Neurobiology of Aging 34 (2013) 2003–2013

Contents lists available at SciVerse ScienceDirect

Neurobiology of Aging

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

Measurements of medial temporal lobe atrophy for prediction of Alzheimer's disease in subjects with mild cognitive impairment

Lies Clerx [a,*], Ineke A. van Rossum [b], Leah Burns [c], Dirk L. Knol [d], Philip Scheltens [b], Frans Verhey [a], Pauline Aalten [a], Pablo Lapuerta [e], Laura van de Pol [b], Ronald van Schijndel [b], Remko de Jong [b], Frederik Barkhof [b], Robin Wolz [f], Daniel Rueckert [f], Martina Bocchetta [g], Magdalini Tsolaki [h], Flavio Nobili [i], Lars-Olaf Wahlund [j], Lennart Minthon [k], Lutz Frölich [l], Harald Hampel [m], Hilkka Soininen [n], Pieter Jelle Visser [a,b]

[a] Department of Psychiatry and Neuropsychology, Maastricht University, School for Mental Health and Neuroscience, Alzheimer Center Limburg, Maastricht, the Netherlands
[b] Department of Neurology, Alzheimer Center, VU University Medical Center, Amsterdam, the Netherlands
[c] Global Health Economics & Outcomes Research, Bristol-Myers Squibb Company, Princeton, NJ, USA
[d] Department of Clinical Epidemiology and Statistics, VU University Medical Center, Amsterdam, the Netherlands
[e] Lexicon Pharmaceuticals, The Woodlands, TX, USA
[f] Department of Computing, Imperial College London, London, UK
[g] IRCCS San Giovanni, Laboratory of Epidemiology and Neuroimaging, Brescia, Italy
[h] Aristotle University of Thessaloniki, Memory and Dementia Centre, 3rd Department of Neurology, G Papanicolaore General Hospital, Thessaloniki, Greece
[i] Clinical Neurophysiology Service, Department of Neurosciences, Ophthalmology and Genetics, University of Genoa, Genoa, Italy
[j] NVS Department, Karolinska Institutet, Karolinska University Hospital, Huddinge, Sweden
[k] Clinical Memory Research Unit, Department of Clinical Sciences, Malmö Lund University, Malmö, Sweden
[l] Department of Geriatric Psychiatry, Zentralinstitut für Seelische Gesundheit, University of Heidelberg, Mannheim, Germany
[m] Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, Goethe-University, Frankfurt, Germany
[n] Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland

ARTICLE INFO

Article history:
Received 24 January 2013
Accepted 9 February 2013
Available online 27 March 2013

Keywords:
Alzheimer's disease
Mild cognitive impairment
Dementia
Magnetic resonance imaging
Medial temporal lobe atrophy
Cerebrospinal fluid (CSF)
Aβ1–42
Tau

ABSTRACT

Our aim was to compare the predictive accuracy of 4 different medial temporal lobe measurements for Alzheimer's disease (AD) in subjects with mild cognitive impairment (MCI). Manual hippocampal measurement, automated atlas-based hippocampal measurement, a visual rating scale (MTA-score), and lateral ventricle measurement were compared. Predictive accuracy for AD 2 years after baseline was assessed by receiver operating characteristics analyses with area under the curve as outcome. Annual cognitive decline was assessed by slope analyses up to 5 years after baseline. Correlations with biomarkers in cerebrospinal fluid (CSF) were investigated. Subjects with MCI were selected from the Development of Screening Guidelines and Clinical Criteria for Predementia AD (DESCRIPA) multicenter study ($n = 156$) and the single-center VU medical center ($n = 172$). At follow-up, area under the curve was highest for automated atlas-based hippocampal measurement (0.71) and manual hippocampal measurement (0.71), and lower for MTA-score (0.65) and lateral ventricle (0.60). Slope analysis yielded similar results. Hippocampal measurements correlated with CSF total tau and phosphorylated tau, not with beta-amyloid 1–42. MTA-score and lateral ventricle volume correlated with CSF beta-amyloid 1–42. We can conclude that volumetric hippocampal measurements are the best predictors of AD conversion in subjects with MCI.

1. Introduction

Alzheimer's disease (AD) is the leading cause of dementia in the elderly, affecting more than 27 million people worldwide. Early detection of AD might prevent irreversible damage by enabling preventative treatment (Masters and Beyreuther, 2006; Vellas et al., 2007). A primary focus of research in AD is identifying which biomarkers are clinically useful for the early diagnosis of AD.

Medial temporal lobe (MTL) atrophy as assessed using structural magnetic resonance imaging (MRI) has proven to be an effective clinical aid in the early diagnosis of AD (Visser et al., 2002a), and this method predicts AD in subjects with mild cognitive impairment (MCI) (DeCarli et al., 2007; Rusinek et al., 2004; Schoonenboom

* Corresponding author at: Department of Psychiatry and Neuropsychology, Maastricht University Medical Center, PO Box 616, 6200 MD Maastricht, the Netherlands. Tel.: +31 43 3881841; fax: +31 43 3884092.
E-mail address: L.Clerx@maastrichtuniversity.nl (L. Clerx).

0197-4580/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neurobiolaging.2013.02.002

et al., 2008; Visser et al., 1999, 2002a). There are several ways to determine the degree of MTL atrophy in the brain including manual delineation, (semi-) automated techniques to measure hippocampal volume, qualitative ratings of MTL atrophy (MTA-score), and assessment of lateral ventricular volume (Aljabar et al., 2009; Apostolova et al., 2006; Chou et al., 2010; Echávarri et al., 2011; Jack et al., 2004; McHugh et al., 2007; Nestor et al., 2008; Scheltens et al., 1992; Varela-Nallar et al., 2010; Wolz et al., 2010a, 2010b). Each of these methods has its strengths and limitations.

Manual volumetry is considered the gold standard (Barnes et al., 2009; Boccardi et al., 2011; van de Pol et al., 2007b) but is time-consuming, which limits routine clinical or large-scale research use. Automated measurements are quick and widely applicable, but might be susceptible to scanner and scan protocol variability. Volumetric measurements of the lateral ventricle require a minimum of rater time with robust automatic segmentations but show a lot of variability and asymmetry between subjects (Nestor et al., 2008). Qualitative ratings are quick to perform but sensitive to interrater variability and show lower accuracy rates compared with volumetric analysis (DeCarli et al., 1990; Galton et al., 2001). Visual rating scales are furthermore insensitive to detect atrophy progression over time (Ridha et al., 2007).

A number of studies have examined differences between various techniques to measure atrophy of the MTL, mostly comparing manual with automated hippocampal volumetry (Lehmann et al., 2010; Sanchez-Benavides et al., 2010b; Shen et al., 2010) or volumetric hippocampal measurements to a visual rating scale (Ridha et al., 2007; Scheltens et al., 1992; Urs et al., 2009; Wahlund et al., 2000; Westman et al., 2011). These studies typically evaluate the diagnostic accuracy of different MRI techniques by comparing AD patients with healthy control subjects. Most studies found that manual hippocampal measurement and automated hippocampal segmentation results were similar (Hsu et al., 2002; Lehmann et al., 2010). However, the performance of automated techniques might be less precise when applied in AD patients suffering from moderate to severe brain atrophy and/or white matter hyperintensities which might lead to false allocations of gray matter, white matter, or cerebrospinal fluid (CSF) (Carmichael et al., 2005; Levy-Cooperman et al., 2008; Sanchez-Benavides et al., 2010a). One study reported that visual rating of MTL atrophy is a quick and clinically useful technique for differentiating AD from control subjects and is quicker and more accurate than volumetry (Wahlund et al., 2000).

To our knowledge, no study has compared the diagnostic performance of manual and atlas-based hippocampal segmentation, lateral ventricle volume, and a qualitative rating. Moreover, no comparative studies have been performed on the predictive accuracy of these different methods to predict AD in subjects with MCI, their relation with CSF biomarkers of AD, and the effect of multicenter settings on diagnostic performance.

The aim of the present longitudinal study was to compare the predictive accuracy of 4 different MTL measurements for the progression to AD-type dementia in patients with MCI over a 2-year follow-up period. Atrophy of the MTL was assessed using manual measurement of the hippocampus, automatically measured hippocampal volume based on atlas registration (learning embeddings for atlas propagation; LEAP), volumetric measurement of the expansion of the lateral ventricle, and a largely used qualitative rating scale. Because subjects might convert at a later point in time, slope analyses were additionally performed with annual cognitive decline up to 5 years as an outcome measure. The correlation of MTL measures with AD biomarkers in CSF was also investigated and the predictive accuracy was tested in a multicenter study with different scan protocols and in a single-center study.

2. Materials and methods

2.1. Subjects

We selected participants with MCI from the Development of Screening Guidelines and Clinical Criteria for Predementia AD (DESCRIPA) study and the Alzheimer Center of the VU University Medical center (VUmc) in Amsterdam. DESCRIPA is a multicenter prospective cohort study from the European Alzheimer's Disease Consortium aimed at developing clinical criteria and screening guidelines for predementia AD (Visser et al., 2008). For this study participants were selected from 9 of the 20 participating centers where MRI scanning was performed as part of clinical practice or as research protocol. The VUmc cohort in the present study included the VUmc subjects enrolled in the DESCRIPA study and an additional sample of subjects that were seen outside the DESCRIPA inclusion period (Supplementary Appendix 1).

Inclusion criteria for both cohorts were: age 54 years or older, diagnosis of MCI, and availability of results for each MRI measure and outcome at follow-up. Exclusion criteria were diagnosis of dementia at baseline or any somatic, psychiatric, or neurological disorder (e.g., epilepsy) that might have caused the cognitive impairment (Visser et al., 2008). At baseline, scans were available for 456 subjects. Visually rated MTL atrophy was available for all subjects. Of these, 54 had no follow-up data and were excluded. Of the remaining 402 subjects, scans were not available in digital format for 21 subjects. Of the remaining 381 scans, manual segmentation of the hippocampus could be performed on 341 scans (reasons missing: technical problem in volumetric measurement [$n = 5$], technical problem in automated intracranial volume estimation [$n = 25$], and logistical [$n = 10$]), LEAP-based volumetry on 357 scans (reasons missing: technical problem in volumetric measurement [$n = 11$] and logistical [$n = 13$]), and lateral ventricle volumetry on 335 scans (reasons missing: technical problem in volumetric measurement [$n = 37$], and logistical [$n = 9$]). Data for all 4 medial temporal lobe measurements were available for 328 subjects; 156 from DESCRIPA and 172 from VUmc. There were no differences between included and excluded subjects with respect to age, sex, educational level, and cognitive test results.

2.2. Clinical and cognitive assessment

All participants underwent a standard diagnostic workup, including clinical history, medical and neurological examination, clinical chemistry, functional evaluation using the Clinical Dementia Rating scale (Morris, 1993), the Mini-Mental State Examination (MMSE), and rating scales for depression and neuropsychiatric symptoms. A neuropsychological battery was performed to evaluate performance in several cognitive domains. In each center a primary test for verbal memory, language, attention, executive functioning, and visuoconstruction was chosen that was identical or similar to tests used in other centers (Visser et al., 2008). Raw scores on neuropsychological tests were corrected for age, education, and sex, in accordance with locally collected or published normative data and expressed as z-scores, which were used for further analysis. Baseline diagnosis of MCI was made according to the criteria of Petersen and colleagues (Petersen, 2004; Petersen et al., 1999). Subjects with a z-score $<$-1.5 SD on any of the following tests: the learning measure or delayed recall of a word list learning test or equivalent memory test, the Trail Making Test part A, Trail Making Test part B, verbal fluency, Rey figure copy test or an equivalent test were defined as having MCI (reference Vos). (Vos et al., 2012). We calculated a composite score as the average z-score of the 6 tests if scores were available for at least 3 tests (Visser et al., 2009).

Follow-up was conducted annually for up to 5 years. The primary outcome measure was conversion to AD-type dementia