Neurobiology of Aging 52 (2017) 153–166

Contents lists available at ScienceDirect

Neurobiology of Aging

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

Roadmap to Alzheimer's Biomarkers in the Clinic

Clinical validity of delayed recall tests as a gateway biomarker for Alzheimer's disease in the context of a structured 5-phase development framework

Chiara Cerami [a,b,c,*], Bruno Dubois [d], Marina Boccardi [e,f], Andreas U. Monsch [g], Jean Francois Demonet [h], Stefano F. Cappa [b,i], for the Geneva Task Force for the Roadmap of Alzheimer's Biomarkers

[a] Vita-Salute San Raffaele University, Milan, Italy
[b] Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy
[c] Clinical Neuroscience Department, San Raffaele Turro Hospital, Milan, Italy
[d] Dementia Research Center and Department of Neurology, Salpêtrière University Hospital, Paris University, Paris, France
[e] Laboratory of Neuroimaging & Alzheimer's Epidemiology, IRCCS Centro S. Giovanni di Dio Fatebenefratelli, Brescia, Italy
[f] LANVIE (Laboratory of Neuroimaging of Aging) – Department of Psychiatry, University of Geneva, Geneva, Switzerland
[g] Memory Clinic, University Center for Medicine of Aging, Felix Platter Hospital, Basel, Switzerland
[h] Leenaards Memory Centre – CHUV, Lausanne University, Lausanne, Switzerland
[i] NeTS Center, Istituto Universitario di Studi Superiori, Pavia, Italy

ARTICLE INFO

Article history:
Received 23 November 2015
Received in revised form 5 March 2016
Accepted 22 March 2016

Keywords:
Episodic memory
Delayed recall
Cued recall tasks
Free word list tests
Biomarker development
Early diagnosis
Biomarker-based diagnosis
Alzheimer's disease
Mild cognitive impairment
5-Phases

ABSTRACT

Although Alzheimer's disease criteria promote the use of biomarkers, their maturity in clinical routine still needs to be assessed. In the light of the oncology framework, we conducted a literature review on measures used to assess delayed recall impairment due to medial temporal lobe dysfunction (i.e., free and cued word list recall tests). Ample evidence is available for phases 1 (rationale for use), 2 (discriminative ability), and 3 (early detection ability) for many of the tests in routine use. Evidence about phase 4 (performance in real world) and phase 5 (quantify impact and costs) is yet to come. Administration procedures have been standardized and cutoff scores are well validated in large Alzheimer's disease and mild cognitive impaired series. Some aspects (e.g., different task formats), however, hamper the comparability of results among different populations and the reproducibility between laboratories. No definite guideline for their use can thus be proposed at the moment. Accordingly, the maturity of such markers is not yet sufficient and requires future investigation to promote the proper use of memory measures in clinical settings.

1. Introduction

The correct identification of Alzheimer's disease (AD) represents a challenge for clinicians especially at the predementia stages. Recent developments in this area of research are specifically devoted to support the early identification of AD pathology in vivo and to the application of reliable biomarkers of disease in clinical settings. The need of more accurate early and differential diagnosis, indeed, prompted the development of new research criteria supporting the use of biomarkers in order to recognize AD in prodromal or even preclinical stages (Albert et al., 2011; Dubois et al., 2007, 2010, 2014; Jack et al., 2011; McKhann et al., 2011; Sperling et al., 2011).

Since the introduction of new clinical criteria, AD research has been mainly focused on the application of distinctive topographic (e.g., 18F-fluorodeoxyglucose positron emission tomography [FDG-PET]; magnetic resonance imaging [MRI]) and pathophysiological (e.g., amyloid-PET or cerebrospinal fluid [CSF]) biomarkers in clinical research. Several methodological problems have emerged about their implementation in clinical routine. Neither a definite diagnostic algorithm nor clear quantitative measures for the use of biomarkers in patients suspected for AD have been clearly outlined.

* Corresponding author at: Division of Neuroscience, Università Vita-Salute San Raffaele and San Raffaele Scientific Institute, Via Olgettina, 60, 20132 Milan, Italy. Tel.: +39 0226435760; fax: +39 0226435738.
E-mail address: cerami.chiara@hsr.it (C. Cerami).

0197-4580/$ – see front matter © 2016 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neurobiolaging.2016.03.034

Thus, different biomarkers and measurement tools are used by researchers according to their availability in community-based and clinical-based studies, obviously resulting in heterogeneous findings.

To overcome similar problems in the field of oncology, Pepe et al. (2001) suggested systematizing the investigation of cancer biomarkers on the basis of the methodology used for pharmacological investigation. Since a formal structure to guide the process of AD biomarker development was lacking so far, an effort has recently been launched to adopt the previously mentioned oncology model to effectively systematize the available scientific evidence for the use of biomarkers in AD diagnosis, with the aim to promote rigor in their application to clinical settings. The present study focuses on the analysis of the maturity of the assessment of episodic memory by means of delayed recall tasks in the framework of this model.

Whether cognitive testing can properly be considered as a marker of disease is an open question. The concept is fully compatible with a broad definition, such as "…a characteristic which can be objectively measured and evaluated as an indicator of a physiological as well as a pathological process or pharmacological response to a therapeutic intervention" (Jain, 2012). In the field of dementia, however, neuropsychological testing is generally considered separately from imaging and CSF biomarkers (e.g., Ewers et al., 2012). Within the MCI and/or prodromal AD context, neuropsychological testing is usually considered as a sort of "gatekeeper" for the application of biomarkers, as the presence of objective impairment is required by the diagnostic criteria to separate these conditions from subjective complaints. In any case, given the central role of cognitive assessment, it is surprising that studies assessing the sensitivity and specificity of neuropsychological tests for diagnosis of AD or the predictive value of the progression from mild cognitive impairment (MCI) to AD are relatively scant and heterogeneous in methodology. In particular, the presence of an early and significant objective deficit of memory and learning has been considered as the main criterion supporting the diagnosis of typical AD condition for decades (APA, 2000; McKhann et al., 1984; Diagnostic and Statistical Manual of Mental Disorders [DSM-IV]), and an impaired memory performance in comparison to a healthy control group is considered as the best cognitive predictor of the development of future AD (Elias et al., 2000; Sarazin et al., 2007; Small et al., 2000). According to Braak and Braak staging of AD (Braak and Braak, 1991), the earliest neuropathological changes involve the entorhinal cortex and the hippocampal structures, disconnecting the Papez circuit and selectively affecting the ability to consolidate new information. This results in an impaired performance on delayed recall memory tasks (Squire et al., 2004).

Moreover, in the context of recent diagnostic criteria for the very early and/or prodromal stages of AD, the performance in specific memory tests has gained a special status. The presence of impaired memory performance on objective testing is required for the definition of MCI (hence "amnesic MCI") or prodromal AD status. If a subject presents with subjective memory complaints, in the absence of objective memory dysfunction, he or she is defined as having a "subjective memory impairment", an inconsistently defined construct (Abdulrab and Heun, 2008), which in clinical practice does not usually lead to further investigation but only to reassurance and long-term follow-up (Berrios et al., 2000; Jessen et al., 2014).

Since cognitive assessment remains a critical component of diagnosis in clinical and research settings, it is vital to determine the capacity of specific memory measures in detecting early disease changes and predicting disease progression, in order to recommend tests having the greatest predictive accuracy. The tasks assessing memory ability may differ with respect to the modality of stimulus presentation, the testing procedure, the structure of the to-be-remembered information or the presence of facilitators to improve encoding and recall (cued paradigms).

Many memory measures are used in clinical and research settings. The most common are measures of delayed free recall of word lists. A variety of standardized verbal learning tasks, such as the Rey Auditory Verbal Learning Test (RAVLT; Rey, 1941), the California Verbal Learning Test (CVLT; Delis et al., 1987, 2000; CVLT-II, Delis et al., 2000), and the Hopkins Verbal Learning Test (Brandt and Benedict, 2001), are commonly used for clinical diagnosis and disease monitoring of AD dementia and mild cognitive impaired (MCI) patients. In addition, some word list tasks, often quicker to administer, that is, with fewer words to learn (10-word list) and less learning trials (2 or 3), are part of neuropsychological batteries, such as the Wechsler Memory battery (http://www.pearsonclinical.com/), Alzheimer's Disease Assessment Scale-cognition (ADAS-cog) (Mohs et al., 1983), Consortium to establish a registry for Alzheimer's disease (CERAD) (Morris et al., 1988), or Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005). Direct measurement of the number of items recalled at the learning trials (i.e., the immediate recall) or after a time delay (i.e., the delayed recall), as well as the difference between immediate and delayed recall (i.e., the savings) are the main measures obtained from the word list free recall tasks used to evaluate patient performances.

Other neuropsychological tests for the assessment of verbal long-term memory are logical memory (short story recall) and associative learning tasks (e.g., in the Wechsler Memory battery). Both immediate and delayed scores are obtained from story recall tests. In this case, the processing of a coherent stream of information typically benefits from intrinsic semantic organization of the material, while the ability to self-generate organizational strategies is required for free recall of word list tasks—with the exception of the CVLT (Randolph et al., 1994). Other tests assess nonverbal memory, such as, the delayed recall of Rey figure or the Cambridge Neuropsychological Test Automated Battery (http://www.cambridgecognition.com/), testing the ability to form and remember associations between the attributes of an experience. These tasks are sensitive to the functional integrity of the medial temporal lobe (Eichenbaum and Cohen, 2001; Meltzer and Constable, 2005), as performance largely depends on the ability to bind and encode arbitrary information.

Though the amnestic syndrome is typical of the onset of AD, impairments on delayed recall tasks may also be present in non-AD neurodegenerative diseases (e.g., behavioral variant of frontotemporal dementia; Hornberger et al., 2010) as well as in other conditions (e.g., vascular MCI, depression), characterized by cognitive deficits that can affect the learning phase or the encoding and recall processes (Dickerson and Eichenbaum, 2010). In AD, the amnestic profile is typically characterized by poor learning and rapid forgetting over relatively short periods, reflecting damage to the hippocampal structures (Squire et al., 2004). To accurately detect memory impairment of the hippocampal type, the design of the test used to assess memory ability, and in particular of the learning phase of the task, is crucial. Free recall is dependent upon intact attentional processing, registration, and retrieval mechanisms. An effective retrieval of the to-be-remembered information can be better achieved with an "encoding specificity", based on the correspondence of the semantic cue during encoding and retrieval. The use of this learning technique produces efficient results in healthy subjects (Ivnik et al., 1997). In agreement to this evidence, the specific neuropsychological features of the memory impairment of AD has been stressed in the International Working Group research diagnostic criteria (Dubois et al., 2010, 2014) and defined as follows: "objective evidence of significantly impaired episodic memory on testing, generally consisting of a recall deficit that does not improve significantly with cueing or recognition testing after

effective encoding of information has been controlled". Test paradigms providing controlled learning conditions (encoding) and distinguishing between the performance on free recall and the effects of cueing nonrecalled items (i.e., the Grober-Buschke paradigm, Buschke, 1984) can distinguish between the encoding and storage impairments typical of AD condition and low memory performances resulting from reduced attentional resources or ineffective recall strategies (Grober and Buschke, 1987). The Free and Cued Selective Reminding Test (FCSRT; Buschke, 1984; Grober and Buschke, 1987) has thus gradually entered standard neuropsychological batteries, partially replacing traditional free recall word list tasks (e.g., RAVLT). Three FCSRT measures are mainly used to detect AD: free recall, total recall (the sum of free and cued recall), and cue efficiency (the ratio of cued recall successes to the number of cued recall attempts). The high negative predictive values of free and total recall FCSRT (Auriacombe et al., 2010; Dierckx et al., 2009) support the utility of these measures for ruling out dementia in both primary and secondary care settings. Low positive predictive values (Auriacombe et al., 2010; Dierckx et al., 2009) may limit their usefulness. This effect is, however, more restricted among patients attending memory clinics (Auriacombe et al., 2010), where disease prevalence is higher.

For the considerations summarized previously, among the large number of measures used in clinical practice to assess episodic memory ability, the present discussion is limited to delayed free and cued recall tests for several reasons. In particular, they are the most widely used tests in clinical and research settings (Maruta et al., 2011). Moreover, there is extensive evidence that they may represent the most sensitive measures of memory decline and that they may be relatively specific for medial temporal lobe dysfunction, that is, to the earliest pathological changes in AD (Braak and Braak, 1991).

2. Methods

2.1. Target

This study was performed with reference to the model imported from the oncology field (Pepe et al., 2001) and adapted to the field of dementia, specifically to the aim of performing the differential diagnosis of AD at the prodromal stage (Boccardi et al., 2017; Frisoni et al., 2017). The terms of this framework are summarized in this section. The target populations are sporadic AD dementia patients and MCI subjects, as defined in the following section. Although studies with pathological confirmation should be considered as the gold standard to evaluate accuracy of memory measures in AD, the criteria chosen to select articles appropriate for this review also included studies in which AD was diagnosed according to the clinical criteria (at baseline or at follow-up) available at the time of the study. We considered progression to dementia at a minimum of 2-year follow-up as the gold standard for the clinical diagnosis of AD in MCI patients.

2.2. Glossary

2.2.1. Alzheimer's disease

Alzheimer pathology consists of brain amyloidosis and neurodegeneration, usually with mediotemporal and temporoparietal distribution. The term is thus independent of the clinical manifestation of the disease.

2.2.2. AD dementia

This clinical syndrome features both cognitive impairments and functional disability as defined by McKhann (1984). Notably, not all cases of AD dementia have AD pathology due to imperfect accuracy of purely clinical criteria.

2.2.3. Mild cognitive impairment (MCI)

The population with cognitive impairment and no functional disability, including, besides cases with prodromal AD, cases with no neurodegenerative disorders (about 35%–40%), and non-AD neurodegeneration (about 10%–15%) (Bennett et al., 2002; Jack et al., 2008; Rowe et al., 2010). The MCI cases with biomarker positivity are defined as prodromal AD in the clinical criteria by Dubois et al. (2010). The focus of our review is the diagnosis of AD at the MCI stage.

2.2.4. Non-AD neurodegenerative diseases

The neurodegenerative disorders entering differential diagnosis considered in this context are suspected nonamyloid pathology, hippocampal sclerosis, frontotemporal lobar degeneration, tauopathies (e.g., progressive supranuclear palsy and corticobasal degeneration), dementia with Lewy body, and other alpha-synucleinopathies such as multiple system atrophy.

2.3. Conceptual framework

Here, we summarize the results of the translation of the conceptual framework from oncology to the AD field described elsewhere (Boccardi et al., 2017; Frisoni et al., 2017), as steps to be covered in the systematic development of disease markers for a proper use in the clinical routine for the diagnosis of AD. The present review focuses on the maturity of selected neuropsychological tasks for the assessment of delayed recall relative to each step. In reference to the model imported from the oncology field (Pepe et al., 2001), we generally referred to the neuropsychological markers and the other AD disease markers as "biomarkers". All aims and subaims are specifically addressed and qualified as "Fully achieved," "Partly achieved," "Preliminary evidences," "Not achieved," or "Not applicable" based on the available evidence. The evaluation terms and assessments are reported in detail in Table 1.

2.3.1. Phase 1

Phase 1 studies aim to identify the rationale of the biomarker, based on pathology findings, and consist of preclinical exploratory studies.

2.3.2. Phase 2

Phase 2 studies aim to define the ability of the biomarker to discriminate patients from healthy subjects. It focuses on defining the clinical essay allowing reliable discrimination and in the assessment of the possible differential effects of covariates that may influence the thresholds for positivity.

2.3.3. Phase 3

Phase 3 studies aim to define the ability of the biomarker to detect the disease in its early phase, namely, in the prodromal stage as set for the present exercise. They consist of prospective longitudinal repository studies aimed to fine-tune the exact thresholds for positivity and to compare the usefulness of the biomarker compared to or differentially combined with the other available biomarkers.

2.3.4. Phase 4

Phase 4 studies aim to estimate the accuracy and usefulness of the biomarker-based diagnosis in real-world patients. It consists of prospective diagnostic studies and on subjects also undergoing treatment following the biomarker-based diagnosis. It assesses the benefit of the biomarker-based detection in terms of early diagnosis, feasibility, and compliance and provides preliminary evidence about mortality, costs, and undetected cases.

Table 1
The table reports evaluation terms and assessments of aims and subaims of each phase

Phase	General aim	Aim	Progress	Evidence in AD
1—Preclinical exploratory studies	Identify the rational of the BM, based on pathology	To identify leads for potentially useful biomarkers and prioritize identified leads	Fully achieved	Adequate scientific evidence is available supporting the fact that (1) medial temporal lobe structures are early affected in AD; and (2) the extent of cognitive impairment parallels the severity of cortical neurofibrillary tangles pathology.
2—Clinical assay development for clinical disease	Define the ability of the BM to discriminate patients from controls	Primary: To estimate true positive rate (TPR) and false positive rate (FPR) or receiver operating characteristic (ROC) curve for the assay and to assess its ability to distinguish subjects with and without disease	Fully achieved	Good/excellent value for the delayed recall tasks emerged in the discrimination of AD patients from cognitively healthy subjects.
		Secondary 1: To optimize procedures for performing the assay and to assess the reproducibility of the assay within and between laboratories	Partly achieved	Neither sufficient evidence is available nor has direct effort been employed to systemically compare different standardized versions of delayed recall tasks.
		Secondary 2: 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)	Not applicable	—
		Secondary 3: 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.	Partly achieved	The weight of demographic variables on the performances at the delayed recall tasks has not been definitely evaluated in all the reference subpopulations. Apart from educational level, neuropsychological task performances are not controlled for other possible factors influencing cognitive performance at baseline, such as socioeconomic status and other factors possibly affecting cognitive reserve.
		Secondary 4: To assess factors associated with biomarker status or level in diseased subjects—in particular, disease characteristics.	Partly achieved	Neuropsychological performances on delayed memory tasks correlate with age at onset. ApoE ε4 allele presence is not directly associated with cognitive status. Episodic memory performances may not be sensitive to atypical AD. Additional studies assessing factors associated with biomarker status in AD subjects are required.
3—Retrospective longitudinal repository studies	Define the ability of the BM to detect the disease in its early phase	Primary 1: To evaluate the capacity of the biomarker to detect the earliest disease stages	Partly achieved	Scientific evidence is generally supporting good predictive values for delayed recall tasks. Some comparative studies among different tasks are available, with sparse and contradictory findings. There is at the moment insufficient evidence for the superiority of one measure compared to the others, as well as of cued paradigms compared to traditional measures of free recall.
		Primary 2: To define criteria for a biomarker-positive test in preparation for phase 4.	Partly achieved	It seems essential to address the impact of different cutoff points for delayed recall tests in future large longitudinal studies on MCI subjects.
		Secondary 1: To explore the impact of covariates on the discriminatory abilities of the biomarker before clinical diagnosis.	Partly achieved	Performances on episodic memory tasks are associated with demographic variables and cognitive reserve. Several factors that may affect the effectiveness of covariates on test scores in discriminating at-risk patients from control subjects (e.g., occupational history, socioeconomic status and contextual factors such as time of the day or metabolic status) have not been systematically investigated.
		Secondary 2: To compare markers with a view to selecting those that are most promising	Not applicable	—
		Secondary 3: To develop algorithms for positivity based on combinations of markers.	Partly achieved	The literature review is generally supporting improved predictive values for the combination of delayed recall tasks with biological and imaging biomarkers, with a particular impact of the latter. The evidence is, however, scanty and mainly focused on the combination of RAVLT performance with imaging or CSF biomarkers.
		Secondary 4: To determine a biomarker testing interval for phase 4 if repeated testing is of interest.	Not achieved	There are several problematic issues in this area. First, the delayed recall tests tend to go to floor very fast, and are thus of limited usefulness to study disease progression. Repeated testing is of interest only in the very early and prodromal disease stages, when subtle cognitive changes may be detected only with repeated measurements. Second, in the very early and prodromal stage an important issue is availability of parallel forms to account for learning effects.
4—Prospective case finding studies	Quantify the BM accuracy and usefulness in patients	Primary: To determine the operating characteristics of the biomarker-based test in a relevant population by determining	Not achieved	—

	diagnosed and treated based on the BM	the detection rate and the false referral rate. Studies at this stage involve testing people and lead to diagnosis and treatment.	Not achieved	—
		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.	Not achieved	—
		Secondary 2: To assess the practical feasibility of implementing the case finding program and compliance of test-positive subjects with workup and treatment recommendations.	Not achieved	—
		Secondary 3: To make preliminary assessments of the effects of biomarker testing on costs and mortality associated with the disease.	Not achieved	—
		Secondary 4: To monitor disease occurring clinically but not detected by the biomarker testing protocol.	Not achieved	—
5—Disease control studies	Quantify the impact of the BM-based diagnosis on clinically meaningful outcomes and costs	Primary: To estimate the reductions in disease-associated mortality, morbidity, and disability afforded by biomarker testing.	Not achieved	—
		Secondary 1: To obtain information about the costs of biomarker testing and treatment and the cost per life saved or per quality-adjusted life year	Not achieved	—
		Secondary 2: To evaluate compliance with testing and workup in a diverse range of settings.	Not achieved	—
		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.	Not achieved	—

Key: AD, Alzheimer's disease; BM, biomarker; CSF, cerebrospinal fluid; MCI, mild cognitive impairment; RAVLT, Rey Auditory Verbal Learning Test.

2.3.5. Phase 5

Phase 5 studies aim to exactly quantify the impact of the biomarker-based diagnosis on clinically meaningful outcomes and costs. It consists of case-control studies, assessing the reduction in mortality, morbidity, disability, and costs as allowed by the biomarker-based diagnosis. It also aims to compare the costs of different biomarkers testing protocols and settings.

2.4. Evidence evaluation

The primary and secondary aims described in Section 2.3 were assessed as Fully achieved, Partly achieved, Preliminary evidence, Not achieved or Not applicable.

2.4.1. Fully achieved

A primary or secondary aim was assessed as Fully Achieved when its corresponding scientific evidence is available and replicated in adequately powered samples in studies without major methodological faults.

2.4.2. Partly achieved

A primary or secondary aim was assessed as Partly Achieved when scientific evidence is available but not yet sufficiently replicated, or samples are not adequately powered, or other significant methodological limitations can be found in the available literature.

2.4.3. Preliminary evidence

A primary or secondary aim was assessed as Preliminary Evidence when only preliminary evidence is available.

2.4.4. Not achieved

A primary or secondary aim was assessed as Not Achieved when no evidence was found and no studies are known to be ongoing at present to cover this aim.

2.4.5. Not applicable

A primary or secondary aim was assessed as Not Applicable when no evidence can be yielded for the considered biomarker.

2.5. Articles search and selection

Articles were selected searching on the PubMed and Medline databases. Trying to apply the framework of Pepe et al. (2001) to the use of tests designed for the assessment of memory as a marker of the clinical presentation of typical AD, we identified studies targeting well-established candidates for the detection of long-term memory disorders in AD dementia patients or MCI subjects at risk for progression to AD dementia. Only articles including patients classified as MCI or AD according to validated clinical diagnostic and research criteria (Albert et al., 2011; Dubois et al., 2010, 2014; McKhann et al., 1984, 2011; Petersen, 2004; Sperling et al., 2011) were included.

The search was limited to the available evidence regarding free or cued word list recall tests. The reasons for the selection of this specific class of tasks are described in the Section 1. We do not aim to rule out other tests or conceptualizations for testing memory related to early medial temporal cortex and/or hippocampal structures damage in AD dementia with new experimental procedures (see the following section). However, evidence is still lacking to support the applicability of new promising cognitive tools to explore AD-related memory disorders. Short story recall (logical memory) and paired associates (associative memory) learning tasks have been excluded from this review due to the different neuropsychological constructs compared to the verbal word list tasks. Some of these neuropsychological measures

(e.g., paired associate learning and name-face learning) actually showed high sensitivity in sporadic and familial AD (see Lowndes and Savage, 2007 for a review on the usefulness of associative memory tasks). A promising paradigm, namely the visual short-term memory binding task (Parra et al., 2009), which showed very high sensitivity and specificity even in the preclinical phase of genetic AD cases (Parra et al., 2010), still needs additional evidence for the proper use in clinical settings. Other potentially interesting candidates for the early AD recognition are the Delayed Matching Sample (DSM48) visual recognition test (Barbeau et al., 2004) and the topographical memory test described by Hartley et al. (2007), both specifically designed to detect the early pathological involvement of the entorhinal-perirhinal cortex. Low performances on the DMS48 task may detect MCI subjects with a pattern of gray-matter loss, usually described in early AD (Barbeau et al., 2008).

Different strings were used to circumscribe the search specifically to the different aims and/or subaims (Supplementary Tables 1 and 2). When aims were notoriously achieved (e.g., evidence that low performance on delayed recall tasks is a hallmark for typical AD), a set of reference articles or, if available, a review was selected by the authors. In cases with lacking or scanty output, more inclusive strings were admitted to extend the research. Relevant evidence from previous personal knowledge and tracking from reference articles was included. Only articles published in English and up to May 2015 were included. The final selection of articles was based on relevance, as judged by the authors.

3. Current maturity of free and cued word list recall tests

3.1. Phase 1—preclinical exploratory studies

The aim of phase 1 is to identify leads for potentially useful biomarkers and to prioritize them. Evidence coming from large preclinical exploratory studies highly supports the rationale for the use of impaired performance at word list recall tests as marker of typical AD. Such tests assessing long-term recall of recently learned items are often tagged as "episodic memory tests," even though they capture only one important feature of the learning episode (i.e., what the subject remembers) (Tulving, 2002). In typical AD, early neuropathological changes occur in the perirhinal and entorhinal cortices, and in the hippocampal structures disconnecting the Papez circuit (Braak and Braak, 1991), and thus selectively affecting the ability to consolidate new information and resulting in low performances on delayed recall tasks (Squire et al., 2004). Medial temporal lobe structures have a fundamental role in the rapid formation of lasting memory information (Squire et al., 2004). Different studies have assessed the link between the neuropathological hallmarks of AD (i.e., amyloid plaques and neurofibrillary tangles) and the severity of antemortem cognitive impairment (see Nelson et al., 2012 for a review). Despite a large variability among clinicopathological correlation studies, mainly due to different study designs, research cohort characteristics, clinical-neuropsychological assessments or operational procedures for the anatomical-pathological analysis, studies with large autopsy series agreed in reporting that postmortem density of neocortical tau-positive neurofibrillary tangles is the pathological feature best correlating with antemortem cognitive status (Nelson et al., 2012). In particular, the neuroanatomical distribution of neurofibrillary tangles in the medial temporal lobe structures (Braak stages III–V) (Braak and Braak, 1991) is closely related to the presence of impaired consolidation and storage of information (Guillozet et al., 2003). In more advanced AD, as the degeneration and the deposition of neurofibrillary tangles spread to neocortical areas (Braak stages V–VI), other nonmemory cognitive domains (e.g., executive function, visuospatial abilities, or language) are affected (Nelson et al., 2012).

Atypical AD cases, however, do not clearly fit into the Braak staging scheme and may not be conform to the previously mentioned medial temporal lobe dysfunction paradigm. In particular, a hippocampal-sparing neuropathological pattern has been described in addition to the typical AD and the limbic-predominant subtypes (Murray et al., 2011). Consequently, atypical AD cases cannot be early detected by episodic memory tests. Thus, free and cued word list recall tests may be poorly useful in the diagnostic algorithm of such cases (see also the Section 3.2.4).

In conclusion, we can consider phase 1 as Fully Achieved. Adequate scientific evidence is indeed available supporting the fact that (1) medial temporal lobe structures are affected early in AD; and (2) the extent of cognitive impairment parallels the severity of cortical neurofibrillary tangles pathology.

3.2. Phase 2—clinical assay development for clinical disease

The aim of phase 2 is to define the ability of the biomarker to discriminate patients from controls. The primary aim of this phase is to estimate true positive rate and false positive rate or receiver operating characteristic curve for the assay and to assess its ability to distinguish subjects with and without the disease.

Several studies provide evidence of good and/or excellent values in the discrimination of AD dementia patients from healthy control subjects for many tasks used to assess episodic memory impairments (Buschke et al., 1997; Delgado et al., 2016; Ivanoiu et al., 2005; Lemos et al., 2015; Ricci et al., 2012; Sotaniemi et al., 2012; Vogel et al., 2007; Welsh et al., 1991, 1992; Wolfsgruber et al., 2014). This is particularly true for the delayed recall score of free and cued word list tasks.

The delayed recall score of the RAVLT task (i.e., learning a 15-word list presented in 5 trials) outperformed (sensitivity 100% and specificity 98.3%–100%) the immediate recall (i.e., the sum of learning scores of trials 1–5) in discriminating AD dementia from normal individuals both in the Italian and the Australian samples of the study by Ricci et al. (2012). Similarly, a large Finnish study (Sotaniemi et al., 2012) on 171 AD dementia patients and 315 cognitively normal individuals using the memory task of the CERAD (i.e., learning a 10-word list presented in 3 trials), a short neuropsychological battery developed in the United States, showed that the delayed recall is the most efficient subtest able to discriminate patients from controls (sensitivity 94% and specificity 85%). This result was further confirmed by Wolfsgruber et al. (2014) who reported excellent cross-sectional diagnostic accuracy for both delayed (area under the curve [AUC] 0.943) and immediate recall (AUC 0.936).

The use of controlled learning conditions and of cues in the recall phase also demonstrated very high diagnostic value for distinguishing AD dementia patients from healthy controls (Buschke et al., 1997; Delgado et al., 2016; Ivanoiu et al., 2005; Lemos et al., 2015; Vogel et al., 2007). Lemos et al. (2015) showed that both the total immediate and delayed free recall measures of the FCSRT task had excellent accuracy for AD discrimination (Lemos et al., 2015). Direct comparison of these 2 measures provided evidence for a better sensitivity of the total free delayed recall (96%) compared to the total free immediate recall (94%) (Lemos et al., 2015). The latter measures showed however a higher specificity value (99% vs. 97%) (Lemos et al., 2015). In addition, Delgado et al. (2016) showed that both the "word" and the "picture" versions of the FCSRT have high accuracy (AUC $>$0.9) in distinguishing mild AD from healthy subjects. Although visual cues were easier to recall than verbal cues, resulting in greater scores particularly at the total recall measure, "word" and "picture" performances were well correlated with each other (Delgado et al., 2016).

Direct comparisons between free and cued recall tasks, however, do not provide definitive evidence for a better discrimination validity of one class of tasks with respect to the other. In particular, Vogel et al. (2007) showed high sensitivity (88.6%–91.4%) and specificity (96.4%) for the immediate and delayed cued subtests of the Grober-Buschke paradigm, that is, values comparable with the immediate recall of the ADAS-cog (sensitivity 88.6% and sensitivity 96.4%). The highest sensitivity (100%) was reported for the delayed recall of ADAS-cog (Vogel et al., 2007). Ivanoiu et al. (2005) reported similar findings, showing that the total delayed recall of the cued paradigm RI-48 and the delayed recall of CERAD word list are both sensitive measures to discriminate AD from healthy subjects. In contrast, the savings measure is clearly less sensitive than both delayed recall and immediate recall taken in isolation (Ivanoiu et al., 2005).

In conclusion, sufficient scientific evidence is available to consider phase 2 main aim as Fully Achieved. No data support at present the superiority of a single free or cued recall word list task in the ability to discriminate AD dementia patients from control subjects. Good and/or excellent discriminating value emerged for all the delayed recall subtests.

3.2.1. Phase 2, secondary aim 1

The first secondary aim of phase 2 is to optimize procedures for performing the assay and to assess the reproducibility of the assay within and between laboratories.

Limitations are present in current literature on the procedures for the administration of free and cued word list tasks, whose reliability is mandatory to ensure reproducibility across different centers and experimenters. Despite the fact that all the neuropsychological measures object of this review are available in standardized versions for the reference populations, different administration procedures and formats exist, preventing the direct comparability of results among samples.

Different formats of the same task may result in dissimilar performances (Crane et al., 2012). In particular, some tests seem to be more prone to this effect (e.g., the 2 RAVLT word lists used in Alzheimer's Disease Neuroimaging Initiative neuropsychological battery are not equivalent to each other and list 2 is systematically harder than list 1, while the 3 versions of the ADAS-cog word list are much more similar to each other) (Crane et al., 2012).

Several versions of the Grober-Buschke paradigm have been developed for the clinical use (Auriacombe et al., 2010; Buschke, 1984; Buschke et al., 1997; Dierckx et al., 2009; Dion et al., 2015; Frasson et al., 2011; Grober and Buschke, 1987; Grober and Kawas, 1997; Ivnik et al., 1997; Lekeu et al., 2003; Lemos et al., 2015; O'Connell and Tuokko, 2002; Petersen et al., 1994; Sarazin et al., 2007; Tounsi et al., 1999). They differ in the number of the to-be-learned items (i.e., 12 or 16), in the physical characteristics of the stimuli (e.g., line drawings, printed or orally spoken words) and in the testing procedures. In addition, the many versions differ in the number of learning trials (i.e., 3 or 6 trials) and in the inclusion of a yes–no recognition test.

The paradigm construction (e.g., presence of low or high number of to-be-remembered items) may considerably influence patient performance, preventing the direct comparison among case series. Very few are, however, the attempts to compare the psychometric properties of different word list recall task versions (e.g., Delgado et al., 2016). To date, neither sufficient evidence is available nor has direct effort been employed to systemically compare different standardized versions. For these reasons, this subaim can be considered as only Partly Achieved.

3.2.2. Phase 2, secondary aim 2

Secondary aim 2 of phase 2 is 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).

This is not applicable to neuropsychological measures. However, since performances on free and cued recall word list tasks may represent indirect measures of neurodegeneration of the medial temporal lobe structures, as reported previously (see phase 1), a predictable correlation between the topographical distribution of AD pathologic changes and the pattern of cognitive impairment appears to exist.

As example, the FCSRT scores have been recently shown to correlate with hippocampal atrophy (Sarazin et al., 2010) and with gray-matter loss of medial temporal lobe (Koric et al., 2013; Rami et al., 2012). Moreover, impaired FCSRT performance can be correlated with the presence of AD pathology as shown by changes of CSF Aβ values (Rami et al., 2012; Wagner et al., 2012; Xie et al., 2014), even at a prodromal stage (Derby et al., 2013).

Since no data addressing this specific issue are available and only indirect evidence is present, we evaluate this subaim as Not Applicable.

3.2.3. Phase 2, secondary aim 3

Secondary aim 3 of phase 2 is 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.

As the majority of cognitive measures, many factors are associated with performances on delayed recall tasks in cognitively healthy subjects. Age, education, and gender have a relevant impact on the cognitive performances in neuropsychological testing. Adjusted data of standardized versions account for such variables in the determination of the threshold for test positivity. For this reason, raw scores should be corrected for all the confounding demographic features according to the normative data of the reference population in order to control their weight on the performance. Thus, normative data are crucial both in clinical settings, to obtain single individual corrected scores, and in research studies aimed at addressing accuracy of test performance. The availability of good normative data, possibly uncontaminated by undiagnosed cognitively impaired control subjects, is thus a necessary prerequisite in order to consider the task as sufficiently reliable.

An important issue to consider when interpreting cognitive test results is the base rate of low performance in the normative sample. For example, 60.6% of the normative sample of the CERAD neuropsychological assessment battery obtained one or more scores at or below the 10th percentile (Mistridis et al., 2015).

In this context, many tasks included in this review have normative data for many different languages and countries. Efforts for providing normative data for the Grober-Buschke paradigm are more recent, and normative data for FCSRT are available for English-, Spanish-, French-, Italian-, and Danish-speaking populations (Dion et al., 2015; Frasson et al., 2011; Grober et al., 1998; Ivnik et al., 1997; Mokri et al., 2013; Vogel et al., 2007).

It must be underlined that the concept of cognitive reserve includes not only the educational level but also to lifestyle and occupation habits (Xu et al., 2015). These factors have significant direct and indirect effects on performances at delayed recall tasks and should therefore be considered in the evaluation and diagnosis of memory complaining subjects (Lojo-Seoane et al., 2014). For example, total recall of FCSRT is more sensitive to cognitive change in MCI subjects at low rather than at high levels of baseline cognition (Mura et al., 2014). The ceiling effect of this task suggests that it should be used with caution to assess memory changes in subjects with high premorbid cognitive performance (Mura et al., 2014).

The weight of demographic variables on the performances at the delayed recall tasks has not been definitely evaluated in all the

reference subpopulations. Moreover, apart from educational level, neuropsychological task performances are not controlled for other possible factors influencing cognitive performance at baseline, such as socioeconomic status and other factors possibly affecting cognitive reserve. For these reasons, this subaim is to consider as Partly Achieved. It is expected that ongoing studies will fill this gap within a few years.

3.2.4. Phase 2, secondary aim 4

Secondary aim 4 of phase 2 is to assess factors associated with biomarker status or level in diseased subjects—in particular, disease characteristics. The relationship between several disease characteristics (e.g., age at onset, genetic factors, comorbidities, atypical presentation) and episodic memory performances has been partly explored in AD dementia patients.

Memory tasks are not uniformly impaired in early- and late-onset AD cases (van der Flier et al., 2011). Multiple studies indicated that early onset of AD is associated with lower performance on different memory tasks (Kaiser et al., 2012; Kalpouzos et al., 2005). This impairment is strictly intersected with the ApoE carrier condition (van der Flier et al., 2011). Early-onset AD carriers of the ApoE ε4 allele had worse results for memory tasks than did noncarriers (van der Flier et al., 2011), and the homozygous ε4/ε4 genotype is frequently seen in patients with amnestic presentations than in those with nonmemory phenotypes (van der Flier et al., 2011). ApoE ε4 allele condition is also a well-established risk factor for late-onset AD affecting the age of onset (Shi et al., 2014). Although this evidence, it is still not clear if the weight of the ApoE ε4 allele carrier condition is the cognitive decline of AD. Results of a meta-analysis on 1700 patients (Allan and Ebmeier, 2011) showed indeed that different ApoE alleles do not modify the speed of clinical progression of dementia in terms of cognitive decline or mortality.

Finally, an atypical topographical localization of the neuronal injury in the course of AD pathology (as in case of posterior cortical atrophy, logopenic variant of primary progressive aphasia, or frontal variant of AD) may result in a relative sparing of memory function in the early phase of disease.

We can thus consider this subaim as only Partly Achieved, since further studies are required. Notably, investigations on additional genetic risk factors and on the influence of comorbidities, such as those with diabetes or sleep apnea, are still lacking.

3.3. Phase 3—prospective longitudinal repository studies

The primary aims of phase 3 consist in evaluating the capacity of the biomarker to detect preclinical disease (primary aim 1) and in defining criteria for a positive biomarker test in preparation for phase 4 (primary aim 2).

Different studies explored the capability of delayed recall tasks to detect the earliest stages of the AD pathology, that is, MCI that will progress to AD dementia. One point in need of an adequate consideration is the criterion for positivity to be used for neuropsychological tests. In the case of MCI, the threshold for positivity is the same as for AD dementia, namely a neuropsychological performance between 1 and 2 standard deviations below an age- and education-matched control population (Petersen, 2004). A reduced performance on memory tasks compared with that of the reference population is required to qualify a cognitive complaint as "objective." Thus, this is a mandatory prerequisite for defining the condition of MCI and distinguishing it from subjective memory complaints (Petersen, 2004; DSM-V, APA, 2000).

The cutoff between normal and pathological findings to be applied on individual basis is crucial for the clinical use of any cognitive marker. Standardization procedure and normative data according to the reference population (see Section 3.2.3) allow us to clearly define cognitive performance abnormality at the individual level. A challenge, however, could be the determination of specific cutoff points for the prodromal AD phase accounting for possible confounding variables (see Gainotti et al., 2014 for a critical review).

Predictive values of both learning and retention measures of free word list tasks have been widely investigated and have been consistently shown to predict progressions from MCI to AD in both community-based and memory clinic samples (Anchisi et al., 2005; Chang et al., 2010; Estevez-Gonzalez et al., 2003; Fleisher et al., 2007; Gallagher et al., 2010; Ivanoiu et al., 2005; Kim et al., 2010; Landau et al., 2010; Tabert et al., 2006).

A meta-analytic study on the cognitive impairments in prodromal AD subjects (Bäckman et al., 2005) reported values of immediate and delayed recall subtests for predicting progression to dementia, with a better performance of delayed recall. In agreement with this, Chang et al. (2010) showed good predictive values both for immediate and delayed recall of RAVLT, with better predictive power when both measures are impaired. Literature findings are, however, scanty and heterogeneous in methodology, failing to confirm or disprove the superiority of one measure above the other. As example, Landau et al. (2010) reported very high sensitivity (93%), high specificity (88%), and overall accuracy (90%) for the immediate recall of RAVLT, but in this study, this was the only RAVLT measure taken into account. In line with these authors (Landau et al., 2010), Ivanoiu et al. (2005) also reported very high sensitivity for the immediate recall (100%) of the CERAD 10-word list, even better than the savings (90%) and delayed recall (86%) measures. Both immediate and delayed recall scores of the CERAD were good predictors of incident AD dementia (AUC = 0.835–0.838) in elderly primary care patients (Wolfsgruber et al., 2014).

Comparable diagnostic accuracy and predictive values have been reported for the delayed recall subtest of other word list paradigms, namely CVLT (Anchisi et al., 2005), Seoul Verbal Learning Test (Kim et al., 2010), ADAS-cog 10-word list (Fleisher et al., 2007), Wechsler Memory Scale (Espinosa et al., 2013), delayed word recall task (Gallagher et al., 2010), and for the percent savings from immediate to delayed recall of the Selective Reminding Test (Tabert et al., 2006). In addition to delayed recall, a more stringent clinical classification (i.e., probable MCI according to Petersen's criteria), the presence of multiple cognitive impairments and of at least one apolipoprotein E (APOE) ε4 allele resulted in an increased risk of progression (Espinosa et al., 2013).

Cued paradigms showed also very high predictive value for progression to AD in large studies on MCI subjects (Belleville et al., 2014; Derby et al., 2013; Grober et al., 2000, 2010; Ivanoiu et al., 2005; Mura et al., 2014; Sarazin et al., 2007). The cued delayed recall of the RI48-test showed a higher accuracy (sensitivity = 100% and specificity = 77%) compared to the CERAD 10-word list (Ivanoiu et al., 2005). Out of 251 MCI subjects tested by Sarazin et al. (2007) at baseline and followed at 6-month intervals for up to 3 years, 59 subjects progressed to AD dementia. Among different neuropsychological tasks, the immediate total recall and the index of cueing of FCSRT had the best accuracy (AUC = 94% and 93%; sensitivity = 79.7% and 78%; specificity = 89.9% and 84.9%, respectively) for the diagnosis of AD in the prodromal phase. Additional results have been reported in a prospective longitudinal study on 122 single-domain or multiple-domain amnestic MCI recruited from memory clinics and followed up to 8 years (Belleville et al., 2014). The strongest predictors were free immediate recall of FCSRT together with the delayed recall of a story (i.e., MEMO-TEXT). The best model, yielding high sensitivity (88%), specificity (87.1%), and very high positive (81.8%) and negative (91.7%) predictive values, included both delayed recall tasks and nonmemory measures,

supporting the utility of combining different neuropsychological tasks to obtain higher accuracy (Belleville et al., 2014).

In conclusion, the literature review is generally supporting good predictive values for both immediate and delayed recall measures. Some comparative studies among different tasks (e.g., one free recall task vs. another; free vs. cued paradigms; different memory paradigms with each other) are available, with sparse and contradictory findings. Word list recall tests with a built-in semantic self-cueing option (i.e., CVLT) has been shown to be more sensitive to subtle episodic memory impairments compared to simple unrelated word list learning (Beck et al., 2012). While it is well established that scores of different free word list tasks (e.g., RAVLT and CVLT) highly correlate with each other in cognitively normal individuals (Crossen and Wiens, 1994), unquestionable similarities or relevant differences have not been clearly shown in neurodegenerative patients. Literature findings are inconsistent, with some authors stressing the lack of critical differences between free or cued tasks in the prediction of MCI conversion to AD dementia (Ivanoiu et al., 2005), whereas others supported better operating characteristics for the cued paradigm compared to free word list learning and/or logical memory measures (Derby et al., 2013; Wagner et al., 2012). In addition, better predictive values have been suggested for the combination of delayed recall tasks with other cognitive measures (i.e., Mini-Mental Status Examination and Mental Control subtest of the Wechsler memory scale) (Pozueta et al., 2011; Tierney et al., 1996). Recently, a composite episodic memory score proved in MCI better performance compared to individual memory scores to improve the power to predict progression to AD (Marra et al., 2015). Unfortunately, no direct comparison between a given format of a recall task versus another one in the same population is available.

There is at the moment insufficient evidence for the superiority of one measure compared to the others, as well as of cued paradigms compared to traditional measures of free recall. Further studies are needed to address the impact of different cutoff points for delayed recall tests in MCI. As reported by the Bruscoli and Lovestone's (2004) in a systematic review of conversion studies, the more stringent the measures of memory impairment, the better the prediction of conversion. Therefore, it seems essential to explore this issue in future research on large longitudinal studies. We can thus consider primary aims of phase 3 studies as only Partly Achieved. Some evidence is in fact available but many methodological limitations (e.g., lack of sufficient direct comparisons among the different task measures) prevent to draw final conclusions.

3.3.1. Phase 3, secondary aim 1

Secondary aim 1 of phase 3 is to explore the impact of covariates on the discriminatory abilities of the biomarker before clinical diagnosis.

This issue has already been discussed previously in relation to age, education, and gender (Section 3.2.3). Several additional factors that may affect the effectiveness of covariates on test scores in discriminating at risk patients from control subjects are occupational history, socioeconomic status as well as contextual factors (time of the day, metabolic status). These have not been systematically investigated.

Since no data addressing the previously mentioned factors are available, we can thus consider secondary aim 1 of phase 3 studies as only Partly Achieved.

3.3.2. Phase 3, secondary aim 2

The secondary aim 2 of phase 3 is to compare markers with a view to selecting those that are most promising.

As the performance on memory test is better considered as a gateway to the application of biomarkers, rather than a biomarker per se, a direct comparison between memory test scores and AD biomarkers in terms of diagnostic contribution is not considered as appropriate.

This subaim is thus evaluated as Not Applicable.

3.3.3. Phase 3, secondary aim 3

The aim is to develop algorithms for positivity based on combination of markers.

No specific diagnostic algorithm for positivity based on combination of delayed recall tests and AD biomarkers has been yet proposed. The relevance of including neuropsychological markers in the diagnostic algorithm is incontestable, as cognitive measures contribute to define the population of interest (e.g., amnestic or nonamnestic MCI subjects; AD or non-AD dementia patients) (see Section 1 for the concept of objective memory performance as a "gatekeeper" to biomarker assessment).

Several studies however tested the combination of memory measures (not directly used for the classification of the patient) with structural and/or functional imaging and CSF markers in early or prodromal AD cases. A number of articles, for example, reported predictive values of cognitive measures combined with MRI (Devanand et al., 2007; Eckerström et al., 2013; Ewers et al., 2012; Geroldi et al., 2006; Heister et al., 2011; Richard et al., 2013; Visser et al., 2002), CSF (Eckerström et al., 2013; Richard et al., 2013), or FDG-PET imaging (Anchisi et al., 2005; Chételat et al., 2005; Landau et al., 2010).

Richard et al. (2013) found no incremental value of adding up MRI and CSF biomarkers to the outcome of a memory test (i.e., immediate recall of RAVLT) in the early differential diagnosis of patients with MCI (average follow-up of 39 months). This study simulates well the clinical reality by first establishing diagnostic accuracy for the neuropsychology task and then quantifying the performance of MRI and CSF biomarkers. All 3 diagnostic instruments (i.e., delayed recall test, MRI imaging, and CSF values) substantially contributed to the differentiation of stable MCI from converters to AD (AUCs $\sim$0.65), while neither MRI nor CSF significantly improved diagnostic accuracy compared to the RAVLT administration alone. Different data have been reported by other authors. For instance, Eckerström et al. (2013) reported that, while delayed recall of RAVLT is the best individual predictor of dementia compared to CSF values and hippocampal volume measurement, an incremental diagnostic value is provided by the combination of these diagnostic instruments. The 3 measures combined (i.e., RAVLT, CSF, and MRI) were, indeed, the most successful combination for predicting dementia (AUC = 0.96) (Eckerström et al., 2013). Additionally, Heister et al. (2011) supported the utility of combining different tools to increase the diagnostic value. These authors reported an increased risk of AD conversion in subjects with joint presence of any 2 risk factors among immediate recall of RAVLT, MRI, and CSF (Heister et al., 2011). The combination of greater learning impairment and increased atrophy was associated with the highest risk (namely, 85% of patients with both risk factors converted to AD within 3 years) (Heister et al., 2011). In further support of the additional value of combining different diagnostic measures are the results of the study of Landau et al. (2010) in which the authors showed that MCI patients with abnormal results on both FDG-PET imaging and RAVLT task were 11.7 times more likely to progress to AD than subjects who had normal results on both measures.

In conclusion, the literature review is generally supporting improved predictive values for the combination of delayed recall tasks with biological and imaging biomarkers, with a particular impact of the latter. The evidence is, however, scanty and mainly focused on the combination of RAVLT performance with imaging or CSF biomarkers. No conclusion can, thus, be drawn on the

superiority of a specific combination of one neuropsychological tool and topographical/biological biomarkers relative to another combination (e.g., RAVLT or CVLT or cued paradigms combined to CSF measures or MRI and PET imaging), since the available evidence is not sufficient and not yet sufficiently replicated with adequately powered methods.

In particular, the lack of evidence on which is the delayed recall task to prefer forces us to consider this subaim of phase 3 as only Partly Achieved.

3.3.4. Phase 3, secondary aim 4

This aims to determine a biomarker testing interval for phase 4 if repeated testing is of interest.

There are several problematic issues in the application of the concept to this area. First, the delayed recall tests tend to go to floor very fast and are thus of limited usefulness to study disease progression (Locascio et al., 1995). Repeated testing is of interest only in the very early and prodromal disease stages, when subtle cognitive changes may be detected only with repeated measurements. Second, in the very early and prodromal stage an important issue is availability of parallel forms to account for learning effects. For example, the same performance at follow-up may indicate a lack of practice effects, that is, a minimal (initial) disorder (Zehnder et al., 2007). Unfortunately, this is still a largely neglected area, with some exceptions (i.e., Repeatable Battery for the Assessment of Neuropsychological Status [RBANS]).

On the basis of these considerations, this subaim is at present to consider as Not Achieved.

3.4. Phase 4—prospective diagnostic studies

The primary aim of phase 4 is 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 clinical populations and leading to diagnosis and treatment.

The main aim of phase 4 is thus still Not Achieved.

3.4.1. Phase 4, secondary aim 1

Secondary aim 1 of phase 4 assesses the characteristics of disease detected by the biomarker test—in particular, with regard to the potential benefit incurred by early detection.

Only preliminary evidence is available in the case of AD. In general, the scientific and medical community agrees upon the fact that earlier diagnosis may be advantageous for patients and caregivers, particularly for optimized medical management, future planning, access to services, risk reduction, and delaying nursing home placement (Banerjiee et al., 2007; Banerjiee and Wittenberg, 2009; Relkin, 2000).

This subaim is thus evaluated as Not Achieved.

3.4.2. Phase 4, secondary aim 2

The secondary aim 2 of phase 4 assesses the practical feasibility of implementing the diagnostic program and compliance of test-positive subjects with workup and treatment recommendations.

This subaim has not been assessed by any specific study and is thus evaluated as Not Achieved.

3.4.3. Phase 4, secondary aim 3

The aim is to make preliminary assessments of the effects of biomarker testing on costs and mortality associated with the disease.

This aspect has not been specifically assessed. This aim is thus evaluated as Not Achieved.

3.4.4. Phase 4, secondary aim 4

This aim is to monitor disease occurring clinically but not detected by the biomarker testing protocol.

This subaim has not been specifically assessed and thus evaluated as Not Achieved.

3.5. Phase 5—disease control studies

This final phase addresses whether using biomarkers for early diagnosis reduces the burden of AD in the general population.

Given the lack of effective pharmacological and/or non-pharmacological interventions for AD, population-level screening for dementia may not be acceptable to either the general public or health care professionals (Martin et al., 2015). Stigma, financial implications, impacts on employment and driving, as well as on person's lifestyle and attitude to health, combined with the knowledge of poor prognosis and few treatment options, play a crucial role in decision-making about population screening.

Thus, primary aim and subaims of phase 5 have not been specifically assessed and should be evaluated as Not Achieved.

4. Conclusions

A comprehensive neuropsychological evaluation of cognitive functions is required in all patients with subjective complaint not otherwise explained in order to identify the early stages of neurodegeneration. The choice of the task to assess the different cognitive domains is critical, in view of reaching sufficient sensitivity and specificity to detect early subjects at risk of developing AD dementia (i.e., prodromal AD cases). The psychometric properties of the test (e.g., sensitivity to cognitive change, specificity for neurodegeneration, floor and ceiling effects) have an important impact. The neuropsychological measures used to determine the MCI condition should be considered as the entry point to appropriate diagnostic algorithm for AD or non-AD neurocognitive disorders, as they make it possible to delineate the target population (in research settings) or the patients (in clinical settings), which may be candidate to undergo other invasive (i.e., CSF), costly (i.e., MRI), or both invasive and costly (i.e., PET imaging) biomarker assessments. The proper use of single neuropsychological tasks and, even better, the development of a combination of tasks are particularly urgent in the diagnosis of neurodegenerative dementias. The objective confirmation of long-term memory impairment, which is the earliest and most prominent feature of typical AD dementia (Salmon, 2012), is in need of further specific investigation. Noteworthy, no recommendation concerning the choice of the memory task to be used to assess this function in MCI subjects is currently available. The use of cued memory paradigms is suggested (Dubois et al., 2010, 2014), but no specific guidelines have been provided so far regarding the format that should be preferred. During the last 15 years, cognitive markers have been extensively used, in spite of the absence of clear-cut guidelines, according to the availability of the tasks in the specific research or clinical setting, mainly in academic memory clinics.

This review is part of a wider effort to outline a roadmap for the correct application of AD biomarkers in research and clinical settings (Chiotis et al., 2017; Garibotto et al., 2017; Mattsson et al., 2017; Porteri et al., 2017; Sonni et al., 2017; Ten Kate et al., 2017). The whole framework is borrowed from drug development and its specific translation for the oncology field (Pepe et al., 2001). We have revised the framework in order to adapt it to the current context and to the needs of the field of neurodegenerative dementias with a focus on the maturity of free and cued word list measures. Impairment on this class of tasks is considered as the core of neuropsychological feature of

typical AD dementia, consistent with the neuroanatomic distribution of histopathological abnormalities in medial temporal regions already present in the mildest stages of AD (Braak and Braak, 1991). Some authors previously reviewed neuropsychological predictors of conversion from MCI to AD dementia (see for examples Gainotti et al., 2014; Li et al., 2016; Modrego, 2006). Here, we included only studies providing evidence for diagnostic accuracy in AD or for the prediction of progression from MCI to AD, with the main aim to unveil gray areas in need of further confirmation.

he limitations of this review reflect the inhomogeneous and nonsystematic use of neuropsychological testing in research and clinical settings. Despite the variability among studies in design, statistical approach, follow-up length, and cognitive tests used, a general pattern seems to emerge. The majority of studies on AD dementia patients reported delayed recall as the most sensitive measure. Higher delayed recall scores on auditory verbal learning tasks seem to be "protective" for the progression to AD dementia, as shown by a recent meta-analysis and review of cohort studies assessing risk factors in MCI subjects (Li et al., 2016). The literature is also in favor of good predictive values for the progression from MCI to AD dementia not only for delayed but also for immediate recall measures. Although this literature review provided no definitive support for the relative advantage of cued recall tasks in comparison with tests measuring free recall in screening of patients suspected for AD dementia, cued paradigms seem to offer better operating characteristics compared to free word list learning (Wagner et al., 2012). There is, however, clearly the need for systematic studies directly comparing delayed recall from free or cued tasks. A clear-cut conclusion on the preferable use of a specific measure versus another cannot be drawn at the moment, this remains a critical topic that should be addressed by future research aimed at comparing several neuropsychological tests (and scores) within the memory domain.

We focused on the assessment of delayed recall since its impairment is a key feature of typical AD dementia. Nevertheless, this will not be an accurate neuropsychological measure in atypical AD cases (e.g., posterior cortical atrophy, logopenic variant of primary progressive aphasia, and frontal variant of AD). Therefore, this measure may not always be the best choice to orient toward a screening with second-level biomarkers in suspicion of an AD condition. A combination of memory and nonmemory neuropsychological measures should thus be preferred as the gateway to the assessment of second-level invasive and costly biomarkers. Accordingly, different neuropsychological measures taken together may provide a better accuracy than the measurement of single cognitive domains (see for examples Belleville et al., 2014; Dierckx et al., 2009; Mitchell, 2009a; Mitchell et al., 2009b; Tabert et al., 2006; Tierney et al., 1996). Further research is required to determine the best cognitive algorithm (namely, a combination of different memory tasks or memory and nonmemory tests) that is expected to have a higher predictive diagnostic value because it reflects the different components of the neurobiological dysfunctions occurring in AD pathology.

This review has several limitations. Notwithstanding our efforts to be as inclusive as possible, the literature search was not a formal systematic review. Moreover, the framework adapted by Pepe et al. (2001) to the oncology field had to be adapted to the current context, that is, to provide a reliable and biomarker-based diagnosis in people with cognitive complaints. This required adaptation of the screening phases into a diagnostic perspective, which is conceptually different and has relevant limitations compared to screening studies. Future development of the field may allow to expand the currently developed framework and to fully assimilate it to the oncology and drug-development framework.

Disclosure statement

Chiara Cerami was funded by Fondazione Eli-Lilly grant 2011 "Imaging of neuroinflammation and neurodegeneration in prodromal and presymptomatic Alzheimer's disease phases." Marina Boccardi received a research grant from Piramal. Andreas U. Monsch and Stefano F. Cappa have nothing to declare.

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 PI 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 Neurological Society of Neuroradiology—ENSNR, 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 (Early diagnosis of Alzheimer's disease with biomarkers: Now despite no cure, or later «only if»? - International Exploratory Workshops IZ32Z0_157953) and unrestricted grants from: Alzheimer Forum Switzerland, Association pour la Recherche sur Alzheimer, Genève; Piramal, Eli Lilly and 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. The authors 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 article 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.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.neurobiolaging.2016.03.034.

References

Abdulrab, K., Heun, R., 2008. Subjective memory impairment. A review of its definitions indicates the need for a comprehensive set of standardised and validated criteria. Eur. Psychiatry 23, 321–330.

Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., Jagust, W.J., Petersen, R.C., Snyder, P.J., Carrillo, M.C., Thies, B., Phelps, C.H., 2011. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 270–279.

Allan, C.L., Ebmeier, K.P., 2011. The influence of ApoE4 on clinical progression of dementia: a meta-analysis. Int. J. Geriatr. Psychiatry 26, 520–526.

American Psychiatric Association, 2000. Diagnostic and Statistical Manual of Mental Disorders IV-TR, fourth ed. American Psychiatric Association, Washington, DC. Text revised.

Anchisi, D., Borroni, B., Franceschi, M., Kerrouche, N., Kalbe, E., Beuthien-Beumann, B., Cappa, S., Lenz, O., Ludecke, S., Marcone, A., Mielke, R., Ortelli, P., Padovani, A., Pelati, O., Pupi, A., Scarpini, E., Weisenbach, S., Herholz, K., Salmon, E., Holthoff, V., Sorbi, S., Fazio, F., Perani, D., 2005. Heterogeneity of brain glucose metabolism in mild cognitive impairment and clinical progression to Alzheimer disease. Arch. Neurol. 62, 1728–1733.

Auriacombe, S., Helmer, C., Amieva, H., Berr, C., Dubois, B., Dartigues, J.F., 2010. Validity of the free and cued selective reminding test in predicting dementia: the 3C study. Neurology 74, 1760–1777.

Bäckman, L., Jones, S., Berger, A.K., Laukka, E.J., Small, B.J., 2005. Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. Neuropsychology 19, 520–531.

Banerjee, S., Willis, R., Matthews, D., Contell, F., Chan, J., Murray, J., 2007. Improving the quality of care for mild to moderate dementia: an evaluation of the Croydon Memory Service Model. Int. J. Geriatr. Psychiatry 22, 782–788.

Banerjee, S., Wittenberg, R., 2009. Clinical and cost effectiveness of services for early diagnosis and intervention in dementia. Int. J. Geriatr. Psychiatry 24, 748–754.

Barbeau, E., Didic, M., Tramoni, E., Felician, O., Joubert, S., Sontheimer, A., Ceccaldi, M., Poncet, M., 2004. Evaluation of visual recognition memory in MCI patients. Neurology 62, 1317–1322.

Barbeau, E.J., Ranjeva, J.P., Didic, M., Confort-Gouny, S., Felician, O., Soulier, E., Cozzone, P.J., Ceccaldi, M., Poncet, M., 2008. Profile of memory impairment and gray matter loss in amnestic mild cognitive impairment. Neuropsychologia 46, 1009–1019.

Beck, I.R., Gagneux-Zurbriggen, A., Berres, M., Taylor, K.I., Monsch, A.U., 2012. Comparison of verbal episodic memory measures: Consortium to Establish a Registry for Alzheimer's Disease-Neuropsychological Assessment Battery (CERAD-NAB) versus California Verbal Learning Test (CVLT). Arch. Clin. Neuropsychol. 27, 510–519.

Belleville, S., Gauthier, S., Lepage, E., Kergoat, M.J., Gilbert, B., 2014. Predicting decline in mild cognitive impairment: a prospective cognitive study. Neuropsychology 28, 643–652.

Bennett, D.A., Wilson, R.S., Schneider, J.A., Evans, D.A., Beckett, L.A., Aggarwal, N.T., Barnes, L.L., Fox, J.H., Bach, J., 2002. Natural history of mild cognitive impairment in older persons. Neurology 59, 198–205.

Berrios, G.E., Markova, I.S., Girala, N., 2000. Functional memory complaints: hypochondria and disorganization. In: Berrios, G.E., Hodges, J.R. (Eds.), Memory disorders in psychiatric practice. Cambridge University Press, Cambridge, pp. 384–399.

Boccardi, M., Gallo, V., Yutaka, Y., Vineis, P., Padovani, A., Mosimann, U., Giannakopoulos, P., Gold, G., Dubois, B., Jack, C.R., Winblad, B., Frisoni, G.B., Albanese, E., 2017. The Geneva Task Force for the Roadmap of Alzheimer's Biomarkers, The Biomarker-based Diagnosis of Alzheimer's Disease. 2 – Lessons From Oncology, Neurobiol. Aging 52, 141–152.

Braak, H., Braak, E., 1991. Neuropathological staging of Alzheimer-related changes. Acta Neuropathol. 82, 239–259.

Brandt, J., Benedict, R.H.B., 2001. Hopkins Verbal Learning Test—Revised. Professional Manual. Psychological Assessment Resources, Inc, Lutz, FL.

Bruscoli, M., Lovestone, S., 2004. Is MCI really just early dementia? a systematic review of conversion studies. Int. Psychogeriatr. 16, 129–140.

Buschke, H., 1984. Cued recall in amnesia. J. Clin. Neuropsychol. 6, 433–440.

Buschke, H., Sliwinski, M.J., Kuslansky, G., Lipton, R.B., 1997. Diagnosis of early dementia by the Double Memory Test: encoding specificity improves diagnostic sensitivity and specificity. Neurology 48, 989–997.

Chang, Y.L., Bondi, M.W., Fennema-Notestine, C., McEvoy, L.K., Hagler Jr., D.J., Jacobson, M.W., Dale, A.M. Alzheimer's Disease Neuroimaging Initiative, 2010. Brain substrates of learning and retention in mild cognitive impairment diagnosis and progression to Alzheimer's disease. Neuropsychologia 48, 1237–1247.

Chételat, G., Eustache, F., Viader, F., De La Sayette, V., Pélerin, A., Mézenge, F., Hannequin, D., Dupuy, B., Baron, J.C., Desgranges, B., 2005. FDG-PET measurement is more accurate than neuropsychological assessments to predict global cognitive deterioration in patients with mild cognitive impairment. Neurocase 11, 14–25.

Chiotis, K., Saint-Aubert, L., Boccardi, M., Gietl, A., Picco, A., Varrone, A., Garibotto, V., Herholz, K., Nobili, F., Nordberg, A., 2017. The Geneva Task Force for the Roadmap of Alzheimer's Biomarkers, Clinical validity of increased cortical uptake of amyloid ligands on PET as a biomarker for Alzheimer's disease in the context of a structured 5-phase development framework, Neurobiol. Aging 52, 214–227.

Crane, P.K., Carle, A., Gibbons, L.E., Insel, P., Mackin, R.S., Gross, A., Jones, R.N., Mukherjee, S., Curtis, S.M., Harvey, D., Weiner, M., Mungas, D. Alzheimer's Disease Neuroimaging Initiative, 2012. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). Brain Imaging Behav. 6, 502–516.

Crossen, J.R., Wiens, A.N., 1994. Comparison of the Auditory-Verbal Learning Test (AVLT) and California Verbal Learning Test (CVLT) in a sample of normal subjects. J. Clin. Exp. Neuropsychol. 16, 190–194.

Delgado, C., Muñoz-Neira, C., Soto, A., Martínez, M., Henríquez, F., Flores, P., Slachevsky, A., 2016. Comparison of the psychometric properties of the "word" and "picture" versions of the free and cued selective reminding test in a Spanish-speaking cohort of patients with mild Alzheimer's disease and cognitively healthy controls. Arch. Clin. Neuropsychol. 31, 165–175.

Delis, D.C., Kramer, J.H., Kaplan, E., Ober, B.A., 1987. The California Verbal Learning Test. New York: The Psychological Corporation, San Antonio, TX.

Delis, D.C., Kramer, J.H., Kaplan, E., Ober, B.A., 2000. The California Verbal Learning Test, second ed. New York: The Psychological Corporation, San Antonio, TX.

Derby, C.A., Burns, L.C., Wang, C., Katz, M.J., Zimmerman, M.E., L'italien, G., Guo, Z., Berman, R.M., Lipton, R.B., 2013. Screening for predementia AD: time-dependent operating characteristics of episodic memory tests. Neurology 80, 1307–1314.

Devanand, D.P., Pradhaban, G., Liu, X., Khandji, A., De Santi, S., Segal, S., Rusinek, H., Pelton, G.H., Honig, L.S., Mayeux, R., Stern, Y., Tabert, M.H., de Leon, M.J., 2007. Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. Neurology 68, 828–836.

Dickerson, B.C., Eichenbaum, H., 2010. The episodic memory system: neurocircuitry and disorders. Neuropsychopharmacology 35, 86–104.

Dierckx, E., Engelborghs, S., De Raedt, R., Van Buggenhout, M., De Deyn, P.P., Verté, D., Ponjaert-Kristoffersen, I., 2009. Verbal cued recall as a predictor of conversion to Alzheimer's disease in mild cognitive impairment. Int. J. Geriatr. Psychiatry 24, 1094–1100.

Dion, M., Potvin, O., Belleville, S., Ferland, G., Renaud, M., Bherer, L., Joubert, S., Vallet, G.T., Simard, M., Rouleau, I., Lecomte, S., Macoir, J., Hudon, C., 2015. Normative data for the Rappel libre/Rappel indicé à 16 items (16-item Free and Cued Recall) in the elderly Quebec-French population. Clin. Neuropsychol. 28, S1–S19.

Dubois, B., Feldman, H.H., Jacova, C., Dekosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P.J., Scheltens, P., 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol. 6, 734–746.

Dubois, B., Feldman, H.H., Jacova, C., Cummings, J.L., Dekosky, S.T., Barberger-Gateau, P., Delacourte, A., Frisoni, G., Fox, N.C., Galasko, D., Gauthier, S., Hampel, H., Jicha, G.A., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Sarazin, M., de Souza, L.C., Stern, Y., Visser, P.J., Scheltens, P., 2010. Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol. 9, 1118–1127.

Dubois, B., Feldman, H.H., Jacova, C., Hampel, H., Molinuevo, J.L., Blennow, K., DeKosky, S.T., Gauthier, S., Selkoe, D., Bateman, R., Cappa, S., Crutch, S., Engelborghs, S., Frisoni, G.B., Fox, N.C., Galasko, D., Habert, M.O., Jicha, G.A., Nordberg, A., Pasquier, F., Rabinovici, G., Robert, P., Rowe, C., Salloway, S., Sarazin, M., Epelbaum, S., de Souza, L.C., Vellas, B., Visser, P.J., Schneider, L., Stern, Y., Scheltens, P., Cummings, J.L., 2014. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol. 13, 614–629.

Eckerström, C., Olsson, E., Bjerke, M., Malmgren, H., Edman, A., Wallin, A., Nordlund, A., 2013. A combination of neuropsychological, neuroimaging, and cerebrospinal fluid markers predicts conversion from mild cognitive impairment to dementia. J. Alzheimers Dis. 36, 421–431.

Eichenbaum, H., Cohen, N.J., 2001. From Conditioning to Conscious Recollection: Memory Systems of the Brain. Oxford University Press, Oxford.

Elias, M.F., Beiser, A., Wolf, P.A., Au, R., White, R.F., D'Agostino, R.B., 2000. The preclinical phase of Alzheimer disease. A 22-year prospective study of the Framingham cohort. Arch. Neurol. 57, 808–813.

Espinosa, A., Alegret, M., Valero, S., Vinyes-Junqué, G., Hernández, I., Mauleón, A., Rosende-Roca, M., Ruiz, A., López, O., Tárraga, L., Boada, M., 2013. A longitudinal follow-up of 550 mild cognitive impairment patients: evidence for large conversion to dementia rates and detection of major risk factors involved. J. Alzheimers Dis. 34, 769–780.

Estevez-Gonzalez, A., Kulisevsky, J., Boltes, A., Otermin, P., Garcia-Sanchez, C., 2003. Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: comparison with mild cognitive impairment and normal aging. Int. J. Geriatr. Psychiatr. 18, 1021–1028.

Ewers, M., Walsh, C., Trojanowski, J.Q., Shaw, L.M., Petersen, R.C., Jack Jr., C.R., Feldman, H.H., Bokde, A.L., Alexander, G.E., Scheltens, P., Vellas, B., Dubois, B., Weiner, M., Hampel, H. North American Alzheimer's Disease Neuroimaging Initiative (ADNI), 2012. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiol. Aging 33, 1203–1214.

Fleisher, A.S., Sowell, B.B., Taylor, C., Gamst, A.C., Petersen, R.C., Thal, L.J. Alzheimer's Disease Cooperative Study, 2007. Clinical predictors of progression to Alzheimer disease in amnestic mild cognitive impairment. Neurology 68, 1588–1595.

Frasson, P., Ghiretti, R., Catricalà, E., Pomati, S., Marcone, A., Parisi, L., Rossini, P.M., Cappa, S.F., Mariani, C., Vanacore, N., Clerici, F., 2011. Free and cued selective reminding test: an Italian normative study. Neurol. Sci. 32, 1057–1062.

Frisoni, G.B., Perani, D., Bastianello, S., Bernardi, G., Porteri, C., Boccardi, M., Cappa, S.F., Trabucchi, M., Padovani, A., 2017. A roadmap to the use of biomarkers for the diagnosis of Alzheimer's disease in clinical practice: the Italian inter-societal consensus, Neurobiol. Aging 52, 119–131.

Gainotti, G., Quaranta, D., Vita, M.G., Marra, C., 2014. Neuropsychological predictors of conversion from mild cognitive impairment to Alzheimer's disease. J. Alzheimers Dis. 38, 481–495.
Gallagher, D., Mhaolain, A.N., Coen, R., Walsh, C., Kilroy, D., Belinski, K., Bruce, I., Coakley, D., Walsh, J.B., Cunningham, C., Lawlor, B.A., 2010. Detecting prodromal Alzheimer's disease in mild cognitive impairment: utility of the CAMCOG and other neuropsychological predictors. Int. J. Geriatr. Psychiatry 25, 1280–1287.
Garibotto, V., Herholz, K., Boccardi, M., Picco, A., Varrone, A., Nordberg, A., Nobili, F., Ratib, O., 2017. The Geneva Task Force for the Roadmap of Alzheimer's Biomarkers, Maturity of FDG-PET as a biomarker for Alzheimer's disease in the context of a structured 5 −phase development framework, Neurobiol. Aging 52, 183–195.
Geroldi, C., Rossi, R., Calvagna, C., Testa, C., Bresciani, L., Binetti, G., Zanetti, O., Frisoni, G.B., 2006. Medial temporal atrophy but not memory deficit predicts progression to dementia in patients with mild cognitive impairment. J. Neurol. Neurosurg. Psychiatry 77, 1219–1222.
Grober, E., Buschke, H., 1987. Genuine memory deficits in dementia. Dev. Neuropsychol. 3, 13–36.
Grober, E., Kawas, C., 1997. Learning and retention in preclinical and early Alzheimer's disease. Psychol. Aging 12, 183–188.
Grober, E., Lipton, R.B., Katz, M., Sliwinski, M., 1998. Demographic influences on free and cued selective reminding performance in older persons. J. Clin. Exp. Neuropsychol. 20, 221–226.
Grober, E., Lipton, R.B., Hall, C.B., Crystal, H., 2000. Memory impairment on free and cued selective reminding predicts dementia. Neurology 54, 827–832.
Grober, E., Sanders, A.E., Hall, C., Lipton, R.B., 2010. Free and cued selective reminding identifies very mild dementia in primary care. Alzheimer Dis. Assoc. Disord. 24, 284–290.
Guillozet, A.L., Weintraub, S., Mash, D.C., Mesulam, M.M., 2003. Neurofibrillary tangles, amyloid, and memory in aging and mild cognitive impairment. Arch. Neurol. 60, 729–736.
Hartley, T., Bird, C.M., Chan, D., Cipolotti, L., Husain, M., Varga-Khadem, F., Burgess, N., 2007. The hippocampus is required for short term topographical memory in humans. Hippocampus 17, 34–48.
Heister, D., Brewer, J.B., Magda, S., Blennow, K., McEvoy, L.K. Alzheimer's Disease Neuroimaging Initiative, 2011. Predicting MCI outcome with clinically available MRI and CSF biomarkers. Neurology 77, 1619–1628.
Hornberger, M., Piguet, O., Graham, A.J., Nestor, P.J., Hodges, J.R., 2010. How preserved is episodic memory in behavioral variant frontotemporal dementia? Neurology 74, 472–479.
Ivanoiu, A., Adam, S., Van der Linden, M., Salmon, E., Juillerat, A.C., Mulligan, R., Seron, X., 2005. Memory evaluation with a new cued recall test in patients with mild cognitive impairment and Alzheimer's disease. J. Neurol. 252, 47–55.
Ivnik, R.J., Smith, G.E., Lucas, J.A., Tangalos, E.G., Kokmen, E., Petersen, R.C., 1997. Free and cued selective reminding test: MOANS norms. J. Clin. Exp. Neuropsychol. 19, 676–691.
Jack Jr., C.R., Lowe, V.J., Senjem, M.L., Weigand, S.D., Kemp, B.J., Shiung, M.M., Knopman, D.S., Boeve, B.F., Klunk, W.E., Mathis, C.A., Petersen, R.C., 2008. 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. Brain 131, 665–680.
Jack Jr., C.R., Albert, M.S., Knopman, D.S., McKhann, G.M., Sperling, R.A., Carrillo, M.C., Thies, B., Phelps, C.H., 2011. Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 257–262.
Jain, K.K., 2012. Biomarkers of neurological disorders. In: Jain, K.K. (Ed.), Applications of Biotechnology in Neurology. Humana Press, Totowa, NJ, pp. 49–53.
Jessen, F., Amariglio, R.E., van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., Dubois, B., Dufouil, C., Ellis, K.A., van der Flier, W.M., Glodzik, L., van Harten, A.C., de Leon, M.J., McHugh, P., Mielke, M.M., Molinuevo, J.L., Mosconi, L., Osorio, R.S., Perrotin, A., Petersen, R.C., Rabin, L.A., Rami, L., Reisberg, B., Rentz, D.M., Sachdev, P.S., de la Sayette, V., Saykin, A.J., Scheltens, P., Shulman, M.B., Slavin, M.J., Sperling, R.A., Stewart, R., Uspenskaya, O., Vellas, B., Visser, P.J., Wagner, M. Subjective Cognitive Decline Initiative (SCD-I) Working Group, 2014. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement. 10, 844–852.
Kaiser, N.C., Melrose, R.J., Liu, C., Sultzer, D.L., Jimenez, E., Su, M., Monserratt, L., Mendez, M.F., 2012. Neuropsychological and neuroimaging markers in early versus late-onset Alzheimer's disease. Am. J. Alzheimers Dis. Other Demen. 27, 520–529.
Kalpouzos, G., Eustache, F., de la Sayette, V., Viader, F., Chételat, G., Desgranges, B., 2005. Working memory and FDG-PET dissociate early and late onset Alzheimer disease patients. J. Neurol. 252, 548–558.
Kim, S.H., Seo, S.W., Yoon, D.S., Chin, J., Lee, B.H., Cheong, H.K., Han, S.H., Na, D.L., 2010. Comparison of neuropsychological and FDG-PET findings between early- versus late-onset mild cognitive impairment: a five-year longitudinal study. Dement. Geriatr. Cogn. Disord. 29, 213–223.
Koric, L., Ranjeva, J.P., Felician, O., Guye, M., de Anna, F., Soulier, E., Didic, M., Ceccaldi, M., 2013. Cued recall measure predicts the progression of gray matter atrophy in patients with amnesic mild cognitive impairment. Dement Geriatr. Cogn. Disord. 36, 197–210.
Landau, S.M., Harvey, D., Madison, C.M., Reiman, E.M., Foster, N.L., Aisen, P.S., Petersen, R.C., Shaw, L.M., Trojanowski, J.Q., Jack Jr., C.R., Weiner, M.W., Jagust, W.J. Alzheimer's Disease Neuroimaging Initiative, 2010. Comparing predictors of conversion and decline in mild cognitive impairment. Neurology 75, 230–238.
Lekeu, F., Van der Linden, M., Chicherio, C., Collette, F., Degueldre, C., Franck, G., Moonen, G., Salmon, E., 2003. Brain correlates of performance in a free/cued recall task with semantic encoding in Alzheimer disease. Alzheimer Dis. Assoc. Disord. 17, 35–45.
Lemos, R., Simões, M.R., Santiago, B., Santana, I., 2015. The free and cued selective reminding test: validation for mild cognitive impairment and Alzheimer's disease. J. Neuropsychol. 9, 242–257.
Li, J.Q., Tan, L., Wang, H.F., Tan, M.S., Tan, L., Xu, W., Zhao, Q.F., Wang, J., Jiang, T., Yu, J.T., 2016. Risk factors for predicting progression from mild cognitive impairment to Alzheimer's disease: a systematic review and meta-analysis of cohort studies. J. Neurol. Neurosurg. Psychiatry 87, 476–484.
Locascio, J.J., Growdon, J.H., Corkin, S., 1995. Cognitive test performance in detecting, staging, and tracking Alzheimer's disease. Arch. Neurol. 52, 1087–1099.
Lojo-Seoane, C., Facal, D., Guàrdia-Olmos, J., Juncos-Rabadán, O., 2014. Structural model for estimating the influence of cognitive reserve on cognitive performance in adults with subjective memory complaints. Arch. Clin. Neuropsychol. 29, 245–255.
Lowndes, G., Savage, G., 2007. Early detection of memory impairment in Alzheimer's disease: a neurocognitive perspective on assessment. Neuropsychol. Rev. 17, 193–202.
Marra, C., Gainotti, G., Fadda, L., Perri, R., Lacidogna, G., Scaricamazza, E., Piccininni, C., Quaranta, D., 2015. Usefulness of an integrated analysis of different memory tasks to predict the progression from mild cognitive impairment to Alzheimer's disease: the episodic memory score (EMS). J. Alzheimers Dis. 50, 61–70.
Martin, S., Kelly, S., Khan, A., Cullum, S., Dening, T., Rait, G., Fox, C., Katona, C., Cosco, T., Brayne, C., Lafortune, L., 2015. Attitudes and preferences towards screening for dementia: a systematic review of the literature. BMC Geriatr. 15, 66.
Maruta, C., Guerreiro, M., de Mendonça, A., Hort, J., Scheltens, P., 2011. The use of neuropsychological tests across Europe: the need for a consensus in the use of assessment tools for dementia. Eur. J. Neurol. 18, 279–285.
Mattson, N., Lönneborg, A., Boccardi, M., Blennow, K., Hansson, O., 2017. The Geneva Task Force for the Roadmap of Alzheimer's Biomarkers, Maturity of Aβ42, tau, and phospho-tau in the cerebrospinal fluid as biomarkers for Alzheimer's disease in the context of a structured 5-phase development framework, Neurobiol. Aging 52, 196–213.
McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34, 939–944.
McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack Jr., C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., Mohs, R.C., Morris, J.C., Rossor, M.N., Scheltens, P., Carrillo, M.C., Thies, B., Weintraub, S., Phelps, C.H., 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 263–269.
Meltzer, J.A., Constable, R.T., 2005. Activation of human hippocampal formation reflects success in both encoding and cued recall of paired associates. Neuroimage 24, 384–397.
Mistridis, P., Egli, S.C., Iverson, G.L., Berres, M., Willmes, K., Welsh-Bohmer, K.A., Monsch, A.U., 2015. Considering the base rates of low performance in cognitively healthy older adults improves the accuracy to identify neurocognitive impairment with the Consortium to Establish a Registry for Alzheimer's Disease-Neuropsychological Assessment Battery (CERAD-NAB). Eur. Arch. Psychiatry Clin. Neurosci. 265, 407–417.
Mitchell, A.J., 2009a. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. J. Psychiatr. Res. 43, 411–431.
Mitchell, J., Arnold, R., Dawson, K., Nestor, P.J., Hodges, J.R., 2009b. Outcome in subgroups of mild cognitive impairment (MCI) is highly predictable using a simple algorithm. J. Neurol. 256, 1500–1509.
Modrego, P.J., 2006. Predictors of conversion to dementia of probable Alzheimer type in patients with mild cognitive impairment. Curr. Alzheimer Res. 3, 161–170.
Mohs, R.C., Rosen, W.G., Davis, K.L., 1983. The Alzheimer's Disease Assessment Scale: an instrument for assessing treatment efficacy. Psychopharmacol. Bull 19, 448–450.
Mokri, H., Avila-Funes, J.A., Meillon, C., Gutiérrez Robledo, L.M., Amieva, H., 2013. Normative data for the mini-mental state examination, the free and cued selective reminding test and the Isaacs set test for an older adult Mexican population: the Coyoacán cohort study. Clin. Neuropsychol. 27, 1004–1018.
Morris, J.C., Mohs, R.C., Rogers, H., Fillenbaum, G., Heyman, A., 1988. Consortium to establish a registry for Alzheimer's disease (CERAD) clinical and neuropsychological assessment of Alzheimer's disease. Psychopharmacol. Bull 24, 641–652.
Mura, T., Proust-Lima, C., Jacqmin-Gadda, H., Akbaraly, T.N., Touchon, J., Dubois, B., Berr, C., 2014. Measuring cognitive change in subjects with prodromal Alzheimer's disease. J. Neurol. Neurosurg. Psychiatry 85, 363–370.
Murray, M.E., Graff-Radford, N.R., Ross, O.A., Petersen, R.C., Duara, R., Dickson, D.W., 2011. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. Lancet Neurol. 10, 785–796.
Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., Chertkow, H., 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J. Am. Geriatr. Soc. 53, 695–699.

Nelson, P.T., Alafuzoff, I., Bigio, E.H., Bouras, C., Braak, H., Cairns, N.J., Castellani, R.J., Crain, B.J., Davies, P., Del Tredici, K., Duyckaerts, C., Frosch, M.P., Haroutunian, V., Hof, P.R., Hulette, C.M., Hyman, B.T., Iwatsubo, T., Jellinger, K.A., Jicha, G.A., Kövari, E., Kukull, W.A., Leverenz, J.B., Love, S., Mackenzie, I.R., Mann, D.M., Masliah, E., McKee, A.C., Montine, T.J., Morris, J.C., Schneider, J.A., Sonnen, J.A., Thal, D.R., Trojanowski, J.Q., Troncoso, J.C., Wisniewski, T., Woltjer, R.L., Beach, T.G., 2012. Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature. J. Neuropathol. Exp. Neurol. 71, 362–381.

O'Connell, M.E., Tuokko, H., 2002. The 12-item Buschke memory test: appropriate for use across levels of impairment. Appl. Neuropsychol. 9, 226–233.

Parra, M.A., Abrahams, S., Fabi, K., Logie, R., Luzzi, S., Della Sala, S., 2009. Short-term memory binding deficits in Alzheimer's disease. Brain 132, 1057–1066.

Parra, M.A., Abrahams, S., Logie, R.H., Méndez, L.G., Lopera, F., Della Sala, S., 2010. Visual short-term memory binding deficits in familial Alzheimer's disease. Brain 133, 2702–2713.

Pepe, M.S., Etzioni, R., Feng, Z., Potter, J.D., Thompson, M.L., Thornquist, M., Winget, M., Yasui, Y., 2001. Phases of biomarker development for early detection of cancer. J. Natl. Cancer Inst. 93, 1054–1061.

Petersen, R.C., Smith, G.E., Ivnik, R.J., Kokmen, E., Tangalos, E.G., 1994. Memory function in very early Alzheimer's disease. Neurology 44, 867–872.

Petersen, R.C., 2004. Mild cognitive impairment as a diagnostic entity. J. Intern. Med. 256, 183–194.

Porteri C., Albanese E., Scerri C., Carrillo M.C., Snyder H.M., Martensson B., Baker M., Giacobini E., Boccardi M., Winblad B., Frisoni G.B., Hurst S., for the Geneva Task Force for the Roadmap of Alzheimer's Biomarkers. 2017. The Biomarker-based Diagnosis of Alzheimer's Disease. 1 – Ethical and Societal Issues, Neurobiol. Aging 52, 132–140.

Pozueta, A., Rodríguez-Rodríguez, E., Vazquez-Higuera, J.L., Mateo, I., Sánchez-Juan, P., González-Perez, S., Berciano, J., Combarros, O., 2011. Detection of early Alzheimer's disease in MCI patients by the combination of MMSE and an episodic memory test. BMC Neurol. 11, 78.

Rami, L., Solé-Padullés, C., Fortea, J., Bosch, B., Lladó, A., Antonell, A., Olives, J., Castellví, M., Bartres-Faz, D., Sánchez-Valle, R., Molinuevo, J.L., 2012. Applying the new research diagnostic criteria: MRI findings and neuropsychological correlations of prodromal AD. Int. J. Geriatr. Psychiatry 27, 127–134.

Randolph, C., Gold, J.M., Kozora, E., Cullum, M., Hermann, B.P., Wyler, A.R., 1994. Estimating memory function: disparity of Wechsler Memory Scale – Revised and California Verbal Learning Test indices in clinical and normal samples. Clin. Neuropsychologist 8, 99–108.

Relkin, N., 2000. Screening and early diagnosis of dementia. Am. J. Manag. Care 6, S1111–S1118 discussion S1119–1124.

Rey, A., 1941. L'examen psychologique dans les cas d'encéphalopathie traumatique. Arch. de Psychol. 28, 21.

Ricci, M., Graef, S., Blundo, C., Miller, L.A., 2012. Using the Rey Auditory Verbal Learning Test (RAVLT) to differentiate Alzheimer's dementia and behavioural variant fronto-temporal dementia. Clin. Neuropsychol. 26, 926–941.

Richard, E., Schmand, B.A., Eikelenboom, P., Van Gool, W.A. Alzheimer's Disease Neuroimaging Initiative, 2013. MRI and cerebrospinal fluid biomarkers for predicting progression to Alzheimer's disease in patients with mild cognitive impairment: a diagnostic accuracy study. BMJ Open 3. http://dx.doi.org/10.1136/bmjopen-2012-002541.

Rowe, C.C., Ellis, K.A., Rimajova, M., Bourgeat, P., Pike, K.E., Jones, G., Fripp, J., Tochon-Danguy, H., Morandeau, L., O'Keefe, G., Price, R., Raniga, P., Robins, P., Acosta, O., Lenzo, N., Szoeke, C., Salvado, O., Head, R., Martins, R., Masters, C.L., Ames, D., Villemagne, V.L., 2010. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. Neurobiol. Aging 31, 1275–1283.

Salmon, D.P., 2012. Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease. Curr. Top Behav. Neurosci. 10, 187–212.

Sarazin, M., Berr, C., De Rotrou, J., Fabrigoule, C., Pasquier, F., Legrain, S., Michel, B., Puel, M., Volteau, M., Touchon, J., Verny, M., Dubois, B., 2007. Amnestic syndrome of the medial temporal type identifies prodromal AD: a longitudinal study. Neurology 69, 1859–1867.

Sarazin, M., Chauviré, V., Gerardin, E., Colliot, O., Kinkingnéhun, S., de Souza, L.C., Hugonot-Diener, L., Garnero, L., Lehéricy, S., Chupin, M., Dubois, B., 2010. The amnestic syndrome of hippocampal type in Alzheimer's disease: an MRI study. J. Alzheimers Dis. 22, 285–294.

Shi, J., Han, P., Kuniyoshi, S.M., 2014. Cognitive impairment in neurological diseases: lessons from apolipoprotein E. J. Alzheimers Dis. 38, 1–9.

Small, B.J., Fratiglioni, L., Viitanen, M., Winblad, B., Backman, L., 2000. The course of cognitive impairment in preclinical Alzheimer disease. Three- and 6-year follow-up of a population-based sample. Arch. Neurol. 57, 839–844.

Sonni, I., Ratib, O., Boccardi, M., Picco, A., Herholz, K., Nobili, F., Varrone, A., 2017. The Geneva Task Force for the Roadmap of Alzheimer's Biomarkers, Maturity of presynaptic dopaminergic imaging with 123I-Ioflupane and noradrenergic imaging with 123I-MIBG in the differential diagnosis between Alzheimer's disease and Dementia with Lewy bodies in the context of a structured 5–phase development framework, Neurobiol. Aging 52, 228–242.

Sotaniemi, M., Pulliainen, V., Hokkanen, L., Pirttilä, T., Hallikainen, I., Soininen, H., Hänninen, T., 2012. CERAD-neuropsychological battery in screening mild Alzheimer's disease. Acta Neurol. Scand. 125, 16–23.

Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack Jr., C.R., Kaye, J., Montine, T.J., Park, D.C., Reiman, E.M., Rowe, C.C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M.C., Thies, B., Morrison-Bogorad, M., Wagster, M.V., Phelps, C.H., 2011. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 280–292.

Squire, L.R., Stark, C.E., Clark, R.E., 2004. The medial temporal lobe. Annu. Rev. Neurosci. 27, 279–306.

Tabert, M.H., Manly, J.J., Liu, X., Pelton, G.H., Rosenblum, S., Jacobs, M., Zamora, D., Goodkind, M., Bell, K., Stern, Y., Devanand, D.P., 2006. Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. Arch. Gen. Psychiatry 63, 916–924.

Ten Kate, M., Barkhof, F., Boccardi, M., Visser, P.-J., Jack, C.R., Lovblad, K.-O., Frisoni, G.B., Scheltens, P., 2017. The Geneva Task Force for the Roadmap of Alzheimer's Biomarkers, Maturity of medial temporal atrophy as a biomarker for Alzheimer's disease in the context of a structured 5-phase development framework, Neurobiol. Aging 52, 167–182.

Tierney, M.C., Szalai, J.P., Snow, W.G., Fisher, R.H., Nores, A., Nadon, G., Dunn, E., St George-Hyslop, P.H., 1996. Prediction of probable Alzheimer's disease in memory-impaired patients: a prospective longitudinal study. Neurology 46, 661–665.

Tounsi, H., Deweer, B., Ergis, A.M., Van der Linden, M., Pillon, B., Michon, A., Dubois, B., 1999. Sensitivity to semantic cuing: an index of episodic memory dysfunction in early Alzheimer disease. Alzheimer Dis. Assoc. Disord. 13, 38–46.

Tulving, E., 2002. Episodic memory: from mind to brain. Annu. Rev. Psychol. 53, 1–25.

van der Flier, W.M., Pijnenburg, Y.A., Fox, N.C., Scheltens, P., 2011. Early-onset versus late-onset Alzheimer's disease: the case of the missing APOE ε4 allele. Lancet Neurol. 10, 280–288.

Visser, P.J., Verhey, F.R., Hofman, P.A., Scheltens, P., Jolles, J., 2002. Medial temporal lobe atrophy predicts Alzheimer's disease in patients with minor cognitive impairment. J. Neurol. Neurosurg. Psychiatry 72, 491–497.

Vogel, A., Mortensen, E.L., Gade, A., Waldemar, G., 2007. The Category Cued Recall test in very mild Alzheimer's disease: discriminative validity and correlation with semantic memory functions. Eur. J. Neurol. 14, 102–108.

Wagner, M., Wolf, S., Reischies, F.M., Daerr, M., Wolfsgruber, S., Jessen, F., Popp, J., Maier, W., Hüll, M., Frölich, L., Hampel, H., Perneczky, R., Peters, O., Jahn, H., Luckhaus, C., Gertz, H.J., Schröder, J., Pantel, J., Lewczuk, P., Kornhuber, J., Wiltfang, J., 2012. Biomarker validation of a cued recall memory deficit in prodromal Alzheimer disease. Neurology 78, 379–386.

Welsh, K., Butters, N., Hughes, J., Mohs, R., Heyman, A., 1991. Detection of abnormal memory decline in mild cases of Alzheimer's disease using CERAD neuropsychological measures. Arch. Neurol. 48, 278–281.

Welsh, K.A., Butters, N., Hughes, J.P., Mohs, R.C., Heyman, A., 1992. Detection and staging of dementia in Alzheimer's disease. Use of the neuropsychological measures developed for the Consortium to Establish a Registry for Alzheimers Disease. Arch. Neurol. 49, 448–452.

Wolfsgruber, S., Jessen, F., Wiese, B., Stein, J., Bickel, H., Mösch, E., Weyerer, S., Werle, J., Pentzek, M., Fuchs, A., Köhler, M., Bachmann, C., Riedel-Heller, S.G., Scherer, M., Maier, W., Wagner, M. AgeCoDe Study Group, 2014. The CERAD neuropsychological assessment battery total score detects and predicts Alzheimer disease dementia with high diagnostic accuracy. Am. J. Geriatr. Psychiatry 22, 1017–1028.

Xie, J., Gabelle, A., Dorey, A., Garnier-Crussard, A., Perret-Liaudet, A., Delphin-Combe, F., Bathsavanis, A., Dauphinot, V., Lehmann, S., Mercier, B., Desestret, V., Roullet-Solignac, I., Vighetto, A., Krolak-Salmon, P., 2014. Initial memory deficit profiles in patients with a cerebrospinal fluid Alzheimer's disease signature. J. Alzheimers Dis. 41, 1109–1116.

Xu, W., Yu, J.T., Tan, M.S., Tan, L., 2015. Cognitive reserve and Alzheimer's disease. Mol. Neurobiol. 51, 187–208.

Zehnder, A.E., Bläsi, S., Berres, M., Spiegel, R., Monsch, A.U., 2007. Lack of practice effects on neuropsychological tests as early cognitive markers of Alzheimer disease? Am. J. Alzheimers Dis. Other Demen. 22, 416–426.