Psychiatry Research: Neuroimaging 233 (2015) 299–305

Contents lists available at ScienceDirect

Psychiatry Research: Neuroimaging

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

Comparison of automated brain segmentation using a brain phantom and patients with early Alzheimer's dementia or mild cognitive impairment

Iven Fellhauer [a,*], Frank G. Zöllner [b], Johannes Schröder [a], Christina Degen [a], Li Kong [a], Marco Essig [c], Philipp A. Thomann [d], Lothar R. Schad [b]

[a] Section of Geriatric Psychiatry and Institute of Gerontology, Department of Psychiatry, Heidelberg University, Germany
[b] Computer Assisted Clinical Medicine, Medical Faculty Mannheim, Heidelberg University, Germany
[c] German Cancer Research Center, Heidelberg, Germany
[d] Department of Psychiatry, Heidelberg University, Germany

ARTICLE INFO

Article history:
Received 15 October 2014
Received in revised form
15 June 2015
Accepted 7 July 2015
Available online 17 July 2015

Keywords:
AD
Brainweb
FreeSurfer
FSL
MCI
MRI
SPM

ABSTRACT

Magnetic resonance imaging (MRI) and brain volumetry allow for the quantification of changes in brain volume using automatic algorithms which are widely used in both, clinical and scientific studies. However, studies comparing the reliability of these programmes are scarce and mainly involved MRI derived from younger healthy controls. This study evaluates the reliability of frequently used segmentation programmes (SPM, FreeSurfer, FSL) using a realistic digital brain phantom and MRI brain acquisitions from patients with manifest Alzheimer's disease (AD, $n=34$), mild cognitive impairment (MCI, $n=60$), and healthy subjects ($n=32$) matched for age and sex. Analysis of the brain phantom dataset demonstrated that SPM, FSL and FreeSurfer underestimate grey matter and overestimate white matter volumes with increasing noise. FreeSurfer calculated overall smaller brain volumes with increasing noise. Image inhomogeneity had only minor, non- significant effects on the results obtained with SPM and FreeSurfer 5.1, but had effects on the FSL results (increased white matter volumes with decreased grey matter volumes). The analysis of the patient data yielded decreasing volumes of grey and white matter with progression of brain atrophy independent of the method used. FreeSurfer calculated the largest grey matter and the smallest white matter volumes. FSL calculated the smallest grey matter volumes; SPM the largest white matter volumes.

Best results are obtained with good image quality. With poor image quality, especially noise, SPM provides the best segmentation results. An optimised template for segmentation had no significant effect on segmentation results. While our findings underline the applicability of the programmes investigated, SPM may be the programme of choice when MRIs with limited image quality or brain images of elderly should be analysed.

Contents

1. Introduction . . . 300
2. Material and methods . . . 300
   2.1. Synthetic data . . . 300
   2.2. Patient data . . . 300
   2.3. Image processing . . . 301
      2.3.1. FreeSurfer . . . 301

* Corresponding author.
E-mail address: iven.fellhauer@med.uni-heidelberg.de (I. Fellhauer).

http://dx.doi.org/10.1016/j.pscychresns.2015.07.011
0925-4927/

2.3.2. FSL . . . 301
2.3.3. Statistical Parametric Mapping . . . 301
2.3.4. Statistical analysis . . . 301
3. Results . . . 301
4. Discussion . . . 303
References . . . 304

1. Introduction

High-resolution structural magnetic resonance imaging enables the investigation of changes that develop in severe neuropsychiatric diseases such as Alzheimer's disease (AD) including mild cognitive impairment (MCI) as its preclinical state, or schizophrenia. These structural changes, which go beyond those typically observed in normal aging, are visible on MRI scans (Pantel and Schröder, 2006; Schröder and Pantel, Submitted for publication) and can be quantified using appropriate segmentation programmes, which segment the brain into grey and white matter.

After the advent of the first manual segmentation programmes such as NMRWin (Friedlinger et al., 1995; Pantel et al., 1997) automatic and semi-automatic methods were introduced. However, even today, despite the high expenditure of time, manual segmentation is still the gold standard. In manual segmentation, the demarcation of grey and white matter is performed along the anatomical boundaries of the regions of interest on each layer. Automated methods assume nowadays this time-consuming method of tissue classification (Shen et al., 2010). The three most commonly used programme packages for the segmentation of brain tissue are FreeSurfer (Martinos Center for Biomedical Imaging[1]), FSL (FMRIB Software Library[2]) and SPM (Wellcome Trust Centre for Neuroimaging[3]).

For technical reasons, interferences like noise or inhomogeneity in the image data frequently occur during the MRI acquisition (Metcalf et al., 2010). These interferences may affect the automatic segmentation methods and thus, alter the results of the volume measurement. Therefore, the segmentation algorithms have to be robust to provide reliable results. However, studies comparing the reliability of these programmes are scarce and mainly involved MRI derived from younger healthy controls (Klauschen et al., 2009). This study will evaluate the impact of variable image quality on the segmentation results and the robustness of the segmentation algorithms in patients with MCI or beginning AD.

To evaluate the segmentation results on variable image quality, the currently most commonly used and freely available automatic segmentation toolboxes FSL, FreeSurfer and SPM were used. Furthermore, it is investigated whether these segmentation algorithms let to reliable results in a study of clinical MRI datasets which were obtained in patients with MCI or AD.

Manual segmentation remains the gold standard for quantifying atrophic brain changes, although it is a time consuming procedure, which places high demands on the skills and anatomic knowledge of the investigator (Pantel and Schröder, 2007; Schröder and Pantel, Submitted for publication). Especially the exact evaluation of initial, discrete pathological changes, important for early diagnosis of MCI and AD is rendered difficult. Accordingly, the algorithms used in the evaluation process must be sufficiently sensitive to allow a quantification of the more discrete changes associated with diagnoses of MCI and early AD. The automated segmentation requires that the contrast of MRI scans is sufficiently large for algorithms to be able to detect grey matter, white matter, and cerebrospinal fluid for correct brain segmentation, including partial volume effects.

[1] http://surfer.nmr.mgh.harvard.edu
[2] http://fmrib.ox.ac.uk/fsl
[3] http://fil.ion.ucl.ac.uk/spm/

2. Material and methods

2.1. Synthetic data

A realistic digital brain phantom was obtained from the Brainweb Simulated Brain Database[4] (Cocosco et al., 1997) to evaluate the robustness of automated brain segmentation methods on images with different image quality. The Brainweb phantom based on an anatomical model of a normal brain that was created by registering and preprocessing 27 MRI scans of the same individual (Collins et al., 1998). The Simulated Brain Database offers the possibility to simulate noise (n) and intensity nonuniformity (rf). The noise can be varied from 0% to 9% and the intensity nonuniformity can be varied from 0% to 40%. To investigate the robustness of the automated segmentation methods image data sets with variable image quality were generated: n0rf0, n0rf20, n0rf40, n1rf0, n1rf20, n1rf40, n3rf0, n3rf20, n3rf40, n5rf0, n5rf20, n5rf40, n7rf0, n7rf20, n7rf40, n9rf0, n9rf20, n9rf40. For the different tissue types (grey matter, white matter, CSF) of this data, reference values are available in the Brainweb database and serve as benchmark for the investigated automated segmentation methods.

2.2. Patient data

MRI datasets from 115 subjects – patients and healthy controls-were obtained from the memory clinic of the Section of Geriatric Psychiatry at the University Hospital Heidelberg. In addition, MRI datasets from 11 additional healthy control subjects were drawn from the Interdisciplinary Longitudinal Study on Adult Development and Aging (Pantel et al., 2003). In total, MRI datasets from 60MCI patients, 34 AD patients and 32 healthy controls were included in this study. All three groups showed only minor, nonsignificant differences with respect to age (F (2,123)=0.774, p=0.463), gender (χ^2 (2)=1.636, p=0.441), education (χ^2 (2)=1.636, p=0.441), marital status (χ^2 (6)=3.618, p=0.728) and employment status (χ^2 (4)=6.516, p=0.164). Severity of cognitive deficits and dementia was assessed on the Mini-Mental-State-Examination-MMSE-(Folstein et al., 1975), the Global Deterioration Scale-GDS-(Reisberg et al., 1982) and the Clinical Dementia Rating-CDR-(Morris, 1993), respectively. MRI was performed on a 1.5 T scanner (Siemens Magnetom Avanto, Siemens Healthcare Sector, Erlangen, Germany) using a 3D MPRAGE sequence with parameters TR/TE=10.0/4.0 ms, 128 sagittal slices, slice thickness, 1.0 mm, matrix=256 × 256, voxel size 1.02 × 1.02 × 100 mm³, FOV=260 mm (Dos Santos et al., 2011).

[4] http://brainweb.bic.mni.mcgill.ca/brainweb/