A Systematic Review of Longitudinal Studies Which Measure Alzheimer’s Disease Biomarkers

Emma Lawrence\textsuperscript{a,}\textsuperscript{*}, Carolin Vegvari\textsuperscript{a}, Alison Ower\textsuperscript{a}, Christoforos Hadjichrysanthou\textsuperscript{a}, Frank De Wolf\textsuperscript{a,b} and Roy M. Anderson\textsuperscript{a}

\textsuperscript{a}Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London, UK

\textsuperscript{b}Janssen Prevention Center, Leiden, The Netherlands

Handling Associate Editor: M. Arfan Ikram

Accepted 14 June 2017

Abstract. Alzheimer’s disease (AD) is a progressive and fatal neurodegenerative disease, with no effective treatment or cure. A gold standard therapy would be treatment to slow or halt disease progression; however, knowledge of causation in the early stages of AD is very limited. In order to determine effective endpoints for possible therapies, a number of quantitative surrogate markers of disease progression have been suggested, including biochemical and imaging biomarkers. The dynamics of these various surrogate markers over time, particularly in relation to disease development, are, however, not well characterized. We reviewed the literature for studies that measured cerebrospinal fluid or plasma amyloid-\textbeta (A\textbeta) and tau, or took magnetic resonance image or fluorodeoxyglucose/Pittsburgh compound B-positron electron tomography scans, in longitudinal cohort studies. We summarized the properties of the major cohort studies in various countries, commonly used diagnosis methods and study designs. We have concluded that additional studies with repeat measures over time in a representative population cohort are needed to address the gap in knowledge of AD progression. Based on our analysis, we suggest directions in which research could move in order to advance our understanding of this complex disease, including repeat biomarker measurements standardization and increased sample sizes.

Keywords: Alzheimer’s disease, biomarker, cross-sectional, dementia, longitudinal

INTRODUCTION

Alzheimer’s disease (AD) is characterized by progressive cognitive decline leading to dementia. It has been estimated that over 35.6 million people have dementia worldwide [1]. Continued high prevalence of AD [1] makes it a major public health issue, due to the high financial and emotional cost. AD is thought to be caused by neuronal death and brain atrophy [2]; this is often present alongside the accumulation of both tau tangles and amyloid-\textbeta (A\textbeta) plaques in the brain. Much debate has occurred about the order of events leading to neuronal death [3]; nevertheless, it is observed that both neurofibrillary tau tangles and A\textbeta plaques are present in the brains of AD patients postmortem [4]. Therefore, they have been widely employed as diagnostic markers of the disease and, concomitantly, possible quantitative measures of progression. Biomarkers, such as A\textbeta and tau monomers and oligomers, can be measured in the blood and cerebrospinal fluid (CSF).
Traditionally, measures of cognition have been used as clinical trial endpoints to assess treatments for AD [5, 6]. Current measures of cognitive decline have shown various degrees of sensitivity and specificity [7–9]. It has, however, been suggested that there is a neuropathological threshold beyond which any treatment will fail to affect cognition given the profound amount of brain atrophy developed [10]. Therefore, much research effort has been invested in preventative treatment, i.e. to stop neurodegeneration before it becomes too severe. Given that such a treatment will have to be administered prior to any signs of cognitive dysfunction in order for it to be effective, alternative clinical endpoints need to be established. The European Medicine Agency (EMA) and US Food and Drug Administration (FDA) have stated that the rate of disease progression could be linked to a biomarker indicative of underlying pathology [6, 11].

Biomarkers were added to the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer’s Disease and Related Disorders Association (ADRDA) (NINCDS-ADRDA) diagnostic criteria in 2011 [12]. Therefore, a variety of quantitative, AD-specific measures have been characterized that it is hoped can aid in diagnosing the disease. In addition to their use as endpoints, it has also been suggested that clinical trial populations can be enriched by including those individuals exhibiting a biomarker profile indicative of future conversion. Defining this preclinical stage itself is a topic of debate in the field [13]. Despite their use in diagnostics and potential use in clinical trial enrichment, knowledge about the dynamics of these biomarkers over time is limited. An understanding of exactly when these become meaningful prognostic markers is imperative in order for them to be clinically useful.

Many biological markers of AD pathology have been characterized to date. AD patients have significantly lower Aβ42 and higher t-tau and p-tau in the CSF compared to controls [14]. Total Aβ burden can also be assessed in the brain via Pittsburgh compound B (PiB)-positron electron tomography (PET) scanning [15]. In addition, marked brain atrophy in AD and all-cause dementia cases can be observed with magnetic resonance imaging (MRI) scanning [16, 17]. Decreased glucose metabolism as assessed by fluorodeoxyglucose (FDG)-PET is also a hallmark of the disease [18]. Various studies have described the differences in these biomarkers between cognitively normal and AD patients. A previous systematic review and meta-analysis by Olsson et al. [14] assessed the utility of CSF and blood based markers in distinguishing between those with AD and controls, in addition to mild cognitive impairment (MCI) due to AD and stable MCI. An earlier systematic review by McGhee et al. [19] used an expanded set of criteria to identify biomarkers of interest, including any biomarker that could be used to describe the progression of AD. The work presented in this paper focuses on the longitudinal use of classical Aβ and tau markers, as well as MRI and PET, in cohort studies. In order to make the selected studies comparable to other investigations, those which have taken cognitive functional measures have also been highlighted.

METHODS

Search terms

In order to ensure that studies relevant to our analysis were identified, we conducted a review of the literature. We searched the US National Library of Medicine National Institutes of Health in 2015 for articles published in English between January 1995 and August 2015. The search terms are listed below:


MeSH indexing is a system which places publications under categories of relevance; therefore, publications selected will be based on how they have been indexed. Cognitive testing was not used as a search term for this review, although methods used in the identified studies are discussed.
**Inclusion criteria**

We included studies which measured brain atrophy with MRI, amyloid levels with PiB-PET, tau levels as assessed by tau PET, CSF tau (phosphorylated tau or total tau), CSF Aβ (1–40, 1–42, or other variants), blood/plasma tau (phosphorylated tau, total tau, or antibodies to tau), blood Aβ (1–40, 1–42, or other variants or antibodies to amyloid-β), or autopsy data with tau or Aβ brain staining. Within this paper, we define “biomarker” as any of these measures, i.e., as a measure of biological state independent from clinical or cognitive measures. We included studies which measured 50 or more people at a minimum of 2 distinct points in time (longitudinal study).

**Exclusion criteria**

We excluded reviews and intervention studies (unless the placebo group fits the inclusion criteria). We excluded papers that were not written in English.

**Analysis**

Studies were identified through reading the methods section of the paper. Analysis featured in this paper was performed on longitudinal studies (cross-sectional studies were omitted from analysis). In certain instances, the study populations were not precisely identified in the paper, in which case the papers were excluded from the analysis (see Fig. 1). Further information about the study, for example, location, sample size, cognitive testing method, diagnosis method, was obtained from reading the methods section of the paper. When more than one paper was published per study, a selection of papers was used for data extraction, including the first published and most recently published papers. For very large studies with more than 10 papers, relevant websites were consulted for information. In this analysis, we defined “number of participants” as the number of participants who had completed at least two biomarker assessments. We categorized the studies into groups, based on the way participants were selected or recruited. These can be summarized as: i) those which randomly recruited from memory clinic admissions; ii) those who selected healthy participants but with an unrepresentative bias toward family history; iii) those which followed patients who had existing white matter changes; iv) those who recruited from an existing population cohort; and v) those who recruited groups of participants based on diagnostic status. We collated the age and age distribution from the studies, noting the mean age and standard deviation within the overall study population, as well as for each diagnostic group (cognitively normal, mild cognitively impaired, and AD individuals). If these values were not presented for the overall population, but rather for the distinct diagnostic groups, the mean age reflects the weighted average from the individuals within the diagnostic groups present in the study.

**RESULTS**

We identified 1,415 records after searching PubMed; of these, 233 met our inclusion criteria (Fig. 1). Given that any one individual study may have been considered in more than one publication, we attempted to identify those studies. In this case, we identified 70 longitudinal studies from the 233 publications. A total of 22 articles were omitted from analysis as the study name was not clearly identifiable. We therefore conducted the present analysis on the remaining 48 identified longitudinal studies as listed in Table 1.

**Study location**

Out of the 48 longitudinal studies identified, the majority took place in North America or Europe. Nineteen studies took place in the USA, many being multi-site studies as, for example, the ADNI study (see insert, Fig. 2c). Only 4 studies were conducted outside of North America/Europe, and these took place in Japan, Taiwan, and South Korea. Seven
<table>
<thead>
<tr>
<th>Study</th>
<th>Biomarkers measured</th>
<th>Number WITH 2 + measures</th>
<th>Follow-up length</th>
<th>No of measures</th>
<th>Location</th>
<th>Reference (identified from review)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAPET (Alzheimer’s Disease and Positron Emission Tomography)</td>
<td>MRI</td>
<td>50</td>
<td>5 y</td>
<td>4</td>
<td>South Korea</td>
<td>[146, 147]</td>
</tr>
<tr>
<td>AddNeuroMed</td>
<td>MRI</td>
<td>378</td>
<td>1 y</td>
<td>3</td>
<td>Europe</td>
<td>[104, 148]</td>
</tr>
<tr>
<td>ADNI 1 (Alzheimer’s Disease Neuroimaging Initiative)</td>
<td>CSF, blood, MRI, FDG, PiB-PET</td>
<td>800</td>
<td>6 y</td>
<td>7–9</td>
<td>USA</td>
<td>[28–110]</td>
</tr>
<tr>
<td>AIBL (The Australian Imaging, Biomarker &amp; Lifestyle Flagship Study of Ageing)</td>
<td>MRI, PiB-PET, blood</td>
<td>1100</td>
<td>4.5 y</td>
<td>3</td>
<td>Australia</td>
<td>[110, 149–159]</td>
</tr>
<tr>
<td>Biobank facilities of the Institute Born-Bunge</td>
<td>CSF</td>
<td>61</td>
<td>minimum 30 days</td>
<td>2</td>
<td>Belgium</td>
<td>[139]</td>
</tr>
<tr>
<td>BIOCARD</td>
<td>CSF, MRI, Plasma</td>
<td>349</td>
<td>17 y max</td>
<td>2.4 MRI pp</td>
<td>USA</td>
<td>[121, 122]</td>
</tr>
<tr>
<td>BLSA (Baltimore Longitudinal Study of Aging)</td>
<td>MRI</td>
<td>152</td>
<td>5.9 y</td>
<td>up to 10</td>
<td>USA</td>
<td>[103, 105, 116–118]</td>
</tr>
<tr>
<td>Bordeaux-3City</td>
<td>MRI</td>
<td>240</td>
<td>4 y</td>
<td>2</td>
<td>France</td>
<td>[160]</td>
</tr>
<tr>
<td>Brain Aging Project at the University of Kansas Alzheimer and Memory Program</td>
<td>MRI</td>
<td>109</td>
<td>2 y</td>
<td>2</td>
<td>USA</td>
<td>[161]</td>
</tr>
<tr>
<td>Cardiovascular Health Study (CHS) Cognition Study</td>
<td>MRI (plasma in subset)</td>
<td>2,101</td>
<td>7 y</td>
<td>2-3</td>
<td>USA</td>
<td>[162–168]</td>
</tr>
<tr>
<td>Cognitive Disorders Clinic at the National Hospital for Neurology and Neurosurgery, London, England</td>
<td>MRI</td>
<td>73</td>
<td>1 y</td>
<td>2</td>
<td>UK</td>
<td>[169]</td>
</tr>
<tr>
<td>Davis Alzheimer’s Disease Center longitudinal cohort</td>
<td>MRI</td>
<td>150</td>
<td>4 y</td>
<td>2</td>
<td>USA</td>
<td>[170]</td>
</tr>
<tr>
<td>DIAN (Dominantly Inherited Alzheimer Network)</td>
<td>CSF, PiB, FDG, PET, MRI</td>
<td>122</td>
<td>1.1–2.1 y</td>
<td>2+</td>
<td>USA, Australia, Europe, Asia, and South America</td>
<td>[111]</td>
</tr>
<tr>
<td>Framingham Cohort study</td>
<td>MRI</td>
<td>408</td>
<td>1-2 y</td>
<td>2</td>
<td>USA</td>
<td>[171–176]</td>
</tr>
<tr>
<td>German Dementia Competence Network</td>
<td>MRI</td>
<td>66</td>
<td>21 months (Average 627 days)</td>
<td>2</td>
<td>Germany</td>
<td>[177]</td>
</tr>
<tr>
<td>Göteborg MCI study</td>
<td>CSF, MRI, SPECT, EEG, Plasma</td>
<td>226</td>
<td>10 y</td>
<td>3</td>
<td>Sweden</td>
<td>[178–180]</td>
</tr>
<tr>
<td>IDADO (Improving the early Diagnosis of Alzheimer’s Disease and Other dementias)</td>
<td>MRI</td>
<td>71</td>
<td>2 y</td>
<td>2</td>
<td>The Netherlands</td>
<td>[181]</td>
</tr>
<tr>
<td>Intervention study – RAGE Aβ</td>
<td>CSF, MRI, plasma</td>
<td>133</td>
<td>18 months</td>
<td>3 MRI</td>
<td>USA</td>
<td>[182]</td>
</tr>
<tr>
<td>Intervention study – Gal-Int-11</td>
<td>MRI</td>
<td>174</td>
<td>2 y</td>
<td>2</td>
<td>The Netherlands</td>
<td>[183]</td>
</tr>
<tr>
<td>Intervention study - Alzheimer’s Disease Cooperative Study MCI Donepezil/Vitamin E trial</td>
<td>MRI</td>
<td>111</td>
<td>3 y</td>
<td>2 North America</td>
<td>[184]</td>
<td></td>
</tr>
<tr>
<td>Intervention study - Donepezil</td>
<td>MRI</td>
<td>93</td>
<td>1 y</td>
<td>2</td>
<td>Japan</td>
<td>[185]</td>
</tr>
<tr>
<td>Intervention study - Donepezil</td>
<td>MRI</td>
<td>88</td>
<td>1.5–2 y</td>
<td>2</td>
<td>USA</td>
<td>[186]</td>
</tr>
<tr>
<td>Intervention study - ELND005</td>
<td>MRI</td>
<td>82</td>
<td>6.5 y (78 months)</td>
<td>5</td>
<td>North America</td>
<td>[187]</td>
</tr>
<tr>
<td>Intervention study - Memantine</td>
<td>MRI</td>
<td>114</td>
<td>1 y</td>
<td>4</td>
<td>France, Germany, Switzerland, and the UK</td>
<td>[188]</td>
</tr>
<tr>
<td>Study</td>
<td>Modality</td>
<td>Duration</td>
<td>Duration Unit</td>
<td>Country</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-----------</td>
<td>----------</td>
<td>---------------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention study - VITACOG</td>
<td>MRI</td>
<td>2 y</td>
<td></td>
<td>UK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention study - Elan</td>
<td>MRI</td>
<td>10.9 months</td>
<td></td>
<td>5 countries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johns Hopkins Alzheimer’s Disease Research Center</td>
<td>MRI</td>
<td>3, 6, 12 months</td>
<td></td>
<td>USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knight Alzheimer Disease Research Center</td>
<td>PiB-PET</td>
<td>2.6 y mean</td>
<td></td>
<td>USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kuopio, Finland</td>
<td>Plasma</td>
<td>3 y</td>
<td></td>
<td>Finland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LADIS (Leukoaraiosis And DIStability in the elderly) prospective study</td>
<td>MRI</td>
<td>3 y</td>
<td></td>
<td>multinational</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malmo</td>
<td>CSF</td>
<td>1-2 y</td>
<td></td>
<td>Sweden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malmo University Hospital Memory clinic</td>
<td>CSF</td>
<td>6 months</td>
<td></td>
<td>Sweden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mayo Clinic Registry</td>
<td>MRI (plasma)</td>
<td>median 3.7 y</td>
<td></td>
<td>USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mayo Clinic Study of Aging (MCSA)</td>
<td>MRI, FDG PET, amyloid PET</td>
<td>median 1.3 y</td>
<td></td>
<td>USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melbourne Healthy Aging Study and Austin Health Memory Disorders Clinic</td>
<td>PiB-PET</td>
<td>3 y max</td>
<td></td>
<td>Australia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory Clinic at Huddinge University Hospital</td>
<td>MRI, SPECT, blood</td>
<td>mean 16 months</td>
<td></td>
<td>Sweden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory Clinic of Higashi Matsudo Municipal Hospital</td>
<td>MRI, MRI</td>
<td>3 y</td>
<td></td>
<td>Japan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIRIAD (Minimum Interval Resonance Imaging in AD)</td>
<td>MRI</td>
<td>1 y</td>
<td></td>
<td>UK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rush Alzheimer’s Disease Center (RADC) Clinic, the Religious Order Study &amp; the Rush Memory and Aging Project</td>
<td>MRI</td>
<td>5 y</td>
<td></td>
<td>USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OASIS (Open Access Series of Imaging Studies)</td>
<td>MRI</td>
<td>2 y (average 719 day)</td>
<td></td>
<td>USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OBAS (Oregon Brain Aging Study)</td>
<td>MRI</td>
<td>6.4 y average</td>
<td>5.8</td>
<td>USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitea River Valley Sweden</td>
<td>CSF</td>
<td>1 y</td>
<td></td>
<td>Sweden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taipei Veterans General Hospital</td>
<td>MRI</td>
<td>1 y</td>
<td></td>
<td>Taiwan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCSF Memory and Aging Center (MAC)</td>
<td>MRI</td>
<td>1 y</td>
<td></td>
<td>USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uppsala Longitudinal Study of Adult Men</td>
<td>Plasma</td>
<td>7 y</td>
<td></td>
<td>Sweden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vu University Medical Centre</td>
<td>CSF, MRI</td>
<td>2 y</td>
<td></td>
<td>The Netherlands</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington Heights-Inwood Columbia Aging Project</td>
<td>MRI</td>
<td>4.5 y</td>
<td></td>
<td>USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin Registry for Alzheimer’s Prevention</td>
<td>MRI</td>
<td>4 y</td>
<td></td>
<td>USA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
E. Lawrence et al. / Longitudinal Studies of Alzheimer’s Disease

Number of studies by country

- Multinational
- Australia
- Belgium
- Finland
- France
- Germany
- Japan
- South Korea
- Sweden
- Taiwan
- The Netherlands
- UK
- USA

Fig. 2. Study locations. Proportion of studies in each country. Countries in multinational studies included: France, Germany, Switzerland, and the United Kingdom. One article did not specify in which countries the study took place. One article specified multinational European study sites, 2 articles specified North American, and 1 study featured USA, Australia, Europe, and Argentina.

studies took place in more than one country, and these were not always specified. Figure 3 shows the countries which were identified in the studies.

Sample size

Another important feature of the identified studies was the size of the sample that had undergone repeat assessments. While there were a number of large studies, over a third (35.4%) had less than 100 participants (studies with under 50 participants were excluded, see methods). Five studies had over 500 participants and of these only 2 had more than 1,000. Results are summarized in Fig. 4. This trend for low numbers in AD studies is most likely due to the cost involved in a number of the procedures being employed, such as MRI scanning [20].

Recruitment and selection

Given the small sample sizes within the identified studies, the recruitment and selection criteria employed were assessed. We categorized the studies into groups, based on the way recruited participants were selected, as summarized in the methods. The results are outlined in Table 2. The majority (68.45%) of the studies identified (33/48) had selectively recruited patients based on their diagnosis status. Seven studies recruited from an existing population cohort, and although this was seemingly
representative of the original cohort, often the sample sizes were small – 5 out of 7 having a sample of under 300 individuals. The remaining methods of study categorization are summarized in Table 2. They refer to different features of the population or address specific scientific questions. In general, there were few studies following a representative population cohort. This is likely due to the costs, length of time, and logistics involved in such studies.

**Diagnosis methods used**

Given that a large proportion of studies recruited participants based on the diagnosis of the participant, we considered which diagnosis criteria were commonly used. There was a good degree of consensus on this, with 37/48 studies using either the original or revised NINCDS-ADRDA diagnostic criteria for AD or probable AD [12]. Other criteria utilized included use of the Clinical Dementia Rating, for example using cut-points in the rating scale; this was done in 8 out of the 48 studies. The Diagnostic and Statistical Manual of Mental Disorders IV and III were referenced in 14/48 studies. A number of studies used different diagnosis methods, such as use of cut-points in cognitive scales and undefined “consensus diagnosis procedure”.

The diagnosis of MCI was described in 12 studies, and the Peterson criteria [21] were cited in 5 instances. In other cases, tailored criteria were described. Various methods for describing a cognitively healthy cohort were used, often featuring cognitive scores in the healthy or normal range. Overall, there was good concordance in the diagnosis techniques employed in the identified study populations.

**Neuropsychological testing**

Neuropsychological and cognitive testing can have numerous limitations in the context of clinical trial endpoints, including insensitivity at early stages [22]. Many tests are designed specifically to diagnose or assess dementia and are not specific to AD. Additionally, the wide variety of tests available are not always comparable, for example, they measure differing aspects of cognition or function. From the 48 studies identified, over 90 different tests were referenced. These tests were done in various combinations in the different studies. There appeared to be no pattern in the selection of the tests, other than the grouping into domains such as memory, executive function or attention.

The Mini-Mental State Examination (MMSE) was the most commonly used test out of all the identified studies (64.6%). This is not surprising, given its ease and speed of administration. It was designed to differentiate dementia from other psychiatric illnesses. It has demonstrated specificity in diagnosing dementia, but is not sensitive enough to diagnose AD [23]. The Alzheimer’s Disease Assessment Scale (ADAS) is a battery of tests which are used to assess cognitive and non-cognitive dysfunctions in people with AD [5]. The cognitive arm (ADAS-Cog) is made up of 11 tasks, which assess memory, language, praxis, attention among other cognitive abilities. Despite its common use in clinical trials of treatments for symptomatic patients, only 9 out of the 48 studies referred to using the ADAS-cog assessment. Other articles referred to a “comprehensive battery” or did not describe the tests in detail in the methods sections of the publications. Surprisingly, 3 out of the 48 studies do not reference use of cognitive tests. We were unable to discern if this was because they were not performed or if they were omitted from the publication.

**Biomarkers measured**

Of primary importance in this review were the different biomarkers which were measured in the studies. CSF and blood biomarker studies have been comprehensively reviewed recently by Olsson et al. [14]; however, these did not discriminate those which had performed repeat measures. It is somewhat surprising, that in total, only 48 named studies had performed repeat biomarker measurements. Monitoring the levels of these markers is highly desirable within clinical trial design, improving diagnosis and developing prognostic tools among other things. Even within the 48 studies, a majority of them (28/48) (Fig. 5) had only used MRI scans of whole brain or hippocampal atrophy, which is not a measure specific
During our search, we observed that many studies had compared biomarker level at one point in time for diagnostic use only. However, these studies do not document the changes that occur over time. We assessed the number of repeat measures which had been taken from an individual (Fig. 6). Out of the 48 identified studies, 28 (58.33%) had performed measurements at just 2 separate times. There was often a lack of clarity in the literature, with a different number of repeats taken in different patients. This was due to various reasons, for example, death, drop out, and illness. Three studies (ADNI [28–110], BLSA [103, 105, 116–118] and MIRIAD [119, 120]) took more than 6 repeat measures. These were, however, limited to MRI measures of brain volume and atrophy, which are not specific measures of AD.

Another important feature of the follow-up measure is the amount of time between measurements. The average time between the initial and last measure taken in the studies is presented in Fig. 6. 26 out of the 48 studies took all measures in less than 2.5 years, representing a fairly short follow-up for a disease with such protracted development, often lasting decades. Only 10 studies return to patients at 5 years or more following the first measure. The longest described follow-up was 17 years in the BIOCARD study [121, 122]. We also found a lack of clarity surrounding repeat measures in the literature, with the precise timing of the protocols not being clearly described, with only average intervals for entire groups being presented. In most of the cases, the follow-up was not uniform between participants. It should be noted, however, that some studies were not complete at the time of writing, so the number of, and length of, follow-up measures may increase over time.

**Average age of participants**

It has been postulated that the preclinical phase of AD can begin as early as 40 years old or younger [13, 123]; however, profiles of AD biomarkers in young individuals have not been characterized frequently. The average age of the participants in the studies we identified is 71 years, with a median age of 73 years (Fig. 7). One study identified included two birth-year cohorts performed in Uppsala, Sweden, where men were recruited for a 70-year-old cohort and 77-year-old cohort [124]. Although 19 out of the 48 studies (39.6%) recruited participants who were 60 years old or younger, it was very rare for any participants to be under 50 years old. One of the only studies to assess young individuals was the DIAN Study, which specifically recruits young healthy individuals, from families with a history of AD, to assess the progression of Early Onset AD (EOAD). Benzinger et al. [111] presented findings for participants with an average age of 39.3 (SD 9.46) and 38.8 (SD 10.4). The other study which clearly utilized under 50 s was
the Wisconsin Registry for Alzheimer’s Prevention. Okonkwo et al. [125] also aim to study younger individuals from families with a history of AD.

**DISCUSSION**

Our aim was to review the literature to identify cohort studies which incorporated longitudinal measurements of AD biomarkers and MRI imaging. The work presented builds upon the previous systematic reviews performed by Olsson et al. [14] and McGhee et al. [19], and focuses on well-established AD biomarkers (Aβ, tau, and PET markers) in cohorts comprised of at least 50 subjects. It has been 6 years since Jack et al. first proposed the hypothetical temporal model of AD progression [126]. Understanding the dynamics of these measures is of obvious importance for planning clinical trials of possible therapies in preclinical patients. This will also facilitate the development of mathematical models for the prediction of AD development and progression. However, finding data to support such models has proven difficult. We have observed that there is a lack of studies which measure AD biomarkers in a representative population cohort over long periods of time. This work adds to the recently published review by McGhee and colleagues [20], with a focus on characterizing the age and geographical location of patients with repeat measurements over time. In general, our findings agree with those of McGhee et al., in that there is not sufficient research to support the adoption of any pathological biomarker, a system for selecting biomarkers, such as that proposed by McGhee et al., should therefore be adopted in the future.
Although there appears to be a bias in the location of AD studies toward Europe and North America, these areas are those shown to be most affected by the increasing prevalence of dementia [127]. It is therefore not surprising that these are the most studied regions. However, while the incidence of AD seems to be stable or perhaps slightly declining in developed countries (possibly due to improved cardiovascular health) [128], the incidence of dementia in developing countries is predicted to increase [1]. More studies in these countries are therefore desirable. The small sample sizes of the cohorts followed was striking, with over a third following just 50–100 patients. This is most likely due to the large costs of biomarker measurement. In the cost analysis presented by Silverman et al., which is estimated from Medicare values, an FDG-PET scan was listed as $1,661, an MRI as up to $1,294.17, and neuropsychological testing as just $84.33 [20]. It is clear that cheaper alternatives to these imaging scans are desperately needed. Indeed, efforts are now focused on the development of blood based biomarkers [129]. If this is achieved, they will enable much larger populations to be screened for preclinical markers.

We found a paucity of representative population cohorts that measure biomarkers over time. This implies that there are limitations to what can be concluded from the identified studies on the temporal dynamics of biomarkers. Such information is highly desirable in any assessment of possible therapies. The
most common method of recruitment for these studies is to select participants based on their diagnosis. This has led to a bias towards studying AD in patients who are either mildly cognitively impaired or those who have progressed to the disease state. There is therefore a lack of knowledge of the entire disease spectrum by age within the general population. Meaning that the process which causes AD may well be a part of aging rather than a pathogenic process. A number of the selected studies did not recruit an ethnically representative sample. For example, in the ADNI study, over 90% of participants had White Caucasian ancestry in comparison to the US population, where 63% have this ethnicity. In particular, Blacks, Asians, and Hispanics were underrepresented in this sample. Studies tended to exclude participants with health conditions other than AD. Given that AD is thought to be a disease of mixed pathology, and occurs concomitantly with other diseases of old age, it can be argued that studies of comorbid individuals are of real value. At present, they are scarce in the literature. It would be of further interest to compare the profile of early against late onset AD, with age as a central stratification in assessing the dynamics of change in biomarkers over time.

Braak et al. demonstrated that tau tangles and Aβ plaques can be identified in the brains of individuals from as young as age 20 and 40, respectively [123]. It is therefore of importance to connect this pathology with CSF, blood, and brain imaging markers. One important current challenge facing the development of preclinical preventative treatments for AD is assessing efficacy in a clinically healthy pre-onset population. This makes the dynamics of change in biomarker measures in younger populations of great importance. We found that the average age studied was 70–75 years old, with individuals under 50 rarely included in the studies identified. A relatively small number of existing studies have measured these biomarkers in young populations. Patenico et al. assessed a group of 21–63-year-olds for CSF biomarkers (Aβ, p-tau, and t-tau) and demonstrated that they were significantly different from older age groups [130]. Blomberg et al. assayed participants as young as 45 and found a positive correlation between CSF tau levels and age [131]. This was also demonstrated by Sjogren et al., who additionally set reference values for CSF tau in the age clusters 21–50, 51–70, and 71–93 years. They demonstrated different profiles in these age groups [132]. In order to build on these findings, larger populations of young cohorts should ideally be studied.

It was clear that the NINCDS-ARDRA criteria have been well adopted in published studies as these were widely used for diagnosis. This consensus in measurement techniques is important in drawing epidemiological comparisons between studies. Studies on variability in different clinician’s diagnosis (the recording of measurement error due to the person making the diagnosis) for the same patient at one point in time is of obvious importance. This has been done for the NINCDS-ARDA criteria; Lopez et al. found “fair” to “substantial” agreement between clinicians when they diagnosed 40 patients with blinded notes [133]. More recently, Khan et al. surveyed 2,618 patients and found high intra-class correlation for ADAS-Cog [134]. Repeated diagnosis by different clinicians at one point in time would also aid in our understanding of variance. Computerized tests would also help in reducing variance due to human subjectivity. A good understanding of variance in measurement (both biomarkers, cognitive tests and brain scans) is critical in evaluating useful clinical trial endpoints.

Progression of AD is frequently quantified by cognitive measures. As such they have been frequently used as an endpoint in published clinical trials [6]. However, we found that there is no consensus on which tests are most useful in monitoring disease progression, as a large variety of different tests were used in the identified studies (over 90). Tests were used in different combinations, making it difficult to compare results from the different studies. Furthermore, overall composite scores were calculated in different ways. The most commonly used test identified was the MMSE; however, this test has limited diagnostic accuracy [135] in addition to being an unreliable predictor of conversion to a disease state [136]. The ADAS-Cog also has a number of limitations including a non-linear relationship with disease progression and ceiling effects [137, 138]. Improvements in the sensitivity and specificity of these measures are therefore needed. At present, there is a lack of standardized cognitive measures in the field. Standardized cognitive measures, as well as a comprehensive understanding of the variance associated with them, is also critical to their use as clinical trial endpoints.

We found that MRI was performed far more frequently than any other biomarker measure. The main caveat associated with it, is that it is not a specific measure for AD, meaning that any abnormalities detected could signal other disease types [24, 25]. Amyloid-PET and FDG-PET scans are more specific...
to AD [26, 27], but were infrequently used in comparison to other measures. This is likely due to the more involved procedure as well as the need for expensive equipment. While CSF and blood were taken in a number of studies, different biomarkers were often measured with different assays, again leading to a lack of comparability between studies. Much work has been done on the diagnostic performance of CSF Aβ and tau. For example, Olsson et al. demonstrated that CSF T-tau, p-tau, Aβ42, NFL and plasma T-tau were strongly associated with AD [14]. However, future work should focus on the timing and order that these biomarkers become abnormal, to enable them to be more useful in a prognostic context. An understanding of the changes over time (i.e., dynamics by age and time) is also important for understanding measurement variance. CSF levels of Aβ1-42, T-tau, and P-tau were shown to be stable during a 30-day study [139], but longer time frames should also be considered. We found that it was most common to measure biomarkers at 2 points in time, with very few studies repeating measures more than twice. This can be attributed to the high costs of the measures used [20] or the invasiveness of CSF sample collection [140]. Earlier and more frequent measures would facilitate more robust conclusions to be drawn on the precise timing and etiology of AD pathology. In addition, they would help to define a pre-clinical AD profile to aid in the design of trials of possible therapies. However, it is recognized that because of the invasiveness of the procedure of sample collection patient compliance to repeated sampling may be understandably problematic.

As this work was not intended to be a full systematic review, some relevant studies may have been omitted from the analysis. More published articles could be screened, to identify studies excluded by our selection criteria. Several studies were omitted from our analysis as the study population was unclear in the published research. The study methods were also unclear in a number of papers, meaning that our findings may not be truly representative of the work completed. The search excluded all cause dementia studies, and this may have resulted in AD cases being missed from the analysis. It should also be noted that as MRI is not specifically an AD-related marker and that studies that have been identified in this review may have incorrectly classified non-AD dementia as AD, there may be further heterogeneity among the results. Tissue such as plasma and CSF may have been stored and not assayed in a number of studies. Future longitudinal analyses could, therefore, be performed on this, while the studies were not identified through reviewing the literature. Indeed, many longitudinal cohort studies freeze samples such as plasma for future analysis.

In general, there was a lack of clarity regarding reporting of the study methods in many articles, including the study location and populations sampled. Other aspects of the protocol such as recruitment criteria, cognitive tests used, and assessment schedules were sometimes not discussed in detail. A more thorough description of methods would help researchers draw comparisons between studies. Standardized reporting methods across cohort studies would help with this.

Despite the numerous limitations that presently exist within the field, there remains many new initiatives that aim to complete gaps in our current knowledge of the epidemiology of AD. The CHARIOT Register aims to enlist cognitively healthy participants from age 60 upwards, to be used as a basis for future research studies [141]. Dementias platform UK have bought together 31 UK based cohorts, in order to see if they can be useful to study dementia retrospectively [142]. The European Medical Information Framework-AD (EMIF-AD) also aims to connect Europe-wide cohorts for the benefit of AD research [143]. Other long term initiatives are taking place in Italy [144] and Japan [145] to follow-up healthy and diseased participants. Many of these new initiatives focus on collaboration between studies and data accessibility, so it is to be hoped that following a period of measurement standardization, the field will be in a good position to support clinical trials of possible therapies.

In summary, we have found that there are few studies that record longitudinal measures of AD biomarkers in well-defined and large cohorts of participants. This is of particular relevance to the development of preventative treatments, given the costs of trials that run over many years given the long period over which disease progression takes place. Therapies that induce small improvements in slowing progression would be better than no therapy at all – which is the current situation. To detect low efficacy in a relatively short time span of a few years in an as yet cognitively unimpaired population will require a much better understanding of the temporal dynamics of biomarker changes. Understanding of these temporal and age related changes will also help us to better understand the true etiology of this disease and therefore aid in the development of future treatments. At present, many new cohorts are being established, and
more data pooling across studies is taking place in order to improve our knowledge. But it is clear from this review that urgent needs include better standardization in measurement, more precise determination of measurement error, better longitudinal follow-up of participants, and larger study population sizes.

ACKNOWLEDGMENTS

This study was funded by the Janssen Prevention Center.

Authors’ disclosures available online (http://j-alz.com/manuscript-disclosures/17-0261r2).

REFERENCES


different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology* **62**, 591-600.


