Biomarkers for Alzheimer’s disease therapeutic trials

Harald Hampel a,e, Gordon Wilcock b, Sandrine Andrieu c,d,e, Paul Aisen f, Kaj Blennow g, K. Broich h, Maria Carrillo i, Nick C. Fox j, Giovanni B. Frisoni k, Maria Isaac l, Simon Lovestone m, Agneta Nordberg n,o, David Prvulovic a, Christina Sampaio p, Philip Scheltens q, Michael Weiner r, Bengt Winblad s, Nicola Coley c,d, Bruno Vellas c,d,t

for the Oxford Task Force Group

* Corresponding author. Tel.: +49 69630187300; fax: +49 69630187303.
E-mail address: harald.hampel@med.uni-frankfurt.de (H. Hampel).

1 Other members listed in Appendix A.

A B S T R A C T

The development of disease-modifying treatments for Alzheimer’s disease requires innovative trials with large numbers of subjects and long observation periods. The use of blood, cerebrospinal fluid or neuroimaging biomarkers is critical for the demonstration of disease-modifying therapy effects on the brain. Suitable biomarkers are those which reflect the progression of AD related molecular mechanisms and neuropathology, including amyloidogenic processing and aggregation, hyperphosphorylation, accumulation of tau and neurofibrillary tangles, progressive functional, metabolic and structural decline, leading to neurodegeneration, loss of brain tissue and cognitive symptoms. Biomarkers should be used throughout clinical trial phases I–III of AD drug development. They can be used to enhance inclusion and exclusion criteria, or as baseline predictors to increase the statistical power of trials. Validated and qualified biomarkers may be used as outcome measures to detect treatment effects in pivotal clinical trials. Finally, biomarkers can be used to identify adverse effects. Questions regarding which biomarkers should be used in clinical trials, and how, are currently far from resolved. The Oxford Task Force continues and expands the work of our previous international expert task forces on disease-modifying trials and on endpoints for Alzheimer’s disease clinical trials. The aim of this initiative was to bring together a selected number of key international opinion leaders and experts from academia, regulatory agencies and industry to condense the current knowledge and state of the art regarding the best use of biological markers in Alzheimer’s disease clinical trials and to propose practical recommendations for the planning of future AD trials.

Progress in Neurobiology xxx (2010) xxx–xxx

ARTICLE INFO

Article history:
Received 30 September 2010
Received in revised form 10 November 2010
Accepted 22 November 2010

Keywords:
Alzheimer’s disease
Biomarkers
Neuroimaging
Early diagnosis

Abbreviations: AD, Alzheimer’s disease; PET, positron emission tomography; CSF, cerebrospinal fluid; MRI, magnetic resonance imaging; fMRI, functional magnetic resonance imaging; EADC, Alzheimer Disease Consortium; ADCS, Alzheimer’s Disease Cooperative Study; ADNI, Alzheimer’s disease neuroimaging initiative; MCI, mild cognitive impairment; SPECT, single photon emission computed tomography.

© 2010 Elsevier Ltd. All rights reserved.

0301-0082/S – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pneurobio.2010.11.005
1. Introduction

Alzheimer’s disease (AD) is the most common form of dementia. It is clinically characterized by progressive deterioration of episodic memory and a global decline of cognitive functions ultimately leading to dependency on custodial care. AD is a multifaceted disease and several mechanistic and pathological substrates contribute to its chronic progression over the course of decades. Characteristic neuropathological hallmarks of AD are extracellular accumulation of fibrillar Aβ peptides, intracellular neurofibrillary tangles comprised of hyperphosphorylated tau protein and progressive reduction in the number of synapses, dendrites and neurons (Selkoe, 1994; Braak and Braak, 1995). Further converging molecular mechanisms and substrates include immunological alterations, inflammation, oxidative stress, microvascular changes, and excitotoxicity.

Despite the fact that currently existing drug therapies for AD cannot substantially improve the clinical and biological progression of the disease, there is still an urgent need to further optimize early detection of AD and to accurately assess biological treatment effects of candidate drugs. Actually, accurate trait – as well as highly sensitive and dynamic state – markers represent essential prerequisites for the labeling and for the development of disease-modifying drugs. Moreover, these markers would allow applying qualification and validation of biomarkers based on epidemiologic, neuroimaging may help to reduce unexplained variance, thereby increasing statistical power to detect treatment effects (Weiner, 2009). Furthermore, some biomarkers might better reflect AD progression or better predict clinical benefits of drug treatments than clinical measures, especially at early stages of the disease, and so could be used as surrogate endpoints for clinical trials. By definition, “a surrogate endpoint is a biomarker that is intended to substitute for a clinical endpoint” (Group, 2001). From a regulatory perspective, however, it is considered a general principle that if it is possible to ascertain effects of a drug on clinical outcome measures in trials of reasonable sizes and duration, biomarkers will not be acceptable as primary outcome measures. So the use of biomarkers as surrogate endpoints might only be considered in settings in which clinical outcomes cannot be practically assessed, e.g. in very early or preclinical stages of AD, when clinical outcomes may not occur for many years after treatment initiation. For such a situation careful qualification and validation of biomarkers based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence is required. It must be taken into consideration that a strong, independent and consistent association between surrogate endpoint and clinical outcome is necessary but not sufficient. There must be a link between a treatment-induced change in the biomarker and the desired clinical outcome measure, as well as a link between the treatment induced change in the biomarker and change of disease process (Katz, 2004; Baker and Kramer, 2003; Fleming and DeMets, 1996). In the past 10–15 years, biomarker development has substantially progressed. The development of highly specific immunoassays which can discriminate between various isoforms of Aβ or the development of functional imaging methods and highly sophisticated statistical analyses have helped to tremendously improve diagnostic sensitivity and specificity of used biomarkers.

1.1. Definition and use of biomarkers

The Biomarkers Definitions Working Group of the National Institutes of Health (Group, 2001) defined a biomarker, as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention”. Biomarkers have many potential uses in clinical trials, for example as outcome measures, as subject selection criteria, or as markers of disease processes. For exhaustive literature review of biomarkers for clinical trials see Hampel et al. (2010), Blennow et al. (2010), Hampel et al. (2008) and Jessen and Hampel (2009).

In addition to traditional clinical assessment combining cognitive and functional (e.g. activities of daily living) or global outcome measures (e.g. clinical global impression of improvement) have been sufficient for the development of symptomatic treatments that directly improve or stabilize cognition and function in individuals with AD in the short-term (Birks and Harvey, 2006; Birks et al., 2009; Loy and Schneider, 2006; McShane et al., 2006). However, the development of disease-modifying treatments, requiring trials with many subjects and long observation periods, is complicated by the considerable variance in cognitive and clinical assessments. The use of biomarker assessments in blood, cerebrospinal fluid or involving neuroimaging may help to reduce unexplained variance, thereby increasing statistical power to detect treatment effects (Weiner, 2009). Furthermore, some biomarkers might better reflect AD progression or better predict clinical benefits of drug treatments than clinical measures, especially at early stages of the disease, and so could be used as surrogate endpoints for clinical trials. By definition, “a surrogate endpoint is a biomarker that is intended to substitute for a clinical endpoint” (Group, 2001). From a regulatory perspective, however, it is considered a general principle that if it is possible to ascertain effects of a drug on clinical outcome measures in trials of reasonable sizes and duration, biomarkers will not be acceptable as primary outcome measures. So the use of biomarkers as surrogate endpoints might only be considered in settings in which clinical outcomes cannot be practically assessed, e.g. in very early or preclinical stages of AD, when clinical outcomes may not occur for many years after treatment initiation. For such a situation careful qualification and validation of biomarkers based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence is required. It must be taken into consideration that a strong, independent and consistent association between surrogate endpoint and clinical outcome is necessary but not sufficient. There must be a link between a treatment-induced change in the biomarker and the desired clinical outcome measure, as well as a link between the treatment induced change in the biomarker and change of disease process (Katz, 2004; Baker and Kramer, 2003; Fleming and DeMets, 1996). In the past 10–15 years, biomarker development has substantially progressed. The development of highly specific immunoassays which can discriminate between various isoforms of Aβ or the development of functional imaging methods and highly sophisticated statistical analyses have helped to tremendously improve diagnostic sensitivity and specificity of used biomarkers.

1.2. Functional imaging candidate biomarkers based on the ‘network paradigm’ to extend the conceptual framework of Alzheimer’s disease

Much AD research is based on a currently preferred hypothetical model positing that AD begins with the pathological initiation of the amyloidogenic cascade leading to Aβ accumulation in the brain, which leads ultimately to synaptic dysfunction, neurode-
generation, and cognitive/functional decline (Jack et al., 2010). This leads to the hypothetical notion that the earliest detectable pathological changes seem to be those related to Aβ (detected in cerebrospinal fluid (CSF) and/or by metabolic PET derived “amyloid imaging”). Along this line, Aβ-related biomarker candidates have gained center-stage for functions such as early detection, prediction and mapping of effects of anti-amyloid treatments on the brain.

Moreover, there is accumulating evidence promoting complimentary concepts for effective AD biomarker discovery including the development and validation of non-linear dynamic early “functional” biomarker candidates. The concept of Aβ-related inception of AD is currently being extended by a hypothesized pathophysiologic role of abnormal functional brain network coordination (based on the neural “network paradigm”). The question has been proposed whether aberrant neuronal coordination and activation in large-scale interconnected networks of the AD brain may not only reflect Aβ-related pathophysiology but also precede and even drive abnormal molecular and metabolic AD related mechanisms, including abnormal APP processing and Aβ accumulation in the brain (Palop et al., 2006; Palop and Mucke, 2010). This hypothesis is supported by accumulating evidence provided by experimental models and by findings in patients with epilepsy which collectively show that elevated levels of neural activation can indeed induce increasing Aβ production (Mackenzie and Miller, 1994; Kamnetz et al., 2003; Cirrito et al., 2005). Intriguingly, brain areas that are part of resting state networks, which generally display chronically high neural activation throughout lifespan, clearly represent AD pathologic predilection areas and are indeed particularly prone to AD-related molecular mechanisms. They include the very first neocortical regions which are affected by Aβ deposition in presymptomatic to clinically manifest AD (Buckner et al., 2005, 2009; Hedden et al., 2009). Studies investigating functional connectivity in resting state brain networks (Fox and Greicius, 2010) demonstrated abnormally high grades of neural connectivity in young subjects harboring increased genetic risk for late-onset AD (LOAD) (Filippini et al., 2009), indicating altered activation and coordination within these networks. These functional changes most likely reflect adaptive neuroplasticity dynamics. However, it has so far not been determined at which point (of duration or magnitude) these mere adaptive changes may turn into maladaptive changes and into disease-propagating factors and finally chronic pathological feedback cycles. Conversely, functional network connectivity successively breaks down to subnormal levels with advanced brain fibrillar Aβ deposition—even in patients affected by presymptomatic (Hedden et al., 2009) and preclinical (Sorg et al., 2007) stages of AD. This currently proposed extended AD pathophysiological concept may represent a paradigm shift in our understanding of the pathophysiology and the biological course of neurodegenerative disorders in general, stressing the interplay between (1) a set of risk factors (e.g. genetic, etc.), (2) non-linear, dynamic functional adaptation and consecutive maladaptation within vulnerable large-scale brain interconnected networks and (3) molecular events of pathophysiological relevance (such as Aβ accumulation), ultimately leading to progressive neurodegenerative and increasingly irreversible damage to brain structures (Fig. 1). The exact interrelations between these pathophysiologically relevant hypothesis-driven strands await further evaluation. The attractive advantage of this extended hypothetical concept is that measures of neuronal coordination may lead to more disease reality-fitting, sufficiently complex biomarker solutions, allowing one to detect earliest (perhaps even pre-amyloidogenic) stages of the disease and to improve early detection and prediction of AD. Because of their true functional nature and inherent dynamics, these functional coordination candidate biomarkers may also be particularly helpful to rapidly assess and track biological effects of symptomatic as well as of disease-modifying compounds on functionally relevant brain networks. However, these functional imaging markers are still in their infancy and need further rigorous development and confirmation by population-based-studies and longitudinal studies, as well as in multicenter validation studies.

Subsequently, neurodegeneration is detected by a rise of CSF tau species, synaptic dysfunction (measured by FDG-PET), and neuron loss indicated by atrophy, most notably in medial temporal lobe (measured with structural MRI). The temporal sequence of changes in Aβ deposition, CSF tau species, and imaging using FDG-PET and MRI remain to be determined. These changes ultimately lead to memory loss, general cognitive decline and eventually dementia. Expression of each element of AD pathology (e.g. Aβ and tau deposits, atrophy) is influenced by many modifying factors including age, APOE genotype, and cerebrovascular disease (white matter lesions detected by fluid attenuated recovery (FLAIR MRI)) and microbleeds (detected by T2* MRI) and there are expected to be wide differences among individuals. It should be emphasized that the above stated model is simply a model which needs to be tested and verified.

1.3. Timing and other influencing factors of biomarker use

Disease modifying drugs are likely to be most effective in the earlier stages of AD, before neurodegeneration is too severe and widespread, so trials for this type of drug will need to include AD cases in the earlier stages of the disease (Blinnow and Zietzerberg, 2009). Validated biomarkers that could enable accurate identification of AD pathology at an early stage would be of great use (Dubois et al., 2007). Alternatively, baseline biomarker measurements can be used for enrichment and stratification in proof-of-concept studies, as well as for supporting go/no-go decision making of phase III trials.

Biomarkers should be used in all stages of drug development including phase I, phase II and phase III. They can be used to enhance inclusion and exclusion criteria, for stratification or as baseline predictors to increase the statistical power of trials. Biomarkers can also be used as outcome markers to detect treatment effects. Particularly, if biomarkers are intended to be used as surrogate endpoints in pivotal studies, they must have been qualified to be a substitute for a clinical standard of truth and as such reasonably predict a clinical meaningful outcome. Finally, biomarkers can be used to identify adverse effects.

Nevertheless there are several pitfalls to be faced in the interpretation of biomarker data in AD drug development, such as the fact that biomarkers may be non-specific to AD, it may not be feasible to measure them in the appropriate system (i.e. the central nervous system) and the risk of over-interpreting biomarker data in phase II trials if statistical significance levels are not adjusted for multiple comparisons (Aisen, 2009). Failure to consider these issues could contribute to false conclusions and costly errors.

The Oxford Task Force continues and expands the work of our previous international expert task forces on disease-modifying trials and on endpoints for Alzheimer’s trials (Vellas et al., 2007, 2008). The aim of this Task Force was to bring together a selected number of experts from academia, regulatory agencies and industry to share experience on the use of biomarkers in AD therapeutic trials, condense the current knowledge and state of the art regarding the use of biological markers in AD therapy trials, and propose practical recommendations for the planning of future AD trials.

2. Methods

Under the auspices of the European Alzheimer Disease Consortium (EADC), a network of expert centers in the field of AD (funded by the European Union: 5th FP
QLAM 2001-00003), and in collaboration with US colleagues from the Alzheimer's Disease Cooperative Study (ADCS), we organised a Task Force to propose an international position paper on the use of biomarkers for Alzheimer's trials. The Oxford Task force members were carefully selected because of their role as key opinion leaders and experts in the academic or regulatory sectors, or due to their experience in the pharmaceutical industry. After an extensive literature search (published papers were identified via a Medline search, clinicaltrials.gov and the authors' experience and contacts in the field), the experts selected by the organizing committee (SA, CS, BV, GW) were asked to write a comprehensive review of methodological aspects relating to biochemical and neuroimaging biomarkers to be taken into consideration for trials in the field of AD.

Papers (Wischik and Staff, 2009; Fox and Kennedy, 2009; Gispen-de Wied et al., 2009; Lovestone and Thambisetty, 2009; Nordberg, 2009; Saumier et al., 2009; Weiner, 2009; Blennow and Zetterberg, 2009; Dubois, 2009; Hampel and Broich, 2009) were circulated to all members before the Task Force meeting that was held in Oxford in January 2009. Each member was also asked to list the main questions that he or she thought should be answered at the meeting. Of the questions that were suggested, the organizing committee selected three main questions to be answered: (i) What is the value of biomarkers for AD Drug Trials? (ii) What have we learned from recent trials? (iii) What can we recommend for future trials?

At the meeting, after general presentations, thematic groups met to consider specific responses for plasma and CSF biomarkers, neuroimaging and cost issues. Recommendations were presented to the Task Force for general discussion. The conclusions that were reached regarding these questions are presented below.

### 3. Value of biomarkers for AD drug trials

Biomarkers can be used in AD trials in a number of different ways. First, they can be used diagnostically, together with clinical and cognitive data, as inclusion and exclusion criteria. For example, in a phase I or phase II study of a treatment aimed at reducing brain amyloid (e.g. passive or active vaccine or secretase inhibitor), amyloid PET scanning or CSF Aβ measurements (low CSF Aβ reflects high brain amyloid load; Fagan et al., 2006) could be used to select subjects with high brain amyloid load. Another example would be a study of subjects who are either normal or who have very mild memory complaints, to identity subjects with brain amyloid who are likely to be at high risk for progression. In such a use, subjects would be enrolled in the study and subjected to a number of biomarkers studies and depending on the result, then patients would then be enrolled into the treatment portion of the study while other subjects would be excluded.

Second, biomarkers could be used as "baseline covariates" or "predictors". It is now established for example (in particular from the Alzheimer's disease Neuroimaging Initiative (ADNI) study) that MRI (brain atrophy), FDG-PET, CSF Aβ measurements and amyloid PET all “predict” future conversion from MCI to dementia (Anchisi et al., 2005; Chetelat et al., 2003; Okello et al., 2009; Risacher et al., 2009; Vemuri et al., 2009). Therefore, baseline biomarker measurements of this type could be incorporated into linear models as baseline covariates, and would thus increase the statistical power to detect treatment effects and either decrease the sample size or decrease the required length of the study.
Thirdly, biomarkers can be used as outcome measures, in particular, to detect disease-modifying treatment effects, as opposed to just symptomatic effects, which would be difficult to demonstrate with clinical measures alone. The use of biomarker outcomes may also enable the detection of treatment effects at an earlier stage than clinical measures alone.

Finally, biomarkers can be used to detect adverse effects such as inflammation, immunological reactions, microbleeds, vasogenic edema or other effects (Gilman et al., 2005; Salloway et al., 2009).

3.1. Structural imaging biomarkers

Structural imaging is used to evaluate structural brain changes related to AD. In particular, cerebral atrophy, which reflects neuronal loss, can be assessed using MRI, but three-dimensional cortical thinning or decreased cortical gray matter density could also be potentially useful structural imaging markers (Thompson et al., 2003).

There are a number of potential uses of structural imaging in AD trials. Firstly, it can be used in a diagnostic capacity in order to enrich trial populations with pure AD cases. Structural imaging has long been part of the entry criteria for AD treatment trials (Scheltens et al., 2002), in particular to exclude conditions such as a brain tumor, a hematoma, and more importantly vascular causes of dementia, thus creating a more homogeneous trial population with a higher proportion of subjects with AD as their primary pathology (Fox and Kennedy, 2009). More recently, MRI has been incorporated in trial inclusion criteria, for example to identify medial temporal lobe atrophy which is predictive of progression from MCI to AD (Desikan et al., 2008; Devanand et al., 2007; Jessen and Hampel, 2009). However, there are disadvantages in using MRI to select the study population, including cost and delays to recruitment. Furthermore, it may be argued that the trial population is no longer representative of typical AD patients if all patients with detectable vascular components are excluded (Fox and Kennedy, 2009).

However, perhaps the most interesting use of structural imaging in AD trials is as an outcome measure, i.e. a marker of disease progression in trials of disease-modifying drugs able to provide a means of identifying disease modification as distinct from symptomatic effects (Fox and Kennedy, 2009; Jessen and Hampel, 2009). To date, MRI measures have not been accepted as surrogate outcomes in AD—and are likely to require multiple results from disease-modifying trials to do so (Fox and Kennedy, 2009; Gispen-de Wied et al., 2009). The most established markers of progression on MRI are hippocampal and whole brain atrophy rates (Fox et al., 2000; Jack et al., 2004). At the group level, both distinguish patients with AD from controls, correlate with clinical decline and predict progression to AD (Henneman et al., 2009; Jack et al., 2005). The hippocampus is a particularly attractive marker of progression because of its early pathological involvement and its early, disproportionate and progressive atrophy on MRI (Braak and Braak, 1998). In mild AD (e.g. MMSE > 20), the hippocampus is reduced in volume by 15–25% relative to controls, and mean hippocampal atrophy rates are typically around 3–6% per year; this contrasts with rates of only 1–2% per year in normal elderly subjects (aged around 70–80 years) (Barnes et al., 2009; Morra et al., 2008; Schott et al., 2005). Interestingly, hippocampal atrophy is believed to be detectable through MRI about 5 years before whole brain atrophy (Ridha et al., 2006). The variance in rates of loss depends on the heterogeneity of the subjects recruited, measurement technique and error and critically on the duration of the study: variance is greater with rates calculated from shorter inter-scan intervals especially for intervals of under 12 months duration (Schott et al., 2005). Typically reported standard deviations of AD hippocampal atrophy rates are around 2.5–3.5% per year for a one year study falling to 2–2.5% per year for an 18–24-month study (Fox and Kennedy, 2009). Most studies to date have used manual outlining of the hippocampus; more recently semi-automated measures have been used or proposed (Schuff et al., 2009; de Pol et al., 2007; Teipel et al., 2010b,c). Fully automated (template-based) measures have also been developed but have not to date been used in large trials (Khan et al., 2008; Morra et al., 2008).

Recently, besides the medio-temporal lobe structures, volume reduction of another central region of interest, the basal forebrain (N. basalis Meynert), one of the most important structures of the AD pathology affected cholinergic system, discriminated between patients with AD, subjects with mild cognitive impairment and healthy controls (Teipel et al., 2010b,c, 2005).

Rates of whole brain atrophy have been incorporated into a number of trials in AD. Whole brain measures have the advantage that all brain structures affected by the disease process are represented by the measure, but at the cost of including areas that may not be subject to the relevant pathology. Global measures thereby avoid focussing on a particular region of interest and potentially missing important therapeutic effects elsewhere. Rates of whole brain atrophy in AD are typically 1.5–2.2% per year, while normal aging rates of atrophy (for a mean age of 70 years) are around 0.3% per year (Fox et al., 2005). As with hippocampal atrophy, it is this differential between disease-related atrophy and normal aging which offers the possibility of providing evidence of a disease-modifying effect of therapy. The standard deviation of the AD brain atrophy rate is typically ~1% per year for a single center study of 1 year or more (Fox and Kennedy, 2009). Sample size for brain and hippocampal measures are similar with 100–200 subjects needed for 90% power to detect a 20% slowing of atrophy (Schott et al., 2005; Schuff et al., 2009). It has been recently proposed that taking the effect of baseline covariates into account (ADAS-Cog, hippocampal volume, and ApoE genotype) might decrease sample size to around 50 per arm (Schuff, 2009). The development of software that can provide automated or semi-automated measurements of region volumes or cortical thickness, has led to these measures being suggested as outcomes in disease-modifying trials (Desikan et al., 2008; Du et al., 2004; Lerch et al., 2005; Morra et al., 2008; Lerch et al., 2008; Querbes et al., 2009). Multiple regions may be measured which avoids the limitations of one global or regional measure; nonetheless this raises issues for trial designs that typically pre-specify a single outcome and may require a step-down approach to several measures.

The use of MRI in more recent trials has also proved to be important for safety outcomes and to be a sensitive marker of effects that may, at least initially, be clinically silent (Salloway et al., 2009). Typically, MRI for safety includes a FLAIR sequence (sensitive to immunological reactions, inflammation, infarction and vasogenic edema) susceptibility-weighted imaging (to detect micro-hemorrhages) and some studies incorporate a diffusion sequence (Fox and Kennedy, 2009). The use of these markers is in their relatively early stages; methods of assessments are still somewhat variable and lack consensus guidelines (Fox and Kennedy, 2009).

The safety, availability, reliability, and relative affordability of MRI make it feasible for large trials and image processing and analysis can be performed at a central site, thus reducing measurement variation (Scheltens et al., 2002; Fox and Kennedy, 2009). Whatever the imaging marker, however, these will enter clinical trials only after an accurate tuning of the procedures for data acquisition and quality control in multi-center settings. Such an effort has been undertaken in the US with the ADNI project (Mueller et al., 2005) and has expanded to Europe (Frisoni et al., 2008), Japan (Iwatsubo et al., 2006), and Australia (Ellis et al., 2009). Other ADNI-independent large-scale multi-center dementia
validation networks provide supportive evidence on the feasibility and reliability of neuroimaging-based AD biomarkers (Ewers et al., 2006; Teipel et al., 2010b,c).

As a promising complimentary future extension to established core structural biomarkers, functional neuroimaging parameters that assess non-linear dynamic functional states of large-scale brain interconnectivity networks are particularly attractive research targets (Bokde et al., 2009). For example, measures of resting state network connectivity have the potential for exceptional diagnostic sensitivity and specificity (up to 95–100%) (Koch et al., 2010) at all but particularly early disease stages. They may also be of particular relevance when combined with measurements of microstructural integrity of white matter fiber tracts which represent the structural basis of large-scale networks in the brain (Teipel et al., 2010a). However, the functional MRI (fMRI) and neuropsychology (EEG, MEG) research field in AD yet requires standardization and validation of methods (Ewers et al., in press).

3.2. Molecular imaging markers

Molecular imaging opens up new possibilities for early diagnosis as well as evaluation of drug mechanisms and treatment efficacy in AD. The rapid development of different positron emission tomography (PET) amyloid imaging ligands has increased the possibility of measuring amyloid plaques in the brain of AD patients, and the recent development of F18-labeled amyloid-binding ligands such as AV-45 allows more widespread utilization of amyloid imaging than is feasible with 11C-PIB. A major drawback of 11C-PIB is its short half-life, which requires advanced equipment of the PET imaging center (cyclotron) and limits its availability. In addition, functional imaging using PET, single photon emission computed tomography (SPECT) or fMRI allow the effects of therapy on brain function to be assessed through measurement and mapping of cerebral glucose utilization (FDG-PET), perfusion (SPECT, fMRI) or activation (fMRI) and has been proposed as outcome measures in AD trials (Alexander et al., 2002).

A number of imaging modalities offer other non-invasive ways to assess brain changes associated with AD: micro-structural changes may be detected with magnetization transfer or diffusion imaging and cerebral metabolite levels may be measured with MR spectroscopy (MRS) (Kantarci et al., 2007; Krishnan et al., 2003; Glodzik et al., 2008).

Functional imaging can be used in a diagnostic capacity in clinical trials for enrichment of study populations. FDG-PET shows a typical pattern of reduced cortical uptake in the region of the temporal and parietal association cortex in AD patients. MCI subjects already show, to a lesser extent, a similar distribution of metabolic deficits which, in one study, predicted conversion from MCI to AD with an accuracy of over 80% (Hampel and Broich, 2009). Therefore, FDG-PET is of particular use in the in vivo diagnosis of early stages of AD (Hampel et al., 2008; Silverman et al., 2002; Teipel et al., 2008). Furthermore, recent results from the ADNI program show that FDG-PET is a powerful predictor of future cognitive decline in MCI (Landau et al., 2009).

Molecular imaging with 11C-PIB, the most studied amyloid PET ligand so far, also shows a robust difference between mild AD patients and healthy controls (Nordberg, 2008, 2009; Nordberg et al., 2010; Forsberg et al., 2010). Amyloid PET scanning could be used to identify subjects with brain amyloid, especially in phases I and II studies. In addition, high PIB retention has been observed in subjects with mild cognitive impairment (MCI) who later will convert to AD (Forsberg et al., 2008) and in cognitive normal elderly subjects who progress to AD (Morris et al., 2009).

Mapping of cerebral glucose utilization (FDG-PET), perfusion (SPECT, fMRI) or activation (fMRI) have been proposed as outcome measures in AD trials (Alexander et al., 2002; Wischik and Staff, 2009; Petrella et al., 2009). In the past, quantitative FDG-PET studies under activation have been successfully used to track effects of anti-dementia treatments on cerebral rate of glucose metabolism (Teipel et al., 2006). FDG-PET appears sensitive to clinical change in the mild to moderate stages of AD (Engler et al., 2006; Reiman et al., 2009) but also to brain functional changes caused by drugs with no disease-modifying effect (Kadir et al., 2008a,b). While this property may make it difficult to distinguish disease-modifying from symptomatic effects over short intervals, sustained changes may provide support for preservation of synaptic function. Furthermore, functional imaging might be useful in trials of drugs with combined symptomatic and disease-modifying effects (Doody et al., 2008); if used in combination with structural MRI, functional imaging changes might support claims of a combined effect.

PIB retention appears to reach a plateau in early stages of AD with no further significant progression over time (Engler et al., 2006; Nordberg, 2008; Jack et al., 2009; Scheinin et al., 2009). Even a 5-year follow-up study found no further increase in brain PIB retention in AD patients (Kadir et al., in press) A stronger correlation has been observed between cognition and cerebral glucose metabolism than between cognition and PET amyloid imaging (Kadir and Nordberg, 2010; Nordberg, 2008). 11C-PIB retention in the brain of AD patients has been demonstrated to negatively correlate with CSF levels of Aβ 1–42 (Forsberg et al., 2008). Only one PET tracer, 18F-FDDNP, has so far been claimed to label neurofibrillary tangles. It is expected that this field will continue to advance. Thus presently PIB may not be suitable as an outcome measure, but its application in treatment studies to investigate amyloid-modifying strategies as a marker of a biological mechanism is conceivable (Hampel et al., 2008; Teipel et al., 2008). Multicenter validation of amyloid-related PET-imaging radiotracers with longer half-life, such as fluoride-labeled compounds, is currently necessary.

11C-PMP PET has been used to assess acetylcholinesterase activity in the brain in AD patients, which is particularly useful for measurement of the effectiveness of acetylcholinesterase inhibiting compounds in patients in vivo (Kadir et al., 2008a).

3.3. Cerebrospinal fluid biomarkers

Biochemical changes in the brain extracellular fluid are reflected in the CSF. Thus, the measurement of AD biomarkers such as beta-amyloid and tau in the CSF should reflect related brain pathology, i.e. amyloid plaques and tangles of hyperphosphorylated tau.

CSF biomarkers may have a key role in AD trials in the baseline evaluation of patients eligible for the trial and as diagnostic markers to enrich the patient sample with pure AD cases. Another use of CSF biomarkers in such trials is for patient stratification. For example, patients with biomarker evidence of a disturbance in Aβ metabolism or deposition, such as low CSF Aβ42, may show a better effect of anti-Aβ disease-modifying drug candidates than those with normal CSF Aβ42 levels (Blinnlow and Zetterberg, 2009). The three core candidate CSF biomarkers T-tau, P-tau and Aβ42 have been extensively evaluated in numerous studies. Very consistently, all studies have found a marked increase in both CSF T-tau and P-tau accompanied by a marked decrease in Aβ42 in AD cases with dementia, for review see (Blinnlow and Hampel, 2003; Blinnlow, 2005; Hampel et al., 2010). The diagnostic performance of these CSF biomarkers to discriminate AD from non-demented aged individuals is high, with sensitivity and specificity figures of 80–90% (Blinnlow and Hampel, 2003). Normal CSF levels are found in several important differential diagnoses, such as depression and Parkinson’s disease, and in particular, P-tau helps to differentiate AD from other dementias, such as frontotemporal dementia and Lewy body dementia (Hampel et al., 2004).
Several studies have similarly evaluated the performance of core CSF biomarkers alone and CSF biomarkers in combination with neuroimaging and genetic information (multi-modal prediction model) to identify incipient AD in patients with MCI (Blinnlow and Hampel, 2003). Recent studies with extended clinical follow-up periods show that the combination of all three core CSF biomarkers (T-tau, P-tau and Aβ42) may have a predictive value as high as 95% to differentiate MCI cases with progression to AD from stable MCI cases and MCI cases with other types of underlying pathology (Hansson et al., 2006). However, a multicenter study found that while a combination of Abeta42/P-tau ratio and T-tau identified incipient AD in MCI subjects with reasonable sensitivity (83%), the specificity and positive predictive value were lower (72% and 62%, respectively) than in single-center studies (Mattsson et al., 2009).

Two population-based studies have found a significant reduction in CSF Aβ42 in cognitively normal elderly people that later developed AD, while there was no significant change in CSF T-tau or P-tau (Gustafson et al., 2007; Skoog et al., 2003). A recent clinical study also found that CSF Aβ42, but not T-tau and P-tau, predict cognitive decline in healthy elderly (Stomrud et al., 2007). These data show that CSF biomarkers, especially CSF Aβ42, may predict preclinical AD in cognitively normal elderly individuals.

The usefulness of emerging CSF biomarker candidates for amyloid-related clinical trials such as CSF activity and concentration of beta-site amyloid precursor protein cleaving enzyme 1 (BACE1) needs to be better elucidated (Zhong et al., 2007; Ewers et al., 2008).

There is variation in CSF biomarkers levels between research centers (Mattsson et al., 2009). This variation is likely due to confounding factors such as absorption of proteins to test tubes or due to loss of proteins during processing of samples, and differences in assays and laboratory procedures. Thus, there is a need for standardization of all procedures, from routines for lumbar puncture to handling and transportation of CSF samples. Importantly, variation can be kept to a minimum if all analyses are performed in one specialized laboratory after the completion of the study, and paired samples are analyzed side-by-side on the same plate. Indeed, if such precautions are taken there is a very low intra-individual variability of T-tau, P-tau and Aβ42 levels in studies with longitudinal CSF samples during 6 months and 2 years (Blinnlow et al., 2007; Zetterberg et al., 2007). This paves the way for the use of CSF biomarkers in AD clinical trials on disease-modifying drugs. International multi-center reliability studies using core, feasible biomarker candidates from CSF have been successfully concluded (Buerger et al., 2009a,b).

Finally, another application of CSF biomarkers in clinical trials is as safety measures, to enable early and specific detection of side-effects of the drug. Non-AD specific CSF biomarkers of inflammation or infection, for example, would be most suited to this purpose. Validation of the diagnostic value of CSF biomarkers (and of biomarkers in general) does benefit from application of markers in cases that allow for later post-mortem histopathological confirmation of diagnosis (Clark et al., 2003).

3.4. Plasma biomarkers

Plasma is a potential source of biomarkers for neurodegenerative changes in the brain, such as beta-amyloid, since CSF is absorbed into the blood on a regular basis. Furthermore, damage to the blood–brain barrier, which occurs in AD, may facilitate this process (Zipser et al., 2007). The main advantage of plasma biomarkers for clinical trials is that samples can be more easily obtained from patients than CSF which requires a lumbar puncture. Plasma biomarkers have often been considered in a diagnostic capacity or for identifying subjects likely to develop AD, with most studies focusing on plasma Aβ. However, results remain contradictory (Ewers et al., 2010; Hampel et al., 2008; Schneider et al., 2009), and the diagnostic value is not considered to be very strong at the current time. Many other hypothesis-driven biomarkers in plasma have also been proposed from studies using pre-determined protein arrays (i.e. biomarkers of microvascular change) (Buerger et al., 2009b; Ewers et al., 2010) or various exploratory proteome-based methodologies (Lovestone et al., 2007), with some, especially acute phase and inflammatory markers (Ray et al., 2007), suggesting high sensitivity and specificity as diagnostic markers and markers of progression, and others (Hye et al., 2006; Cutler et al., 2008; Akuffo et al., 2008), suggesting potential uses as outcome measures.

Generally, plasma biomarkers have thus far not been shown to be useful in AD studies (Lammfelt et al., 2008), and research is still mainly at an exploratory phase (DeMattos et al., 2002). Nonetheless, the development and validation of plasma-based biomarkers would be a very significant advance, allowing more widespread and repeated use in both trials and clinical practice, and so collection, and duration of biological material including blood and cells for RNA, should be incorporated wherever possible into trial protocols.

Finally, a number of genotypes has been found to correlate with AD progression and can thus be used in addition to plasma, CSF, and imaging biomarkers in order to improve homogeneity of diagnostic groups. Moreover, genetic (and epigenetic) markers may independently interact with other biomarkers (Vemuri et al., 2010) and with effects of disease-modifying compounds (Salloway et al., 2009). While the assessment of the ApoE genotype status is generally useful in clinical trials, other genetic factors are still under investigation (Zetzschke et al., 2010).

3.5. Lessons learned from recent trials

Despite a wide range of possible uses of biomarkers in AD clinical trials, much of the published data reports the use of biomarkers as outcome measures (Table 1). They have been employed as an adjunct to the more traditional outcome measures that are now long-established in drug evaluation protocols. For a successful drug, surrogate outcomes could be helpful in routine treatment by providing evidence for a drug effect in a biochemical or other parameter before clinical evidence of disease modification is apparent.

Although most AD clinical trials have centered on amyloid load reduction strategies, there are other therapeutic approaches, including inhibitors of tau aggregation, e.g. methylthioninium chloride (MTC) (Citron, 2010). A phase II study of MTC in mild to moderate AD patients reported a correlation between the level of disease severity and SPECT scan deficits at baseline, and also a correlation between imaging response and ADAS-Cog response following treatment (Wischik and Staff, 2009). In addition, the distribution of MTC effects on both PET and SPECT scan maps related well to those areas of the brain known to be particularly affected by tau pathology in AD. Thus, positive clinical effects in this phase II trial were mirrored on functional imaging outcomes.

Tramiprosate is a compound that inhibits the aggregation of amyloid fibrils and consequently lowers the rate of plaque production (Gervais et al., 2007). Volumetric MRI (vMRI) was one of the outcome measures employed (Saumier et al., 2009). At baseline, there was a statistically significant correlation between hippocampal volume and CDR-SB scores (p < 0.05), but not ADAS-cog scores (p = 0.11). At 78 weeks respective mean changes of hippocampal volume, ADAS-cog and CDR scores were ~ 192 mm³ (SD 223), 7.8 (SD 92) and 2.0 points (SD 2.6) in the placebo group. There was a correlation approaching statistical significance between change in hippocampal volume and change in ADAS-
Table 1

<table>
<thead>
<tr>
<th>Therapeutic agent</th>
<th>Class/mechanism</th>
<th>Phase</th>
<th>Biomarker</th>
<th>Biomarker purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rosiglitazone (Wyeth/Elan)</td>
<td>PPARγ-Agonist</td>
<td>III</td>
<td>Exploratory proteomics</td>
<td>Disease modification</td>
</tr>
<tr>
<td>Banpinezumab (Wyeth/Elan)</td>
<td>Passive immunotherapy</td>
<td>III</td>
<td>MRI</td>
<td>Disease modification</td>
</tr>
<tr>
<td></td>
<td>(amyloid protein deposition inhibitor)</td>
<td></td>
<td></td>
<td>(vMRI), safety (clinical MRI)</td>
</tr>
<tr>
<td>IVIG</td>
<td>Passive immunotherapy (anti-Aβ)</td>
<td>III</td>
<td>APOE</td>
<td>Stratification (for safety purposes)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FDG-PET</td>
<td>CNS functional effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MRI</td>
<td>Disease modification (vMRI),</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>safety (clinical MRI)</td>
</tr>
<tr>
<td>Solanezumab (Lilly)</td>
<td>Passive immunotherapy (anti-Aβ)</td>
<td>III</td>
<td>Amyloid imaging</td>
<td>CNS amyloid reduction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Plasma Aβ</td>
<td>Peripheral sequestration</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AV45</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CSF Aβ</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FDG-PET</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>III</td>
<td>AV45-PET</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FDG-PET</td>
<td></td>
</tr>
<tr>
<td>LY-450139 Semagacestat (Lilly)</td>
<td>Gamma secretase inhibitor</td>
<td>I–II</td>
<td>CSF Aβ sampling (radiolabeled leucine study)</td>
<td>CNS pharmacodynamic effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>vMRI</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AV45-PET</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FDG-PET</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>III</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACC-001, CAD106, MK (Wyeth/Elan)</td>
<td>Active immunotherapy</td>
<td>II</td>
<td>Amyloid imaging</td>
<td>Disease amyloid reduction</td>
</tr>
<tr>
<td></td>
<td>(amyloid protein deposition inhibitor)</td>
<td></td>
<td>MRI</td>
<td></td>
</tr>
<tr>
<td>PF-04494700 (TTP-488) (Pfizer)</td>
<td>RAGE inhibition</td>
<td>II</td>
<td>vMRI, plasma Aβ (CRP, IL6, TNFα, TNFβ)</td>
<td>Disease-modification</td>
</tr>
<tr>
<td>PBT-2 (Prana Biotechnology)</td>
<td>Metal interaction</td>
<td>II</td>
<td>Plasma Aβ</td>
<td>Pharmacodynamic signal</td>
</tr>
<tr>
<td>AAV-NGF (Ceregene)</td>
<td>Neurotrophin gene delivery</td>
<td>II</td>
<td>MRI</td>
<td>Disease modification (vMRI),</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>safety (clinical MRI)</td>
</tr>
<tr>
<td>H0220 (Newron)</td>
<td>Cytoprotective endogenous steroid</td>
<td>II</td>
<td>“biochemical markers relevant to AD”</td>
<td>Disease modification</td>
</tr>
<tr>
<td>AD-02 (Affiris)</td>
<td>Anti-amyloid immunotherapy</td>
<td>II</td>
<td>vMRI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anti-amyloid immunotherapy</td>
<td>I</td>
<td>CSF biomarkers Amyloid PET</td>
<td>Disease modification</td>
</tr>
<tr>
<td>R-1450 (Roche)</td>
<td></td>
<td></td>
<td></td>
<td>Pharmacodynamic signal, disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>modification</td>
</tr>
<tr>
<td>GSI-953 (Wyeth)</td>
<td>Gamma-secretase inhibitor</td>
<td>I</td>
<td>CSF Aβ40/42</td>
<td>Pharmacodynamic signal, disease</td>
</tr>
<tr>
<td>GSK-933776 (GSK)</td>
<td>Anti-amyloid immunotherapy</td>
<td>I</td>
<td>Plasma &amp; CSF biomarkers Exploratory PET</td>
<td>Pharmacodynamic signal,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenserine</td>
<td>Dual: AChE inhibition and β-APP inhibition</td>
<td>I</td>
<td>FDC-PET</td>
<td>Disease modification</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amyloid imaging</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CSF Aβ</td>
<td>Pharmacodynamic signal, disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CSF &amp; plasma Aβ40</td>
<td></td>
</tr>
<tr>
<td>PAZ-417 (Wyeth)</td>
<td>Plasminogen activator inhibitor</td>
<td>I</td>
<td></td>
<td>pharmacodynamic signal, disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>modification</td>
</tr>
</tbody>
</table>

MRI: magnetic resonance imaging; vMRI: volumetric MRI; APOE: apolipoprotein E; FDG-PET: 18 flurodeoxyglucose-positron emission tomography; Aβ: amyloid beta; CSF: cerebrospinal fluid; RAGE: receptor for advanced glycation end products; information from www.clinicaltrials.gov.

cog score (−0.16, p = 0.07), but not between hippocampal volume change and change in CDR-SB score (0.08, p = 0.43). Thus, there was some evidence to suggest that vMRI may have some clinical validity and value in monitoring disease progression (Gauthier et al., 2009; Saumier et al., 2009). However, it is probably not surprising that the correlation between ADAS-cog scores and hippocampal volume changes was not of greater magnitude or statistical significance since the ADAS-cog measures skills across many cognitive functions, while hippocampal volume change is probably more specific to learning and memory. There is therefore a lesson to be learnt here when planning the outcome measures and the biomarkers to support them: we should not expect to see correlations between two outcomes which are not measuring the same disease mechanism or process. In addition, it is important to bear in mind that the timing of change in a biomarker may not mirror that of change in an clinical outcome measure used to represent the impact of the disease on normal function. Indeed, one of the desired properties of a biomarker surrogate endpoint is that it should predict future clinical benefit or decline (Group, 2001), so we may well not expect to see a correlation between change in biomarkers and clinical outcomes that are measured over the same time period if the biomarker changes are hypothesised to precede the related clinical changes.

Furthermore, one must be cautious in extrapolating from natural history studies directly to a therapeutic trial. Unexpected results can occur, as shown in the AN1792 Aβ immunization study (Fox et al., 2005). Those patients who produced higher antibody titers (antibody responders) showed greater atrophy than those who did not, which represented the opposite outcome of what intuitively was expected. However, there are possible explanations, such as the clearance of amyloid from the brain, or reduction in inflammation associated with plaques. This result demonstrates that we need to carefully consider the expected effects of anti-amyloid therapies on clinical and biomarker outcomes in both the short and the long term when planning future clinical trials.

The relationship of plaque load to clinical efficacy has been questioned by the phase I AN1792 immunotherapy trial follow-up report in which a small number of participants who died several years after the trial was completed showed apparent reduction in amyloid load despite progression of their dementia at autopsy (Holmes et al., 2008). Nevertheless, the recent follow-up report from the larger phase II AN1792 trial indicated possible long-term functional benefits associated with AN1792 immunization in antibody responders (Vellas et al., 2009). The clinical utility of targeting aggregated Aβ, soluble amyloid oligomers or multiple different Aβ species remains an active area of investigation, and a number of clinical trials will be evaluating the value of amyloid imaging modalities in the future (Table 1). These unexpected results underscore, why regulatory bodies insist on careful qualification and validation of biomarkers before they can be used as surrogate endpoints.

A recent clinical trial has also shown the value of neuroimaging techniques as indicators of unanticipated adverse events. In the banpinezumab study of targeted beta amyloid immunotherapy...
MRI changes were found (Grundman and Black, 2008; Salloway et al., 2009) which suggested that safety monitoring might be improved by including MRI sequences sensitive to vasogenic edema and micro-hemorrhage. In transgenic mice, passive Aβ immunization results in reduction of brain amyloid (Pfeifer et al., 2002), removal of vascular Aβ (Schroeter et al., 2008) and cerebral micro-hemorrhages associated with amyloid-laden vessels (Pfeifer et al., 2002; Schroeter et al., 2008). The potential relationship of these preclinical findings to the vasogenic edema observed on MRI in some patients treated with passive Aβ immunization is a subject of investigation; however, MRI clearly has an important role as a safety monitoring tool.

A number of CSF biomarkers have been employed in clinical trials in both phase II and phase III studies (Fig. 2). Some are of less value as in non-treatment studies they do not change over time in a way that would correlate with clinical scores. A similar picture has emerged from clinical trials: e.g. in the AN1792 study a small subgroup showed a decrease in CSF Tau without a change in CSF Aβ (Giffin et al., 2005). Similarly, a gamma secretase inhibitor, semagacestat (LY450139), in a phase II clinical trial showed a relatively immediate impact on plasma and CSF Aβ levels, but no parallel change in cognitive function even after 3 months of treatment (Fleisher et al., 2008). Similar findings have been reported with a monoclonal antibody against Aβ (Solanezumab, LY2062430), over a 12-week period despite changes in plasma and CSF Aβ (Siemers et al., 2010). The lack of correlation between biomarker changes and clinical outcome measures in early phase studies may well partly reflect the short duration and insufficient power of the studies, or dissociation between the different measures in terms of their timing within the disease course. The tramiprosate phase III study was partly based on the outcome of the phase II study in which there was a significant dose-dependent reduction in CSF Aβ42: the highest dose of tramiprosate reduced Aβ42 by approximately 25% (Aisen et al., 2006). The negative results of the phase III study, despite an observed biomarker change at phase II, emphasize that the use of a single biomarker at this stage may lead to an over-optimistic interpretation of the value of the compound.

Bateman et al. (2009) utilized a recently developed method of stable-isotope labeling combined with CSF sampling to directly measure Aβ-metabolism during treatment with the gamma-secretase inhibitor semagacestat (LY450139). Using this method they demonstrated that the study drug dose-dependently decreased Aβ production, whereas previous studies using different methods failed to detect a drug effect on CSF Aβ (Siemers et al., 2007, 2006). Thus, biomarker results for the same study drug can vary depending on the analytical method used. It is therefore unwise to rely on one biomarker or one technique alone for go/no-go decisions for phase III trials. The seemingly increased sensitivity of this new method of measuring Aβ metabolism combined with its lower variability (Bateman et al., 2009) suggest that it may be of value for future AD clinical trials.

Much has been learned therefore from recent disease-modifying trials: (i) biomarker outcomes may or may not correlate with clinical outcomes during a trial, but this may have to be expected if the two are not measuring the same process or are not measuring it at the same stage; (ii) biomarkers may behave differently in clinical trials than in natural history studies in response to drug effects; (iii) biomarkers can play an important role as safety measures; and (iv) it is probably not wise to base go/no-go decisions on the results of one type of biomarker alone.

Another important lesson from biomarker research with special relevance to AD trials with disease modifiers, is that amyloid accumulation in the brain is a very early event that emerges already during preclinical stages of AD and that there is only little further increase of fibrillar Aβ in the brain after the onset of clinical dementia. Along with the consistent failures of Aβ-targeting approaches in phase III clinical trials in mild-moderate AD so far, these findings congruently suggest that therapeutic approaches targeting Aβ may be more effective when applied at the time when rate of Aβ formation and deposition is highest. Following available information, this time appears to be in pre-dementia stages of AD, i.e. in selected MCI subjects or even in pre-MCI subjects with abnormal accumulation of Aβ as measured by amyloid PET or CSF Aβ42 levels.

It is now generally accepted that it is important to include a relevant biomarker outcome in all clinical trials, even if only on a sub-group of subjects since clinical outcomes alone may not provide sufficient evidence of the benefit or lack thereof of potential new AD drugs. However, much work remains to be done in order to validate biomarkers as outcome measures for AD trials, especially if they are to be used as surrogate endpoints (Fig. 3). For example, imaging biomarkers that can be used in animal models as well as in clinical trials will need to be developed to improve the predictability of the effect of drugs in humans (Frisoni and Delacourte, 2009).

Neurochemical and imaging candidate biomarkers generally have to undergo certain stages of development before they can be established as biomarkers that can be used for clinical trials. One approach to exploring neurochemical candidate biomarkers is scanning of large arrays of various proteins and molecules (proteomics) and then testing a small number of most promising candidates in dedicated confirmatory studies. Currently, manual hippocampus volumetry, 18F-FDG-PET, amyloid PET ligands as well as core AD CSF biomarkers (T-tau, P-tau, Aβ_{1-40}, Aβ_{1-42}) have reached the most advanced stage of development (stage III) and are being used in large multicenter controlled clinical trials.

Extended hypothetical temporal trajectories of candidate biomarkers in presymptomatic AD, MCI and clinical AD. The earliest changes that occur are hypothesized to be abnormal neural interconnectivity network function (fMRI) and slightly impaired glucose metabolism (FDG-PET). They are followed by brain Aβ accumulation, a drop in CSF Aβ levels and a larger impairment of glucose metabolism. Later on, T-tau levels increase, indicating neuronal damage, paired with clinically relevant cognitive decline and brain structural changes.
**Table 2**

<table>
<thead>
<tr>
<th>Recommended uses of biomarkers</th>
<th>Measurement recommended at the following time points</th>
<th>Inclusion/exclusion criteria</th>
<th>Stratification</th>
<th>Evaluation of treatment effect</th>
<th>Safety monitoring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B</td>
</tr>
<tr>
<td><strong>“WET”</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aβ42 + Aβ40</td>
<td>++</td>
<td>**</td>
<td>***</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Total tau</td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phos tau</td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BACE 1 activity</td>
<td>++</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasma</td>
<td>APOE status</td>
<td>NA</td>
<td>***</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aβ42 + Aβ40</td>
<td>–</td>
<td>**</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td><strong>“Dry”</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imaging</td>
<td>APOE status</td>
<td>NA</td>
<td>***</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aβ42 + Aβ40</td>
<td>–</td>
<td>**</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Tracers (PET)</td>
<td>Amyloid</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDC-PET</td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imaging</td>
<td>APOE status</td>
<td>NA</td>
<td>***</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aβ42 + Aβ40</td>
<td>–</td>
<td>**</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Tracers (PET)</td>
<td>Amyloid</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDC-PET</td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>–</td>
<td>***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***Recommended for this use, strong supportive evidence base (multiple studies, consistent results). 
**Appears suitable for this use, some supportive evidence (several studies, trend for consistency). 
*May be of use, but requires further study (few studies, still unsolved inconsistencies). 
(−) Not recommended (no data at the current time). 
NA: not applicable for this purpose.

(○ ○ ○ ○) Recommended for safety monitoring.
(○ ○ ○) Suitable for safety monitoring.
(○ ○) Not necessarily recommended for safety monitoring, but data could always be analyzed descriptively if available.

Time points for measurement: B = baseline; D = during trial (several time points); E = endpoint.

(+) Recommended to measure biomarker at this time point.

Aβ: amyloid beta; BACE 1: beta amyloid cleaving enzyme 1; APOE: apolipoprotein E; PET: positron emission tomography; FDG-PET: 18 fluorodeoxyglucose positron emission tomography; MRS: magnetic resonance spectroscopy; fMRI: functional magnetic resonance imaging; SPECT: single photon emission computed tomography.

Biomarkers may also be more widely used in future trials to identify patients at earlier stages of the disease (Fig. 4). The recently proposed new research criteria for the diagnosis of AD (Dubois et al., 2007) present a path for early trials in patients not yet demented; such trials will likely rely on CSF measures and other biomarkers. However, it must be remembered that if a patient population is defined by a specific biomarker that the cost of using the biomarker will be added to the cost of the drug. Furthermore, the more specific the biomarker to define the population, the more conditional the labeling of the indication will be and thus the biomarker is likely to be a condition for reimbursement. Therefore, further analysis ongoing of multi-site biomarker validation trial data need to implement additional cost-effectiveness analyses for application of single markers and/or combinations as well.

### 4. Recommendation for future trials

From the data published and the experience acquired in recent trials, the Oxford Task Force makes the following recommendations:

1. Whatever the biomarker chosen, there is a need to standardize and validate (performance) technical aspects of acquisition, measurement and analysis in both animal models and human studies. It is acknowledged that both in the field of the biochemistry for CSF and plasma biomarkers, and in the field of neuroimaging, variations attributable to non-biological factors such as the equipment, data acquisition and analysis are major sources of error.

2. No single biomarker should be used in isolation in a trial to guide decision-making. Biomarkers are more informative when used as guides in sets of complementary data.

3. When a biomarker is directly linked to a therapeutic mechanism, it can be used as a key measure to guide go/no-go decisions in phase II; in many other cases, biomarker analysis remains exploratory.

4. The biomarkers that can be recommended for use in clinical trials at the current time, along with their potential uses and timing of measurements, are listed in Table 2.

5. Although there are a number of other biomarkers, both CSF and blood-based, that are currently under intense scrutiny, they cannot presently be recommended for clinical trials, since they have not been validated in human studies. They may, however, be useful for exploratory purposes, and so clinical trials should, wherever possible, collect biological samples and make them widely available. This is the route most likely to result in biomarkers of utility for clinical trials of the future.

6. The use of valid biomarkers must be envisaged for the demonstration of disease-modifying effects. Regulatory authorities will have to accept and incorporate biomarker data in their decisions regarding the licensing of disease-modifying drugs for AD. It is clear that as of 2009 it is not foreseeable to license a medicinal product based on non-clinical biomarker outcomes alone. Thus, as recommended by the European Medicines Agency (EMA), it is clearly highly important to incorporate non-clinical biomarkers in the clinical development of products intended for neurodegenerative disorders.

### 5. Conclusions

The real value of biomarker use in clinical trials will only be determined by awaiting the outcomes of present and future clinical phase III studies where biomarkers are measured over a relatively long timescale, and correlated with changes in clinical outcome measures. Reliance on biomarkers as surrogate endpoints...
in pivotal studies for regulatory bodies will require thorough validation and assurance that, indeed, they are reliable indicators of clinically meaningful benefit. It will be necessary to demonstrate that the impact of multiple therapeutic interventions results in biomarker changes that are associated with standard cognitive and clinical effects but predict their magnitude of change. Beside their use in clinical trials on pharmacological interventions, neuroimaging as well as neurochemical biomarkers represent promising endpoints in non-pharmacological interventions such as physical or cognitive stimulation paradigms in cognitively impaired subjects (Buschert et al., 2010).

Conflict of interest

GW, NF, PS, MC, MW, MI, KBr, KB, DP, and PA report no disclosures or conflicts of interest.

Acknowledgements

GW was partly funded by the NIHR Biomedical Research Center Programme, Oxford, UK. KBI has served at Advisory Boards for Innogenetics, Ghent, Belgium; NF’s research group has received payments for consultancy or for image analysis services from Bristol Myers Squibb, Elan/Janssen AI, GE Healthcare, Lundbeck, Teva and Wyeth/Pfizer. PS has consulted for Pfizer, Elan, Lundbeck, Danone, Nutricia, Novartis, Genentech and Roche AG. Renumerations for consultancy are paid to his institution only. SI has in the past five years provided consultancy to Pfizer, Noscira, GEHC, have received honoraria for speaking from the Lundbeck Neuroscience Institute, have grant funding jointly with GSK and am named as an inventor on patents related to biomarkers held by my employers. Simon GF: Scientific AB and consultancies: served in Abs for Lilly, BMS, Bayer, Lundbeck, Elan, Astra Zeneca, Pfizer, Baxter, Taurx, Wyeth. AN is an investigator in clinical trials sponsored by Novartis AB, Jansen-Cilag, Torreyepines Therapeutics, GSK, Wyeth and Bayer, has been consulted for Astra-Zeneca, served in advisory board for Elan, Pfizer, GSK, Novartis AB, Lundbeck AB, and GE Health Care, received honorarium for lectures from Novartis AB, Pfizer, Jansen-Cilag, Merck AB, Lilly, Bayer, and received research grants from Novartis AB, Pfizer, GE Health Care, and Johnson & Johnson. HH received consulting fees from B.R.A.H.M.S. Biotech, Pfizer/Eisai, Glaxo-Smith-Kline, Novartis and Bristol-Myers-Squibb. He has also received lecture fees from Pfizer/Eisai, Glaxo-Smith-Kline, Lundbeck and Sanofi-Aventis. His institution has received grants/grants pending from Hoffmann-La Roche, GEHC, and Novartis and Sanofi-Aventis. BW serves on scientific advisory boards for Diagenic, Elan Corporation, Janssen, Lundbeck Inc., Medivation Inc., Novartis, Pfizer Inc. and Merz Pharmaceuticals, LLC; serves on the editorial advisory boards of the American Journal of Alzheimer’s Disease & Other Dementias, the Journal of Cellular and Molecular Medicine, Alzheimer’s Care Today, Alzheimer’s & Dementia, Dementia and Other Dementias, Aging, and Neurology. GW received research support to his institution from Dainippon Sumitomo Pharma Co., Ltd.; and receives research support from the Swedish Medical Research Council and the Swedish Brain Power programme. BW has no stocks in pharmaceutical industry. BV is a principal investigator in trials funded by: Exonihit, Ipsen, Lilly, Pierre-Fabre, Servier, TauRx Therapeutics; member of a steering committee, member of an advisory board, or equivalent body: Eisai, Elan, Exonhit, GSK, Lilly, Medivation, Nestlé, Nutricia, Pfizer, Pierre-Fabre, Roche, Sanofi, Servier, TauRx Therapeutics, Wyeth; investigator (not principal) for the development of a product: BMS, Elan, Lilly, Medivation, Pfizer, Roche, Wyeth. NC received PhD funding from Ipsen Pharma and the French Association of Technical Research (ANRT) in the form of a CIFRE PhD studentship.

SA serves on a scientific advisory board: for Beaufour Ipsen Pharma SAS, Esai, Pierre Fabre, Pfizer, Lilly; has received funding for travel from Esai, Lundbeck, Beaufour Ipsen Pharma SAS, Pierre Fabre, Nestlé, Pfizer, Novartis, Pfizer; - has received honoraria from Roche, Esai Inc., Pfizer; Pierre Fabre Laboratories, Eli Lilly and Company, Lundbeck Inc., and Nestlé S.A., Chiesi, Servier, Janssen. The authors are grateful to Mrs. Kirsten Duggan for technical support in the preparation of the manuscript.

Appendix A. Task Force Members

Abadie E, France; Abu-Shakra S, USA; Aisen P, USA; Andrieu S, France; Antoun Z, France; Ashwood T, Sweden; Banzet S, Switzerland; Black R, USA; Blennow K, Molnai, Sweden; Bogdanovic N, UK; Booij J, Norway; Boswel D, UK; Broich K, Germany; Cantillon M, USA; Carrillo M, USA; Cedarbaum J, USA; Del Signore S, UK; Douillet P, France; Dubois B, France; Duveau F, France; Fox N, UK; Frisoni G, Italy; Gispelin-Wied C, Switzerland; Graf A, Switzerland; Grundman M, USA; Hample H, Germany; Heisterberg J, Denmark; Hendriks S, USA; Hennessy T, UK; Hoerr R, Germany; Hulme A, USA; Hutton M, USA; Imbert G, Switzerland; Ingvarsson A, France; Isaac M, UK; Keime-Guibert F, France; Kohler S, Germany; Krempien S, Germany; Langenberg A, USA; Langstrom B, Sweden; Larsen S, Denmark; Lonneborg A, Norway; Lovestone S, UK; Matusевичius D, Sweden; Mo Yi, USA; Nordberg A, Sweden; Nordgren I, Sweden; Ostrowski S, Switzerland; Palmantier R, Canada; Ryan J, USA; Sampaio C, Portugal; Saumier D, Canada; Schindler R, USA; Seely L, USA; Siemens E, USA; Sol O, France; Swartz J, UK; Therasse P, Belgium; Touchon J, France; Van der Flier W, Netherlands; Vellas B, France; Visser PJ, Netherlands; Von Raison F, Switzerland; Wilcock G, UK; Winblad B, Sweden; Wischik C, UK; Zvartau-Hind M, UK.

References


Neurology 73, 2061–2070.


