CHAPTER 2

Clinical validity of medial temporal atrophy as a biomarker for Alzheimer’s disease in the context of a structured 5-phase development framework


Neurobiology of Aging 2017; 52:167-182
Chapter 2

ABSTRACT

Research criteria for Alzheimer’s disease recommend the use of biomarkers for diagnosis, but whether biomarkers improve the diagnosis in clinical routine has not been systematically assessed. Aim is to evaluate the evidence for use of medial temporal lobe atrophy (MTA) as biomarker for Alzheimer’s disease at the MCI stage in routine clinical practice, with an adapted version of the 5-phase oncology framework for biomarker development. A literature review on visual assessment of MTA and hippocampal volumetry was conducted with other biomarkers addressed in parallel reviews. Ample evidence is available for Phase 1 (rationale for use) and Phase 2 (discriminative ability between diseased and control subjects). Phase 3 (early detection ability) is partly achieved: most evidence is derived from research cohorts or clinical populations with short follow-up but validation in clinical MCI cohorts is required. In phase 4, only the practical feasibility has been addressed for visual rating of MTA. The rest of phase 4 and phase 5 have not yet been addressed.
INTRODUCTION

Biomarker assessment at the time of clinical dementia is important for differential diagnosis and thereby also for prognosis and potential treatment. New clinical criteria have introduced the use of biomarkers for the diagnosis of Alzheimer’s disease (AD) [1–4]. These criteria also recommend the use of biomarkers for the diagnosis of AD in non-demented subjects with mild cognitive impairment (MCI) [5]. Early diagnosis offers an opportunity for early intervention, improved guidance for caregivers and more accurate prognosis.

Several biomarkers for AD have been developed; however, there is insufficient systematically addressed evidence to implement them for a diagnosis of AD at the MCI stage in routine clinical practice [6]. To overcome a similar problem in the field of oncology, Pepe and colleagues [7] suggested to systematize the investigation of cancer biomarkers based on a framework borrowed from drug development. A similar approach may boost the adoption of AD biomarkers in clinical practice. An effort has recently been launched to adapt the oncology framework to suit the current goal of diagnosis of AD at the MCI stage, as described in the accompanying summarizing paper [8]. The present study fulfils a specific part of this wider plan: the analysis of the available evidence for medial temporal lobe atrophy (MTA) on MRI in the context of this framework. The other studies from this effort assessed within the common framework the following biomarkers: episodic memory assessment [9], cerebrospinal fluid measures [10], amyloid PET [11], [18F]FDG-PET [12], and [123I]-Ioflupane and [123I]-MIBG imaging [13].

In a research setting, there are three methods commonly used to assess atrophy of the medial temporal lobe: visual rating, manual volumetry and automated volumetry. In visual rating, atrophy is assessed on an ordinal scale. The most widely adopted visual rating scale in research, used in more than 100 publications, is the five-point Scheltens scale which was developed over twenty years ago [14]. MTA is visually assessed on coronal images taking into account the width of the choroid fissure, width of temporal horn and height of hippocampus. Although visual rating scales can be broadly applied to a range of imaging acquisition methods and performed by a trained radiologist, it is still only sparsely adopted in clinical practice [15]. In manual volumetry, hippocampal atrophy is quantified by drawing multiple regions of interest by an experienced rater on adjacent coronal MRI slices, typically 1-3 mm thick. This requires trained raters and is time-consuming. Several automated methods, which estimate volumes of structures by means of a computer algorithm, have been developed. Most of these algorithms involve an automated segmentation and classification of hippocampal tissue. Volume is then calculated as the sum of all voxels classified as hippocampal tissue. These techniques require specialist software and expertise,
are also time-consuming and results can vary across scanners and acquisition protocols. Manual and automated volumetry are currently only used in research settings and not yet implemented for clinical use.

In this paper, we will review evidence for the maturity of visual rating of MTA (vMTA) and hippocampal volume (HCV) as biomarkers for AD at the MCI stage, where evidence is interpreted under the light of an adapted version of the oncology framework [7,8].

**METHODS**

This review was performed with reference to the oncology framework, which was adapted to the field of dementia, specifically to the aim of performing the diagnosis of sporadic AD at the MCI stage [8]. The lexicon of this framework is extensively described in Frisoni et al. [8] and is summarized in this section. Only sporadic, not familial AD, is considered. The standard reference for diagnosis was AD neuropathology or the development of incident Alzheimer’s dementia at follow-up.

**Glossary**

*Alzheimer’s disease* (AD) is defined as the presence of Alzheimer’s pathology consisting of cerebral amyloid plaques and tangles, supposedly leading to a pattern-specific neurodegeneration (medio-temporal and temporo-parietal distribution). The term is thus independent of the clinical manifestation of the disease.

*Alzheimer’s dementia* is the clinical syndrome featuring acquired and progressive cognitive impairment associated with functional disability as defined by the NINCDS-ADRDA criteria [16]. Notably, not all cases of clinically diagnosed Alzheimer’s dementia have AD pathology due to the imperfect accuracy of purely clinical criteria.

*Mild cognitive impairment (MCI)* is used to indicate the clinical condition between normality and dementia in which patients experience acquired cognitive impairment of greater severity than expected by age, but no functional disability. This population includes cases with AD (about 50%), cases with other neurodegenerative pathologies (10-15%) and cases without a neurodegenerative disorder (35-40%) [17–20]. The MCI cases with AD biomarker positivity are defined as *Prodromal AD* in the clinical criteria by Dubois et al. [21]. The focus of this review is the use of biomarkers in the MCI stage.

*Non-Alzheimer neurodegenerative disease* includes neurodegenerative disorders that are not primarily due to AD pathology. These belong to a large pathological spectrum including...
hippocampal sclerosis, frontotemporal lobar degeneration, progressive supranuclear palsy, corticobasal degeneration, argyrophilic grain disease, Lewy Body disease and other alpha-synucleinopathies such as multiple system atrophy.

**Conceptual framework**

The main phases of the present framework for the development of biomarkers for early diagnosis reflect the phases covered in the original oncology framework by Pepe et al. [7] and in turn inspired by pharmaceuticals development. The shift of the reference methodological model from the field of oncology to that of dementia implies a shift of aims from screening to diagnosis, as the examined biomarkers are validated for use in the clinical population of MCI [8,22]. The present review assesses the clinical validity of medial temporal lobe atrophy within a translated framework, consisting of five phases with a main aim and various sub-aims. These (sub-)aims and evidence reported are summarized in Table 1, together with the most pertinent references. There are five consecutive phases of development that should be completed before clinical use of the biomarker for prediction of AD at the MCI stage.

**Phase 1:** Phase 1 studies are preclinical exploratory studies, in which the aim is to find leads for potential biomarkers by identifying characteristics specific to the disease, based on pathology findings, which could be detected with clinical tests.

**Phase 2:** Evaluation of the biomarker’s ability to discriminate patients with AD from controls. Ideally, evidence is based on studies in which a diagnosis of AD is also supported by autopsy findings. Sub-aims of phase 2 focus on defining and optimizing the clinical assay allowing reliable discrimination between patients and controls (sub-aim 1), determining the relation between pathological measurement and biomarker measurement (sub-aim 2) and the assessment of possible differential effects of factors in patients and controls that may influence the thresholds for positivity (sub-aims 3 and 4). Relevant factors can be for example age, apolipoprotein E (APOE) ε4 carrier status and educational attainment.

**Phase 3:** Phase 3 studies consist of prospective longitudinal studies and the main aim is to assess the ability of the biomarker to detect the disease at the MCI stage by evaluating the biomarker’s ability to predict the development of incident Alzheimer’s dementia at follow-up. In other words, to distinguish MCI progressors from non-progressors. We included studies examining the biomarker at baseline in subjects with MCI and sufficiently long follow-up, ideally over three years. Second main aim is to fine-tune the threshold for positivity. Sub-aims are to assess the impact of covariates on the discriminatory ability of the biomarker in the MCI stage (sub-aim 1), to compare the usefulness of the biomarker in comparison to or in combination with other available biomarkers (sub-aims 2 and 3) and to determine a biomarker testing interval (sub-aim 4).
Phase 4: Main aim of phase 4 is to estimate the accuracy and usefulness of the biomarker-based early diagnosis in real world patients. It consists of prospective diagnostic studies in which the biomarker is used for an early diagnosis of AD, affecting decision-making regarding patient management and treatment. Sub-aims are to assess the benefit of the biomarker-based early diagnosis (sub-aim 1) and the feasibility of the biomarker assessment (sub-aim 2), and provides preliminary evidence on impact on mortality, morbidity and costs (sub-aim 3), and undetected cases (sub-aim 4).

Phase 5: Phase 5 studies aim to quantify the impact of the biomarker-based early diagnosis on clinically meaningful outcomes and costs. They consist of disease-control studies assessing the reduction in mortality, morbidity and disability allowed by the biomarker-based diagnosis. Sub-aims are to assess cost-effectiveness (sub-aim 1), evaluate compliance in different settings (sub-aim 2) and to compare different biomarker testing protocols (sub-aim 3). However, this phase can be properly carried out only in the context of an effective and accepted treatment available. For the AD field, only mortality and quality of life may be properly considered within this phase.

Literature search, article selection and evidence evaluation
References for this review were selected searching the PubMed/Medline database in June 2015. A different search algorithm was used for the aims of the 5-phases, each one comprising an aim-specific key word string and a biomarker specific key-word string: "MTA" OR "medial temporal" OR "hippocamp*". The aim-specific key word strings can be found in Supplementary Table 1. Only papers published in English were included. References were also selected on the basis of the authors' personal knowledge and by screening references from retrieved articles. When aims were unequivocally achieved, a reference paper or review was selected by the authors. The final selection of articles was based on relevance to topics covered in this review, as judged by the authors.
Table 1: Available or required evidence indicating full, partial or lack of achievement of the phases adapted from the oncology framework (Pepe et al., 2001) for visual rating of medial temporal lobe atrophy (vMTA) and hippocampal volume (HCV).

<table>
<thead>
<tr>
<th>Phase Design</th>
<th>General aim</th>
<th>Specific aim</th>
<th>Progress</th>
<th>Evidence in Alzheimer’s disease</th>
<th>Important references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1—Preclinical Exploratory Studies</td>
<td>Identify the rational of the biomarker, based on pathology</td>
<td>Primary: To identify leads for potentially useful biomarkers and prioritize identified leads.</td>
<td>Fully achieved</td>
<td>Pathological studies show early medial temporal lobe involvement with neuronal loss in hippocampus.</td>
<td>23-26</td>
</tr>
<tr>
<td>Phase 2—Clinical Assay Development for Clinical Disease</td>
<td>Define the ability of the biomarker to discriminate patients from controls</td>
<td>Primary: To estimate TPR and FPR or ROC for the assay and to assess its ability to distinguish subjects with and without disease</td>
<td>Fully achieved</td>
<td>vMTA and HCV separate clinical Alzheimer’s dementia patients from cognitively healthy subjects with good sensitivity and specificity. Few studies have examined pathologically verified samples. vMTA and HCV are less useful in differential diagnosis of dementia patients.</td>
<td>28-36</td>
</tr>
<tr>
<td></td>
<td>Secondary 1: To optimize procedures for performing the assay and to assess the reproducibility of the assay within and between laboratories.</td>
<td>Fully achieved for vMTA; Partly achieved for HCV</td>
<td>vMTA has good reproducibility in trained raters. Standardized method of collecting MRI (ADNI protocol) and manual volumetry (EADC-ADNI harmonized segmentation protocol) are sparsely implemented in memory clinics. Much variability between different automated HCV measurement algorithms.</td>
<td>38-41; 46; 48; 55-56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Secondary 2: To determine the relationship between biomarker tissue measurements made on tissue and the biomarker measurements made on the noninvasive clinical specimen.</td>
<td>Fully achieved</td>
<td>Good correlation between hippocampal size on MRI and histological measurements and severity of neurodegenerative changes on pathology.</td>
<td>31; 60-67; 69</td>
<td></td>
</tr>
<tr>
<td>Phase Design</td>
<td>General aim</td>
<td>Specific aim</td>
<td>Progress</td>
<td>Evidence in Alzheimer’s disease</td>
<td>Important references</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Phase 2—Clinical Assay Development for Clinical Disease</td>
<td>Define the ability of the biomarker to discriminate patients from controls</td>
<td>Secondary 3: To assess factors associated with biomarker status or level in control subjects. If such factors affect the biomarker, thresholds for test positivity may need to be defined separately for target subpopulations.</td>
<td>Fully achieved</td>
<td>Well-known age-associated HCV loss (higher vMTA scores). Amount of atrophy also dependent on APOE ε4 genotype, vascular pathology, education; impact of these latter variables only relevant for volumetry, not for visual rating.</td>
<td>75–77; 84-86; 89-90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary 4: To assess factors associated with biomarker status or level in diseased subjects—in particular, disease characteristics.</td>
<td>Fully achieved</td>
<td>Similar as for healthy subjects: influence of age, APOE ε4, vascular pathology. Additionally, clinical presentation is relevant: subjects with early onset and primary non-memory presentation have relatively spared medial temporal lobes.</td>
<td>94-96; 100; 102-112</td>
</tr>
<tr>
<td>Phase 3—Prospective Longitudinal Repository Studies</td>
<td>Define the ability of the biomarker to detect the disease in its early phase</td>
<td>Primary 1: To evaluate the capacity of the biomarker to detect the earliest disease stages</td>
<td>Partly achieved</td>
<td>No evidence in clinical MCI cohorts with long follow-up. In research populations or clinical populations with shorter follow-up, there is a reasonably good specificity but lower sensitivity to predict clinical progression in subjects with MCI.</td>
<td>28; 119-123</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Primary 2: To define criteria for a biomarker positive test in preparation for phase 4.</td>
<td>Partly achieved</td>
<td>Age and APOE ε4 related cut-offs for vMTA based on discrimination of controls from Alzheimer’s dementia. Validation needed in clinical MCI cohorts. No universal cut-offs for volumetry; substantial variability between acquisition protocols and measurement algorithms.</td>
<td>66; 94; 132; 137</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary 1: To explore the impact of covariates on the discriminatory abilities of the biomarker before clinical diagnosis.</td>
<td>Partly achieved</td>
<td>Impact of age, APOE ε4 genotype and clinical presentation.</td>
<td>132-133; 138</td>
</tr>
<tr>
<td>Phase Design</td>
<td>General aim</td>
<td>Specific aim</td>
<td>Progress</td>
<td>Evidence in Alzheimer's disease</td>
<td>Important references</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Phase 3—Prospective Longitudinal Repository Studies</td>
<td>Define the ability of the biomarker to detect the disease in its early phase</td>
<td>Secondary 2: To compare markers with a view to selecting those that are most promising</td>
<td>Partly Achieved</td>
<td>Various studies on association of two or more core biomarkers; usually best predictive value for combination of amyloid marker with an injury marker.</td>
<td>18; 141-152; 155</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary 3: To develop algorithms for positivity based on combinations of markers.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary 4: To determine a biomarker testing interval for phase 4 if repeated testing is of interest.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 4—Prospective Diagnostic Studies</td>
<td>Quantify the biomarker accuracy and usefulness in patients diagnosed and treated based on the BM</td>
<td>Primary: To determine the operating characteristics of the biomarker-based test in a relevant population by determining the detection rate and the false referral rate. Studies at this stage involve testing people and lead to diagnosis and treatment.</td>
<td>Not achieved</td>
<td>Not applicable for vMTA. Preliminary evidence on added value of hippocampal atrophy rates for predicting clinical progression to Alzheimer’s dementia in patients with MCI.</td>
<td>72; 157; 162-163; 166-169</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary 1: To describe the characteristics of disease detected by the biomarker test—in particular, with regard to the potential benefit incurred by early detection.</td>
<td>Not achieved</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase Design</td>
<td>General aim</td>
<td>Specific aim</td>
<td>Progress</td>
<td>Evidence in Alzheimer's disease</td>
<td>Important references</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Phase 4—Prospective Diagnostic</td>
<td>Quantify the biomarker accuracy and usefulness in patients diagnosed and</td>
<td>Secondary 2: To assess the practical feasibility of implementing the diagnostic</td>
<td>Preliminary evidence for vMTA; Not achieved for HCV</td>
<td>Imaging integrated in standard work-up for dementia patients in most memory clinics. vMTA and</td>
<td>40; 131</td>
</tr>
<tr>
<td>Studies</td>
<td>treated based on the BM</td>
<td>program and compliance of test-positive subjects with work-up and treatment</td>
<td></td>
<td>quantitative assessment not yet widely implemented.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>recommendations.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary 3: To make preliminary assessments of the effects of biomarker</td>
<td>Not achieved</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>testing on costs and mortality associated with the disease.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary 4: To monitor disease occurring clinically but not detected by the</td>
<td>Not achieved</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>biomarker testing protocol.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 5—Disease Control Studies</td>
<td>Quantify the impact of the biomarker-based diagnosis on clinically</td>
<td>Primary: To estimate the reductions in disease-associated mortality, morbidity,</td>
<td>Not achieved</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>meaningful outcomes and costs</td>
<td>and disability afforded by biomarker testing.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary 1: To obtain information about the costs of biomarker testing and</td>
<td>Not achieved</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>treatment and the cost per life saved or per quality-adjusted life year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary 2: To evaluate compliance with testing and work-up in a diverse</td>
<td>Not achieved</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>range of settings.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase Design</td>
<td>General aim</td>
<td>Specific aim</td>
<td>Progress</td>
<td>Evidence in Alzheimer's disease</td>
<td>Important references</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------</td>
<td>----------</td>
<td>---------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Phase 5—Disease Control Studies</td>
<td>Quantify the impact of the biomarker-based diagnosis on clinically meaningful outcomes and costs</td>
<td>Secondary 3: To compare different biomarker testing protocols and/or to compare different approaches to treating test positive subjects in regard to effects on mortality and costs.</td>
<td>Not achieved</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ADNI = Alzheimer's Disease Neuroimaging Initiative; APOE = apolipoprotein E; FPR = false positive rate; MCI = mild cognitive impairment; TPR = true positive rate; ROC = receiver operating characteristic.
Figure 1: Synopsis of the maturity of a visual rating of medial temporal lobe atrophy as borrowed from the oncology framework [7].

AD = Alzheimer’s disease; HC = healthy controls; MCI = mild cognitive impairment; vMTA = visual rating of medial temporal lobe atrophy.

Figure 2: Synopsis of the maturity of hippocampal volumetry as borrowed from the oncology framework [7].

AD = Alzheimer’s disease; HC = healthy controls; HCV = hippocampal volume; MCI = mild cognitive impairment.

For all phases, available literature was assessed and used to evaluate whether each aim was considered as Achieved, Partly Achieved, presenting with Preliminary Evidence, or Not Achieved for vMTA and HCV. Results of this assessment are visualized in Figure 1 and Figure 2. **Fully Achieved:** scientific evidence is available and replicated in representative samples in studies without major methodological faults. **Partly Achieved:** scientific evidence is available but not yet sufficiently replicated, or samples are not representative, or other
significant methodological limitations can be found in the available literature. Preliminary evidence: only preliminary evidence is available. Not Achieved: no evidence was found and no studies are known to be ongoing at the time of the writing of this review.

**EVIDENCE FOR CLINICAL VALIDITY OF MEDIAL TEMPORAL LOBE ATROPHY ON MRI**

**Phase 1 – Pilot studies**

The aim of phase 1 studies is to find leads for potential biomarkers based on pathological findings.

AD is characterized by extracellular amyloid beta depositions and intraneuronal or extraneuronal neurofibrillary changes, eventually leading to neuronal destruction. Autopsy studies have shown that these changes already start many years before the onset of clinical symptoms with early and prominent neurofibrillary tangles in medial temporal lobe structures [23–26]. These neuropathological changes in the medial temporal lobe are accompanied by atrophy, which can be visualized on structural MRI [27].

The first aim can be considered Fully Achieved for vMTA and HCV.

**Phase 2 – Clinical assay development for clinical disease**

The purpose of the second phase is to find a clinical measurement based on the findings from phase 1, which can be easily obtained and sufficiently distinguishes subjects with and without AD. Secondary aims in this phase are to optimize the procedures for performing the measurement, assess reproducibility, validate the measurement against pathological measurements and assess factors associated with the measurement in controls and diseased subjects.

**Phase 2, Primary aim: ability to distinguish patients from controls.**

The primary aim of phase 2 is to assess the ability of vMTA and HCV to distinguish patients with AD from healthy controls.

Many case-control studies have evaluated the use of MTA in differentiating subjects with clinically diagnosed Alzheimer’s dementia from healthy controls, which have recently been reviewed [28]. Average specificity of a visual read of MTA to distinguish clinical Alzheimer’s dementia from healthy controls is 79% (C.I. 75–83) with average-good sensitivity of 70% (C.I. 65–74). In a pathology verified sample of young AD subjects (mean age 59), the sensitivity of vMTA was high (92%) with a specificity of 62% [29]. Specificities and sensitivities of
manual hippocampal volumetry are on average 82% (C.I. 78 - 85) and 79% (C.I. 76 - 82) respectively [28]. Specificities and sensitivities for automated measurements of HCV are on average 81% (C.I. 77 - 85) and 72% (C.I. 67 - 77) respectively [28]. The specificity and sensitivity of HCV for detecting AD were in a similar range in two neuropathological studies: 0.80 - 0.87% and 0.75 - 0.82% [30,31].

Although MTA reasonably well distinguishes subjects with dementia from healthy controls, it is less useful in the differential diagnosis of Alzheimer’s dementia. Atrophy of the medial temporal lobe and hippocampus is also seen in other neurodegenerative diseases, as well as in vascular dementia [29,30,32–36].

The primary aim of phase 2 can be considered Fully Achieved for vMTA and HCV.

Phase 2, Secondary aim 1: optimize procedures for biomarker assessment.
The first secondary aim of phase 2 is to optimize procedures for measuring medial temporal lobe atrophy and to assess the reproducibility of this measurement.

Visual rating of MTA is quick and easy and can be performed by any trained rater, usually a radiologist. In contrast to volumetric methods, vMTA is relatively independent of acquisition protocol. Merely a good quality anatomical scan is required, which is usually a T1-weighted MRI with coronal reconstructions, but can also be a high-resolution CT scan [37]. The Scheltens scale has reasonable inter-rater agreement and reproducibility. A study shortly after the development of the scale reports inter-rater agreements with kappa values of 0.59 - 0.62 for dichotomized vMTA into present or absent [38]. In more recent publications, the inter-rater kappa is usually higher, up to 0.90 [39]. This may be due to advances in imaging acquisition techniques with higher field strengths and better display methods or more experience of the raters. Several studies have shown that the reproducibility and accuracy of vMTA is higher when scoring is performed by trained investigators [40,41].

Volumetric methods may be influenced by scanner type and acquisition protocol [42–45]. To increase uniformity of MRI acquisition methods between different sites, the Alzheimer’s Disease Neuroimaging Initiative (ADNI), developed a standardized protocol for MRI acquisition [46]. Although this sequence has been adopted by some scanner manufacturers, it is not yet widely implemented for clinical use.

Over the years, various methods for manual segmentation of the hippocampus and medial temporal lobe have been developed [47]. Due to the complexity of the hippocampal region and different definitions of anatomical landmarks across research groups, manual volumetry has varying reproducibility rates. Recently, effort has been put into the
development of a standardized method for manual segmentation of the hippocampus: the EADC-ADNI Harmonized Protocol (HarP) [48]. Compared to local protocols, using the harmonized protocol for hippocampal segmentation results in higher intra- and interrater agreement [49]. Given the time-consuming and expensive nature of manual outlining, automated volumetry has higher potential to be broadly used in clinical setting. Automated methods approach performance of manual outlining in distinguishing healthy controls, MCI and subjects with Alzheimer’s dementia [28,50]. Several (semi)automated algorithms have been developed for the automated measurement of hippocampal volume. Non-commercial, widely used in research, algorithms include FIRST (FMRIB’s integrated registration and segmentation tool, FSL) [51,52] and FreeSurfer [53]. Several commercially available algorithms have been developed, such as Assessa* (IXICO) based on the LEAP algorithm [54] and NeuroQuant* (CorTechs Labs). These various methods use different a priori anatomical information on the hippocampus, as well as different computational strategies for volume estimation and therefore also provide different volumes for same subjects [55,56]. Although steps towards the standardization of automated methods have been undertaken [57–59], their generalizability across acquisition methods and centres is still insufficient for their routine application in clinical practice.

This sub-aim can be considered Fully Achieved for vMTA and manual HCV. For automated HCV, this aim is considered Partly Achieved.

Phase 2, Secondary aim 2: relationship between pathology and biomarker measurement

Secondary aim 2 of phase 2 consists of determining the relationship between the biomarker measurement and actual pathology.

Medial temporal lobe atrophy assessed on MRI correlates well with neuropathological findings. In post-mortem studies, HCV measured on MRI correlates strongly with histological volume measurements [60]. Furthermore, hippocampal atrophy measurements on MRI are indicative of Braak neurofibrillary tangle stage at autopsy, even in clinically non-demented subjects [31,61–63], and HCV on MRI correlates strongly with histopathological measures of neuron count and neurofibrillary tangle density in the hippocampus [64,65]. Good correlations between vMTA and AD pathology at autopsy (Braak staging and plaques and tangles in hippocampus) have been demonstrated by a few studies [66,67]. However, high vMTA scores and decreased HCV are not exclusive to AD pathology [30,66,68]. Good correlations between Braak stage, neuronal count and HCV on MRI has also been demonstrated for the HarP manual outlining protocol [69].

Several studies have compared visual rating of MTA with volumetric methods and found that vMTA ratings correlate well with MRI volumetric measurements [40,70–74].
This sub-aim is considered *Fully Achieved* for vMTA and HCV.

**Phase 2, Secondary aim 3: impact of covariates on biomarker measurement in healthy controls**

Phase 2, secondary aim 3 assesses the impact of covariates on the biomarker level in control subjects.

Several factors that influence medial temporal lobe atrophy in cognitively healthy subjects have been identified: age, *APOE* ε4 genotype, vascular risk factors and co-morbid brain disease (mostly psychiatric). In studies examining cognitively healthy elderly subjects, a confounding effect of preclinical AD cannot be ruled out since only very few studies have examined pathology verified healthy controls.

Numerous studies have demonstrated age-related decline in HCV in cognitively healthy adults [75–78]. Some studies have shown non-linear effects of aging on HCV with increased decline with advancing age [75,76,79,80]. Age-related decline in HCV is reflected in age-related increase in vMTA scores in cognitively healthy subjects [14,81]. Older aged subjects are also more prone to have non-Alzheimer pathology leading to medial temporal lobe atrophy, such as hippocampal sclerosis [66]. Age-related decline in HCV is mediated by a protective effect of higher education [82]. Other studies have also shown a protective effect of education on HCV [83].

Various studies report decreased HCV in cognitively healthy subjects with an *APOE* ε4 allele compared to *APOE* ε4 non-carriers [84,85]. However, this effect is not undisputed with some researchers finding no effect of *APOE* ε4 on cross-sectional HCV [76,86–88]. A possible explanation for these differences could be the inclusion of different study populations, where some of the studies that failed to find an effect included younger populations.

Studies on associations between vascular risk factors, vascular brain lesions and HCV have yielded inconsistent results [77]. In the population-based Rotterdam study, researchers found an association between longitudinal hippocampal atrophy rates, vascular white matter lesions and diastolic blood pressure [84]. Another large scale study showed an effect of smoking and diastolic blood pressure on HCV [89]. Others did not find effects of vascular risk factors and vascular white matter hyperintensities on HCV [90].

Several psychiatric diseases are also associated with decreased HCV in non-demented subjects, such as depression [91,92] and post-traumatic stress disorder [93].

This sub-aim is considered *Fully Achieved* for vMTA and HCV.
Phase 2, Secondary aim 4: impact of covariates on biomarker measurement in patients with Alzheimer’s disease

Phase 2, secondary aim 4 assesses the impact of covariates on the biomarker measurement in subjects with AD.

Many of the same factors that influence medial temporal lobe atrophy in healthy subjects also apply to subjects with Alzheimer’s dementia. Additionally, several disease-related factors have been shown to affect the amount of medial temporal lobe atrophy: age-of-onset, disease stage and clinical presentation. A large scale study combining data from the ADNI and AddNeuroMed databases has shown that in subjects with Alzheimer’s dementia, visually rated MTA scores are influenced by age, gender and disease duration [94]. Similarly, lower HCV is found with increasing age and increasing cognitive impairment in patients with Alzheimer’s dementia [78,95,96].

Compared to the rest of the brain, medial temporal atrophy is most pronounced at an early disease stage; at later stages other cortical areas also become more affected [97–99]. Subjects with early-onset (age ≤ 65 years) Alzheimer’s dementia have different atrophy patterns than subjects with late-onset Alzheimer’s dementia. Subjects with early-onset Alzheimer’s dementia have more pronounced parietal and precuneal atrophy, whereas subjects with late-onset Alzheimer’s dementia have more prominent medial temporal atrophy [100–103]. Moreover, early onset-subjects have more frequently a non-memory clinical presentation, which is also associated with relative hippocampal sparing [104–106]. In Alzheimer’s dementia, carriership of an APOE ε4 allele is associated with decreased volume of medial temporal lobe structures [107–112]. This effect may be limited to younger subjects: sensitivity of vMTA to detect early-onset Alzheimer’s dementia is high in APOE ε4 carriers (82%), but only 47% in APOE ε4 non-carriers [94], whereas in late-onset the sensitivity was around 80%, regardless of APOE genotype. Taken together, these studies suggest that medial temporal lobe atrophy may not be useful as a biomarker for early diagnosis of Alzheimer’s dementia in young onset patients, especially those without an APOE ε4 allele or non-memory presentation.

Studies on the effects of vascular pathology on medial temporal lobe atrophy have inconsistent results, with some finding increased medial temporal lobe atrophy associated with white matter hyperintensities [113,114], whereas other studies did not find such a relation [115].

This sub-aim is considered Fully Achieved for vMTA and HCV.
Chapter 2

Phase 3 – Prospective longitudinal repository studies
In phase 3, the ability of medial temporal lobe atrophy to detect subjects with MCI who will progress to Alzheimer's dementia is assessed based on prospective clinical studies. Evidence is presented from studies examining medial temporal lobe atrophy as a predictor for clinical decline from the stage of MCI. In this phase, criteria for a positive biomarker test are defined and factors influencing the abilities of the biomarker to detect pre-clinical disease are assessed. Additionally, in this phase different biomarkers are compared and possibly combined for optimal detection.

Phase 3, Primary aim 1: ability of biomarker to predict progression to Alzheimer's dementia
The primary aim of phase 3 is to assess the capacity of the biomarker to detect subjects with MCI who will progress to Alzheimer's dementia.

Compared to cognitively healthy controls, subjects with MCI have reduced HCV in the following order: control > MCI > Alzheimer's dementia [50,116–118]. A considerable amount of studies support the ability of medial temporal lobe atrophy to predict progression to dementia from the MCI stage. These studies have recently been summarized, resulting in an overall specificity of 75% (C.I. 67 - 82) and average sensitivity of 60% (C.I. 51 - 68) for the visual rating of MTA on MRI [28]. Most of the studies reported by Frisoni et al. [28] have relatively short clinical follow-up in the order of 1-2 years. deCarli et al. [119] assessed data from a clinical trial with three year longitudinal follow-up in 190 subjects with amnestic MCI and found an average specificity of 98% and sensitivity of 14% for progression to Alzheimer's dementia with a vMTA cut-off score of ≥ 2 (average of left and right) [119]. With a less stringent cut-off (≥ 1), specificity was 69% and sensitivity 51%. Liu et al. [120] analysed MCI subjects from the ADNI cohort with three year follow-up and found specificity of 82% and sensitivity of 32% for the prediction of progression to Alzheimer's dementia with a vMTA cut-off of ≥ 3 on any side. Which cut-off scores should be used remains a matter of debate and is elaborated upon in the next section.

Specificities and sensitivities of manual HCV for prediction of progression to dementia at the MCI stage are on average 81% (C.I. 73 - 87) and 58% (C.I. 47 - 68) respectively [28]. Specificities and sensitivities for automated measurements of HCV are on average 66% (C.I. 61 - 71) and 70% (C.I. 63 - 76) respectively [28]. Studies in samples with long follow-up (≥ 3 years) have reported similar specificities of 80-87% and sensitivities of 60-67% for manual volumetry [121,122] and 50% and 83% for automated HCV [123].
Medial temporal lobe atrophy performs reasonably well in the prediction of cognitive decline from the MCI stage but is not specific for AD [30,33,35,36,124]. Other imaging markers such as parietal or frontal atrophy may be more useful in differentiating between AD and other underlying pathologies [68,125–128].

In MCI subjects who progress to Alzheimer’s dementia, vMTA rating is associated with time to dementia and can therefore also serve as a prognostic marker [129]. Similar results have been found for volumetric assessments of the hippocampus in MCI [121] and amyloid positive MCI [129,130].

Much evidence has already been gathered on the ability of medial temporal lobe atrophy to predict progression to Alzheimer’s dementia. However, to fully complete this aim, sensitivities and specificities for progression will also need to be assessed in clinical MCI populations, with longer follow-up. Therefore, we consider this aim Partly Achieved for both vMTA and HCV. Data to fully achieve this aim could be readily available from memory clinic samples with standardized use of MRI [131].

Phase 3, Primary aim 2: define criteria for positive biomarker test

The second primary aim of phase 3 is to define criteria for a positive biomarker test.

Since there is a considerable influence of age on medial temporal lobe atrophy in both healthy subjects and patients with Alzheimer’s dementia, age-related cut-offs for vMTA have been proposed [14,66,94,132]. A recent large-scaled study combining data from the ADNI and AddNeuroMed studies examined the effects of APOE ε4 on vMTA and demonstrated that early-onset Alzheimer’s dementia subjects without APOE ε4 have lower vMTA scores than APOE ε4 carriers [132]. This has led to the proposition of age-related cut-offs, stratified by APOE genotype. Using this same dataset, Ferreira and colleagues found the highest performance of vMTA to distinguish Alzheimer’s dementia from healthy controls when using the following general age-adjusted cut-offs (average scores): 45-74: ≥ 1.5; 75-84: ≥ 2 and 85-94: ≥ 2.5 (sensitivity 80%, specificity 77%). For early-onset (≤ 65 years) APOE ε4 non-carriers a vMTA cut-off ≥ 2 had better performance [94]. These cut-offs will need to be validated for use in prediction of progression from MCI in a prospective clinical setting.

The proposed cut-offs were derived using average vMTA scores of left and right hemisphere. Other researchers have used the highest vMTA score on either side to define abnormality [14,133]. Pereira et al. [132] have examined the performance of different vMTA cut-off scores for differentiating patients with Alzheimer’s dementia from healthy controls, including cut-offs based on average and highest scores. They found a better performance of age-adjusted average cut-off scores compared to age-adjusted highest cut-off scores,
especially in the group of older patients. Another argument for using average scores is the relatively symmetrical distribution of Alzheimer’s pathology, in contrast to some other neurodegenerative diseases [134,135].

For volumetric analysis, several methods to define cut-offs have been used in research settings, depending on the intended use and need for either higher sensitivity or specificity [136]. One strategy that is frequently used involves taking the 95% percentile of a reference population, with the implication that everything below this is abnormal. Another commonly used method is the creation of covariate-corrected Z-scores, also called W-scores [137]. W-scores represent where a hippocampal volume would fall on the normal distribution of healthy controls, corrected for covariates. W-scores require the availability of a normative data set. Advantages of this method are the possibility to include covariates and the relative robustness of derived cut-offs against the use of different measurement algorithms or acquisition methods.

A big challenge in defining universal cut-offs in automated volumetric analysis is the influence of image acquisition method (such as scanner type and acquisition protocol) on obtained results [42–45]. Due to the large variability in HCV obtained from different automated methods [55,56], universal absolute cut-off points cannot be defined. With the development of the HarP [48], and of certified labels that may be used to train algorithms for automated segmentation [57], results from automated volumetry may become more consistent between methods in the future.

This sub-aim is considered Partly Achieved for vMTA and in the stage of Preliminary Evidence for HCV.

**Phase 3, Secondary aim 1: impact of covariates on biomarker measurement in subjects with MCI**

Secondary aim 1 of phase 3 is to explore the impact of covariates on the discriminatory ability of the biomarker to predict progression from the MCI stage.

Few studies have addressed this issue directly. Factors that may impact the ability of medial temporal lobe atrophy to predict AD at the MCI stage are age, clinical presentation, and APOE genotype. In very old subjects there is age-related hippocampal atrophy in healthy subjects as well as in Alzheimer’s dementia patients [78], which could affect the discriminatory ability between normal and abnormal. In early-onset Alzheimer’s dementia, there is less prominent involvement of the medial temporal lobe, making this biomarker less suitable for young subjects [101,103]. Different clinical subtypes of Alzheimer’s dementia are associated with specific brain atrophy patterns, which may already be visible in the MCI
Clinical validity of MTA in MCI

stage. Compared to amnestic MCI, subjects with non-amnestic MCI have relatively spared medial temporal lobes [133,138], which reduces the sensitivity of medial temporal lobe atrophy for prediction of progression to Alzheimer-type dementia in non-amnestic MCI.

Using ADNI and AddNeuroMed data, it has been shown that APOE ε4 carriership and early-onset disease before the age of 65 affect the performance of vMTA for prediction of clinical progression from the MCI stage [132]. Further replication of these findings in clinical cohorts and extension to volumetric methods are required.

This secondary aim is considered *Partly Achieved* for vMTA and HCV.

**Phase 3, Secondary aims 2 & 3: comparison and combination of biomarkers**

Secondary aims 2 and 3 of phase 3 are to compare biomarkers and develop algorithms for positivity based on combinations of markers.

Research criteria developed by the National Institute of Aging and the Alzheimer Association (NIA-AA) have incorporated different stages of likelihood of developing Alzheimer’s dementia based on the combination of amyloid markers (either CSF or PET) and injury markers, such as medial temporal atrophy [139]. Neuroimaging and CSF markers are currently widely accepted biomarkers for AD in research settings [140]. There is, however, not yet enough knowledge about their use in clinical practice. There is no consensus on which biomarkers, in what combination and in which order should be used in the work-up of clinical MCI patients. Various studies have examined the combination of multiple imaging markers for prediction of decline in patients with MCI, which have recently been summarized [141]. Compared to FDG-PET, measures of hippocampal atrophy seem to be less accurate at predicting conversion to Alzheimer’s dementia from MCI [28,142–145] but many studies found highest accuracies for a combination of biomarkers [143,144]. Studies combining information from amyloid PET and structural MRI have also reported highest accuracies for combinations of both markers rather than a single biomarker [18,146]. Multiple studies have also examined the combined use of CSF and MRI markers for the prediction of clinical progression in subjects with MCI. Some have found better prediction for CSF and others for MRI but nearly all of them show added benefit of a combination of both [147–153]. Only the study from Bouwman et al. assessed vMTA in a clinical cohort, the others assessed HCV in research settings. Differences in findings between the studies may largely be explained by use of highly selected samples, different definitions of MCI (or only inclusion of amnestic MCI) and variations in methods used for deriving the biomarker measures.
Some studies have shown that high vMTA scores and decreased HCV may be better predictors of time to clinical progression than evidence of amyloid pathology [130,154]. Taken together, these studies suggest that different biomarkers provide complementary information and that combinations of biomarkers may be needed for accurate prediction of clinical progression to Alzheimer’s dementia at the MCI stage. Although recent efforts are focussing on devising algorithms taking into account all available biomarker data to aid in the prediction of progression to Alzheimer’s dementia in clinical settings [155], more research is still needed to determine which biomarkers should be used, in which order they should be assessed and what to do in the presence of conflicting results.

These sub-aims can be considered Partly Achieved for vMTA and HCV.

**Phase 3, Secondary aim 4: biomarker testing interval**
Secondary aim 4 of phase 3 is to determine a biomarker testing interval.

Progressive hippocampal atrophy is an important imaging finding in Alzheimer’s dementia. Compared to cognitively healthy subjects, hippocampal atrophy rates are 2-4 times greater in subjects with Alzheimer’s dementia [156–160]. For repeated testing to be useful for diagnosis or prognosis in clinical practice, a biomarker should be able to detect changes over short intervals. A visual rating scale is not sensitive enough to detect changes over short term follow-up evaluations in the order of one year [72]. Repeated testing may be valuable in the case hippocampal volumetry becomes available for routine clinical application. Hippocampal atrophy rates can be measured reliably over a period of one year and some studies have even show that volumetric atrophy rates can be measured over periods as short as 6 months [161–164]. Several longitudinal studies have found that increased hippocampal atrophy rates are associated with rapid progression to Alzheimer’s dementia in MCI [165–168] and may perform better in prediction of cognitive decline than baseline hippocampal volumes alone [157,169].

This sub-aim is not applicable for vMTA. For HCV this sub-aim can be considered at the stage of Preliminary Evidence.

**Phase 4 – Prospective diagnostic studies**
The primary aim of phase 4 is to determine the operating characteristics of the biomarker in a clinical setting.

Most studies that have examined the use of medial temporal atrophy as a predictor for clinical progression in subjects with MCI have used highly selected populations in terms of MCI subtypes, scan quality and comorbidities and are therefore not generalizable to
Clinical validity of MTA in MCI

In order to assess the value of medial temporal atrophy as a biomarker for early diagnosis in MCI, it is important to also examine clinical populations using methods that are feasible for broad implementation. In phase 4 studies, use of the biomarker leads to early diagnosis and the effects on patient management and outcome are assessed. We are not aware of any prospective clinical studies examining the systematic use of medial temporal atrophy for the prediction of AD at the MCI stage.

**Phase 4, Secondary aim 1: characteristics of disease detected by biomarker in early stage**
Secondary aim 1 of phase 4 assesses the characteristics of the disease identified in an early stage by the biomarker in a clinical setting, specifically with regard to potential benefit for the patient incurred by early detection.

This aim is *Not Achieved* for vMTA and HCV.

**Phase 4, Secondary aim 2: feasibility of biomarker measurement**
The secondary aim of phase 4 assesses the practical feasibility of implementing the biomarker measurement in a clinical setting and the compliance of test-positive subjects with work-up and treatment recommendations.

Although not formally assessed, a visual rating of MTA should be feasible in clinical setting. Most memory clinics already use imaging in the work-up of memory disorders to exclude other (treatable) pathologies such as tumours, vascular damage as well as give direction to underlying neurodegenerative pathology such as frontotemporal dementia or progressive supranuclear palsy. Tertiary memory clinics or centres associated with a research facility often use standardized MRI protocols and may also adopt structured radiology reporting, including a visual rating of MTA [40,131]. The application of manual or automated volumetry is still distant from implementation in daily clinical practice. To be widely implemented in clinical setting, volumetric analysis should be sufficiently standardized, fully automated and easy to use.

This aim can be considered at the stage of *Preliminary Evidence* for vMTA and *Not Achieved* for HCV.

**Phase 4, Secondary aim 3: impact on costs and mortality**
The aim is to evaluate the effects of biomarker testing on costs and mortality associated with AD.

This aim is *Not Achieved* for vMTA and HCV.
**Phase 4, Secondary aim 4: monitor undetected cases**
This aim includes monitoring disease occurring clinically but not detected by the biomarker testing. In other words, this subaim assesses how many subjects with MCI show clinical progression in the absence of medial temporal lobe atrophy at baseline and the clinical trajectories of these subjects.

This aim is *Not Achieved* for vMTA and HCV.

**Phase 5 – Disease control studies**
This final phase addresses whether using biomarkers for early diagnosis reduces the burden of AD in the general population. There are currently no studies assessing changes in mortality and morbidity, impact on economic costs or overdiagnosis associated with use of medial temporal lobe atrophy as diagnostic biomarker for AD at the MCI stage.

This entire phase is considered *Not Achieved* for vMTA and HCV.

**CONCLUSIONS AND FUTURE PERSPECTIVES**
In this paper we reviewed the evidence for medial temporal lobe atrophy as a biomarker for prediction of AD at the MCI stage. We performed this review in the context of a wider effort, aiming to accelerate the use of AD biomarkers at the predementia stage, where differentiating AD from normal aging and other causes of cognitive impairment is of huge clinical and societal relevance. The effort has borrowed a biomarker validation framework developed for oncology biomarkers [7] and ultimately taking inspiration from the traditional 4-phase drug development framework. Our working group has adapted the oncology framework to the predementia context, highlighting those issues sufficiently investigated and those in need of more research. This will allow funders of biomedical research to prioritize research topics towards the achievement of the ultimate aim of appropriate, effective and efficient use of AD biomarkers in the clinic.

In this review, we examined the available evidence for use of medial temporal lobe atrophy for prediction of AD at the MCI stage in the light of this framework. The first phase has been achieved. There is ample evidence to support medial temporal lobe atrophy as a characteristic feature of AD (phase 1). The second phase, focusing on the ability of the biomarkers to distinguish subjects with AD from healthy controls has also been achieved for a visual rating of MTA. For HCV, phase 2 has not yet been fully completed, with insufficient progress on the optimization and standardization of measurement algorithms.
There are still some steps to be taken in the third phase. The current evidence does not support the use of vMTA rating or HCV in isolation for the prediction of Alzheimer's dementia at the MCI stage. Accuracy for the prediction of progression to Alzheimer's dementia from the MCI stage has not reached clinically acceptable levels with, on average, sensitivities and specificities below 80% [28]. It should be noted that most studies have been performed on research samples with short follow-up. Prospective studies on clinical MCI populations with sufficiently long follow-up are lacking in the literature. Studies with short follow-up, in the order of 1-2 years, may underestimate the predictive ability of the biomarker for clinical progression. The predictive ability of medial temporal lobe atrophy is dependent on age, clinical presentation and APOE ε4 genotype [94,132,138]. However, clear guidelines on cut-offs for clinical (sub)populations have not yet been established.

Given the insufficient accuracy of medial temporal lobe atrophy alone for the prediction of AD at the MCI stage in a clinical setting, phase 4 and phase 5 studies should not be undertaken on this single biomarker but rather focus on assessing the impact of combinations of biomarkers [28,170]. Multiple studies have supported evidence for a model in which abnormal amyloid markers are present in early disease; whereas neuronal injury markers, such as medial temporal atrophy, may be more useful in predicting advancing pathology and thereby serve as a prognostic marker, rather than diagnostic marker [129,130,171,172]. Large size clinic-based studies assessing which combinations of biomarkers should be used, and in which order and what to do in the case of conflicting biomarker results are needed. Related to its potential role as a prognostic marker, HCV may also be valuable as a monitor of disease progression and could be used as a biomarker outcome measure in clinical trials [173,174]. For the latter, algorithms will need to have very low measurement errors (below 1.5%) to detect effects over a one year follow-up period, as yearly hippocampal atrophy rates are in the order of 2.9% - 5.6% per year for AD and 0.3% - 2.2% in healthy aging [175], resulting in an average difference of 3% atrophy per year between the groups.

Clinical implementation of a vMTA rating should be achievable, since this is a quick and accessible tool, which can be easily learned with adequate training. As structural imaging is already integrated in the work-up of patients with dementia in many clinics, it should be practically feasible to implement neuroimaging at the MCI stage as well. Unfortunately, a routine visual assessment of MTA is not yet widely adopted, especially in non-specialized clinics [15]. Adequate training may have a significant impact on the performance and application of visual rating scales in clinical practice. There might be a role for scientific societies to improve knowledge and know-how on visual ratings through the development of guidelines with reference images and training programs. Hippocampal volumetry using sophisticated analysis methods, rather than a visual read, may ultimately be a more powerful predictor of progression from MCI, but little progress has been made towards the
integration of these in clinical work streams [71]. The EMA has approved measurements of HCV for use in clinical trials and similar efforts are being undertaken to get approval from the FDA [59,176]. To be widely adopted in clinical setting, volumetric analysis should be fully automated and easy to use, for example through implementation on the scanner console.

In this paper, we have discussed the assessment of medial temporal lobe atrophy on MRI. Although MRI is the preferred imaging modality in the work-up of dementia, some patients are unable to undergo MRI for various reasons (e.g. claustrophobia, pacemaker, unavailability) or MRI is not included in national guidelines on dementia care [177]. In such cases a high resolution CT-scan with multiplanar reconstruction may also be used for the visual assessment of MTA, as has been validated by one study [37]. Future research may focus on further validating the use of vMTA on CT. Due to the lower contrast resolution, CT has only sporadically been used for volumetric measurements [178].

This review has several limitations. We reviewed the performance of medial temporal lobe atrophy in MCI patients; however, the definition of MCI is not homogeneous across different studies. This issue is addressed by Cerami et al. [9]. A further limitation of this study is that, notwithstanding our efforts to be as inclusive as possible, the literature search was not conducted as a formal systematic review. A number of PubMed research strings were proposed centrally for the whole project; however, the literature databases and some selection criteria for included papers were chosen by the authors of each review, who additionally added papers from personal knowledge or other papers reference lists. Second, the original Pepe and colleague's framework was developed to screen cancer in asymptomatic populations and has been further adapted to the early diagnosis of AD in symptomatic memory clinic patients. Future developments of the field of AD and in drug development may require and allow to extend this framework to asymptomatic preclinical patients. Therefore, the nature of this whole effort is liable to change in the near future, but still necessary to proceed in a fruitful way to improve clinical practice in the Alzheimer’s field.

Acknowledgements

The Geneva Task Force for the Roadmap of Alzheimer’s Biomarkers includes the participants to a workshop held in Geneva on December 8-9, 2014. The P.I. of the Geneva Roadmap effort is Giovanni B Frisoni, with Bengt Winblad and Clifford R Jack Jr as co-PIs. The task force includes experts in biomarker development from the oncology community; experts on diagnostic AD biomarkers from Switzerland and Europe; representatives of pertinent scientific societies (Federation of European Societies of Neuropsychology - FENS, European Society of Neuroradiology - ESNR, International Foundation of Clinical Chemistry and
Laboratory Medicine - IFCCLM, European Association of Nuclear Medicine - EANM, and Swiss Federation of Clinical Neuro Societies - SFCNS); representatives of patient advocates, bioethicists and regulatory agencies, and early career researchers. The Geneva Task Force has been endorsed by the EADC – European Alzheimer’s Disease Consortium. The workshop was funded thanks to a competitive grant by the Swiss National Science Foundation and unrestricted grants from: Alzheimer Forum Switzerland, Association pour la Recherche sur Alzheimer, Genève; Piramal, Eli Lilly & Company, General Electric, Guerbet, TEVA Pharma; Academie Suisse de Sciences Médicales, Vifor Pharma Switzerland., Novartis, Siemens, and IXICO. The Alzheimer’s Association hosted the first follow-up meeting of the initiative at the 2015 AAIC congress in Washington. We acknowledge the help from Margherita Mauri and Daria Gennaro (IRCCS Fatebenefratelli, Brescia, Italy) and Agnese Picco (Università di Genova, Genova, Italy) who took care of the logistics of the workshop. The following scientific societies took part to the Geneva Workshop for the Roadmap of Alzheimer’s Biomarkers on December 8-9 2014. Flavio Nobili was delegate from the European Association of Nuclear Medicine (EANM) Neuroimaging Committee. Kaj Blennow was delegate and Chair of the International Federation of Clinical Chemistry and Laboratory Medicine Working Group for CSF proteins (IFCC WG-CSF). Frederik Barkhof was delegate from the European Society of Neuroradiology (ESNR). Stefano Cappa was delegate and Chair of the Federation of European Societies of Neuropsychology (FENS). Urs Mosimann was delegate from the Swiss Federation of Clinical Neuro Societies (SFCNS). The content of this paper represents the opinion of the individual authors and is not necessarily endorsed by the scientific societies which took part to the Geneva Workshop for the Roadmap of Alzheimer’s Biomarkers, except the ESNR which formally endorsed it on September 5th 2015.
REFERENCES


Clinical validity of MTA in MCI


47. Konrad C, Ukas T, Nebel C, Arlt V, Toga AW, Narr KL. Defining the human hippocampus in cerebral


114. Leeuw F-E de, Korf E, Barkhof F, Scheltens P. White Matter Lesions Are Associated With Progression of
Clinical validity of MTA in MCI


SUPPLEMENTARY DATA

Supplementary Table 1: Search criteria.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Aim</th>
<th>Aim-specific key words string</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1: preclinical exploratory studies</td>
<td><strong>Primary aim:</strong> To identify leads for potentially useful biomarkers and prioritize identified leads.</td>
<td>NO STRINGS USED</td>
</tr>
<tr>
<td></td>
<td><strong>Primary aim:</strong> To estimate TPR and FPR or ROC curve for the assay and to assess its ability to distinguish subjects with and without disease.</td>
<td>(“accuracy” OR “sensitivity” OR “specificity” OR “ROC” OR “predictive value”) AND (a) AND (b) AND (d).</td>
</tr>
<tr>
<td></td>
<td><strong>Secondary aim 1:</strong> To optimize procedures for performing the assay and to assess the reproducibility of the assay within and between laboratories.</td>
<td>(“standardization” OR “visual” OR “measure” OR “assessment” OR “reading” OR “quantification” AND (“reproducibility” OR “reliability” OR “agreement”) AND (a) AND (d).</td>
</tr>
<tr>
<td></td>
<td><strong>Secondary aim 2:</strong> To determine the relationship between biomarker tissue measurements made on tissue (phase 1) and the biomarker measurements made on the noninvasive clinical specimen (phase 2).</td>
<td>(“autopsy” OR “autoptic” OR “patholog*” OR “neuropatholog*” OR “istopathol*”) AND MRI AND (a) AND (d).</td>
</tr>
<tr>
<td></td>
<td><strong>Secondary aim 3:</strong> To assess factors (e.g. sex, age, etc.), associated with biomarker status or level in control subjects. If such factors affect the biomarker, thresholds for test positivity may need to be defined separately for target subpopulations.</td>
<td>- (“APOE” OR “Apolipoprotein E”) AND (b) AND (d) - (“vascular risk factors” OR “white matter hyperintensities) AND (b) AND (d)</td>
</tr>
<tr>
<td></td>
<td><strong>Secondary aim 4:</strong> To assess factors associated with biomarker status or level in diseased subjects—in particular, disease characteristics.</td>
<td>- (“APOE” OR “Apolipoprotein E”) AND (a) AND (d) - (“vascular risk factors” OR “white matter hyperintensities) AND (a) AND (d) - (“early-onset” OR “late-onset”) AND (a) AND (d)</td>
</tr>
<tr>
<td>Phase 2: clinical assay development for clinical disease</td>
<td><strong>Primary aim:</strong> To evaluate the capacity of biomarkers to detect pre-clinical disease and define criteria for a positive biomarker test in preparation for phase 4.</td>
<td>- (“follow-up” OR “followup” OR “conversion” OR “progression” OR “decline” OR “predict”) AND MRI AND (“visual” OR “rating”) AND (c) AND (d) - (“cut-off” OR “cut-point”) AND (d)</td>
</tr>
<tr>
<td></td>
<td><strong>Secondary aim 1:</strong> To explore the impact of covariates on the discriminatory abilities of the biomarker before clinical diagnosis.</td>
<td>- (“APOE” OR “Apolipoprotein E”) AND (c) AND (d) - (“amnestic” OR “non-amnestic”) AND (c) AND (d)</td>
</tr>
<tr>
<td>Phase</td>
<td>Aim</td>
<td>Aim-specific key words string</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>------------------------------</td>
</tr>
<tr>
<td><strong>Chapter 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Phase 3: Prospective repository studies</strong></td>
<td>Secondary aim 2: To compare markers with a view to selecting those that are most promising.</td>
<td>(“follow-up” OR “followup” OR “conversion” OR “progression” OR “decline” OR “predict*”) AND (“combinat*” OR “associat*” OR “compar*”) AND MRI AND (a) AND (c)</td>
</tr>
<tr>
<td></td>
<td>Secondary aim 3: To develop algorithms for positivity based on combinations of markers.</td>
<td>(“follow-up” OR “followup” OR “conversion” OR “progression” OR “decline” OR “predict*”) AND (“combinat*” OR “associat*” OR “compar*”) AND MRI AND (a) AND (c)</td>
</tr>
<tr>
<td></td>
<td>Secondary aim 4: To determine a biomarker testing interval for phase 4 if repeated testing is of interest.</td>
<td>(“atrophy rates”) AND (a) AND (c) AND (d)</td>
</tr>
<tr>
<td><strong>Phase 4: Prospective Diagnostic Studies</strong></td>
<td>Primary aim: To determine the operating characteristics of the biomarker-based test in a relevant population by determining the detection rate and the false referral rate. Studies at this stage involve testing people and lead to diagnosis and treatment.</td>
<td>(“diagnosis” OR “treatment”) AND (a) AND (c) AND (d)</td>
</tr>
<tr>
<td></td>
<td>Secondary aim 1: To describe the characteristics of disease detected by the biomarker test—in particular, with regard to the potential benefit incurred by early detection.</td>
<td>(“clinical diagnosis” OR “treatment” OR “memory clinic”) AND (“benefit*” OR “outcome” OR “improve”) AND (a) AND (c) AND (d)</td>
</tr>
<tr>
<td></td>
<td>Secondary aim 2: To assess the practical feasibility of implementing the case finding program and compliance of test-positive subjects with work-up and treatment recommendations.</td>
<td>(“clinical diagnosis” OR “treatment” OR “memory clinic”) AND (“benefit*” OR “compliance” OR “mortality” OR “morbidity” OR “Qol” OR “quality of life”) AND (a) AND (d)</td>
</tr>
<tr>
<td></td>
<td>Secondary aim 3: To make preliminary assessments of the effects of biomarker testing on costs and mortality associated with the disease.</td>
<td>(“clinical diagnosis” OR “treatment” OR “memory clinic”) AND (“benefit*” OR “compliance” OR “mortality” OR “morbidity” OR “Qol” OR “quality of life” OR “cost”) AND (a) AND (d)</td>
</tr>
<tr>
<td></td>
<td>Secondary aim 4: To monitor disease occurring clinically but not detected by the biomarker testing protocol.</td>
<td>(“clinical diagnosis” OR “memory clinic” OR “criteria” OR “recommendation”*) AND (“accuracy” OR “sensitivity” OR “specificity” OR “ROC” OR “predictive value” OR “concordance” OR “confirm” OR “negative detection rate” OR “negative referral rate” OR “false negative rate”) AND (a) AND (d)</td>
</tr>
<tr>
<td><strong>Phase 5: Disease Control Studies</strong></td>
<td>Primary aim: To estimate the reductions in disease-associated mortality, morbidity, and disability afforded by biomarker testing.</td>
<td>(“diagnosis” OR “detection”) AND (“benefit*” OR “compliance” OR “mortality” OR “morbidity” OR “Qol” OR “quality of life” OR “financial impact” OR “cost” OR “effectiveness”) AND (a) AND (d)</td>
</tr>
<tr>
<td>Phase</td>
<td>Aim</td>
<td>Aim-specific key words string</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Phase 5: Disease Control Studies</td>
<td><strong>Secondary aim 1:</strong> To obtain information about the costs of biomarker testing and treatment and the cost per life saved or per quality-adjusted life year</td>
<td>(&quot;diagnosis&quot; OR &quot;detection&quot;) AND (&quot;benefit*&quot; OR &quot;compliance&quot; OR &quot;mortality&quot; OR &quot;morbidity&quot; OR &quot;QoL&quot; OR &quot;quality of life&quot; OR &quot;outcome&quot;) AND (&quot;financial impact&quot; OR &quot;cost*&quot; OR &quot;effectiveness&quot;) AND (a) AND (d)</td>
</tr>
<tr>
<td></td>
<td><strong>Secondary aim 2:</strong> To evaluate compliance with testing and work-up in a diverse range of settings.</td>
<td>(&quot;diagnosis&quot; OR &quot;treatment&quot;) AND (&quot;benefit*&quot; OR &quot;compliance&quot; OR &quot;mortality&quot; OR &quot;morbidity&quot; OR &quot;QoL&quot; OR &quot;quality of life&quot;) AND (&quot;primary care&quot; OR &quot;second level&quot; OR &quot;third level&quot;) AND &quot;memory clinic&quot; AND &quot;cost*&quot; AND (a) AND (d)</td>
</tr>
<tr>
<td></td>
<td><strong>Secondary aim 3:</strong> To compare different biomarker testing protocols and/or or to compare different approaches to treating test positive subjects in regard to effects on mortality and costs.</td>
<td>&quot;diagnosis&quot; OR &quot;treatment&quot;) AND (&quot;protocol&quot; OR &quot;recommendation*&quot; OR &quot;criteria&quot;) AND (&quot;benefit*&quot; OR &quot;compliance&quot; OR &quot;mortality&quot; OR &quot;morbidity&quot; OR &quot;QoL&quot; OR &quot;quality of life&quot;) AND (&quot;financial impact&quot; OR &quot;cost*&quot; OR &quot;effectiveness&quot;) AND (a) AND (d)</td>
</tr>
</tbody>
</table>

(a) ("Alzheimer")
(b) ("Healthy controls" OR "Cognitively normal" OR "controls" OR "normal")
(c) ("MCI" OR "mild cognitive impairment" OR "prodromal")
(d) ("MTA" OR "medial temporal" OR "hippocamp")