaluation of the output. All resulting segmentations of brain volumes were visually inspected and considered to be of good quality. Brain volumes for each subject were calculated by multiplying the probabilities by the voxel volume, and volume changes were calculated by taking the difference between normalized BL and FU brain volumes. Depending on the types of tissues classified in the training data other volumes can also be determined separately, including gray matter, white matter, CSF, lateral ventricular, and white matter hyperintensity volume.

Manual segmentations

In the implementation of the methods described in the respective validation studies, SIENA and US use T1 images and kNN mainly uses IR images to determine the brain CSF border (Anbeek et al., 2005; Ashburner and Friston, 2005; Smith et al., 2002). Therefore, the intra-rater reliability of both sequences for manual segmentation of brain volume was first determined voxelwise and as a mean difference, to select the optimal sequence for manual segmentation. Brain volumes were manually segmented twice on IR images and twice on T1 images by a trained researcher (MP) blinded for subject number, on ten randomly selected half slices from the data in this study. The intra-rater reliability of these manual segmentations for the T1 and IR images showed a sensitivity of 0.98 and 0.97, specificity of 0.99 and 0.99, similarity index of 0.96 and 0.96 and a mean difference ± SD of 0.53 cc ± 0.53 and 0.19 cc ± 0.69, respectively (Dice, 1945). Because of the slightly superior intra-rater reliability, the IR images were used for further manual segmentations. Brain volume was manually segmented by two trained researchers (JB, MP) on six additional randomly selected slices, which showed a good inter-rater reliability with a sensitivity of 0.95, specificity of 0.99 and similarity index of 0.96 (Dice, 1945).

Finally, total brain volumes of all BL and FU1 scans were manually segmented by the same trained researcher (MP). Intra-cranial nerves, such as the optic nerve, and large intra-cranial vessels were not included in the brain segmentations. All segmentations were inspected by a neurologist experienced in neuroimaging (GJB) and segmentations were adjusted accordingly. Brain volume change was calculated by taking the difference of the BL and FU1 measurements.