Disease-modifying trials in Alzheimer’s disease: a European task force consensus

Bruno Vellas, Sandrine Andrieu, Cristina Sampaio, Gordon Wilcock for the European Task Force group

After symptomatic treatments, the new target for therapeutic approaches in Alzheimer’s disease is the development of disease-modifying drugs. The concept of disease modification in Alzheimer’s disease is controversial and the design of these trials raises many questions. Which populations should be studied? For how long? With which principal and secondary endpoints? Are surrogate markers available? Here, we present a European consensus on disease-modifying trials in Alzheimer’s disease, agreed under the auspices of the European Alzheimer’s Disease Consortium and based on the European perspective of the concept of disease modification, study designs, the role for biomarkers, risk benefit, and pharmacoeconomic issues.

Introduction

After the introduction of symptomatic treatment, the next target for therapeutic approaches in Alzheimer’s disease is the development of disease-modifying drugs. The aim of these drugs should not only be to have symptomatic effects but, more importantly, to delay progression of the disease by acting on pathophysiological processes—for example, many current drugs use strategies that are intended to reduce the load of amyloid β in the brain. This is a crucial time in the fight against Alzheimer’s disease because many of these new potential anti-amyloid treatments (eg, γ-secretase inhibitors, amyloid-β immunotherapy, and amyloid-β antagonists) are now undergoing or are about to enter clinical trials. The design of these trials raises many questions. Which populations should be studied? For how long? How should treatment effects be measured in terms of principal and secondary endpoints? Are surrogate markers available to assess disease-modifying outcomes?

Any flaws in the methodology of these trials will be costly, render them ineffective, and will delay the introduction of more effective treatments for clinical practice. Potential problems that can compromise a trial include a high dropout rate due to trial duration and inadequate sensitivity or specificity of the outcome measures. For these reasons, under the auspices of the European Alzheimer’s Disease Consortium, we decided to organise a task force to propose a European consensus paper in this new area. Task force members were chosen because of their academic, regulatory, or pharmaceutical experience in the area.

Methods

The experts selected by the organising committee (BV, CS, GW) were asked to write a comprehensive review of methodological aspects—ie, biological, neuroimaging, cognitive, and non-cognitive assessment methods for use in trials of potential disease-modifying treatments for Alzheimer’s disease. Papers were circulated to all task force members at the end of November—3 weeks before the task force meeting, held in Toulouse, France, on December 15, 2005. At the same time each task force member was asked to list the main questions that they thought needed to be discussed. More than 50 questions were suggested, and the organising committee selected five main themes to be considered (panel 1). At the meeting, after general presentations, four thematic groups met to consider the specific responses. Recommendations were presented to the task force for general discussion later in the afternoon. The conclusions that were reached regarding these questions are presented in this Review.

The concept of disease-modifying treatment

Disease modification is difficult to define and is the subject of much debate. For a neurodegenerative disorder, such as Alzheimer’s disease, a disease-modifying intervention is typically considered to be one that can reduce progression rate. To some extent, this intuitively implies an effect on the pathophysiological mechanism of the disease. However, from the patient’s perspective, a disease-modifying treatment should result in long-lasting changes in disability, regardless of the drug’s mechanism of action. Thus, we believe that a disease-modifying intervention should be defined as one that has a long-lasting (ie, at least 18 months) effect on disability. This contrasts with interventions that are only able to relieve impairments as measured by the signs and symptoms of the disease.

The clinical trials that are needed to reveal a disease-modification effect will have different features depending on whether a mechanistic or consequentiality approach is used. For the US Food and Drug Administration (FDA), an effect on the pathophysiological process (mechanistic approach) must be shown, whereas, for the
European Medicines Agency, the potential approval of a disease-modifying drug must be based on sound clinical outcomes (consequentiality approach); the Canadian authorities have the same view. A mechanism of action that interferes with the pathological process will add plausibility to the clinical results from trials. There is no public statement from either agency on the matter, but through feedback from several of the attendees at the consensus meeting and their interactions with those institutions, we believe this a fair summary.

Because no such drug is approved at the time of writing this Review, the situation will probably change as trial results come to light. Nevertheless, we are clearly aware of the trade-off between keeping Alzheimer’s disease as a drug target and motivating competition without giving unrealistic expectations to patients and funding agencies. This approach will enable us to set standards and single out those interventions that have long-term effects from those that do not in a short window of opportunity. The ability to identify potentially useful drugs depends on the scientific development of the field, which is still in its infancy, until the development of biomarkers and validated surrogates.

**Trial design**

**Definition of the population to be studied**

Even though disease-modifying drugs would be most useful in the very early stages of the disease, we agreed that patients with mild cognitive impairment are not good targets for disease-modifying trials because of the lack of diagnostic consensus, the difficulties in defining the population, and the poor efficacy of recent trials in this area. Also, mild cognitive impairment is a heterogeneous disorder in which only a subset of patients will have Alzheimer’s disease, and usually the progression of the disease is slower than in Alzheimer’s disease. The subtype of mild cognitive impairment that involves amnestic patients, so-called pre-Alzheimer’s disease mild cognitive impairment, might however be a disease-modifying target.

Use of the amnestic mild cognitive impairment criteria, perhaps coupled with biomarkers, might lead to highly specific outcomes and if effective therapies were available, this could be a relevant target in the future. Patients with mild Alzheimer’s disease are however an ideal target for such trials, although cognitive and non-cognitive assessment tools will have to be adapted to each specific stage of the disease. Trials in patients with very mild Alzheimer’s disease will require very sensitive cognitive tools and long-term follow-up because of the associated cognitive and clinical stability reported during this stage of the disease.

Patients in the moderate stage of the disease can be included in some treatment studies and therefore it may be necessary to increase the slope difference between the placebo and the therapeutic arm. Because patients with severe Alzheimer’s disease are more likely to be lost to follow-up or to drop out, it is probably best to target those with a mini-mental state examination (MMSE) score of 16–27 at entry and living at home with a carer. Those with an MMSE score of 24–27 would also have to meet standard research criteria for probable Alzheimer’s disease to distinguish them from patients with mild cognitive impairment.

Elderly people, including those who have subjective memory complaints, could also be included in disease-modifying trials because the pathological process probably starts decades before the onset of clinical symptoms. Preventive trials will need a long period of assessment because these patients usually have a relatively slow rate of decline. This group of elderly people with subjective memory complaints is very heterogeneous and trials in this group are subject to similar constraints to those involving patients with mild cognitive impairment.

There were concerns about the need to have clear and strict assessment of the patients and homogeneous groups. Neuroimaging (eg, MRI) can be very helpful in defining the population to be studied for a specific drug.

**Factors that affect study design**

Design choices are not limited. Disease modification is a long-term effect and this makes it impractical to adapt trial designs according to response. Such trial designs can only be considered when a biomarker or cluster of biomarkers is established and sufficiently validated to form the basis of the outcome of a trial. When this becomes possible, adaptive designs will be extremely important in a dose-finding context because they might allow a reduction in the sample size or duration of follow-up, and therefore the process will be more efficient. For the same reasons, crossover options are of limited potential. The randomised start or the randomised withdrawal designs might be necessary as a means to address the FDA’s mechanistic concerns. These designs are theoretical constructs to show that the effect of an intervention is maintained beyond its period of administration. They are elegant rationalisations but their actual interpretation will be hampered by the need for long follow-up periods (probably no less then 12 months per period); the effect of dropouts, particularly if this is different in the two periods, which is very likely to happen; placebo effects; and risks of “unblinding” if complex titration phases are needed. So far only one such trial has been published. This was done with Parkinson’s disease and raised more speculation than firm conclusions.

Given the limitations discussed above, the classic randomised, parallel, two-arm placebo-controlled trial is still favoured as a possible approach to establish disease modification in Alzheimer’s disease.

The duration of follow-up should be a trade-off between allowing sufficient time for the intervention to
produce a clinically meaningful effect, not just the minimal period necessary to discriminate from placebo, the need to compensate for the threat to validity that a high attrition rate always presents, and the economic costs of longer follow-up periods. 18 months was suggested as a possible compromise. Nonetheless it might be necessary to go beyond 18 months if the progression rate in the cohort is lower than expected (<4 points on the Alzheimer’s disease cognitive assessment subscale per year) or the drug effect has a very long latency. Given these uncertainties, a sequential design might be worth considering.11

The design will likely be an add-on trial where patients participate in a randomised controlled trial and are allowed to take acetylcholinesterase inhibitors or memantine; or possibly a comparative trial where the new intervention is compared to the approved ones. Such trials would explore the possibility that in the long run (eg, 18 months or more), the test drug, if disease-modifying, would outperform the comparators. So far such a trial has not been done.

A randomised controlled trial could be designed either as a time-to-event comparison or could compare progression rates. If set as a time-to-event study, the event must be a clinically significant milestone of the disease and the delay achieved must also be clinically relevant. It is unreasonable to expect that an intervention will postpone a disease landmark forever. The effect might wear off or be overcome by complementary pathogenic mechanisms. Therefore, in a time-to-event trial, the key issue will be how long a delay in disease progression is needed to indicate disease modification. As a rule of thumb, given the natural span of the disease is about 10 years, such a delay should not be less than 3–6 months and ideally should be more than 6 months.

The difference at endpoint that will be considered clinically relevant should be defined at the outset. A reduction in the progression rate of 30–50% is a reasonable goal. However, the control group should have a progression rate within the expected rate of progression for the baseline characteristics of the patients. If it is less than that typically expected, a 50% reduction in those patients that were treated might not be sufficient to show a meaningful effect. It is also worth mentioning that these two options, time-to-event or comparison of progression rates, do not require the use of an imputation method for missing values, which is an advantage in a disorder like Alzheimer’s disease.

However the statistical models applied to compare progression rates and to model the progression of patients who discontinued early use a process that has similar implications to the “last observation carried forward” imputation method. Therefore the rate of attrition and the possibility of differential attrition are critical for the interpretation of these comparisons. The retrieval of dropouts should always be intensively pursued.

Endpoints
The endpoints for disease-modifying trials should be clinically relevant. They must involve multiple domains because disease modification implies protection of the overall neuronal function, rather than a more limited outcome, such as improvement of a specific neurotransmitter system. The endpoints should include cognitive functions, functional status (activities of daily living), neuropsychiatric symptoms, and cost-effectiveness for the benefit of regulators and health-care providers. Cognitive functions are usually assessed in Alzheimer’s trials by the Alzheimer’s disease cognitive assessment subscale. However, it may be very useful to optimise composite measures according to disease stage (mostly early dementia) and trial duration. Such composite measures may improve sensitivity and power. Other outcomes such as the clinical dementia rating scale10 may be appropriate. Activities of daily living are measured by many scales including the ADCS-ADL (Alzheimer’s Disease Cooperative Study-activities of daily living scale),21 basic activities of daily living,22 and instrumental activities of daily living.23 It is really important to have some effect on activities of daily living in trials on Alzheimer’s disease. Neuropsychiatric symptoms are well assessed by the neuropsychiatry inventory18 but this can be affected by the cultural environment. In addition, it might be useful to study informal carer burden with a measure such as the Zarit scale.24 The resource utilisation in dementia20 scale is useful for assessment of use of care and costs in patients with Alzheimer’s disease.

One important aim in disease-modifying trials is the possibility of delaying the transition from one stage to a more severe stage of the disease. For example, change from a clinical dementia rating scale score of one to two, or change from mild-to-moderate dementia to moderate-to-severe dementia (MMSE score <16). Another approach could be to decrease the percentage of patients with rapid disease progression—eg, based on a loss on the MMSE of greater than three points in 1 year.

Once an effect on clinical outcome measures has been shown, it will be important to explore whether a disease-modifying drug has an effect on surrogate markers—eg, neuroimaging (brain atrophy) or biological markers (tau isoforms, amyloid β).

Analysis
To determine the number of patients needed for a trial, it is essential to understand the natural rate of decline; it has been suggested that patients in the placebo arm of recent randomised controlled trials have typically declined less than those in natural history studies (data from recent observational studies are relevant).11 If the study is done using an add-on design, we must also know the rate of decline under these conditions in long-term treatment use.

We reached a consensus that proposed analysis of the slope of progression as a valid measure of disease
Biomarkers need to be distinguished from clinical endpoints, which can be defined as characteristics or variables that indicate how a patient feels, functions, or survives. A surrogate endpoint is a measurement that substitutes for clinical endpoints and is expected to predict clinical benefit, or the opposite, based on one or other types of scientific evidence.

Biomarkers have been adopted as surrogate endpoints in clinical trials in several other areas of clinical neuroscience and medicine—eg, in assessment of interventions for multiple sclerosis and HIV/AIDS. However, for Alzheimer’s disease, we have yet to show that a biomarker is adequate to predict clinical endpoints or measure the effect of treatment on these endpoints. Biomarkers are competing with the current, albeit unsatisfactory, clinical gold standards that measure change in cognition, behaviour, and functional ability. The latter are moving into new trial designs that have not so far been adequately tested—eg, the random start or withdrawal paradigms.

A practical and ideal biomarker would be simple to obtain—eg, from blood or urine—rather than more complex procedures such as neuroimaging or tissue or cerebrospinal fluid examination. An exception to the latter is the proof-of-concept stage in a clinical trial situation.

Potential biomarkers

Biomarkers may be useful for selection of the study population (fixed biomarkers) and as surrogate outcomes measures (dynamic biomarkers).

With increasing knowledge of the molecular basis of Alzheimer’s disease, several potential biomarkers have been suggested, including proteins, structural changes visible on neuroimaging, and genes. Although APOE is a trait marker that could help enrich studies, albeit with a subset of patients with Alzheimer’s disease, it is inadequate as a biomarker. However it is possible that other genetic candidates will emerge.

Searching for biomarkers in the cerebrospinal fluid seems a logical starting point, and many reports have confirmed the pattern of increased tau and reduced amyloid-β peptides. In general or in combination, sensitivity and specificity levels of about 85% have been reported, and also perhaps some ability to discriminate between dementias. Tau is an intracellular protein and as such is possibly a non-specific marker for neuronal lysis; if so it is unlikely to have adequate sensitivity and specificity for general use. Phospho-tau may be a better biomarker. Amyloid peptide species do not seem any more reliable and also have demanding technical requirements.

Various other cerebrospinal fluid biomarkers have been explored—eg, isoprostanes and glycation, inflammation, and oxidative stress markers. Although some biomarkers show promise, much more work is needed to confirm their potential value. Various plasma markers are being

---

Panel 2: Potential uses of biomarkers in trials of Alzheimer’s disease

- Early diagnosis—eg, at the stage of mild cognitive impairment
- Differential diagnosis
- Monitoring the rate of disease progression
- Monitoring response to therapy—particularly important in association with disease-modifying treatments

modication. This would involve longitudinal analysis of repeated measurements of cognitive tests and other outcome measures, rather than just the difference between placebo and treated groups at the study endpoint. Various problems could result, however, due to the type of cognitive or other tool used (non-linearity, floor and ceiling effects). Another major problem associated with clinical trials with a long duration of follow-up is the probability of missing data. In the case of Alzheimer’s disease, missing data are usually associated with the severity of the disease. Traditionally, several methods have been proposed to manage incomplete data.

For instance, an intention-to-treat analysis must be the primary analysis, but it is then necessary to find the most appropriate imputation method.

The last observation carried forward procedure is the most discussed method. This procedure assumes that outcome remains constant at the last reported value after dropout, but this is unlikely in a long-term trial of Alzheimer’s disease and may bias the outcome in favour of a positive result if more patients drop out from the treated group than placebo group early in the trial. There are alternatives, and a solution usually suggested by statisticians is to model these missing values by increasingly sophisticated methods. However, we must remember that some of the data may be missing because of the evolution of the disease. One recommendation to obtain maximum data in trials could be to plan a retrieved dropout visit at the end of the intended treatment period for each patient.

Role of biomarkers

The term biomarker is usually applied loosely in the context of assessing outcomes of clinical trials in Alzheimer’s disease, but is best defined in accordance with the definitions from the US National Institutes of Health—the result of a working group report published in 2001. This group created a standard reference base for terminology that we can now use in trials of disease-modifying drugs. Thus a biomarker is defined as a characteristic that is objectively measured and assessed as an indicator of normal biological or pathogenic processes or pharmacological responses to a therapeutic intervention. A biomarker should be more than 80% sensitive in detection of Alzheimer’s disease and have a specificity of greater than 80% for distinguishing other dementias (panel 2).
Panel 3: The European Alzheimer’s Disease Consortium’s recommendations for disease-modifying trials

**Target population**
- Early Alzheimer’s disease
- Mild to moderate Alzheimer’s disease

**Study design**
- Randomised, parallel, two-arm, placebo-controlled trial

**Follow-up**
- 18 months

**Statistical analysis proposed**
- Slope analysis

**Primary and secondary outcomes**
- Endpoints should be clinically relevant and include cognitive functions (composite measures), functional status (activities of daily living), neuropsychiatric symptoms (NPI) and cost-effectiveness (RUD, Zarit)\(^19\)
- Biomarkers (biological and neuroimaging)

**Surrogate markers**
- Not recommended for primary outcome at this time

NPI=neuropsychiatry inventory; RUD=resource utilisation in dementia

**Ethical and risk–benefit issues**

Inherent within the design of any clinical trial is an acceptance of a particular level of risk–benefit ratio. This raises important questions for trials as well as for ethics committees and regulators. Significant issues include who decides what levels of risk are acceptable: should it be the professional or should it be the patient and their carers? Is there a role for patient and carer organisations? How can one obtain informed consent from a patient with more than mild dementia?

In general terms it would be reasonable to assume that a higher level of risk is acceptable for a possible curative treatment for a fatal disorder than would be the case for a less serious illness. However, the same is probably not true for short-term limited benefit from a symptomatic drug. Do we need a sliding scale of competence to address this spectrum—eg, greater competence to agree to more risky treatments and less competence to agree to inclusion in a trial of a treatment with less potential for harm? If so, trials of potentially more harmful drugs, but with substantial disease-modifying potential, may only be ethical in patients with mild dementia.

Such consideration of the limits of acceptable risk against possible benefit is central to the assessment of all clinical trials.\(^{35}\) However, it is particularly relevant in situations where individuals have limited abilities to make autonomous and informed decisions. It is over a decade since High and colleagues\(^6\) noted that there was no consensus about such issues, and it would seem this is still the case. It is clearly an issue that needs to be urgently considered, and patient and carer organisations may like to address this together with other stakeholders.

**Pharmacoeconomic issues**

Alzheimer’s disease puts a heavy economic burden on western societies. Assuming societies are willing to pay to alleviate that burden, the price paid for each intervention must be commensurate to the mitigation that is produced. There are different ways to approach these calculations. Whichever cost-effectiveness model is used it must include actual data. Furthermore, apparently small benefits in impairments and disabilities might have a substantial effect on handicap. Thus, there are two good reasons why economic outcomes should be collected in randomised trials: to assess whether there is a scale-up effect on handicap or carer burden from the effect seen in the primary endpoints, and also to collect raw data on resource consumption to be later used in cost-effectiveness models.

**Conclusion**

In conclusion, there was much agreement on the approach to designing trials to assess putative disease-modifying drugs, despite the breadth of European countries that participated in the discussion. The main explored, including amyloid peptide proteins identified by proteomic technology, and similar molecules to those researched in cerebrospinal fluid. However to date, none have achieved sufficient reliability for use as a diagnostic or surrogate marker.

Use of structural neuroimaging in clinical trials is well established, more so in subject selection than reliably assessing disease progression. In the latter context, changes in whole brain and hippocampal size and ventricular volume have been suggested, as well as other changes—eg, in the entorhinal cortex and amygdala. In particular, slowing of whole brain atrophy is an intuitively attractive surrogate outcome for a diffuse neurodegenerative disorder such as Alzheimer’s disease. However, a recent amyloid-β immunisation trial unexpectedly found that antibody responders had a greater brain volume decrease than placebo patients, although increased losses in brain volume were not indicated in worsening cognitive performance.\(^{16}\) In general it would be fair to conclude that MRI is a useful diagnostic screen in clinical trials, but that its place as a surrogate marker has still to be proven. A review by Scheltens and responses by Fox and Frisoni provide further detail.\(^{12,13}\) Imaging of amyloid plaques in the brains of people with Alzheimer’s disease seems to be promising as a marker but further research is needed.\(^{14}\)

In summary, biomarkers are increasingly being sought with the move towards trials of potentially disease-modifying drugs. There is clearly hope on the horizon, but even for candidates such as cerebrospinal fluid tau and amyloid-β levels, much work remains to be done.
points are summarised in panel 3. This is compelling because such treatments are increasingly coming to clinical assessment, and this consensus may help to inform the response of regulatory and funding bodies as the results of trials become available.

Task force group
E Abadie (Paris), G Van Kan Abellan (Toulouse), K Albert (New York), S Andrieu (Toulouse), H Aslanides (Södertälje), R Aisen (New York), B Alafuzoff (Montreal), D Alves (Lisbon), J C M Alves (Lisbon), C A Asplund (Stockholm), L A Asplund (Stockholm), H Assirah (London), B B Andersson (Stockholm), M A. A. Andersson (Stockholm), L. A. Andersson (Stockholm), R. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockholm), M. Andersson (Stockhol


