Endpoints for trials in Alzheimer's disease: a European task force consensus

*Bruno Vellas, Sandrine Andrieu, Cristina Sampaio, Nicola Coley, Gordon Wilcock, for the European Task Force Group**

Lancet Neurol 2008; 7: 436–50

*Members listed in full at end of review

Inserm U558, Toulouse, France (B Vellas PhD, S Andrieu PhD, N Coley MSc)**; University of Toulouse III, Toulouse, France** (B Vellas, S Andrieu, N Coley)**; Gerontopole, Department of Geriatric Medicine** (B Vellas) **and Department of Epidemiology and Public Health** (S Andrieu)**, Centre Hospitalier Universitaire de Toulouse, Toulouse, France; Laboratory of Clinical Pharmacology and Therapeutics and Instituto de Medicina Molecular, Faculty of Medicine, Lisbon, Portugal** (C Sampaio PhD)**; and Nuffield Department of Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK** (G Wilcock DM)

Correspondence to:
Gordon Wilcock
Nuffield Department of Medicine, University of Oxford, Level 7, John Radcliffe Hospital, Oxford OX3 9DU, UK
gordon.wilcock@ndm.ox.ac.uk

Harmful consequences in health status caused by disease are referred to as outcomes, and in clinical studies the measures of these outcomes are called endpoints. A major challenge when deciding on endpoints is to represent the outcomes of interest accurately, and the accuracy of such representation is assessed through validation. Complex diseases like Alzheimer's disease have many different and interdependent outcomes. We present a consensus for endpoints to be used in clinical trials in Alzheimer's disease, agreed by a European task force under the auspices of the European Alzheimer Disease Consortium. We suggest suitable endpoints for primary and secondary prevention trials, for symptomatic and disease-modifying trials in very early, mild, and moderate Alzheimer's disease, and for trials in severe Alzheimer's disease. A clear and consensual definition of endpoints is crucial for the success of further clinical trials in the field and will allow comparison of data across studies.

Introduction

Endpoints are used to measure disease outcomes in clinical studies, and the selection of endpoints that represent the outcomes of interest accurately is a major challenge. Definition of the primary outcome is probably the most important decision in the design of a clinical trial. The size and type of population to be targeted, the clinical relevance of the drug therapy, the rationale for its use in clinical practice, and cost-effectiveness all depend on the primary outcome.

Outcomes for trials in Alzheimer's disease (AD) are still subject to discussion, partly because patients are increasingly being diagnosed during the earliest stages of the disease.[1] For example, the cognitive subscale of the AD assessment scale (ADAS-cog), traditionally thought of as the standard primary cognitive outcome for symptomatic trials, is probably not appropriate for trials in very early AD. Likewise, subjective measures of global improvement, which have long been recommended by regulatory agencies, are difficult to assess in trials of 18 months or longer, and are probably not suitable in such studies. Clinical assessment methods should also be reconsidered because many trials now add a new treatment to an existing standard-of-care regimen, so characteristics of patients are likely to differ from those in previous trials.

The most suitable outcomes should therefore be redefined on the basis of better knowledge of AD, in light of its well characterised stages, the large number of trials in development, and the negative results from some recent trials.[2] These results sometimes contrast with positive effects in animal models,[3] which could mean that the trial methods are at fault or that effects in a subgroup of patients are being missed.

Because the AD phenotype is complex, a single type of measurement is unlikely to capture adequately all the domains of the disease in its different stages.[4] The solution is to use more than one measurement in a particular context and to attempt an integrated interpretation of the results. Once a set of measurements is obtained, it must be assigned a meaning in terms of relevance to the patient's life and daily living abilities.

In this Review, we provide critical analysis of the endpoints that are judged to be valid and have already been used in clinical studies to evaluate outcomes at different stages of AD. We also present recommendations from an international task force on outcomes, which follows on from our task force on disease-modifying trials.[5] We do not make recommendations about the specific development of new endpoints, although we do mention the need for such development where appropriate.

Methods

Under the auspices of the European Alzheimer Disease Consortium (EADC), a network of excellence in the field of AD financed by the European Commission (5th FP QLAM 2001-00003), the organising committee (SA, CS, BV, and GW) set up a task force to propose a European consensus on endpoints for trials in AD. Task force members were chosen for their academic, regulatory, or pharmaceutical experience. The task force included researchers from the USA in addition to Europe because many research and development departments in the pharmaceutical industry are based in the USA. The organising committee asked selected members of the task force to write a comprehensive review of biological, neuroimaging, cognitive, and non-cognitive assessment methods for trials in patients with AD. The resulting 12 papers[4,6–16] were circulated to all members of the task force at the end of March, 2007, before the meeting in Lisbon, Portugal, in April, 2007. Each member was also asked to list the main questions that he or she thought should be answered at the meeting. Of the topics that were suggested, the organising committee selected three to be discussed: outcomes for prevention trials; outcomes for trials in very mild to moderate AD; and outcomes for trials in severe AD. At the meeting, after general presentations, thematic groups met to discuss the specific responses. Recommendations were presented to the task force for general discussion, and here we present the conclusions that were reached with respect to these topics. The

outcome measures used in previous trials are listed according to the type of trial, together with a critique of the strengths and limitations of individual primary outcomes in each context.

Primary prevention trials

Prevention is currently a major issue in AD, and it will continue to be so until there is a cure. Preventive strategies have the potential to decrease the incidence of AD substantially.[17,18] Primary prevention trials in AD[19–36] are relatively new (table 1), and the most widely used outcomes have been developed on the basis of experience from other contexts and expert judgment. Usually, primary prevention trials for dementia involve participants thought to be at greater than average risk of dementia, such as people who are elderly (minimum age 60–75 years)[19,20,24,38] or have another defined risk factor (eg, memory complaints or family history of AD).[19,39] Frailty (eg, walking speed) and cognitive decline have also been linked to AD,[40] and in future frailty might be used to enrich study populations. Patients who meet prespecified criteria for dementia are excluded from primary prevention trials, and the nature of the intervention must also be taken into account when the age and type of target population is defined (eg, frail, healthy, old, very old, or the general adult population). There are two possible primary outcomes for preventive trials: conversion to dementia or cognitive decline.

Conversion to the clinical stage of cognitive impairment

Several large primary prevention trials have measured dementia or incidence of AD as a primary objective.[19–24,38]

	Intervention*	Duration	Primary outcomes	Criteria and measurements	Selected other outcomes	Results
Vellas and co-workers (GuidAge)[19]	Ginkgo biloba extract (240 mg)	5 years	AD incidence†	DSM-IV, NINCDS-ADRDA, NINDS-AIREN	Cognitive decline, functional decline, falls, one-leg balance	Not yet available
DeKosky and co-workers (GEM)[20]	Ginkgo biloba extract (240 mg)	5 years	Dementia incidence†	DSM-IV	Cognitive decline, functional decline	Not yet available
ADAPT research group (ADAPT)[21]	Naproxen (220 mg twice a day) or celecoxib (200 mg twice a day)	5 years	AD incidence†	DSM-IV, NINCDS-ADRDA	Cognitive decline	Negative results (after early termination of trial)‡
Shumaker and co-workers (WHIMS)[22]	Oestrogen (0·625 mg) and progestin (2·5 mg)	Stopped at 5·6 years	Dementia† and MCI incidence	DSM-IV, MCI[37]	..	Greater risk of dementia in treatment group than with placebo
Shumaker and co-workers (WHIMS)[23]	Oestrogen (0·625 mg)	5 years	Dementia† and MCI incidence	DSM-IV, MCI[37]	..	Greater risk of combined endpoint (dementia or MCI) in treatment group than with placebo
Kryscio and co-workers (PREADVISE)[24]	Selenium and vit E	9–12 years	AD incidence†	..	..	Not yet available
Kang and co-workers (WHS cognitive substudy)[25,26]	Vit E (600 IU on alternate days) and aspirin (100 mg on alternate days)	4 years	Cognitive function	Telephone cognitive battery (TICS, East Boston memory—immediate and delayed recall, TICS 10-word list delayed recall, category fluency)	..	No effect on cognitive function
McMahon and co-workers[27]	Folate (1000 µg), vit B6 (10 mg), and vit B12 (500 µg)	2 years	Cognitive function	MMSE, RAVLT†, Wechsler, Reitan TMT, category word fluency, Raven's progressive matrices, COWAT†	..	No effect on cognitive function
Dangour and co-workers (OPAL)[28]	n-3 polyunsaturated fatty acids (0·5 g DHA and 0·2 g EPA)	2 years	Cognitive function	CVLT†	Cognitive performance (immediate and delayed recall, verbal fluency, DS, symbol-digit modalities, reaction time, special memory)	Not yet available
Eussen and co-workers[29]	Vit B12 (1000 µg) *vs* vit B12 (1000 µg) and folic acid (400 µg)	6 months	Cognitive function	MMSE, motor planning, Raven, Stroop, similarities (WAIS), word fluency†, figure of Rey (copy), immediate and delayed recall, DS forwards and backwards, 15-word learning, delayed recall and recognition, finger tapping	..	No difference between groups
Wolters and co-workers[30]	Multivitamins	6 months	Cognitive function	Symbol search and pattern recognition, WAIS-III, KAI	..	No difference between groups
Bryan and co-workers[31]	Folate (750 µg) *vs* vit B12 (15 µg) *vs* vit B6 (75 mg)	1 month	Cognitive function	Verbal fluency, TMT, Stroop, SOPT, excluded letter fluency, WAIS-III letter-number sequencing, digit symbol, verbal fluency, DS backwards, vocabulary, symbol search, activity recall, RAVLT, boxes, spot the word, uses for common objects	..	Supplementation had a significant positive effect on some measures of memory only, and no effect on mood. Dietary status was associated with speed of processing, recall, recognition, and verbal ability

(Continues on next page)

	Intervention*	Duration	Primary outcomes	Criteria and measurements	Selected other outcomes	Results
(Continued from previous page)						
Almeida and co-workers[32]	Oestradiol (0·5–2·0 mg)	5 months	Cognitive function, depression, QoL	CAMCOG, verbal fluency, block design, test faces, CVLT-II	..	No difference between groups
Polo-Kantola and co-workers[33]	Oestradiol (gel 2·5 g or patch 50 µg)	7 months	Cognitive function	Digit-symbol, Stroop, DS forwards, Benton visual retention, subtraction, statement verification, paced auditory serial addition, vigilance, 10-choice reaction time, subtraction	..	No difference between groups
Binder and co-workers[34]	Conjugated oestrogens (0·625 mg) plus medroxyprogesterone acetate (5 mg for 13 days every third month)§	9 months	Cognitive function	TMT A and B, verbal fluency, word fluency, Wechsler associate learning and 20 min delayed recall, cancellation random letter and random figure	..	No difference between groups
Stott and co-workers[35]	Folic acid (2·5 mg), vit B12 (500 µg), vit B6 (25 mg), and riboflavin (25 mg) (alone or in combination)	1 year	Cognitive function	Modified TICS, digit coding	..	No difference between groups
Lewerin and co-workers[36]	Cyanocobalamin (0·5 mg), folic acid (0·8 mg), and vit B6 (3 mg)	4 months	Cognitive function	Block design, digit-symbol, DS forwards and backwards, visual reproduction, Thurstone's picture memory, identical forms, synonyms, figure classification	..	No difference between groups

DSM-IV=Diagnostic and Statistical Manual of Mental Disorders 4th edition. NINCDS-ADRDA=National Institute of Neurological Disorders and Communicative Disorders, Alzheimer's Disease and Related Disorders. NINDS-AIREN=National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherché et l'Enseignement en Neurosciences. GEM=Ginkgo Evaluation of Memory study. ADAPT=Alzheimer's Disease Anti-inflammatory Prevention Trial. WHIMS=Women's Health Initiative Memory Study. PREADVISE=Prevention of Alzheimer's Disease by Vitamin E and Selenium. Vit=vitamin. WHS=Women's Health Study. TICS=telephone interview for cognitive status. RAVLT=Rey auditory verbal learning test. TMT=trail making test. COWAT=controlled oral word association test. OPAL=Older People and n-3 Long-chain Polyunsaturated Fatty Acids study. EPA=eicosapentaenoic acid. DHA=docosahexaenoic acid. CVLT=California verbal learning test. WAIS=Wechsler adult intelligence scale. KAI=Kurztest für Allgemeine intelligenz (ie, short test for general intelligence). SOPT=self-ordered pointing test. DS=digit span. QoL=quality of life. CAMCOG=Cambridge cognitive examination. *Doses are per day unless otherwise indicated. †Primary outcomes or measurements used for power calculations. ‡No difference between groups was detected in primary analyses. Secondary analyses that excluded seven patients with dementia who were previously included showed increased hazard ratios for AD with both treatments. Treatment was discontinued because of an increased cardiovascular risk associated with active treatment. §Medroxyprogesterone acetate given only to women who had not undergone hysterectomy.

Table 1: **Outcomes for primary prevention trials**

One of these studies[38] also measured incidence of mild cognitive impairment (MCI). The continuum between normal ageing, very early AD, and AD makes a threshold for conversion between the stages difficult to identify,[41] and the definition of conversion to dementia is largely subjective. However, there are few viable alternatives, and these endpoints are usable if control mechanisms are in place. For example, an independent attribution committee could decide whether conversion has occurred, and if the study investigator and committee disagree patients could be followed up for a further 6 months.

Cognitive decline

Because the conversion to dementia or AD is difficult to assess, changes in cognitive performance or cognitive decline might be a suitable alternative outcome. This type of outcome might mean that fewer patients need to be studied: to detect a meaningful difference in terms of dementia incidence, around 3000 participants are needed,[42] whereas to detect a significant change in cognitive decline, the number needed is 200–800 (table 1).[25–36] Studies of dementia incidence generally last for 3–5 years, whereas studies of cognitive change are generally shorter, usually 6–12 months. Reduction of study duration might be an effective way to reduce attrition.[9]

Changes in the slope of such outcome measures are being increasingly discussed as a way to show stabilisation or slowing of deterioration compared with the control group. Activities of daily living (ADL) or quality-of-life assessments can be used as co-primary endpoints with the cognitive decline outcome to establish the clinical relevance of differences in the primary endpoint between treatment groups.

Recommendations

In view of the advantages in terms of sample size, duration of follow-up, and stability of the effect, we recommend the use of progression of cognitive decline instead of conversion to dementia, with changes in the slope of cognitive tests as a primary outcome and changes in instrumental ADL as a co-primary outcome.

The choice of measures of cognitive decline is an additional hurdle that was not discussed in detail by the task force. Overall, the knowledge base is not sufficiently well developed to enable us to make recommendations, but changes in memory with cued recall (eg, measured with the Grober Buschke test) seem particularly related to changes that occur in AD.[13]

	Intervention	Duration	Primary outcomes	Criteria and measurements	Selected other outcomes	Results
Thal and co-workers (rofecoxib protocol 078)[43]	Rofecoxib (25 mg/day)	4 years	Cumulative incidence of AD*	NINCDS-ADRDA	Cognitive change (SRT summed and delayed recall, MMSE, ADAS-cog, CDR, BDRS)	Conversion rate per year 6·4% in treatment group and 4·5% in placebo group (p=0·01)
Petersen and co-workers[44]	Vit E plus multivitamin or donepezil plus multivitamin	3 years	Time to development of possible or probable AD*	NINCDS-ADRDA	Changes in cognitive, functional and global measures (MMSE, ADAS-cog, CDR, ADCS-ADL/MCI, GDS, paragraph recall, symbol-digit modalities, category fluency, number cancellation, Boston naming, digits backward, clock-drawing, maze-tracing task)	Conversion rate per year 16% in both groups. No differences between groups at 3 years
Feldman and co-workers (InDDEx)[45]	Rivastigmine (3–12 mg/day)	4 years	Rate of progression to AD*, change in cognitive performance	NINCDS-ADRDA, cognitive battery (New York University paragraph recall [immediate and delayed], delayed word list recall, letter–number sequencing, Buschke free and cued selective reminding, symbol-digit modalities, digit-cancellation task, maze, verbal fluency categories subtest, clock drawing)	Changes in cognitive, functional, and global measures (ADAS-cog, MMSE, CDR, GDS, ADCS-ADL, NPI, QOL-AD, HAM-D, MRI)	No differences between groups for conversion rate (17·3% in treatment group *vs* 21·4% in placebo group across 3–4 years)

SRT=selective reminding test. BDRS=Blessed dementia rating scale. ADSC-ADL=Alzheimer Disease Cooperative Study of Activities of Daily Living inventory. ADCS-ADL/MCI=ADCS-ADL adapted to MCI. InDDEx=Investigation into the Delay to Diagnosis of AD with Exelon. GDS=global deterioration scale. QOL-AD=quality of life scale in Alzheimer's disease. HAM-D=Hamilton rating scale for depression. *Primary outcomes or measurements used for power calculations.

Table 2: **Outcomes for secondary prevention trials**

Secondary prevention trials

Secondary prevention trials in AD involve patients with a certain degree of cognitive impairment (eg, MCI; table 2).[43–45] Patients with MCI are often divided into two groups: those with amnestic MCI, who are more likely to develop AD, and those with non-amnestic MCI. Here, we focus on amnestic MCI.[46,47]

Conversion to dementia

The rate of conversion from amnestic MCI to dementia is variable and depends on the criteria used to select the population. Most trials in patients with amnestic MCI so far have been negative,[44] which is almost certainly due to the difference in definitions of MCI between reversible cognitive decline and early dementia. This factor is important in the interpretation of results from secondary prevention trials. The call for a new definition of early AD[1] underlines the difficulties with use of conversion to dementia as an outcome for trials in patients with MCI.

Cognitive decline

Symptomatic changes in cognitive functions are potentially interesting outcomes for MCI trials. Many tests that assess memory, attention, executive functions, and psychomotor speed have been used in secondary prevention trials. The free and cued selective reminding test might also be valuable in this type of trial.[13,48] Recently, a combination of tests that cover multiple cognitive domains has been developed and validated in the form of a neuropsychological test battery called the NTB.[49] This battery was designed for use in mild to moderate AD but might also be useful in secondary prevention trials. However, there is not yet consensus about the specific cognitive tests or combinations of tests that should be used in secondary prevention trials.

Recommendations

Because multiple cognitive domains can be impaired in MCI, we recommend the use of compound scores that combine performance on different validated tests (eg, episodic memory, working memory, and executive function) and use of z-scores.[7,10,50]

Symptomatic trials in very early, mild, and moderate AD

About 75–80% of people with an established diagnosis of probable AD fall into the mild or moderate category.[51] A significant proportion of patients with MCI have very early AD and go on to fulfil the criteria for probable AD.[1] At this stage of the disease, symptomatic trials probably require the development of more sensitive outcome paradigms than those currently available (table 3).[52–65] The low efficacy in recent trials could be as much due to the quality of the assessment methods as to the effect of the treatment.

Issues that need further study include the importance of subgroups (eg, different genotypes)[66] and what constitutes a clinically relevant benefit in the context of the sigmoidal deterioration pattern reported when disease progression is measured over a period of time (eg, the slow decline in cognition in the early stages, and the faster deterioration during the moderate phase). However, the sigmoidal pattern of deterioration seems likely to be a test artefact because the NTB declines linearly, and the ADAS-cog seems to create the sigmoidal shape from differential sensitivity.[49]

Cognitive outcome

The ADAS-cog is the scale most frequently used to assess cognitive outcome. The 11-item version of the scale validated in the 1980s[67] does not include adequate

	Intervention*	Duration	Primary outcomes	Criteria and measurements	Selected other outcomes	Results
Winblad and co-workers[52]	Donepezil (5–10 mg)	1 year	Global dementia assessment	GBS	Measures of cognition (MMSE), ADL (PDS), behaviour (10-item NPI and disease severity, GDS)	Significant effect at 1 year on MMSE and PDS. Borderline effect on GBS (p=0·054)
Wilcock and co-workers (GAL-GBR-2)[53]	Galantamine (24 mg) *vs* donepezil (10 mg)	1 year	Function, cognition behaviour, caregiver burden	BADL, MMSE, ADAS-cog, NPI, screen for caregiver burden	..	No significant difference between groups
Courtney and co-workers (AD2000)[54]	Donepezil (5–10 mg)	>60 weeks planned	Entry to institutional care, progression of disability	Entry to institutional care, progression of disability (loss of 2 of 4 basic, or 6 of 11 instrumental, activities on BADLS)	Functional ability (BADLS), presence and severity of behavioural and psychological symptoms and signs of dementia (NPI), cognition (MMSE), progress to severe cognitive disability (MMSE <10), psychological wellbeing of the principal caregiver (GHQ-30), death	No difference between groups in entry to institutional care or progression to disability. Significantly better MMSE and BADLS at 2 years in treatment group than with placebo
Mohs and co-workers[55]	Donepezil (5–10 mg)	1 year	Functional decline	ADFACS	CDR, MMSE	38% efficacy reduction in the risk of functional decline in donepezil group compared with placebo group
Rogers and co-workers[56]	Donepezil (5–10 mg)	24 weeks	Cognitive and global function	ADAS-cog, CIBIC-plus	MMSE, CDR, QoL	Significantly better ADAS-cog and CIBIC-plus scores with donepezil than with placebo
Rogers and co-workers[57]	Donepezil (5–10 mg)	12 weeks	Cognitive and global function	ADAS-cog, CIBIC-plus	MMSE, CDR, QoL	Significantly better ADAS-cog, CIBIC plus, and MMSE scores with donepezil than with placebo
Burns and co-workers[58]	Donepezil (5–10 mg)	24 weeks	Cognitive and global function	ADAS-cog, CIBIC-plus	CDR, modified IDDD, QoL	Significantly better cognitive (ADAS-cog) and global (CIBIC-plus) function with donepezil than with placebo
Homma and co-workers[59]	Donepezil (5 mg)	24 weeks	Cognitive and global function	ADAS-cog, CIBIC-plus (Japanese versions)	CDR, MENFIS, CMCS	Significantly better cognitive (ADAS-cog) and global (CIBIC-plus) function with donepezil than with placebo
Rosler and co-workers[60]	Rivastigmine (1–4 or 6–12 mg)	26 weeks	Cognitive and global function	ADAS-cog, CIBIC-plus, PDS	..	Significantly better cognitive (ADAS-cog) and global (CIBIC-plus, PDS) function with high-dose rivastigmine than with placebo
Raskind and co-workers (GAL-USA-1)[61]	Galantamine (24–32 mg)	26 weeks	Cognitive and global function	ADAS-cog/11, CIBIC-plus	DAD	Significantly better cognitive (ADAS-cog/11) and global (CIBIC-plus) function with galantamine than with placebo
Wilcock and co-workers (GAL-INT-1)[62]	Galantamine (24–32 mg)	26 weeks	Cognitive and global function	ADAS-cog/11, CIBIC-plus	DAD	Significantly better cognitive (ADAS-cog/11) and global (CIBIC-plus) function with galantamine than with placebo. Higher-dose group had significantly better DAD scores at the end of treatment than did placebo group
Tariot and co-workers (GAL-USA-10)[63]	Galantamine (8, 16, or 24 mg)	5 months	Cognitive and global function	ADAS-cog/11, CIBIC-plus	ADCS-ADL, NPI	Significantly better outcomes for all measures with galantamine than with placebo
Wilkinson and co-workers[64]	Galantamine (18, 24, or 36 mg)	3 months	Cognitive function	ADAS-cog	CGIC, PDS	Significantly better outcomes for all measures with galantamine than with placebo
Rockwood and co-workers[65]	Galantamine (24 or 32 mg)	3 months	Cognitive and global function	ADAS-cog/11, CIBC-plus	ADAS-cog/13, NPI, DAD	Significantly better outcome on cognitive (ADAS-cog) and global (CIBIC-plus) function with galantamine than with placebo, and significant benefits on DAD. Behavioural symptoms (NPI) did not change significantly from baseline in either group

GBS=Gottfries-Brane-Steen scale. PDS=progressive deterioration scale. GAL-GBR=UK galantamine trial. BADL=Bristol activities of daily living. GHQ-30=general health questionnaire, 30 items. ADFACS=functional assessment and change scale. IDDD=interview for deterioration in daily living activities in dementia. MENFIS=mental function impairment scale. CMCS=caregiver-rated modified Crichton scale. GAL-USA=US galantamine trial. GAL-INT=international galantamine trial. ADAS-cog/11=11-item version of ADAS-cog. CGIC=clinical impression of global change. ADAS-cog/13=13-item version of ADAS-cog. *Doses are per day.

Table 3: **Outcomes for symptomatic trials for mild to moderate AD**

assessment of executive function, attention, and working memory, particularly for patients in the early stages of AD. Thus, the ADAS-cog often needs to be supplemented at the milder end of the disease spectrum (eg, with the trail-making test[68] or the Stroop test[69]). In the 1990s, the researchers who devised the original ADAS-cog reviewed such adaptations and added some executive measures in a 13-item version of the scale.[70] The ADAS-cog will probably continue to be a recommended outcome measure. The NTB includes six supplementary tests to measure attention, working memory, and executive function,[49] and when used in its first trial[71] it captured subtle cognitive changes that were not picked up by the ADAS-cog. If the sensitivity of this battery is confirmed in further studies, it might be increasingly used in future trials.

Some investigators believe that computerised assessment protocols are also important in trials. Many of the

tests used in clinical drug trials have few, if any, parallel forms; therefore, although they have reasonable temporal reliability, they are prone to learning effects after repeated testing that might obscure drug effects. However, most patients with AD do not show major learning effects on psychological tests owing to their memory impairment. The key providers of computerised cognitive testing typically offer multiple validated and genuinely parallel versions of their tests. This reduces the risk that drug effects will be obscured by practice and facilitates the use of prebaseline testing, which is a useful way to reduce postbaseline error variance. Computerised assessments also offer the advantage of precise measurement of timing. However, these assessments are appropriate only if the domains measured are clinically relevant to the disease process and can support a reasonable claim for efficacy.

The ADAS-cog can sometimes show a clinically significant benefit in cognition—eg, a 4-point difference in the cognitive change from baseline to the study end between treated and placebo populations over a 6-month trial period. A 7-point decline at 6 months has been associated with an increased risk of severe dementia or death at 2 years after study entry,[72] which confirms the value of the ADAS-cog scale as a good surrogate marker of long-term prognosis. However, changes in patients with mild AD are becoming increasingly difficult to identify with the ADAS-cog over 6 months, because more of the patients who enter trials have mild AD (mini-mental state examination [MMSE] >20) and are in good general health than in previous studies, and changes in the ADAS-cog in a 6-month period are often small.[6] A non-significant change in the ADAS-cog could therefore be due to a very small or null effect of the drug being tested, to the absence of decline in the placebo group, or to insensitivity of the outcome measures for the study population.[73] Links between baseline severity and change from baseline with the ADAS-cog should therefore also be taken into account.

Functional outcome

Deterioration in day-to-day functional ability is a major feature of most dementias, and any change in performance in this context is an important outcome measure. Scales are based mainly on an interview with the patient, their caregiver, or both, and in general cover both basic ADL (eg, eating and dressing) and more complex or instrumental ADL (eg, ability to use the telephone, other household equipment, or transport, or ability to go shopping). The Alzheimer's Disease Cooperative Society (ADCS)-ADL scale[74] is the scale most widely used to assess functional outcome in patients with AD; the disability assessment for dementia (DAD) scale[75] is also frequently used and can show linear change over time.[76] The Bristol ADL scale was also developed for use in clinical trials.[77,78] In early dementia, assessment should include instrumental ADL items,[79] but whether specific functional measures offer distinct advantages in terms of scaling is unclear.

Behavioural and psychological symptoms of dementia outcomes

Diverse behavioural and psychological symptoms of dementia (BPSD) are used as outcomes. These important outcome measures, sometimes called neuropsychiatric features, are thought of as non-cognitive symptoms of AD and have been shown to predict entry into institutional care. The most widely used scale in this context is the neuropsychiatric inventory (NPI),[80] which covers a range of BPSD, but the use of an overall NPI score might mask important changes in individual BPSD items. The BPSD spectrum is so wide that trials are needed to evaluate its different aspects.[81] The effects of a drug on behavioural disturbances that were present at baseline should also be distinguished from those on disturbances that emerge during a clinical observation period, because these effects might differ.

Global outcome

Global assessment measures often still include the clinician's interview-based impression of change with caregiver input (CIBIC-plus). This measurement has been criticised because it was devised via a regulatory route rather than through basic scientific development; its application has since been standardised[82] but changes in function have now been recommended for regulatory purposes.[83] 10 years ago, a case was made for such global measures to be replaced by appropriate combinations of tests to cover a range of domains.[84] Other measures include the clinical dementia rating (CDR)[85] and the global deterioration scale,[86] in which clinicians decide which statement from a list best describes the patient's ability. The CDR is one of the most frequently used and recommended measures in dementia drug trials, especially the sum of boxes version. However, this rating does not have a behavioural component, and several other global measures, such as the functional assessment staging tool, are also well established.[87]

Health economics outcome

Health economics outcomes have become increasingly important over the past few years. Two approaches are used: statistical modelling, and analysis of data from trials. In analysis approaches, such as the resource utilisation in dementia instrument (RUD), resource use is measured and incremental costs are compared with incremental effects to produce a cost–benefit analysis.[88]

Biomarker outcome

We discuss biomarkers in the section on disease-modifying agents, because they are more likely to be of value in that context than in trials that evaluate symptomatic outcome.

Recommendations

We recommend that cognition and function are the two primary outcomes for trials in mild to moderate dementia. We recommend that cognition is assessed with the ADAS-cog or a neuropsychological test battery in early and mild dementia, and with the ADAS-cog in moderate dementia; function should be assessed with the instrumental ADL at these stages. For moderate dementia, a more global assessment (CDR) can be used as a co-primary outcome with cognition.

Disease-modifying trials in very early, mild, and moderate AD

Disease modification implies arrest or retardation of the processes that led to neuronal loss or malfunction, to produce clinical benefit. This approach, which is being explored mainly in mild and moderate rather than severe AD, has been discussed in detail previously.[5] By definition, symptomatic effects can be difficult to distinguish from disease modification, and some drugs might have both symptomatic and disease-modifying properties. The outcome measures chosen will depend on the trial design and stage of the disease.

Trial design

Clinical trials can continue for 18 months or more. Therefore, outcome measures should be valid at different stages of the disease, because participants might move from one disease stage to another during the study. Additionally, the trial process should not be burdensome for patients and their carers, because the long duration of these studies often results in a high drop-out rate. In mild AD, outcome measures need to be sensitive to the fact that cognition and instrumental ADL scores might both deteriorate slightly over a period such as a year, whereas basic ADL items are usually more stable. In moderate AD, there will be a more obvious decline in all areas over a year, and neuropsychiatric symptoms are more likely to emerge.

Randomised start or withdrawal designs[89] have been advocated, but they require longer follow-up, and are likely to involve higher drop-out rates and costs than other types of trial. An alternative is the delay to milestones approach (ie, a survival-type analysis), which measures the proportion of patients who reach specific endpoints over time. The endpoints chosen will need face-validity in terms of clinical relevance, such as a delay of at least 6 months for an important milestone (eg, CDR). A parallel group design to assess longitudinal data, sometimes called a slope analysis, is another valid and useful approach. One of the strengths of this analysis is that it looks at all available data for any period of time for all participants. Slope analysis can include elements of a survival-type analysis, and a pseudo-staggered start design if there is a subsequent open-label study when patients from the placebo group start to take the drug. This method allows missing data to be taken into account with less bias than with the last observation carried forward and observed cases approaches. The recently proposed natural history staggered start analysis compares the slopes of decline of patients who receive study drug with those of patients who receive placebo, and corrects for the severity of disease at baseline.[90]

Clinical outcome

Outcome measures for disease-modifying trials are currently based on those traditionally used in symptomatic trials[91] (table 4).[92–100] As disease-modifying trials concentrate increasingly on the mild or very early stages of AD, sensitivity issues will need to be addressed; hence, the supplementation of the ADAS-cog with batteries such as the NTB. The primary outcome measures will need to cover cognition, functional ability, and global assessment, and secondary outcome measures will also resemble those used in symptomatic trials. In the Real.fr study,[6] mean changes in the ADAS-cog over 18 months were only 3·02 points (SD 5·63) in patients with a stable dose of cholinesterase inhibitors from at least 4 months before study entry. However, a 2-point change at 18 months on the ADAS-cog scale was related to a further loss of ADL.[101] Knowledge of the inherent difficulties with assessment methods has resulted in further developments and new methods are under evaluation.[102]

Biomarker outcome

Increasing use of biomarkers (eg, in CSF and MRI) in trials will be an inevitable consequence of the pressure to capture disease-modifying properties of new drugs. Although no biomarkers predict clinical decline sufficiently well to be used as a surrogate endpoint instead of clinical measures, many correlate with diagnosis and some correlate with decline. These biomarkers can be used to supplement clinical results to support the effect of an agent on the underlying disease process. However, biomarkers might not always behave as predicted in assessments of imaging outcomes. For example, in the AN1792 trial,[103] more cerebral atrophy was recorded in the treated group than in the placebo group, without an increase in cognitive decline.

To be used as surrogate outcome measures, biomarkers need to be easily obtainable, to be objectively measurable, to relate to disease processes, to be expected to respond to or be stabilised by treatment, and if possible to predict clinical response. Smaller standard deviations than with clinical measures would also be valuable, because this could reduce numbers of patients needed and possibly trial duration. Biomarkers might also improve diagnosis, and possibly eventually indicate which patients would best be treated with specific therapies. Change in a biomarker might also reassure patients and their physicians that a drug is having an effect before a clinical change becomes apparent.

	Intervention*	Duration	Primary outcome	Criteria and measurements	Selected other outcomes	Results
Sanofi-Aventis trials EFC2724 and EFC2946[92]	Xaliproden	18 months	Cognitive function, global rating, brain physiopathology	ADAS-cog, CDR, hippocampal atrophy (subset)	..	No significant effects on primary clinical endpoints in either trial; significantly less hippocampal atrophy with xaliproden than with placebo in EFC2724[90]
NIA (VITAL)[93]	B vitamins	18 months	Cognitive function, global rating	ADAS-cog, CDR	..	Not yet available
Neurochem North American trial[94]	Tramiprosate	18 months	Cognitive function, global rating	ADAS-cog, CDR	Brain volume change (MRI)	No significant effects†
Neurochem European trial[95]	Tramiprosate	18 months	Cognitive function, global rating	ADAS-cog, CDR	..	Trial discontinued‡
NIA (CLASP)[96]	Simvastatin (20–40 mg)	18 months	Cognitive function, global rating	ADAS-cog, ADCS-CDIC	Measures of clinical global change (ADCS-CGIC), mental status, functional ability, behavioural disturbances, quality of life, economic indicators	Not yet available
Feldman and co-workers[97]	Atorvastatin (80 mg) plus donepezil (10 mg)	18 months	Cognitive function, global rating	ADAS-cog, CDR	..	Not yet available
Myriad US trial[98]	R-flurbiprofen	18 months	Cognitive outcome, functional outcome	ADAS-cog, ADCS-ADL	..	Not yet available
Myriad international trial[99]	R-flurbiprofen	18 months	Cognitive outcome, functional outcome	ADAS-cog, ADCS-ADL	..	Not yet available
NIA, ADCS, Martek Biosciences[100]	DHA	18 months	Cognitive change, global rating	ADAS-cog, CDR	..	Not yet available

NIA=National Institute on Aging. VITAL=Vitamins to Slow Alzheimer's Disease Study. CLASP=Cholesterol-owering Agent to Slow Progression of Alzheimer's Disease Study. ADCS-CDIC=ADCS clinical global impression of change. *Doses are per day. †B Vellas, unpublished. ‡Discontinued because of negative results in the North American trial.

Table 4: **Outcomes for disease-modifying trials for mild to moderate AD**

However, no marker to reliably measure a change in response to therapy in a predictable direction has been validated. Feldman and colleagues[45] reported that the rate of change in hippocampal volume did not correlate with changes on any clinical outcome measure. Future trials should include assessment of biomarkers because this will advance our knowledge about their usefulness.

Recommendations

For disease-modifying trials, we recommend adoption of the most widely used outcomes for symptomatic trials in mild to moderate AD: the ADAS-cog, ADCS-ADL scale, and CDR sum of boxes. No available data suggest suitable alternatives. The consensus was that a 2-point difference between groups at 18 months on the ADAS-cog should be the minimal clinically important change (MCIC).

Trials in severe AD

Around 20% of patients with AD are estimated to be at the severe dementia stage (MMSE <10).[51] Patients with severe dementia have both severe cognitive impairment and functional loss in basic ADL. Aggressive forms of the disease, with rapid cognitive decline and functional loss, must be assessed in a different way to severe AD and are not discussed in this Review. However, reduction of the percentage of patients with rapid progression could be an important outcome for new drugs. Severe dementia is a substantial part of the burden for patients, caregivers, and society. For these reasons, clinical trials in patients with severe AD need to be strongly encouraged. Many such patients continue to walk and eat with assistance, communicate in short phrases or single words, interact with others, and complete several basic ADL with help. Even when the disease becomes very severe, they are still able to interact with their environment. The goals of treatment for patients with severe AD are to improve symptoms, preserve residual abilities, and slow symptom progression.

Trial design

The typical design to show symptomatic improvement in severe AD is a randomised, double-blind study to compare differences between drug and placebo in two co-primary outcomes, one for cognition and one preferably for functional impairment[104] (table 5).[105–119] The changes must be robust and clinically meaningful in favour of active treatment versus placebo. The MMSE and the modified MMSE are useful in stratification of patients for trial entry.

Cognitive outcome

The most widely used assessment for patients with advanced dementia is the severe impairment battery (SIB),[120] an objective, performance-based evaluation of cognitive functions by use of a structured, interview-like assessment. The domains tested are analogous to those assessed by the ADAS-cog. Data on reliability and validity are available but more data are needed on the links between the SIB and functional measurements.

Functional outcome

Function is probably the most suitable outcome domain for trials in severe dementia, and tests of functional outcome such as the Bristol ADL scale,[78] the ADCS-ADL scale modified for severe dementia (ADCS-ADL-sev),[121] and the DAD scale[75] have been validated. Assessment of the loss of basic ADL by functional outcome measures is particularly important at this stage of AD.

Global outcome

A global outcome must be included. The CIBIC-plus and clinical global impression of improvement (CGI) scores have been used in trials in severe dementia.[105,107–111,115–118] However, global outcomes are particularly difficult to assess in patients with severe dementia, for example when patients live in nursing homes.

Health economic outcome

The cost of dementia increases as the disease becomes more severe.[122] Some assessments, such as the RUD, have been developed specifically to measure resource use in clinical trials of AD, with a focus on key resource use items: accommodation, community care services, inpatient care, and informal care. The RUD[88] has been

	Intervention*	Duration†	Primary outcomes	Criteria and measurements	Selected other outcomes	Results
Black and co-workers[105]	Donepezil (10 mg)	24 weeks	Global function	SIB‡, CIBIC-plus‡	Cognitive, functional, behavioural, caregiver and resource utilisation outcomes (MMSE, ADCS-ADL-sev, NPI, CBQ, RUSP)	Donepezil had a significant effect on change in SIB scores, CIBIC-plus, and MMSE scores at end of treatment
Winblad and co-workers[106]	Donepezil (5–10 mg)	6 months	Change in global and functional measures	SIB‡, ADCS-ADL-sev‡	..	Patients treated with donepezil improved more in SIB scores and declined less in ADCS-ADL-sev scores compared with placebo
Winblad and Poritis[107]	Memantine (10 mg)	12 weeks	Global change; behavioural outcome	CGI-C‡, BGP-CDS‡	Modified D-scale (Arnold/Ferm)	Memantine had a significant effect on changes in CGI-C and BGP-CDS scores
Tariot and co-workers[108]	Memantine (5–20 mg) in patients already on donepezil	24 weeks	Cognitive change, functional change	SIB‡, ADCS-ADL19‡	CIBIC-plus, NPI, BGP-CDS	Memantine had a significant effect on all outcome measures
Feldman and co-workers (MSAD)[109,110]	Donepezil (5–10 mg)	24 weeks	Global change	CIBIC-plus‡	Cognitive outcomes (sMMSE and SIB), functional outcomes (DAD, IADL-plus, PSMS), behavioural outcomes (NPI), global functioning (FRS), caregiver quality of life (SF-36, CSS), healthcare resource utilisation (CAUST)	Patients receiving donepezil showed significant benefits on CIBIC-plus and all other outcome measures after 24 weeks
Reisberg and co-workers[111]	Memantine (20 mg)	28 weeks	Global change, functional outcome	CIBIC-plus; ADCS-ADL-sev	Measures of cognition, function, and behaviour (SIB, MMSE, GDS, FAST, NPI, RUD)	Patients on memantine had a better outcome than those on placebo, according to CIBIC-plus, ADCS-ADL-sev, SIB
Eisai[112]	Donepezil (23 mg sustained release *vs* 10 mg immediate release)	..	Cognitive, global and functional outcomes	SIB, CIBIC-plus, ADCS-ADL	MMSE	Not yet available
Eisai[113]	Donepezil	..	Global and cognitive function	..	ADL and caregiver burden	Not yet available
Forest Laboratories[114]	Memantine	..	Behavioural outcomes	NPI	CMAI, CGI, ADCS-ADL, agitation/aggression domain of NPI	Study completed. Results not yet available
NYU, Forest Laboratories, Fisher Center for AD Research[115]	Memantine and comprehensive individualised management of AD patients and caregiver training	28 weeks	Global and functional outcomes	CIBIC-plus, ADCS-ADL-sev	SIB, MMSE, FAST, GDS, behavioural pathology in AD, memory and behaviour problems checklist	Not yet available
East Kent Hospitals Trust (MAGD)[116]	Memantine	<12 weeks	Agitation	CMAI	NPI, CGI, SIB	Not yet available
Eisai[117]	E2020 (donepezil 5 and 10 mg)	..	Global and cognitive function	CIBIC-plus, SIB	Behave-AD, ADCS-ADL-sev	Not yet available
Forest Laboratories[118]	Modified-release memantine with concurrent AChEI	..	Global and cognitive function	SIB, CIBIC-plus	NPI, ADCS-ADL	Not yet available
Janssen[119]	Galantamine (8–24 mg)	26 weeks	Cognition, function	SIB, minimum dataset ADL test	NPI, behavioural, social, and physical functioning, level of caregiver support, effect on caregiver, MMSE	Not yet available

CBQ=caregiver burden questionnaire. RUSP=resource utilization for severe AD patients. BGP-CDS=behavioural rating scale for geriatric patients, care-dependency subscale. ADCS-ADL19=modified 19-item ADCS-ADL. MSAD=Moderate to Severe AD study. sMMSE=screening standardised MMSE. IADL-plus=modified instrumental activities of daily living scale. PSMS=physical self-maintenance scale. Behave AD=behavioral pathology in AD. CGI-I=clinical global impression of change. FRS=functional rating scale. SF-36=short form 36. CSS=caregiver stress scale. CAUST=Canadian utilization of services tracking. NYU=New York University. CMAI=Cohen Mansfield agitation inventory. FAST=functional assessment staging. MAGD=Memantine for Agitation in Dementia study. AChEI=acetylcholinesterase inhibitors. *Doses are per day unless otherwise indicated. †Where no duration is given this is because no duration was listed on ClinicalTrials.gov when the trial page was accessed. ‡Primary outcomes or measurements used for power calculations.

Table 5: **Outcomes for symptomatic trials for moderate to severe AD**

validated against caregiver diaries and direct observation. The burden interview[123] is the most widely used assessment in social research to assess the burden for caregivers, but is not very sensitive to change; the relative's stress scale[124] is more sensitive. Informal care should be assessed because much time is spent on it. However, the real value of such health economics data is relative, because it involves a specific selection of patients and results cannot be generalised to the whole population.[125–128]

Other clinically relevant outcomes

Assessment of quality of life is important, but no measures are yet validated for use in patients with severe AD. Any measure is likely to need proxy completion, which might not be representative of the patient's opinion of his or her quality of life. Carer quality of life is also important; some studies are now using the Logston scale,[129] and more data will be soon available. Owing to the day-to-day variability of AD, both bad days and good days in a month could be assessed.

Behavioural disorders are common in severe AD and can be measured by the NPI.[80] The NPI total score must be used with caution: in disease-modifying studies, reductions in frequency and severity of individual NPI domains and the emergence of new NPI symptoms should be assessed separately. The analysis of single NPI items or of subgroups of items, or the use of other specific behavioural scales (eg, to measure apathy or depression) will often be necessary.

Recommendations

We recommend that the time to reach a specified value on an incapacity scale, entry to institutional care, number of admissions to hospital, need for home help, level of satisfaction, loss of a basic ADL, and onset of behavioural disorders are clinically relevant outcomes for trials in patients with severe AD. The SIB, ADCS-ADL-sev, and NPI are the most widely used measurements in trials in patients with severe AD, and they remain recommended by the task force.

Discussion

According to the guidelines recently released for consultation by the European Committee for Human Medicinal Products (CHMP),[83] outcomes for symptomatic trials need to address the following domains: cognition, as measured by objective tests (cognitive endpoint); ADL

	Strengths	Limitations
Prevention trials		
Conversion to dementia or AD[19–24]	Most clinically relevant outcome	Relatively subjective; difficult to determine the precise moment of conversion; incidence can remain low in healthy older patients; independent committee necessary to validate diagnoses
Cognitive decline (eg, memory testing with cued recall)[25–36]	More sensitive and objective than conversion outcome; can detect small cognitive changes; individual cognitive domains can be studied separately	No standardised cognitive test battery for prevention trials; some well known instruments (eg, MMSE) might not be sensitive enough to detect early signs of cognitive decline; clinical relevance and suitability of cognitive decline as a surrogate marker for dementia still to be established
Symptomatic trials: mild to moderate AD		
ADAS-cog[53,56–65]	Widely used and standardised; can show some symptomatic effects in trials of cholinesterase inhibitors; sensitive to change in moderate AD	Inadequate assessment of some cognitive domains, especially in very early AD; unclear definition of what constitutes a clinically important change, especially in mild AD
NTB[71]	More sensitive than the ADAS-cog and covers more cognitive domains	Requires further validation in therapeutic trials
CIBIC-plus[56–63,65]	Significant change reported in clinical trials*	High inter-rater variability hinders comparison between different studies
Symptomatic trials: moderate to severe AD		
Cognition: SIB[105,106,108,112,117–119]	Adapted for patients with severe AD who might not be able to complete other measures (eg, MMSE or ADAS-cog); scoring based on correct responses rather than errors, and partial responses are credited	Might be difficult to show clinically important differences
Function: ADCS-ADL-sev[106,111,115]	Adapted for patients with severe dementia	Might be difficult to show clinically important differences
ADCS-ADL[108,112]	Significant change reported in clinical trials*	Not adapted for severe dementia
CGI-C[107]	Global assessment	Might be subject to inter-rater variability, thus hindering comparison between different studies
CIBIC-plus[105,110–112,116,118]	Global assessment	High inter-rater variability hinders comparisons between different studies
Disease-modifying trials: mild to moderate AD		
ADAS-cog[92–100]	Widely used in symptomatic trials	No significant differences between groups in trials so far.* Only evidence for disease-modifying effects comes from long trial duration
NTB	More sensitive than the ADAS-cog and covers more cognitive domains	Requires further validation in therapeutic trials†
ADCS-ADL[98,99]	Widely used in symptomatic trials	Negative in trials so far*
CDR[92–97,100]	Suggested as a co-primary outcome (global outcome)	No behavioural component
CIBIC	Global assessment	Difficult to interpret in long-term trials

*Most of these trials were carried out several years ago versus placebo. However, significant changes now seem more difficult to detect in add-on trials with patients who have milder AD. †There is still discussion about the clinical relevance of changes in scores on this assessment.

Table 6: **Strengths and limitations of primary outcomes in AD trials**

(functional endpoint); and overall clinical response (global endpoint). The global endpoint should not be one of the two co-primary outcomes, because this position should be reserved for the cognitive and functional outcomes. To show true disease modification there needs to be an effect in a clinical outcome accompanied by strong supportive evidence from a biomarker. To avoid the difficulties posed by the absence of an adequately validated biomarker, a two-step approach is recognised in the CHMP guidelines: in a first step, a delay in the natural course of the disease can be based on clinical signs and symptoms; if these results are supported by a convincing package of biological data, neuroimaging data, or both (eg, to show delay in the progression of brain atrophy), a full claim for disease modification could be considered.[83]

Table 6 shows the main endpoints that are currently used in AD clinical trials. Those for mild to moderate AD are more satisfactory than those for very mild or severe AD. Further development of outcome measures for severe AD and for trials of preventive treatments will probably be informed by current trials. Almost all endpoints to evaluate AD outcomes are scoring systems based on a continuous or categorical scale. As such, the extent of change in these endpoints that is clinically relevant should be clarified. The MCIC for any given measurement is defined as the smallest difference between two assessments that has a perceived effect on disability.[130] Calculation of how big this change should be is difficult: on a properly constructed categorical scale, each step is by definition a MCIC, but on a continuous scale such a change has to be classified empirically. Only a few studies to identify MCIC have been published.[131,132]

Despite previous research, the MCIC has not been empirically established for any of the currently available endpoints. The commonly quoted MCIC for the ADAS-cog of 4 points is the result of a post-hoc agreement rather than prospective findings. Furthermore, it applies only to the patients who are categorised as mild to moderate. Classic randomised placebo-controlled studies are no longer possible in many countries because of the availability of existing medication. We have to use the model of add-on therapy—thus, the natural history of the disease will be different to that in previous studies. Patients with AD in developed countries have a milder form of the disease at the time of diagnosis, a less progressive disease course, and fewer comorbidities than was previously the case.[133] The small decline in outcome that is likely to occur in the placebo group is an important factor in choice of outcome measures generally, and in particular means that clinically relevant benefits might be difficult to show at 6 months or even 18 months. Larger trials and longer follow-up periods are now necessary because we still do not have sufficiently validated biomarkers. Many of the patients with prominent behavioural disorders cannot be recruited into studies, or are more likely to drop out when behaviour deteriorates, which potentially introduces a bias that further complicates study design and interpretation—hence the importance of the extensive neuroimaging and biomarker programmes that are in progress, which will hopefully increase the feasibility of clinical trials in AD and provide proof of disease modification.

One important difficulty in interpretation of present studies of AD is their multicentre nature. Owing to difficulties with recruitment of patients, most trials now involve many centres from North America and Europe, and more recently China and India. Outcome measures have to take this fact into account. Because of the educational and cultural aspects of dementia trials, we recommend the use of more homogeneous target populations, and use of collaborative networks to recruit at least 10–20 patients per centre, and no fewer than five as has been common previously. Site-related effects and smaller than expected changes in placebo groups have been important limiting factors in recent trials (eg, the North American phase III trial of tramiprosate).[2] The effects of a drug that cured AD would be easy to see with currently used outcomes—this scenario is unfortunately not probable, but we can aim to have more drugs with a moderate effect and to reach a clinically relevant endpoint through the association of these drugs. We hope that the outcomes presented in this Review will help with the achievement of this goal in the right target population.

Search strategy and selection criteria

An extensive literature search was undertaken in MEDLINE with the search terms "Alzheimer and clinical trial", "Alzheimer and disease modifying", "Alzheimer and therapeutic trials", "Alzheimer and study design", "MCI and treatment", "dementia and therapeutic trials", "Alzheimer and biomarkers", "Alzheimer and economic aspects", "Alzheimer and neuroimaging", "prevention and Alzheimer", "prevention and cognitive decline", "prevention and dementia", and "prevention and cognitive impairment". Initial searches of electronic databases were done in March, 2007, and were updated in January, 2008. No date restrictions were used. Articles were also identified by members of the task force, through searches of their files and through their experience and contacts in the field. No language restrictions were applied. Clinical trials were identified from the ClinicalTrials.gov online database with the search term "Alzheimer's". The final reference list was generated on the basis of relevance to the topics covered in the Review.

Contributors

BV, SA, CS, and GW planned the meeting, identified and invited participants, and prepared the meeting agenda. All authors contributed to writing of the report, and members of the task force group commented on the draft.

European Task Force Group
Belgium—G Preuveneers (Brussels); E Salmon (Liege). *Denmark*—K Ahrend Petersen, S Larsen (Valby). *France*—J M Orgogozo (Bordeaux); J Gardette, M Zaim (Castres); F Portet, J Touchon (Montpellier); P Robert (Nice); E Abadie, C Deguines, A Delgado, F Duveau, S Goni, P Gustovic, F Keime-Guibert, H Matheix-Fortunet, C Picot, O Sol (Paris); S Andrieu, N Coley, E Reynish, P J Ousset, B Vellas (Toulouse). *Germany*—K Broich (Bonn); A Grazf, R Lane (East Hanover); S Kreimpien, A Stöffler (Frankfurt); R Hoerr (Karlsruhe); L Frolich (Mannheim); A Merdes (München). *Italy*—GB Frisoni (Brescia). *Netherlands*—C Gispen-Wied (The Hague); F Verhey (Maastricht); P Kampuis (Wageningen). *Sweden*—H Basun, I Nordgren (Södertälje); L Jonsson, B Winblad (Stokholm). *Portugal*—C Sampaio (Lisbon). *Switzerland*—J Durga (Lausanne); Y Guigoz (Konolfingen); E Schiffrin (Vevey); H J Möbius (Zurich). *UK*—R Jones (Bath); M Egan (North Wales); N Fox, S Lovestone, J Swartz (London); G Maher-Edwards (Middlesex); G Wilcock (Oxford); J Harrison (Warminster). *USA*—S Abu-Shakra (Irvine, CA); E Liu, R Townsend (San Diego, CA); R Black, J M Ryan (Collegeville, PA); M Carrillo (Chicago, IL); R Mohs (Indianapolis, IN); D Deptula, V Logovinsky, A Ruiz (Whitehouse Station, NJ); R Schindler (New York, NY); S Hendrix, K Zavitz (Salt Lake City, UT).

Conflicts of interest
BV has received research grants from Ipsen, Lilly, Lundbeck, Sevier, Esai, Wyeth, Nestle, and Neurochem. SA has received research grants from Ipsen, Lilly, Lundbeck, Sevier, Esai, and Nestle. CS is an alternate member of the CHMP and the Scientific Advice Working Party (both scientific committees of the European Medicines Evaluation Agency). CS's department has received grants from Kyowa, Servier, Bial, Astellas, Xytis, and Neurobiotech. NC is the beneficiary of a CIFRE PhD studentship that is jointly financed by the French Ministry of Research and Ipsen. GW is a consultant to Myriad and Neuropharm, and has received honoraria from Lundbeck, Astra-Zeneca, Applied Neurodiagnostics, Roche, Lilly, and GSK for participation in advisory boards about choice of potential new drugs and their evaluation.

Acknowledgments
We are grateful to the European Alzheimer Disease Consortium (QLK6-CT-2001-30003).

References
1 Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. *Lancet Neurol* 2007; **6:** 734–46.
2 Butcher J. The hunt for drugs to modify Alzheimer's disease. *Lancet Neurol* 2007; **6:** 1038–39.
3 Rockenstein E, Crews L, Masliah E. Transgenic animal models of neurodegenerative diseases and their application to treatment development. *Adv Drug Deliv Rev* 2007; **59:** 1093–102.
4 Sampaio C. Clinical Relevance on Alzheimer's disease endpoints. *J Nutr Health Aging* 2007; **11:** 316–17.
5 Vellas B, Andrieu S, Sampaio C, Wilcock G. Disease-modifying trials in Alzheimer's disease: a European task force consensus. *Lancet Neurol* 2007; **6:** 56–62.
6 Cortes F, Portet F, Touchon J, Vellas B. Six and 18-month changes in mild to moderate Alzheimer's patients treated with acetylcholinesterase Inhibitors: what can we learn for clinical outcomes of therapeutic trials? *J Nutr Health Aging* 2007; **11:** 330–37.
7 Durga J, Van Boxtel MPJ, Schouten EG, et al. What can we learn from the FACIT trial: a randomized, double blind, controlled trial. *J Nutr Health Aging* 2007; **11:** 320–24.
8 Frisoni GB, Caroli A. Neuroimaging outcomes for clinical trials. *J Nutr Health Aging* 2007; **11:** 348–52.
9 Gardette V, Coley N, Toulza O, Andrieu S. Attrition in geriatric research: how important is it and how should it be dealt with? *J Nutr Health Aging* 2007; **11:** 265–71.
10 Harrison JE. Measuring cognitive change in Alzheimer's disease clinical drug trials. *J Nutr Health Aging* 2007; **11:** 327–29.
11 Jönsson L. Assessing health economic outcome in Alzheimer's disease clinical trials. *J Nutr Health Aging* 2007; **11:** 353–56.
12 Robert P, Verhey FRJ, Aalten P, Cortes F, Byrne EJ. Neuropsychiatric outcome for clinical trials. *J Nutr Health Aging* 2007; **11:** 345–47.
13 Salmon E. Outcomes for secondary preventive trials. *J Nutr Health Aging* 2007; **11:** 325–26.
14 Vellas B. Outcomes for Alzheimer's trials. *J Nutr Health Aging* 2007; **11:** 313.
15 Verhey FRJ, de Vugt ME, Aalten P, Vernooij-Dassen MJ, Byrne EJ, Robert P. Caregiver outcomes in disease modifying trials. *J Nutr Health Aging* 2007; **11:** 342–44.
16 Wilcock G. Outcomes for disease modifying trials. *J Nutr Health Aging* 2007; **11:** 318–19.
17 Brookmeyer R, Gray S, Kawas C. Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. *Am J Public Health* 1998; **88:** 1337–42.
18 Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement* 2007; **3:** 186–91.
19 Vellas B, Andrieu S, Ousset PJ, Ouzid M, Mathiex-Fortunet H. The GuidAge study: methodological issues. A 5-year double-blind randomized trial of the efficacy of EGb 761 for prevention of Alzheimer disease in patients over 70 with a memory complaint. *Neurology* 2006; **67** (suppl 3)**:** S6–11.
20 DeKosky ST, Fitzpatrick A, Ives DG, et al. The Ginkgo Evaluation of Memory (GEM) study: design and baseline data of a randomized trial of *Ginkgo biloba* extract in prevention of dementia. *Contemp Clin Trials* 2006; **27:** 238–53.
21 ADAPT research group. Naproxen and celecoxib do not prevent AD in early results from a randomized controlled trial. *Neurology* 2007; **68:** 1800–08.
22 Shumaker SA, Legault C, Rapp SR, et al. Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. *JAMA* 2003; **289:** 2651–62.
23 Shumaker SA, Legault C, Kuller L, et al. Conjugated equine estrogens and incidence of probable dementia and mild cognitive impairment in postmenopausal women: Women's Health Initiative Memory Study. *JAMA* 2004; **291:** 2947–58.
24 Kryscio RJ, Mendiondo MS, Schmitt FA, Markesbery WR. Designing a large prevention trial: statistical issues. *Stat Med* 2004; **23:** 285–96.
25 Kang JH, Cook N, Manson J, Buring JE, Grodstein F. A randomized trial of vitamin E supplementation and cognitive function in women. *Arch Intern Med* 2006; **166:** 2462–68.
26 Kang JH, Cook N, Manson J, Buring JE, Grodstein F. Low dose aspirin and cognitive function in the women's health study cognitive cohort. *BMJ* 2007; **334:** 987.
27 McMahon JA, Green TJ, Skeaff CM, Knight RG, Mann JI, Williams SM. A controlled trial of homocysteine lowering and cognitive performance. *N Engl J Med* 2006; **354:** 2764–72.
28 Dangour AD, Clemens F, Elbourne D, et al. A randomised controlled trial investigating the effect of n-3 long-chain polyunsaturated fatty acid supplementation on cognitive and retinal function in cognitively healthy older people: the Older People And n-3 Long-chain polyunsaturated fatty acids (OPAL) study protocol [ISRCTN72331636]. *Nutr J* 2006; **5:** 20.
29 Eussen SJ, de Groot LC, Joosten LW, et al. Effect of oral vitamin B-12 with or without folic acid on cognitive function in older people with mild vitamin B-12 deficiency: a randomized, placebo-controlled trial. *Am J Clin Nutr* 2006; **84:** 361–70.
30 Wolters M, Hickstein M, Flintermann A, Tewes U, Hahn A. Cognitive performance in relation to vitamin status in healthy elderly German women-the effect of 6-month multivitamin supplementation. *Prev Med* 2005; **41:** 253–59.
31 Bryan J, Calvaresi E, Hughes D. Short-term folate, vitamin B-12 or vitamin B-6 supplementation slightly affects memory performance but not mood in women of various ages. *J Nutr* 2002; **132:** 1345–56.
32 Almeida OP, Lautenschlager NT, Vasikaran S, Leedman P, Gelavis A, Flicker L. A 20-week randomized controlled trial of estradiol replacement therapy for women aged 70 years and older: effect on mood, cognition and quality of life. *Neurobiol Aging* 2006; **27:** 141–49.
33 Polo-Kantola P, Portin R, Polo O, Helenius H, Irjala K, Erkkola R. The effect of short-term estrogen replacement therapy on cognition: a randomized, double-blind, cross-over trial in postmenopausal women. *Obstet Gynecol* 1998; **91:** 459–66.

34 Binder EF, Schechtman KB, Birge SJ, Williams DB, Kohrt WM. Effects of hormone replacement therapy on cognitive performance in elderly women. *Maturitas* 2001; **38:** 137–46.

35 Stott DJ, MacIntosh G, Lowe GD, et al. Randomized controlled trial of homocysteine-lowering vitamin treatment in elderly patients with vascular disease. *Am J Clin Nutr* 2005; **82:** 1320–26.

36 Lewerin C, Matousek M, Steen G, Johansson B, Steen B, Nilsson-Ehle H. Significant correlations of plasma homocysteine and serum methylmalonic acid with movement and cognitive performance in elderly subjects but no improvement from short-term vitamin therapy: a placebo-controlled randomized study. *Am J Clin Nutr* 2005; **81:** 1155–62.

37 Petersen RC, Doody R, Kurz A, et al. Current concepts in mild cognitive impairment. *Arch Neurol* 2001; **58:** 1985–92.

38 Shumaker SA, Reboussin BA, Espeland MA, et al. The Women's Health Initiative Memory Study (WHIMS): a trial of the effect of estrogen therapy in preventing and slowing the progression of dementia. *Control Clin Trials* 1998; **19:** 604–21.

39 Martin BK, Meinert CL, Breitner JC. Double placebo design in a prevention trial for Alzheimer's disease. *Control Clin Trials* 2002; **23:** 93–99.

40 Fitzpatrick AL, Buchanan CK, Nahin RL, et al. Associations of gait speed and other measures of physical function with cognition in a healthy cohort of elderly persons. *J Gerontol A Biol Sci Med Sci* 2007; **62:** 1244–51.

41 Petersen RC. Conversion. *Neurology* 2006; **67** (suppl 3)**:** S12–13.

42 Green RC, DeKosky ST. Primary prevention trials in Alzheimer disease. *Neurology* 2006; **67** (suppl 3)**:** S2–5.

43 Thal LJ, Ferris SH, Kirby L, et al. A randomized, double-blind, study of rofecoxib in patients with mild cognitive impairment. *Neuropsychopharmacology* 2005; **30:** 1204–15.

44 Petersen RC, Thomas RG, Grundman M, et al. Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med* 2005; **352:** 2379–88.

45 Feldman HH, Ferris S, Winblad B, et al. Effect of rivastigmine on delay to diagnosis of Alzheimer's disease from mild cognitive impairment: the InDDEx study. *Lancet Neurol* 2007; **6:** 501–12.

46 Grundman M, Petersen RC, Bennett DA, et al. Alzheimer Association research roundtable meeting on mild cognitive impairment: what we have learned? *Alzheimer Dement* 2006; **2:** 220–33.

47 Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med* 2004; **256:** 183–94.

48 Tounsi H, Deweer B, Ergis AM, et al. Sensitivity to semantic cuing: an index of episodic memory dysfunction in early Alzheimer disease. *Alzheimer Dis Assoc Disord* 1999; **13:** 38–46.

49 Harrison J, Minassian SL, Jenkins L, Black RS, Koller M, Grundman M. A neuropsychological test battery for use in Alzheimer disease clinical trials. *Arch Neurol* 2007; **64:** 1323–29.

50 Ferris SH, Lucca U, Mohs R, et al. Objective psychometric tests in clinical trials of dementia drugs. Position paper from the International Working Group on Harmonization of Dementia Drug Guidelines. *Alzheimer Dis Assoc Disord* 1997; **11** (suppl 3)**:** 34–8.

51 Hebert LE, Scherr PA, Bienias JL, Bennett DA, Evans DA. Alzheimer disease in the US population: prevalence estimates using the 2000 census. *Arch Neurol* 2003; **60:** 1119–22.

52 Winblad B, Engedal K, Soininen H, et al. A 1-year, randomized, placebo-controlled study of donepezil in patients with mild to moderate AD. *Neurology* 2001; **57:** 489–95.

53 Wilcock G, Howe I, Coles H, et al. A long-term comparison of galantamine and donepezil in the treatment of Alzheimer's disease. *Drugs Aging* 2003; **20:** 777–89.

54 Courtney C, Farrell D, Gray R, et al. Long-term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): randomised double-blind trial. *Lancet* 2004; **363:** 2105–15.

55 Mohs RC, Doody RS, Morris JC, et al. A 1-year, placebo-controlled preservation of function survival study of donepezil in AD patients. *Neurology* 2001; **57:** 481–88.

56 Rogers SL, Farlow MR, Doody RS, Mohs R, Friedhoff LT. A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. *Neurology* 1998; **50:** 136–45.

57 Rogers SL, Doody RS, Mohs RC, Friedhoff LT. Donepezil improves cognition and global function in Alzheimer disease: a 15-week, double-blind, placebo-controlled study. *Arch Intern Med* 1998; **158:** 1021–31.

58 Burns A, Rossor M, Hecker J, et al. The effects of donepezil in Alzheimer's disease—results from a multinational trial. *Dement Geriatr Cogn Disord* 1999; **10:** 237–44.

59 Homma A, Takeda M, Imai Y, et al. Clinical efficacy and safety of donepezil on cognitive and global function in patients with Alzheimer's disease. A 24-week, multicenter, double-blind, placebo-controlled study in Japan. *Dement Geriatr Cogn Disord* 2000; **11:** 299–313.

60 Rosler M, Anand R, Cicin-Sain A, et al. Efficacy and safety of rivastigmine in patients with Alzheimer's disease: international randomised controlled trial. *BMJ* 1999; **318:** 633–38.

61 Raskind MA, Peskind ER, Wessel T, Yuan W. Galantamine in AD: a 6-month randomized, placebo-controlled trial with a 6-month extension. *Neurology* 2000; **54:** 2261–68.

62 Wilcock GK, Lilienfeld S, Gaens E. Efficacy and safety of galantamine in patients with mild to moderate Alzheimer's disease: multicentre randomised controlled trial. *BMJ* 2000; **321:** 1445–49.

63 Tariot PN, Solomon PR, Morris JC, Kershaw P, Lilienfeld S, Ding C. A 5-month, randomized, placebo-controlled trial of galantamine in AD. *Neurology* 2000; **54:** 2269–76.

64 Wilkinson D, Murray J. Galantamine: a randomized, double-blind, dose comparison in patients with Alzheimer's disease. *Int J Geriatr Psychiatry* 2001; **16:** 852–57.

65 Rockwood K, Mintzer J, Truyen L, Wessel T, Wilkinson D. Effects of a flexible galantamine dose in Alzheimer's disease: a randomised, controlled trial. *J Neurol Neurosurg Psychiatry* 2001; **71:** 589–95.

66 Risner ME, Saunders AM, Altman JF, et al. Efficacy of rosiglitazone in a genetically defined population with mild-to-moderate Alzheimer's disease. *Pharmacogenomics J* 2006; **6:** 246–54.

67 Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry* 1984; **141:** 1356–64.

68 Reitan R. Validity of the trail making test as an indicator of brain damage. *Percept Mot Skills* 1958; **8:** 271–76.

69 Stroop JP. Studies of interference on serial verbal reaction. *J Exp Physiol* 1935; **18:** 643–62.

70 Mohs RC, Knopman D, Petersen RC, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's disease assessment scale that broaden its scope. *Alzheimer Dis Assoc Disord* 1997; **11** (suppl 2)**:** S13–21.

71 Gilman S, Koller M, Black RS, et al. Clinical effects of Abeta immunization (AN1792) in patients with AD in an interrupted trial. *Neurology* 2005; **64:** 1553–62.

72 Helmer C, Andrieu S, Peres K, Orgogozo JM, Vellas B, Dartigues JF. Predictive value of 6-month decline in ADAS-cog for survival without severe Alzheimer's disease. *Dement Geriatr Cogn Disord* 2007; **23:** 168–74.

73 Feldman HH, Gauthier S, Chertkow H, Conn DK, Freedman M, Chris M. Progress in clinical neurosciences: Canadian guidelines for the development of antidementia therapies: a conceptual summary. *Can J Neurol Sci* 2006; **33:** 6–26.

74 Galasko D, Bennett D, Sano M, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord* 1997; **11** (suppl 2)**:** S33–39.

75 Gelinas I, Gauthier L, McIntyre M, Gauthier S. Development of a functional measure for persons with Alzheimer's disease: the disability assessment for dementia. *Am J Occup Ther* 1999; **53:** 471–81.

76 Feldman H, Sauter A, Donald A, et al. The disability assessment for dementia scale: a 12-month study of functional ability in mild to moderate severity Alzheimer disease. *Alzheimer Dis Assoc Disord* 2001; **15:** 89–95.

77 Byrne LM, Wilson PM, Bucks RS, Hughes AO, Wilcock GK. The sensitivity to change over time of the Bristol Activities of daily living scale in Alzheimer's disease. *Int J Geriatr Psychiatry* 2000; **15:** 656–61.

78 Bucks RS, Ashworth DL, Wilcock GK, Siegfried K. Assessment of activities of daily living in dementia: development of the Bristol activities of daily living scale. *Age Ageing* 1996; **25:** 113–20.

79 Barberger-Gateau P, Fabrigoule C, Helmer C, Rouch I, Dartigues JF. Functional impairment in instrumental activities of daily living: an early clinical sign of dementia? *J Am Geriatr Soc* 1999; **47:** 456–62.

80 Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric inventory: comprehensive assessment of psychopathology in dementia. *Neurology* 1994; **44:** 2308–14.

81 Aalten P, Verhey FRJ, Boziki M, et al. Consistency of neuropsychiatric syndromes across dementias. Results from the European Alzheimer Disease Consortium (EADC): part II. *Dementia Geriatric Cogn Disord* **25:** 1–8.

82 Schneider LS, Olin JT, Doody RS, et al. Validity and reliability of the Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change. *Alzheimer Dis Assoc Disord* 1997; **11** (suppl 2)**:** S22–32.

83 European Medicines Agency. Committee for medicinal products for human use (CHMP) draft guideline on medicinal products for the treatment of Alzheimer's disease and other dementias. http://www.emea.europa.eu/pdfs/human/ewp/055395endraft.pdf (accessed Nov 13, 2007).

84 Reisberg B, Schneider L, Doody R, et al. Clinical global measures of dementia. Position paper from the International Working Group on Harmonization of Dementia Drug Guidelines. *Alzheimer Dis Assoc Disord* 1997; **11** (suppl 3)**:** 8–18.

85 Morris JC. The clinical dementia rating (CDR): current version and scoring rules. *Neurology* 1993; **43:** 2412–14.

86 Auer S, Reisberg B. The GDS/FAST staging system. *Int Psychogeriatr* 1997; **9** (suppl 1)**:** 167–71.

87 Reisberg B. Functional assessment staging (FAST). *Psychopharmacol Bull* 1988; **24:** 653–59.

88 Wimo A, Nordberg G. Validity and reliability of assessments of time. Comparisons of direct observations and estimates of time by the use of the resource utilization in dementia (RUD)-instrument. *Arch Gerontol Geriatr* 2007; **44:** 71–81.

89 Leber P. Observations and suggestions on antidementia drug development. *Alzheimer Dis Assoc Disord* 1996; **10** (suppl 1)**:** 31–35.

90 Hendrix S, Horton S, Orgogozo JM. Modification of the "randomized withdrawl" and "staggered start" clinical trial designs: toward a practical demonstration of disease modification in Alzheimer's disease (AD). http://www.myriad.com/downloads/EFNS-2007-Disease-Modification-in-AD-Hendrix.pdf (accessed Feb 18, 2008).

91 Schneider L. Prevention therapeutics of dementia. *Alzheimer Dement* 2008; **4:** S122–30.

92 Sanofi-Aventis. Study of Xaliproden (SR57746A) in patients with mild-to-moderate dementia of the Alzheimer's type. http://clinicaltrials.gov/ct2/show/NCT00104013 (accessed Jan 9, 2008).

93 NIA. High dose supplements to reduce homocysteine and slow the rate of cognitive decline in Alzheimer's disease (vitamins to slow Alzheimer's—VITAL). http://clinicaltrials.gov/ct2/show/NCT00056225 (accessed Jan 9, 2008).

94 Neurochem. Evaluation of 3APS in patients with mild to moderate Alzheimer's disease. http://clinicaltrials.gov/ct2/show/NCT00088673 (accessed Jan 9, 2008).

95 Neurochem. European study of 3APS in mild to moderate Alzheimer's disease patients. http://clinicaltrials.gov/ct2/show/NCT00217763 (accessed Jan 9, 2008).

96 NIA. Cholesterol lowering agent to slow progression (CLASP) of Alzheimer's disease study. http://clinicaltrials.gov/ct2/show/NCT00053599 (accessed Jan 9, 2008).

97 Feldman H, Sparks D, Doody R, et al. Design and baseline data for a study of lipitor's effect in Alzheimer's dementia. http://www.kenes.com/adpd2007/program/session1.asp?SessionId=PO2N&SSessionDate=3/16/2007 (accessed Jan 9, 2008) (abstr).

98 Myriad. Flurizan: Alzheimer's disease phase III clinical trial (USA). http://www.myriad.com/research/alzheimers-disease-clinical-trial.php (accessed Jan 9, 2008).

99 Myriad. Flurizan: Alzheimer's disease phase III clinical trial (global). http://www.myriad.com/research/alzheimers-disease-clinical-trial-global.php (accessed Jan 9, 2008).

100 NIA. A randomized double-blind placebo-controlled trial of the effects of docosahexaenoic acid (DHA) in slowing the progression of Alzheimer's disease. http://clinicaltrials.gov/ct2/show/NCT00440050 (accessed Jan 9, 2008).

101 Vellas B, Andrieu S, Cantet C, Dartigues J, Gauthier S. Long-term changes in ADAS-cog: what is clinically relevant for disease modifying trials in Alzheimer? *J Nutr Health Aging* 2007; **11:** 338–41.

102 Ferris SH, Aisen PS, Cummings J, et al. ADCS prevention instrument project: overview and initial results. *Alzheimer Dis Assoc Disord* 2006; **20** (suppl 3)**:** S109–23.

103 Fox NC, Black RS, Gilman S, et al. Effects of Abeta immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. *Neurology* 2005; **64:** 1563–72.

104 Broich K. Outcome measures in clinical trials on medicinal products for the treatment of dementia: a European regulatory perspective. *Int Psychogeriatr* 2007; **19:** 509–24.

105 Black SE, Doody R, Li H, et al. Donepezil preserves cognition and global function in patients with severe Alzheimer disease. *Neurology* 2007; **69:** 459–69.

106 Winblad B, Kilander L, Eriksson S, et al. Donepezil in patients with severe Alzheimer's disease: double-blind, parallel-group, placebo-controlled study. *Lancet* 2006; **367:** 1057–65.

107 Winblad B, Poritis N. Memantine in severe dementia: results of the 9M-Best Study (benefit and efficacy in severely demented patients during treatment with memantine). *Int J Geriatr Psychiatry* 1999; **14:** 135–46.

108 Tariot PN, Farlow MR, Grossberg GT, Graham SM, McDonald S, Gergel I. Memantine treatment in patients with moderate to severe Alzheimer disease already receiving donepezil: a randomized controlled trial. *JAMA* 2004; **291:** 317–24.

109 Feldman H, Gauthier S, Hecker J, et al. Efficacy of donepezil on maintenance of activities of daily living in patients with moderate to severe Alzheimer's disease and the effect on caregiver burden. *J Am Geriatr Soc* 2003; **51:** 737–44.

110 Feldman H, Gauthier S, Hecker J, Vellas B, Subbiah P, Whalen E. A 24-week, randomized, double-blind study of donepezil in moderate to severe Alzheimer's disease. *Neurology* 2001; **57:** 613–20.

111 Reisberg B, Doody R, Stoffler A, Schmitt F, Ferris S, Mobius HJ. Memantine in moderate-to-severe Alzheimer's disease. *N Engl J Med* 2003; **348:** 1333–41.

112 Eisai Medical Research. Comparison of 23 mg Donepezil sustained release (SR) to 10 mg donepezil immediate release (IR) in patients with moderate to severe Alzheimer's disease. http://clinicaltrials.gov/ct2/show/NCT00478205 (accessed Jan 9, 2008).

113 Eisai Medical Research. Efficacy and safety of aricept in the treatment of severe Alzheimer's disease http://clinicaltrials.gov/ct2/show/NCT00096473 (accessed Jan 9, 2008).

114 Forest Laboratories. An evaluation of the safety and efficacy of memantine in agitated patients with moderate to severe Alzheimer's disease. http://clinicaltrials.gov/ct2/show/NCT00097916 (accessed Jan 9, 2008).

115 New York University School of Medicine. Memantine and comprehensive, individualized management of Alzheimer's patients and caregiver training. http://clinicaltrials.gov/ct2/show/NCT00120874 (accessed Jan 9, 2008).

116 East Kent Hospitals Trust. Memantine for agitation in dementia (MAGD). http://clinicaltrials.gov/ct2/show/NCT00371059 (accessed Jan 9, 2008).

117 Eisai Medical Research. A multi-center, randomized, double-blind, parallel group study with 3 groups receiving placebo, 5 mg of E2020 and 10 mg of E2020. http://clinicaltrials.gov/ct2/show/NCT00165659 (accessed Jan 9, 2008).

118 Forest Laboratories. A study of the safety and efficacy of memantine in moderate to severe Alzheimer's disease. http://clinicaltrials.gov/ct2/show/NCT00322153 (accessed Jan 9, 2008).

119 Janssen Pharmaceutica. Treatment of severe Alzheimer's disease: evaluation of efficacy and safety of galantamine hydrobromide in a controlled study. http://clinicaltrials.gov/ct2/show/NCT00216593 (accessed Jan 9, 2008).

120 Saxton J, McGonigle-Gibson K, Swihart A, Miller M, Boller F. Assessment of severely impaired patients: description and validation of a new neuropsychological test battery. *Psychol Assess* 1990; **2:** 298–303.

121 Galasko D, Schmitt F, Thomas R, Jin S, Bennett D. Detailed assessment of activities of daily living in moderate to severe Alzheimer's disease. *J Int Neuropsychol Soc* 2005; **11:** 446–53.

122 Leon J, Cheng CK, Neumann PJ. Alzheimer's disease care: costs and potential savings. *Health Aff (Millwood)* 1998; **17:** 206–16.

123 Zarit SH, Reever KE, Bach-Peterson J. Relatives of the impaired elderly: correlates of feelings of burden. *Gerontologist* 1980; **20:** 649–55.

124 Greene JG, Smith R, Gardiner M, Timbury GC. Measuring behavioural disturbance of elderly demented patients in the community and its effects on relatives: a factor analytic study. *Age Ageing* 1982; **11:** 121–26.

125 Baltussen R, Leidl R, Ament A. Real world designs in economic evaluation. Bridging the gap between clinical research and policy-making. *Pharmacoeconomics* 1999; **16:** 449–58.

126 Limat S, Woronoff-Lemsi MC, Menat C, Madroszyk-Flandin A, Merrouche Y. From randomised clinical trials to clinical practice: a pragmatic cost-effectiveness analysis of paclitaxel in first-line therapy for advanced ovarian cancer. *Pharmacoeconomics* 2004; **22:** 633–41.

127 Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world. Effectiveness versus efficacy studies. *Pharmacoeconomics* 1999; **15:** 423–34.

128 Wolfe F, Michaud K, Dewitt EM. Why results of clinical trials and observational studies of antitumour necrosis factor (anti-TNF) therapy differ: methodological and interpretive issues. *Ann Rheum Dis* 2004; **63** (suppl 2)**:** 13–17.

129 Logston RG, MCCurry SM, Teri L. Quality of life in Alzheimer's disease: patient and caregiver reports. *J Mental Health Aging* 1999; **5:** 21–32.

130 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; **10:** 407–15.

131 Hsieh YW, Wang CH, Wu SC, Chen PC, Sheu CF, Hsieh CL. Establishing the minimal clinically important difference of the Barthel Index in stroke patients. *Neurorehabil Neural Repair* 2007; **21:** 233–38.

132 Schrag A, Sampaio C, Counsell N, Poewe W. Minimal clinically important change on the unified Parkinson's disease rating scale. *Mov Disord* 2006; **21:** 1200–07.

133 Gillette-Guyonnet S, Andrieu S, Cortes F, et al. Outcome of Alzheimer's disease: potential impact of cholinesterase inhibitors. *J Gerontol A Biol Sci Med Sci* 2006; **61:** 516–20.